

Article

Data-Driven Decision Support for Adult Autism Diagnosis Using Machine Learning [†]

Sotirios Batsakis ^{1,2}, Marios Adamou ³, Ilias Tachmazidis ², Sarah Jones ³, Sofya Titarenko ⁴, Grigoris Antoniou ^{3,*} and Thanasis Kehagias ⁵

¹ School of Production Engineering and Management, Technical University of Crete, 73100 Chania, Greece; s.batsakis@hud.ac.uk

² Department of Computer Science, School of Computing and Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK; i.tachmazidis@hud.ac.uk

³ South West Yorkshire Partnership NHS Foundation Trust, Wakefield WF1 3SP, UK; marios.adamou@swyt.nhs.uk (M.A.); sarah.jones1@swyt.nhs.uk (S.J.)

⁴ School of Mathematics, University of Leeds, Leeds LS2 9JT, UK; s.titarenko@leeds.ac.uk

⁵ Department of Electrical and Computer Engineering, Aristotle University, 54124 Thessaloniki, Greece; kehagiat@auth.gr

* Correspondence: g.antoniou@hud.ac.uk

† This paper is an extended version of our paper published in Batsakis, S.; Adamou, M.; Tachmazidis, I.; Antoniou, G.; Kehagias, T. Data-Driven Decision Support for Autism Diagnosis using Machine Learning. In Proceedings of the 13th International Conference on Management of Digital EcoSystems (MEDES'21), Hammamet, Tunisia, 1–3 November 2021.



Citation: Batsakis, S.; Adamou, M.; Tachmazidis, I.; Jones, S.; Titarenko, S.; Antoniou, G.; Kehagias, T. Data-Driven Decision Support for Adult Autism Diagnosis Using Machine Learning. *Digital* **2022**, *2*, 224–243. <https://doi.org/10.3390/digital2020014>

Academic Editors: Mirjana Ivanović, Richard Chbeir, Yannis Manolopoulos and Markos G. Tsipouras

Received: 16 February 2022

Accepted: 2 May 2022

Published: 11 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Adult referrals to specialist autism spectrum disorder diagnostic services have increased in recent years, placing strain on existing services and illustrating the need for the development of a reliable screening tool, in order to identify and prioritize patients most likely to receive an ASD diagnosis. In this work a detailed overview of existing approaches is presented and a data driven analysis using machine learning is applied on a dataset of adult autism cases consisting of 192 cases. Our results show initial promise, achieving total positive rate (i.e., correctly classified instances to all instances ratio) up to 88.5%, but also point to limitations of currently available data, opening up avenues for further research. The main direction of this research is the development of a novel autism screening tool for adults (ASTA) also introduced in this work and preliminary results indicate the ASTA is suitable for use as a screening tool for adult populations in clinical settings.

Keywords: machine learning; autism diagnosis; decision support

1. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by a pervasive impairment in reciprocal social interaction and communication, alongside restricted interests and repetitive behaviors [1,2]. Thus far, no biological markers are evident. It is estimated to affect 9.8 per 1000 adults in England [3]. ASD is usually diagnosed in childhood; however, it is recognized as a lifelong condition [4–8]. In recent years there has been a marked increase in the number of adults referred for autism assessment [9], consequently placing greater demands on health services. Because of this pressure, the time for diagnosis is lengthy with one report finding 29% of adults with autism and 46% of those with Asperger's disorder did not receive a diagnosis until adulthood [10].

NICE (National Institute for Health and Care Excellence, UK) guidelines recommend diagnosis of ASD in adulthood is reached on a consensus of expert opinion made by observations from a variety of assessments, including detailed history taking, current behavioral factors, and cognitive abilities [11]. This means that ASD diagnosis is expensive in time and resources; typically, assessments are lengthy, and subjective. Observations undertaken by multidisciplinary teams should be usual diagnostic procedure [12], which are comprised of

evaluation of current functioning and behaviors, together with a detailed history taking [13]. This process can be complex as the ASD phenotype presents with a range of severities, language ability, and intellects [11]. Furthermore, pertinent to adult ASD populations, issues may occur due to (1) difficulties acquiring an accurate early history; (2) differentiating autistic symptoms from learned behavior or compensation strategies; and (3) differentiating from other conditions, or mental health disorders, specifically schizophrenia [14,15]. These factors may lead to misdiagnosis [12,14–23]. Diagnosing autism is resource intensive because of the quantity of information which is required, ideally from a variety of sources. If information from a caregiver is not available, it can be a challenge to obtain an accurate account of the neurodevelopmental period, as self-insight from the service user may be inaccurate [17,18].

There is a necessity to relieve the pressures on specialist diagnostic services by screening waiting lists to identify and prioritize referrals that are at a greater probability of receiving an autism diagnosis [24]. Employing screening tools can facilitate a timely and economical approach for specialist services if they can identify patients who are more likely to have autism, using a standardized method [24].

Whilst a varied collection of ASD screening measures is available for both developmental and adulthood populations, for ASD in adulthood, the most generally used screening measures for ASD is the autism questionnaire presented in [25], which forms the basis of the analysis in the first part of this work. The objective of this part of the work is to apply machine learning for analyzing autism questionnaire results and investigating the components of the assessment, in relation to diagnostic outcome in a clinical setting. In turn, analysis results can offer insights for decision support for autism diagnosis. This is followed by the introduction of novel assessment tools: the first is completed by the clinician and the second is completed by the patient.

The remainder of this paper is organized as follows. Background and related work are presented in Section 2. Assessment data and analysis over current data is presented in Section 3. Novel screening tools are presented in Section 4 and conclusions and directions of future work are presented in Section 5.

2. Background and Related Work

Numerous screening tools are available for quantifying childhood and adulthood ASD [26–32], yet issues of validity are apparent. Recommended clinical screening measures for quantifying ASD in adulthood include the autism-spectrum quotient (AQ) [33] the Ritvo Autism and Asperger Diagnostic Scale-Revised (RAADS-R) [9,13,34]. The AQ was developed to quantify high functioning autism (HFA) and Asperger’s syndrome (AS) in adult populations. It serves as a standardized measure which can aid clinicians to identify patients that would benefit from a full ASD assessment [25]. Generally, the AQ boasts high sensitivity and specificity [25,33–36]. However, clinically the AQ has shown to be problematic [37–40].

In a clinical sample of 132 patients referred for clinical diagnostic assessment, Kenny and Stansfield [37] reported no difference in scores on the AQ, regardless of ASD/non-ASD diagnosis after full assessment. More recently, Adamou et al. [39] explored the predictive efficacy of the AQ compared to final diagnostic formulation by an expert multidisciplinary team, in a sample of adults referred to a specialist diagnostic service. The AQ measured 74% sensitivity and 30.3% specificity, respectively. No significant association between scores on the AQ and diagnostic outcome was evident. Similar levels of sensitivity (77%) and specificity (29%) have been reported by Ashwood et al. [40] in an ASD sample of 476 patients. In a study which explored AQ scores in adults diagnosed with ASD with average and below average intelligence, only 17% of the sample scored above the diagnostic cutoff of the AQ which again indicates a significantly lower sensitivity than in the original study [41]. Furthermore, AQ scores have failed to correlate with other popular measures of ASD, such as the Autism Diagnostic Interview-Revised or Vineland scores [41].

In studies employing control samples, the AQ has shown discriminative ability between ASD profiles and neurotypical profiles [32,36,42–46], yet it remains uncertain as to how well the AQ performs in those who do not have a clinical diagnosis of ASD, but display ASD traits [46]. In a systematic review of screening tools for ASD populations it was concluded that even though the AQ is commonly utilized in clinics, it is considerably under researched, therefore no recommendations on its use could be put forth by the review [47].

Validation issues are also evident for different measures of ASD that are often used in clinics [48]. The Ritvo Autism Asperger’s Diagnostic Scale-Revised (RAADS-R) was developed for adults, based on the ICD-10 and DSM-5 diagnostic criteria. It covers four areas of neurodevelopment (language, sensory motor, circumscribed interest, and social cognition). The RAADS assessment has a reported sensitivity of 97% and specificity of 100% [49,50]. However other studies have questioned its validity.

In a recent study, Jones et al. [51] found RAADS failed to differentiate between ASD/non-ASD patients after full clinical assessment. Levels of false positives were high, with the assessment only having a 3.03% chance of detecting the absence of ASD in the sample. Other studies have found the assessment (including RAADS-14 [52]) is likely to result in high levels of false positives [34], is unable to differentiate between ASD/non-ASD groups [53], and has significantly reduced specificity in psychiatric control groups [54]. Due to the high levels of false positives, it has been recommended the cut-off threshold score is too low to be clinically valuable [55], and that the assessment fails to cover a full range of behavioral issues, particularly those relevant to milder forms of ASD [56].

The concept of concurrent validity is appropriate here. The RAADS-R has shown a strong positive correlation with AQ scores [52,57] and with validity issues surrounding the AQ [37,39,40], this is problematic for both assessments. It is important to note that a potential justification for the low levels of specificity reported in these studies may be due to the high levels of comorbidity demonstrated in ASD profiles [6,58–63]. For instance, anxiety and depression, may imitate particular ASD symptoms [40,64] thereby leading to false positives. However, until these issues are fully resolved, such assessments are not reliable gauges of which patients should receive full ASD assessment as priority [65]. Notice also that extensive work has been done on ASD diagnosis for children using machine learning [66–70], but adult ASD diagnosis, which is the topic of this work, is a much less studied topic.

3. Data Analysis Using Machine Learning

The dataset used in the machine learning based analysis initially presented in [71] consists of autism assessment results for 192 patients, from Adult ADHD and Autism Service, South West Yorkshire Partnership NHS Foundation Trust, in the South and West Yorkshire geographical area, between 2017 and 2018. The Adult ADHD and Autism Service is a specialist Service in diagnosing ADHD and autism in adulthood. Patients are referred to the service by health care professionals, whom deem it appropriate based on patient’s history and current difficulties. Inclusion criteria dictated that participants were over the age of 18 years (no cut of), had a good comprehension of the English language, and IQ within normal range. The assessment is designed to identify adults who may benefit from a full diagnostic assessment for autism spectrum disorder.

The assessment procedure adopts the procedure proposed in [25] and consists of two parts. The first part consists of a test that the examined individual completes based on AAA AQ and AAA EQ parts (the RAADS AQ, EQ, RQ questionnaires presented in Section 2). The second part (AAA RQ score) is the result of answers of persons familiar with the examined individual, typically close relatives. Related to the diagnosis are social aspects, communication, imagination and obsessions of the examined individual (these are features CLASS SOCIAL, CLASS OBSESSIONS, CLASS COMMUNICATION and CLASS IMAGINATION) and they are defined from responses to AAA AQ, EQ and RQ and clinician’s input. These parts of the AAA examination in turn are the Autism-Spectrum Quotient (AQ) score [33] and the Empathy Quotient (EQ) score [25], in addition to Relatives

Quotient (RQ). Given the AAA AQ, AAA EQ and AAA RQ responses clinicians confirm answers (Yes = 1), which count towards CLASS classification. Thus, CLASS classification is a function of AAA responses and clinician's assessment. The last feature of the dataset is the diagnostic outcome which is a binary categorical feature that the machine learning model has to predict. Overall, the dataset is unbalanced with 28 out of 192 examined patients (14.58%) being diagnosed with autism after a full assessment is completed. Thus, in total the dataset consists of seven numerical input features (three consisting solely of questionnaire's results and four based on questionnaire's results and clinician's input) and an output categorical feature.

The objective of data analysis is to create a model for predicting the diagnostic outcome given the AAA test data [25] as input. Specifically, the input data are AAA test results consisting of AAA AQ, AAA EQ and AAA RQ scores. The AAA AQ has numerical values ranging from 4 to 50 with a mean 34.74 and a standard deviation 8.47, for EQ the corresponding values are 0, 80, 19.99 and 11.38 and for RQ the values are 0, 31, 18.21 and 6.31. In addition, the input data include the features CLASS SOCIAL, CLASS OBSESSIONS, CLASS COMMUNICATION and CLASS IMAGINATION derived from AAA test responses as defined in [34]. The CLASS SOCIAL values range from min = 0 to max = 11 with mean value 2.39 and standard deviation 1.46, for the CLASS OBSESSIONS the corresponding values are 0, 9, 2.30 and 1.22, for CLASS COMMUNICATION values are 0, 5, 2.17 and 1.32 and for CLASS IMAGINATION min, max, mean and standard deviation values are 0, 4, 1.05 and 0.87. The dataset consists of exam results of 192 individuals, with 85.24% of diagnostic outcomes being negative. In this work, various classification methods have been used for the analysis.

3.1. Analysis Using Weka

The first part of the analysis consisted of the application of six machine learning algorithms using Weka [72] over the dataset as presented in [71]. Three of the algorithms are non-interpretable and three are interpretable. The non-interpretable algorithms are multilayer perceptron (the neural network implementation in Weka), SMO (sequential minimal optimization algorithm for training a support vector classifier) and random forest. The interpretable algorithms are the decision tree (J48), logistic regression and semantic artificial neural networks (SANN) [73]. SANN is a variant of neural networks with labeled hidden layer nodes which can be interpreted as logistic regression over each layer given the previous one. In all experiments, pre-processing has been applied by replacing missing values with the average value, while performance estimation and model selection was based on 10-fold cross validation.

The results of experiments using the non-interpretable classification algorithms of Weka and the default hyperparameters are presented in Table 1 (optimal values as marked in bold). Although Table 1 presents some basic results using the non-interpretable algorithms, the imbalance of the dataset and the relative importance of the different diagnostic outcomes and corresponding consequences make the overall precision of algorithms one—but not the only—factor to take into account in the analysis. Thus, a detailed examination is required in order to assess the true usability of a data driven analysis in the decision process. Specifically, the cost of error varies given its type, typically it is a more serious error to predict a negative diagnostic outcome when it is actually positive resulting in the patient not receiving the needed treatment, compared to predicting a positive diagnosis when in fact it is negative with the cost being that of that of conducting a full assessment that eventually leads to a negative diagnosis. This observation in turn changes the use of a machine learning model in practice.

Typically, when each class is considered equally important and having similar costs for all types of errors a classifier selects the class having the higher probability. However, when classes have different importance and also different costs in case of classification errors, then the selection threshold of an algorithm must be adjusted accordingly. Data driven analysis may help making such policies more accurate and efficient. In practice,

up to a certain degree, it is better to make an additional assessment of positive diagnosis to the patient rather than to select a negative diagnostic outcome (which could actually be positive).

Table 1. Classification results using non interpretable algorithms of Weka.

Model	Total Positive Rate	ROC Area
Multilayer Perceptron	0.885	0.805
SMO	0.854	0.500
Random Forest	0.859	0.870

After taking the above observations into account the detailed results for each algorithm are the following: SMO actually assigns all instances as having negative diagnostic outcome where the total positive rate is 0.854 (percentage of instances with negative diagnostic outcome) and the receiver operating characteristic (ROC) curve (or area under the curve—AUC) is 0.500, corresponding to a random classification, thus this model cannot be used in practice. Random forest achieved better results with total positive rate 0.859 and the ROC curve is 0.870. In this case, the classifier can be useful in practice. For example, given a policy that assigns much higher cost to a false negative error than to a false positive, the diagnostic outcome can be classified as positive even if the probability is low, in order to avoid false negative errors. Subsequently, if an assessment result is positive even if the probability of such outcome is according to classifier just 1% then all 28 positive cases will be classified correctly and so are 47 of the negative ones, with the cost of having to provide full assessment in the 117 remaining negative cases. Thus, the classifier can be used for making a decision for filtering out some cases, but also providing full assessment to all cases that have a positive diagnosis. By increasing the threshold to 2% the classification is correct for 26 out of the 28 positive cases and 69 out of the 164 negative cases (95 negative cases will still have full assessment). Thus, reduction of false positives is combined with increase of false negatives and the relative cost of errors is used for defining the proper threshold and decision policy rather than the threshold value that maximizes classification accuracy, that is reported in Table 1. In case of multilayer perceptron (neural network), the total positive rate is 0.885 and the ROC curve is 0.805, thus offering the possibility of implementing a selection policy minimizing the cost of errors, but without creating an interpretable model.

Even though non-interpretable algorithms can assist in decision making by producing models that can predict the probability (given the results of an assessment) of a specific diagnostic outcome, thus facilitating the definition of a decision policy given the relative costs of errors, interpretability of the prediction model is often an important issue. Compliance to legal requirements and regulations means that specific rules have been taken into account when applying an AI-based system and this in turn means that the system's functionality is transparent and interpretable. A proposed approach is to employ interpretable machine learning algorithms, such as logistic regression and decision trees [74]. These algorithms are often efficient but do not always perform as non-interpretable ones, such as support vector machines (SVM) and neural networks.

In the case of neural networks, using existing knowledge for building neural networks was first proposed in [75] and further developed in [76], introducing the knowledge-based artificial neural networks (KBANN). These networks are constructed based on knowledge represented using logic rules, and in [73] a variant of KBANN called semantic artificial neural networks (SANNs) is proposed. SANNs are neural networks with labeled hidden layer nodes as KBANNs, but the construction of such neural networks is based on knowledge graphs rather than rules. In this work the interpretable algorithms applied to the autism assessment dataset are: logistic regression, J48 decision tree and SANN. The SANN is constructed by introducing to the hidden layer nodes representing the AAA score (combining AAA AQ, AAA EQ and AAA RQ scores) and the CLASS score

(combining the CLASS SOCIAL, CLASS OBSESSIONS, CLASS COMMUNICATION and CLASS IMAGINATION scores). The resulting network is presented in Figure 1.

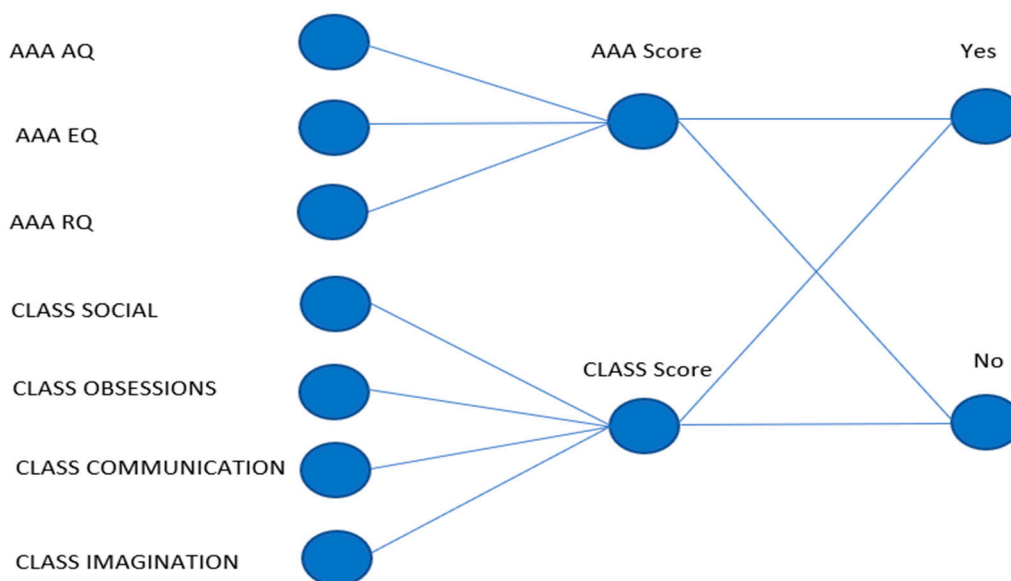


Figure 1. Semantic artificial neural network for classification on the autism dataset.

The results using the interpretable algorithms of Weka are presented in Table 2 (optimal values as marked in bold). In medical diagnosis, interpreting the models is significant for decision making, thus we choose to present the two categories of algorithms separately, since in case that interpretability is not an option but a strict requirement then only the corresponding algorithms can be used. Decision tree (J48) achieved a total positive rate of 0.870 and ROC curve of 0.775.

Table 2. Classification results using interpretable algorithms of Weka.

Model	Total Positive Rate	ROC Area
Logistic Regression	0.844	0.814
Decision Tree (J48)	0.870	0.775
SANN	0.875	0.870

In the case of logistic regression, the coefficients for predicting a negative diagnosis result are AAA AQ: 0.0381, AAA EQ: -0.0064 , AAA RQ: -0.1282 , CLASS SOCIAL: -0.585 , CLASS OBSESSIONS: -0.2791 , CLASS COMMUNICATION: -0.371 , CLASS IMAGINATION: -0.6105 and Intercept: 7.344. These coefficients indicate factors correlated positively or negatively with negative diagnosis and the degree of this correlation (with CLASS features and AAA RQ having more weight).

The third algorithm, SANN, (using the network of Figure 1) achieved a total positive rate of 0.875 and ROC curve of 0.870, outperforming the other two interpretable algorithms. There are two hidden layer nodes in the SANN, the AAA Score node representing the cumulative AAA score and CLASS Score node representing cumulative CLASS score. The output node representing negative diagnostic output has weights of 3.21 at input from the AAA Score node and 4.84 at input from CLASS Score node, while the corresponding weights at positive diagnostic outcome node are -3.21 and -4.48 , respectively. Thus, the positive diagnostic outcome has lower probability when cumulative AAA and CLASS scores are higher. The AAA Score in turn has weights of 5.07 from AAA AQ input, -10.10 from AAA EQ and -12.39 from AAA RQ indicating that overall the higher the AAA AQ the lower the probability of a positive diagnosis and that lower AAA EQ and AAA RQ scores increase the probability of positive diagnostic outcome. Furthermore, AAA

EQ and AAA RQ scores have more weight than AAA AQ. The corresponding weights for the cumulative CLASS Score are for CLASS SOCIAL: -12.70 , CLASS OBSESSIONS: -3.24 , CLASS COMMUNICATION: -3.81 and CLASS IMAGINATION: -2.81 indicating that lower CLASS scores increase probability of positive diagnostic outcome.

Depending on the relative cost of classification errors, by setting a low threshold for accepting a positive diagnosis, the created model can be used to filter out cases which have a negative diagnostic outcome with very high probability. For example, when setting a threshold for classifying a case as positive to 1% then 26 out of 28 positive cases are classified correctly and so are 86 out of 164 negative cases (thus a full assessment is applied for 78 negative cases). Thus, practically more than half of negative cases can be exempted from further examination while keeping almost all of positive cases. This is actually similar to the clinical assessment practice. For example, in this dataset, out of the 192 cases, 28 are positive and 164 are negative. In the screening process, 125 cases went through full assessment and 67 did not. Finally, of these 125 cases, 26 were positive and 99 were negative. Out of the 67 cases, not further assessed, 65 were negative and 2 were positive. Thus, the policy adopted in clinical practice corresponds to that of applying a low threshold classifier, minimizing false negatives for the positive diagnosis class. Notice that, although SANN achieved high performance and is interpretable, a disadvantage of this method is that the construction of network topology must be done manually, thus this algorithm is incompatible with a fully automated data analysis process.

3.2. Analysis Using JADBio

Even though tools such as Weka can be used whether interpretability is required or not, when using a tool such as Weka there are two disadvantages; first the user must be familiar with machine learning which is not always the case in an environment such as the medical domain and second the analyst must apply various algorithms and also has to tune their hyperparameters in order to achieve optimal results. Overall, this is a time-consuming process, and in addition to this it is also uncertain, especially in the case of a large search space for hyperparameter’s values, with respect to the optimal selection of hyperparameters. This is the reason why systems automating machine learning are very important for wide scale adoption of machine learning for data analysis and decision support in the medical domain.

In this work, in addition to the analysis done manually using Weka, the automated analysis tool called JADBio [77] was used as well as in [71]. By using JADBio, users simply upload their data and provide their preferences, subsequently the system selects the optimal model. In an application domain such as medical diagnosis where expertise on machine learning may not be available and a series of trials with many algorithms and their hyperparameters may not be an option due to limitations over resources such as time, the use of tools that automate machine learning tasks is expected to be widespread. JADBio allows for setting user preferences related to feature selection (optional or required), interpretability (optional or required) and time preference (preliminary, typical and extensive). Results using the above preferences are summarized in Table 3.

Table 3. Area under the curve (AUC) results using JADBio.

Analysis Type	Interpretability Required		Interpretability Not Required	
	Feature Selection	No Feature Selection	Feature Selection	No Feature Selection
Preliminary	0.756	0.794	0.750	0.833
Typical	0.778	0.807	0.798	0.830
Extensive	0.794	0.806	0.833	0.823

When using the JADBio system, in the case that interpretability is not required, a support vector machine (SVM) is the optimal model selected when combined with feature selection (and extensive time preference) and classification random forests training 100 trees

is the optimal algorithm when feature selection is not applied. In case the algorithm must be interpretable then ridge logistic regression is the best performing algorithm when combined with feature selection (and extensive time preference) and without feature selection (and typical time preference). Feature selection, pre-processing and hyperparameter selection is performed automatically by the JADBio system and results are presented below.

Specifically, after examining various possible settings the JADBio system applied in pre-processing is constant removal and standardization. Then in feature selection the algorithm applied is the statistically equivalent signature (SES) algorithm with hyper-parameters: $\max K = 2$ (i.e., the maximum conditioning set to use in the conditional independent sets), and $\alpha = 0.1$ (i.e., threshold for assessing p -value significance). JADBio selected three out of the total number of features in the original dataset: CLASS SOCIAL, AAA RQ and CLASS COMMUNICATION. Performance when using all features instead of only these three remained almost identical. The feature selection was applied by estimating the performance decrease when the feature was removed.

The best predictive model was support vector machines (SVM) of type C-SVC with polynomial kernel and hyper-parameters: $\text{cost} = 0.001$ (cost parameter trades off correct classification of training examples against maximization of the decision function's margin), $\gamma = 10.0$ (gamma parameter defines the degree of the influence of a single training example), $\text{degree} = 3$ (degree of the polynomial that SVM returns) having an area under the curve (AUC) of 0.833. Notice that the corresponding algorithm using Weka (SMO) has lower performance because of the different hyperparameter selection. The ROC curve of the best performing model using JADBio is presented in Figure 2. Using the diagram, the user can specify the true positive rate for a specific class (in the case its class 2 indicating a positive diagnostic outcome) given the threshold selected.

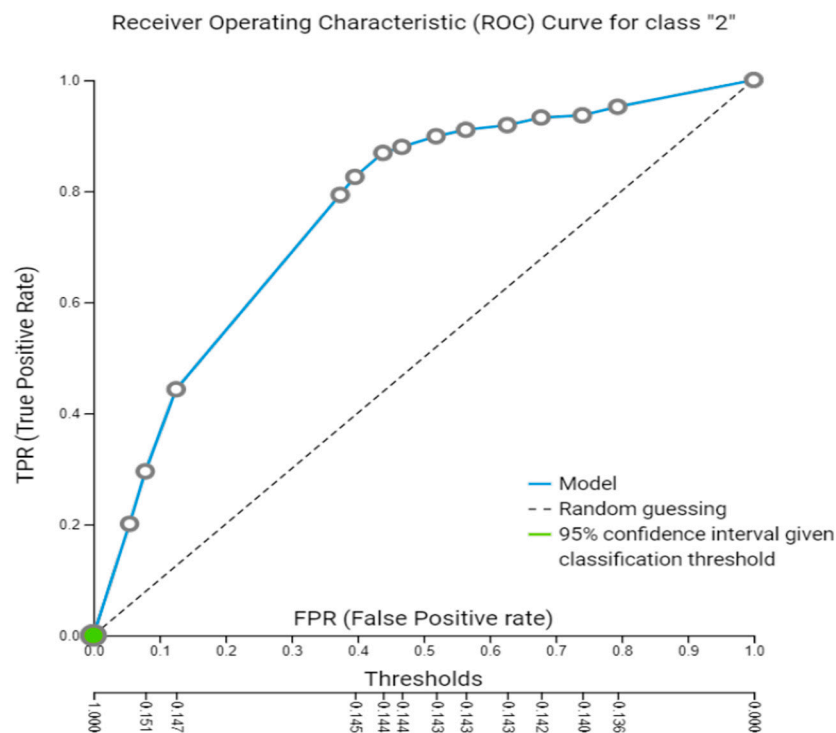


Figure 2. ROC curve of best performing model using JADBio.

The best interpretable model with feature selection was ridge logistic regression with penalty hyper-parameter $\lambda = 100.0$ (λ defines the amount of regularization used in the model produced by the algorithm), with AUC (ROC) 0.794. The ROC curve for ridge logistic regression is presented in Figure 3. Based on the curve, we can see that when setting the threshold to 9.4%, the true positive rate for the positive diagnostic outcome class

is 0.969 and false negatives rate is 0.005. Taking into account the trade-off between false positive error rate and false negative error rate and the corresponding costs the optimal threshold can be defined for cost minimization.

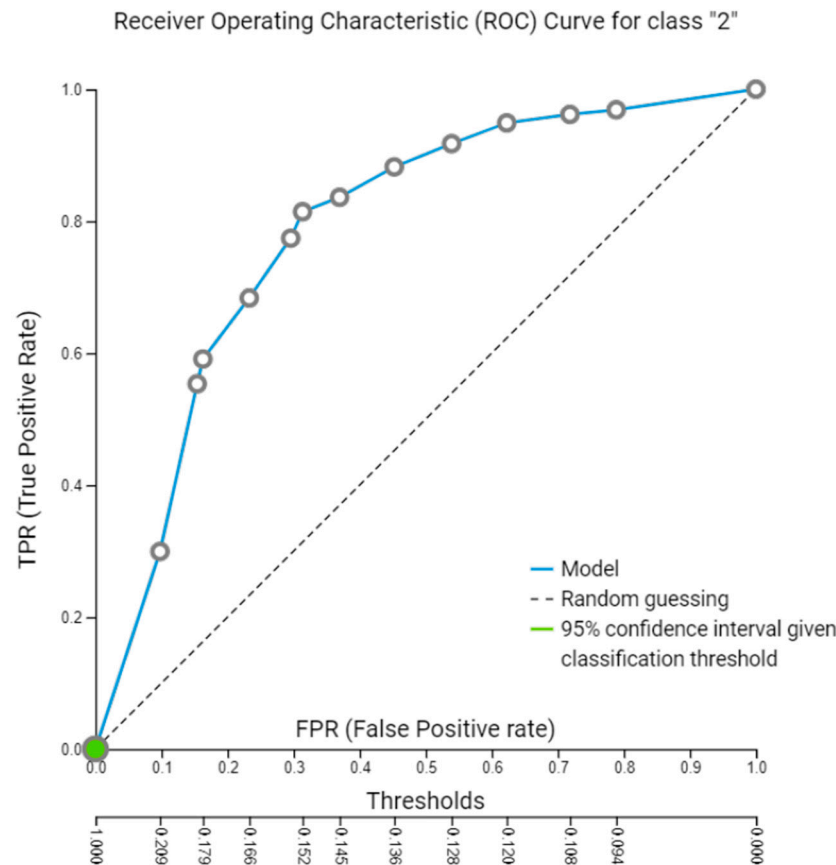


Figure 3. ROC curve of best performing interpretable model with feature selection using JAD-Bio.

Notice that JADBio adopts the bootstrap corrected cross validation performance estimation protocol presented in [78]. The objective of bootstrap corrected cross validation is to overcome the optimistic bias of cross validation, that is the typical method for performance estimation and model selection in machine learning (notice that 10-fold cross validation was used as performance metric in the experiments using Weka). The performance estimation is a task both difficult and critical, especially in medical applications where the reliability of the prediction model is a crucial parameter in decision making. This means that the performance metric of JADBio is less optimistic than that of Weka, but this stricter performance evaluation is also desirable in critical applications.

Overall, the JADBio system produced models (including interpretable models) that offered high performance in addition to fully automating the analysis process which is a great advantage over traditional systems such as Weka. Although the dataset was not balanced and the two classes were difficult to separate (this is illustrated by the poor performance of SMO algorithm using Weka), by carefully selecting the threshold value of the classification model, after taking into account corresponding costs, the performed analysis can assist the decision-making process. Notice also that depending on the cost estimation, a cost benefit analysis, when combined with an examination of the classification models, may lead to a decision to revise the assessment or even discontinue it in case there is no benefit of applying this assessment before the full assessment. This, for example, can be the case when the cost of making a false negative prediction regarding the diagnostic outcome is far greater than that of false positives. Overall, when using Weka the best performing algorithms were typical dense neural networks and random forest and the best

performing interpretable algorithm was SANN. In the case of JADBio, the best performance was achieved using SVM and the best performing interpretable algorithm was ridge logistic regression.

4. Autism Screening Questionnaire

The analysis using machine learning demonstrated the potential but also the limitations of machine learning applications using current datasets, since either the user adopts a low selection threshold which eliminates the false negative results but also allows many false positives or increases the threshold risking having false negative cases. In order to overcome these limitations a novel autism screening questionnaire is proposed (consisting of two parts, one for clinicians and another for patients) and although related datasets are still in development, an analysis based on machine learning is thus not yet feasible, preliminary statistical results are also presented. The new questionnaire can be used in order to add future autism datasets with more data points per case and subsequently improve the performance of machine learning methods over these datasets. The first tool (first part of the questionnaire) consists of 15 questions, where only binary answers are allowed. The questionnaire was undertaken by 30 patients; 8 of them have been diagnosed with autism.

The data contain binary answers, demographic information (age, gender) and the final score calculated based on the answers. Figure 4 shows the distribution of ages according to the diagnosis.

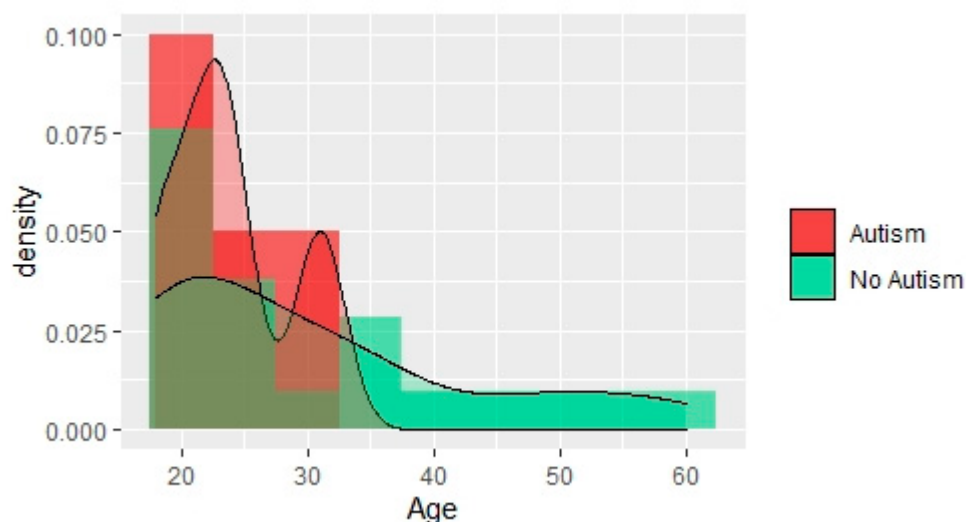


Figure 4. Distribution of ages according to the diagnose.

It can be noticed that ages are strongly skewed towards the younger age. Ages of patients affected by autism lay in the range of 18–31 with the mean being 23.88 years. Ages of patients without autism are strongly skewed to the left and are between 18 and 60 years. The majority of ages are distributed between 20 and 34 years with the mean value of 30.33. The results might be affected by sample size.

Figure 5 demonstrates gender balance for patients affected by autism. It can be concluded that while the collected data shows an equal gender balance for healthy patients, there is a gender imbalanced for patients affected by autism. This result might be affected by sample size.

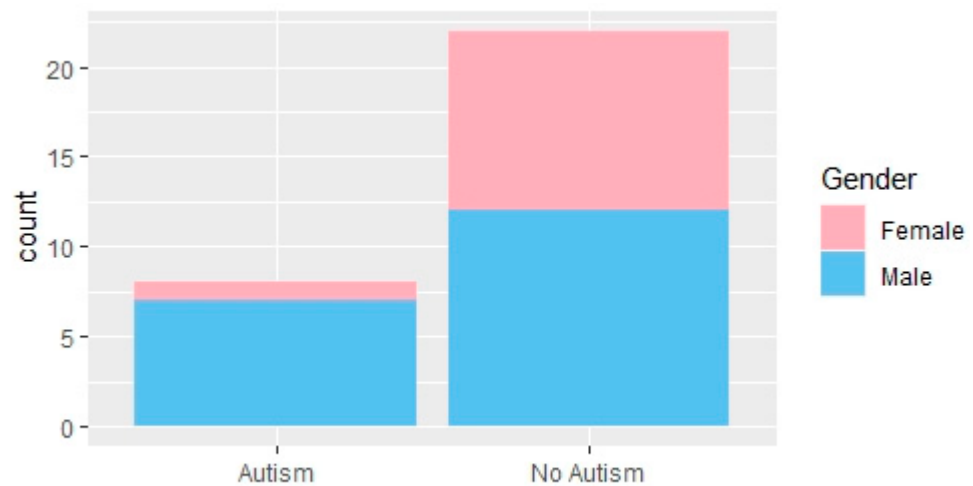


Figure 5. Gender-based frequencies for two diagnoses.

To test the questions for importance the Barnard’s exact test was applied. The test gave the indication that questions Q6, Q7, Q10 and Q14 might be important (unadjusted p -values are 0.23, 0.13, 0.23, 0.23). However, it can be noticed that p -values are higher than 0.05 and additional data are required to prove or disprove the claim. Figures 6 and 7 demonstrate how answers are distributed amongst two categories of patients: healthy and those affected by autism.

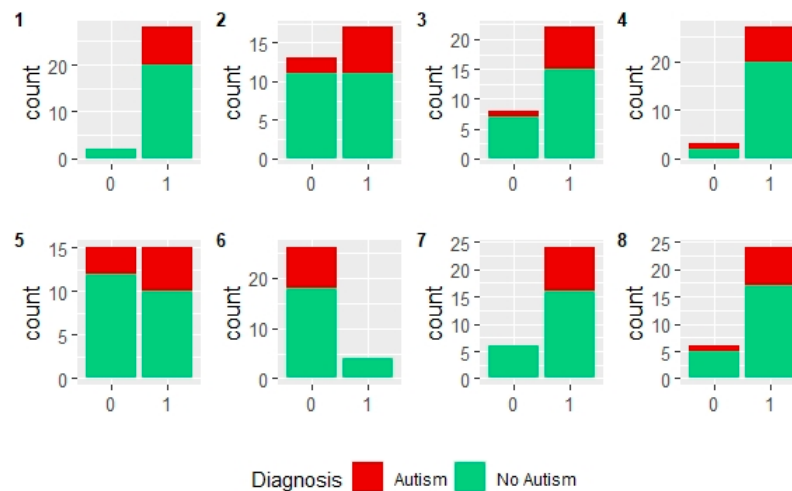


Figure 6. Bar plots demonstrating frequencies of “yes”-“no” answers to questions 1–8; 0 corresponds to “no”, 1 corresponds to “yes”.

Figure 8 shows that in the Dim1–Dim2 factor space the diagnosis “Autism” is strongly associated with Q5-yes and Q2-yes. The diagnosis “No Autism” does not have any strong associations: Q15-no and Q5-no are the closest points. Q7-no, Q3-no and Q8-no are close to each other. Questions Q4-yes, Q10-yes, Q14-yes, Q1-yes and Q11-yes are strongly associated with male gender while Q6-no and Q15-no are closer to female. It should be noted that the contribution of these questions into variability of data is rather low.

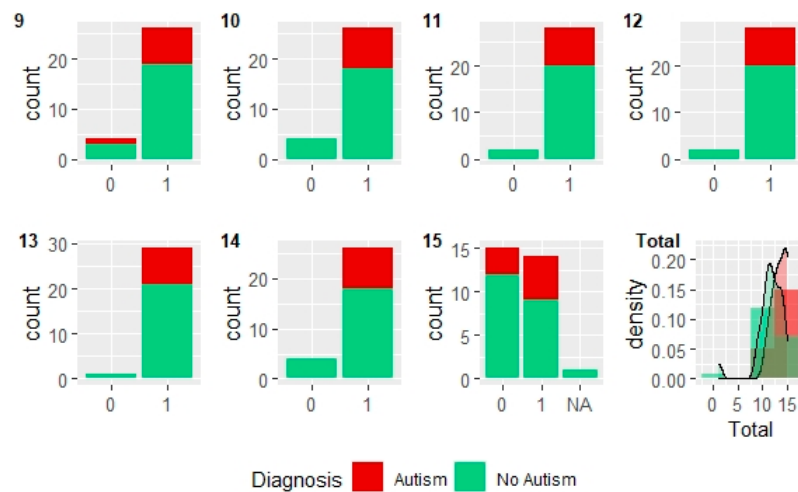


Figure 7. Bar plots demonstrating frequencies of “yes”-“no” answers to questions 9–15; 0 corresponds to “no”, 1 corresponds to “yes”. The bottom-right plot shows a grouped histogram for total score.

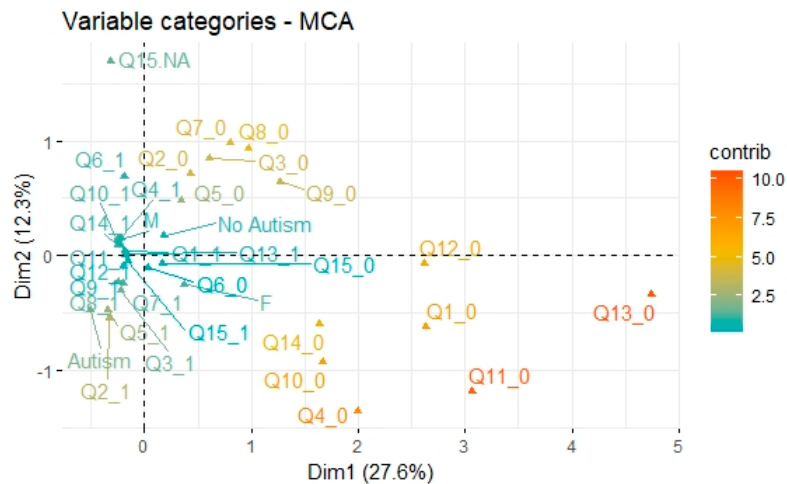


Figure 8. MCA shows the most contributing variables in factor plane Dim1–Dim2.

It can be noticed that Q13-no and Q11-no have a rather strong contribution to variability along Dim1 and Dim2. However, they are far from diagnosis autism/no autism and do not have an association with them.

In factor plane Dim2–Dim3 (see Figure 9) we can see the association of autism and Q14-no as well as no autism and Q11-yes, Q14-yes and Q11-yes. On the other hand, Q6-yes has little association with diagnosis autism/no autism.

In factor plane Dim3–Dim4 (see Figure 10), Q6-yes is still far from diagnosis. Q8-no and Q3-yes are close to no autism while Q3-no and Q9-no seem to have some association with autism.

It can be concluded that being diagnosed with or without autism has little contribution to variability in all five dimensions. Therefore, despite the associations of certain answers with certain diagnoses, at this stage we cannot select them as strong contributing factors.

The second proposed tool consists of 20 questions, where only binary answers are allowed. The questionnaire was undertaken by 18 patients; 4 of them have been diagnosed with autism.

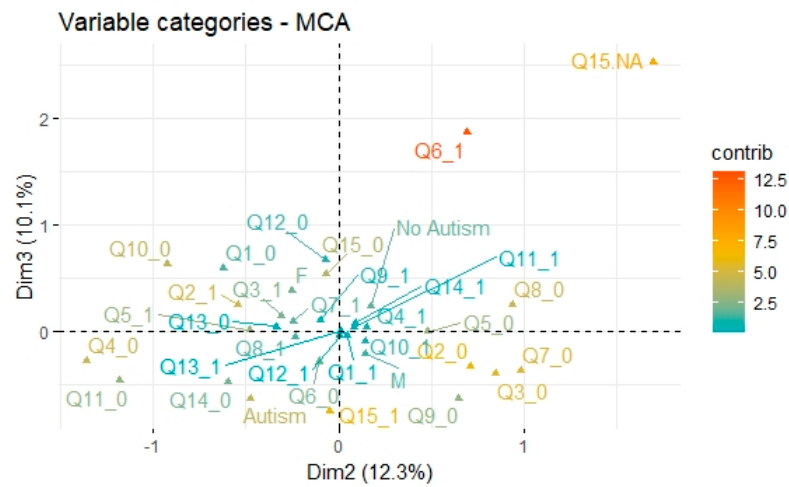


Figure 9. MCA shows the most contributing variables in factor plane Dim2–Dim3.

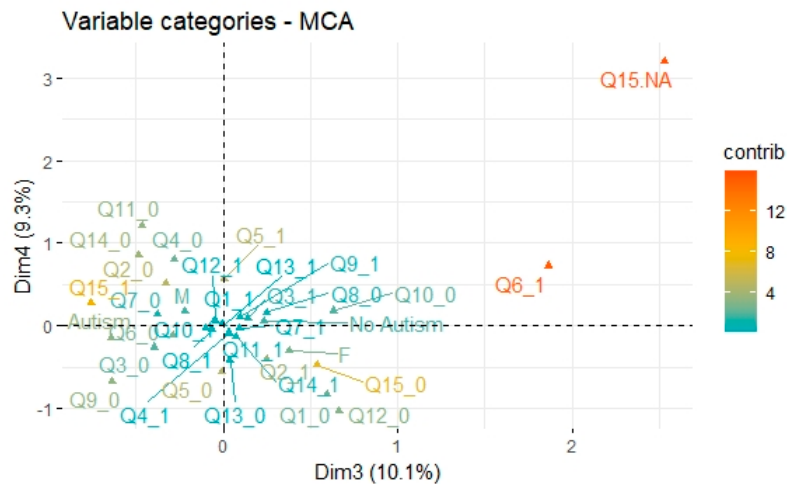


Figure 10. MCA shows the most contributing variables in factor plane Dim3–Dim4.

The data contain binary answers, demographic information (age, gender) and the final score calculated based on the answers. Figure 11 shows distribution of ages according to the diagnosis. It can be noticed that the ages of patients affected by autism are generally older (this can be affected by small sample size). Both samples contain outliers which represent older patients. The mean value of ages of patients affected by autism is 32.25 while the mean value of ages of healthy patients is 27.21. The majority of ages for all patients lay between 20 and 31.

Figure 12 demonstrates gender balance for patients affected by autism. It can be concluded that while the collected data maintains an equal gender balance for healthy patients, there is a gender imbalanced for patients affected by autism. This result might be affected by sample size.

To test the questions for importance the Barnard’s exact test was applied. The test gave the indication that questions Q19 and Q15 might be important (unadjusted *p*-values are 0.0068, 0.1553). Q19 looks particularly promising. However, it can be noticed that additional data are required to prove or disprove the claim.

Figures 13–15 demonstrate how answers are distributed amongst two categories of patients: healthy and those affected by autism.

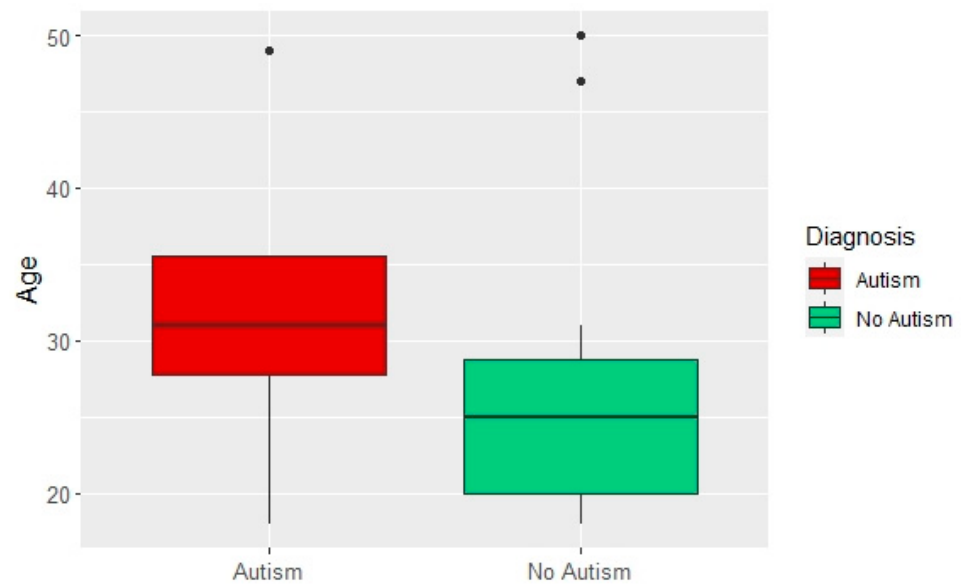


Figure 11. Distribution of ages according to the diagnose.

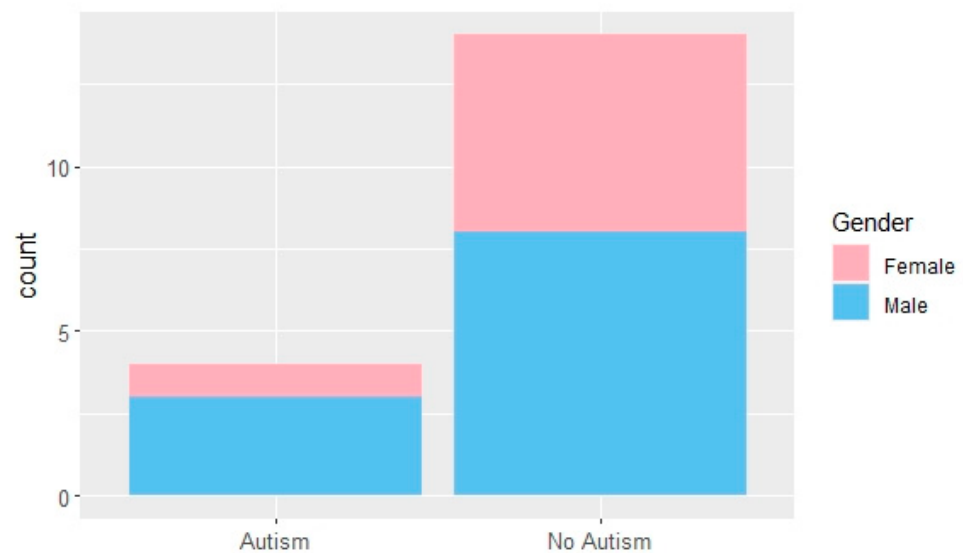


Figure 12. Gender-based frequencies for two diagnoses.

It is interesting to see that all patients answered in the same way to Q18. There seems to be no large difference in ratios autism/no autism in answers to Q6 and Q17 (0.33/0.27). However, it may be worth analyzing answers on the larger sample before making a final decision to discard Q18, Q6 and Q17.

Figures 16 and 17 show that autism mostly contributes to data variability in factor planes Dim3–Dim4 and Dim4–Dim5. It confirms the association of diagnosis “Autism” with Q15-no and Q19-yes as well as Q15-yes and Q19-no and “No Autism”. The plots also show associations of “No Autism” with Q3-no and Q16-no.

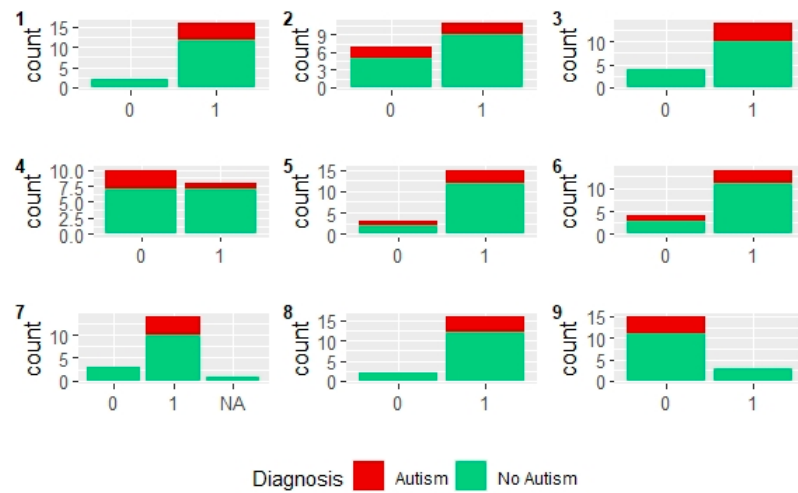


Figure 13. Bar plots demonstrating frequencies of “yes”-“no” answers to questions 1–9; 0 corresponds to “no”, 1 corresponds to “yes”.

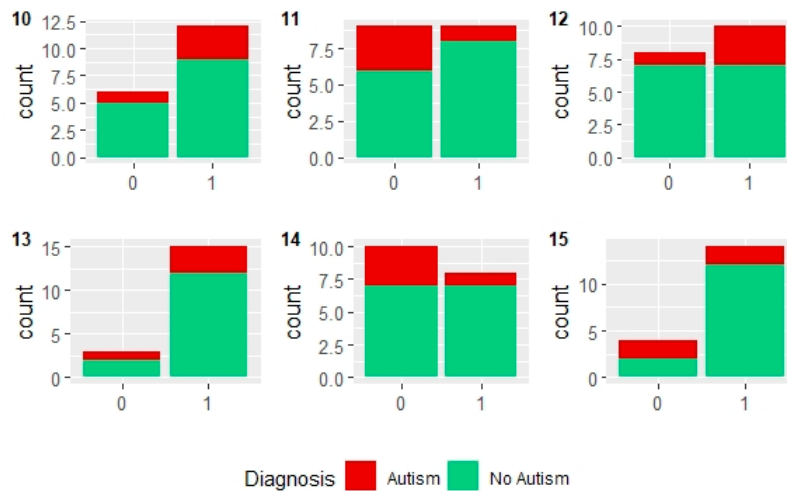


Figure 14. Bar plots demonstrating frequencies of “yes”-“no” answers to questions 10–15; 0 corresponds to “no”, 1 corresponds to “yes”.

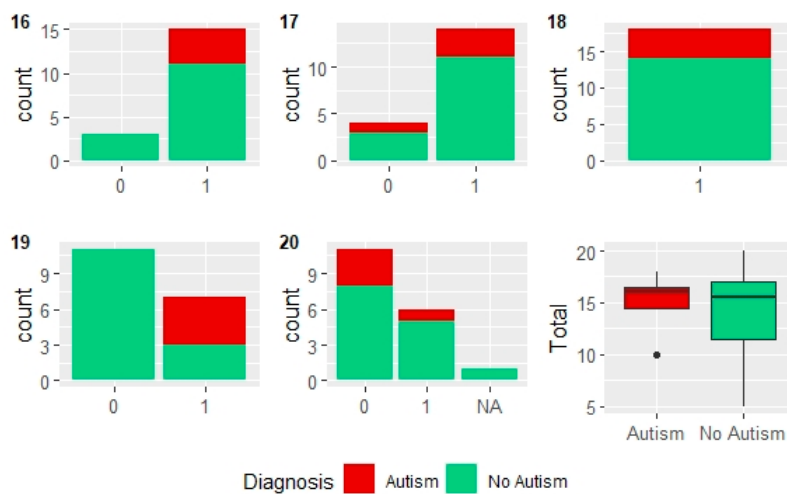


Figure 15. Bar plots demonstrating frequencies of “yes”-“no” answers to questions 16–20; 0 corresponds to “no”, 1 corresponds to “yes”. The bottom-right plot shows the grouped boxplot for total score.

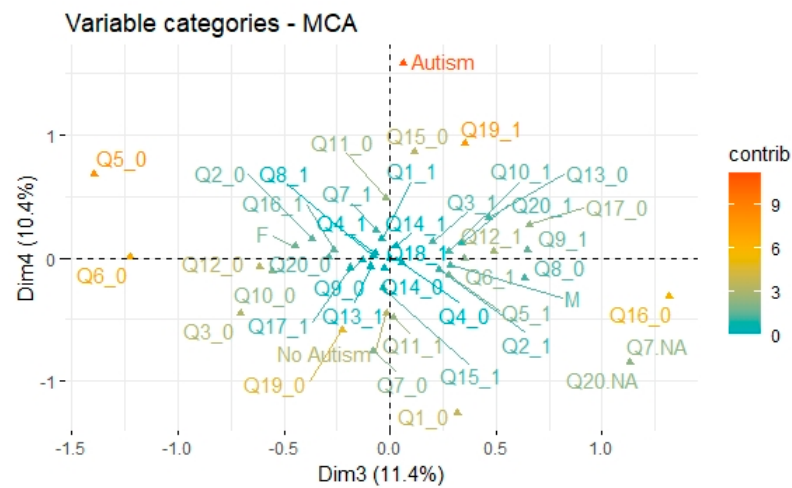


Figure 16. MCA shows the most contributing variables in factor plane Dim3–Dim4.

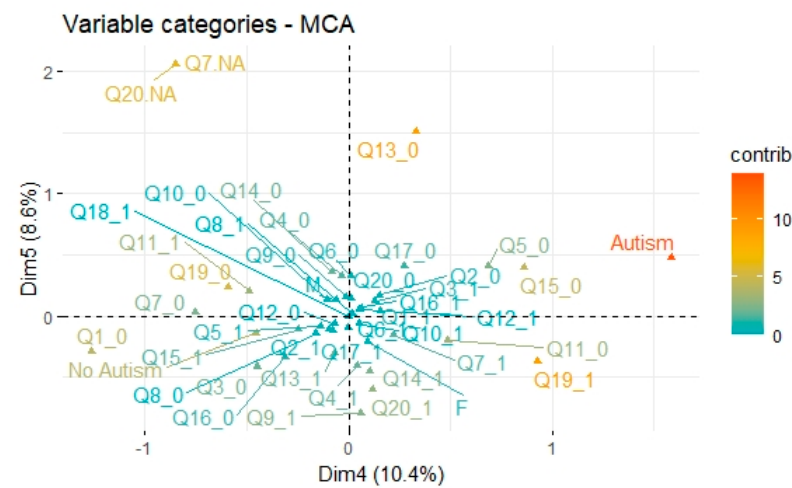


Figure 17. MCA shows the most contributing variables in factor plane Dim4–Dim5.

It is also interesting to note that Q18-yes has zero contribution to variability, which confirms with calculated *p*-values and grouped bar plot shown above.

Overall, our analysis shows that the novel tool contains information that is statistically relevant for identifying people with autism. This is a promising result, but of course more data is needed for gaining further insights and confidence in the tool’s accuracy.

Another interesting question for future exploration is to establish if the new tools will improve prediction accuracy of machine learning methods when applied to dataset containing both the information included in Section 3 and in Section 4.

5. Conclusions and Future Work

This paper presented a data driven analysis over a dataset for autism assessment. Preliminary results showed that various algorithms achieved high performance although the diagnostic outcome classification was not an easy task because of the dataset characteristics (unbalanced, having some features that were not useful and not easily separable i.e., in a linear way). Furthermore, when applying such an analysis in practice, there are other crucial factors besides the total performance, such as the requirement of interpretability and automation of the analysis process, in addition to optimal performance for specific classes and the relative cost of various types of errors when specifying the decision process. In addition, in order to overcome the limitations demonstrated in the performed analysis this work aims to evaluate the validity of an easy to administer new scale, to be used as a

self-report screening tool for adult patients referred for an ASD. This study prioritizes the importance of investigation within a clinical environment similar to where it is intended for use. Data was gathered from patients referred to the Adult ADHD and Autism Service, South West Yorkshire Partnership NHS Foundation Trust.

Future work will proceed in various directions. A particular direction will be to consider richer clinical data; there are even ideas to capture either neurological data, facial expressions through video or in combination. Another interesting idea is to expand the AI technologies used by capturing and representing explicitly, through declarative rules, medical knowledge about how clinical data should be interpreted. Such a knowledge model could be used in conjunction with a machine learning model as discussed in this paper, thus deploying a hybrid AI approach.

Author Contributions: Data analysis and writing, S.B.; data gathering and analysis, M.A.; data analysis, I.T.; background and related work analysis, S.J.; novel tools analysis, S.T.; overview and editing G.A.; writing—review and editing, T.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: Authors would like to thank Adult ADHD and Autism Service, South West Yorkshire Partnership NHS Foundation Trust for its support during this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorder*, 5th ed.; American Psychiatric Press: Washington, WA, USA, 2013.
2. Lord, C.; Cook, E.H.; Leventhal, B.L.; Amaral, D.G. Autism spectrum disorders. *Neuron* **2000**, *28*, 355–363. [[CrossRef](#)]
3. Brugha, T.S.; McManus, S.; Bankart, J.; Scott, F.; Purdon, S.; Smith, J.; Bebbington, P.; Jenkins, R.; Meltzer, H. Epidemiology of autism spectrum disorders in adults in the community in England. *Arch. Gen. Psychiatry* **2011**, *68*, 459–465. [[CrossRef](#)]
4. Murphy, C.M.; Wilson, C.E.; Robertson, D.M.; Ecker, C.; Daly, E.M.; Hammond, N.; Galanopoulos, A.; Dud, I.; Murphy, D.G.; McAlonan, G.M. Autism spectrum disorder in adults: Diagnosis, management, and health services development. *Neuropsychiatr. Dis. Treat.* **2016**, *12*, 1669–1686. [[CrossRef](#)]
5. Harris, J.; Brugha, T.; McManus, S.; Meltzer, H.; Smith, J.; Scott, F.J.; Purdon, S.; Bankart, J. *Autism Spectrum Disorders in Adults Living in Households Throughout England: Report from the Adult Psychiatric Morbidity Survey 2007*; The NHS Information Centre for Health and Social Care: Teddington, UK, 2009.
6. Fombonne, E. Epidemiology of autistic disorder and other pervasive developmental disorders. *J. Clin. Psychiatry* **2005**, *66* (Suppl. S10), 3–8.
7. Kan, C.; Buitelaar, J.K.; Van Der Gaag, R.J. Autism spectrum disorders in adults. *Ned. Tijdschr. Geneesk.* **2008**, *152*, 1365–1369.
8. Wing, L.; Potter, D. The epidemiology of autistic spectrum disorders: Is the prevalence rising? *Ment. Retard. Dev. Disabil. Res. Rev.* **2002**, *8*, 151–161. [[CrossRef](#)]
9. Ritvo, R.A.; Ritvo, E.R.; Guthrie, D.; Ritvo, M.J.; Hufnagel, D.H.; McMahon, W.; Tonge, B.; Mataix-Cols, D.; Jassi, A.; Attwood, T.; et al. The Ritvo Autism Asperger Diagnostic Scale-Revised (RAADS-R): A scale to assist the diagnosis of autism spectrum disorder in adults: An international validation study. *J. Autism Dev. Disord.* **2011**, *41*, 1076–1089. [[CrossRef](#)]
10. Barnard, J.; Harvey, V.; Potter, D.; Prior, A. *Ignored or Ineligible? The Reality for Adults with Autism Spectrum Disorders*; National Autistic Society: London, UK, 2001.
11. Le Couteur, A.; Haden, G.; Hammal, D.; McConachie, H. Diagnosing autism spectrum disorders in pre-school children using two standardised assessment instruments: The ADI-R and the ADOS. *J. Autism Dev. Disord.* **2008**, *38*, 362–372. [[CrossRef](#)]
12. Molloy, C.A.; Murray, D.S.; Akers, R.; Mitchell, T.; Manning-Courtney, P. Use of the Autism Diagnostic Observation Schedule (ADOS) in a clinical setting. *Autism* **2011**, *15*, 143–162. [[CrossRef](#)]
13. National Institute for Health and Care Excellence. *Autism Spectrum Disorder in Adults: Diagnosis and Management (Guideline CG142)*; National Institute for Health and Care Excellence: London, UK, 2016.
14. Barlati, S.; Deste, G.; Gregorelli, M.; Vita, A. Autistic traits in a sample of adult patients with schizophrenia: Prevalence and correlates. *Psychol. Med.* **2019**, *49*, 140–148. [[CrossRef](#)]

15. De Crescenzo, F.; Postorino, V.; Siracusano, M.; Riccioni, A.; Armando, M.; Curatolo, P.; Mazzone, L. Autistic symptoms in schizophrenia spectrum disorders: A systematic review and meta-analysis. *Front. Psychiatry* **2019**, *10*, 78. [[CrossRef](#)] [[PubMed](#)]
16. Bastiaansen, J.A.; Meffert, H.; Hein, S.; Huizinga, P.; Ketelaars, C.; Pijnenborg, M.; Bartels, A.; Minderaa, R.; Keyzers, C.; de Bildt, A. Diagnosing autism spectrum disorders in adults: The use of Autism Diagnostic Observation Schedule (ADOS) module 4. *J. Autism Dev. Disord.* **2011**, *41*, 1256–1266. [[CrossRef](#)] [[PubMed](#)]
17. Berthoz, S.; Hill, E.L. The validity of using self-reports to assess emotion regulation abilities in adults with autism spectrum disorder. *Eur. Psychiatry* **2005**, *20*, 291–298. [[CrossRef](#)] [[PubMed](#)]
18. Frith, U. *Autism: Explaining the Enigma*, 2nd ed.; Blackwell Publishing: Oxford, UK, 2003.
19. Leyfer, O.T.; Folstein, S.F.; Bacalman, S.; Davis, N.O.; Dinh, E.; Morgan, J.; Tager-Flusberg, H.; Lainhart, J.E. Comorbid psychiatric disorders in children with autism: Interview development and rates of disorders. *J. Autism Dev. Disord.* **2006**, *36*, 849–861. [[CrossRef](#)] [[PubMed](#)]
20. Gillberg, C. *A Guide to Asperger Syndrome*; Cambridge University Press: Cambridge, UK, 2002.
21. Fusar-Poli, L.; Brondino, N.; Politi, P.; Aguglia, E. Missed diagnoses and misdiagnoses of adults with autism spectrum disorder. *Eur. Arch. Psychiatry Clin. Neurosci.* **2020**, *272*, 187–198. [[CrossRef](#)]
22. Gould, J.; Ashton-Smith, J. Missed diagnosis or misdiagnosis? Girls and women on the autism spectrum. *Good Autism Pract.* **2011**, *12*, 34–41.
23. Hull, L.; Petrides, K.V.; Allison, C.; Smith, P.; Baron-Cohen, S.; Lai, M.-C.; Mandy, W. “Putting on my best normal”: Social camouflaging in adults with autism spectrum conditions. *J. Autism Dev. Disord.* **2017**, *47*, 2519–2534. [[CrossRef](#)]
24. Glascoe, F.P. Screening for developmental and behavioral problems. *Ment. Retard. Dev. Disabil. Res. Rev.* **2005**, *11*, 173–179. [[CrossRef](#)]
25. Arun, P.; Chavan, B.S. Development of a screening instrument for autism spectrum disorder: Chandigarh autism screening instrument. *Indian J. Med. Res.* **2018**, *147*, 369–375. [[CrossRef](#)]
26. Chakraborty, S.; Bhatia, T.; Sharma, V.; Antony, N.; Das, D.; Sahu, S.; Sharma, S.; Sharma, V.; Brar, J.S.; Iyengar, S.; et al. Protocol for development of the Indian autism screening questionnaire: The screening version of the Indian scale for assessment of autism. *Indian J. Psychol. Med.* **2020**, *42* (Suppl. S6), S63–S67. [[CrossRef](#)]
27. Thabtah, F. An accessible and efficient autism screening method for behavioural data and predictive analyses. *Health Inform. J.* **2018**, *25*, 1739–1755. [[CrossRef](#)] [[PubMed](#)]
28. Thabtah, F.; Peebles, D. Early autism screening: A comprehensive review. *Int. J. Environ. Res. Public Health* **2019**, *16*, 3502. [[CrossRef](#)] [[PubMed](#)]
29. Eaves, L.C.; Wingert, H.D.; Ho, H.H.; Mickelson, E.C.R. Screening for autism spectrum disorders with the social communication questionnaire. *J. Dev. Behav. Pediatrics* **2006**, *27*, S95–S103. [[CrossRef](#)] [[PubMed](#)]
30. Woodbury-Smith, M.R.; Robinson, J.; Wheelwright, S.; Baron-Cohen, S. Screening adults for asperger syndrome using the AQ: A preliminary study of its diagnostic validity in clinical practice. *J. Autism Dev. Disord.* **2005**, *35*, 331–335. [[CrossRef](#)]
31. Sappok, T.; Heinrich, M.; Underwood, L. Screening tools for autism spectrum disorders. *Adv. Autism* **2015**, *1*, 12–29. [[CrossRef](#)]
32. Baron-Cohen, S.; Hoekstra, R.A.; Knickmeyer, R.; Wheelwright, S. The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* **2001**, *31*, 5–17. [[CrossRef](#)]
33. Eriksson, J.M.; Andersen, L.M.J.; Bejerot, S. RAADS-14 screen: Validity of a screening tool for autism spectrum disorder in an adult psychiatric population. *Mol. Autism* **2013**, *4*, 49. [[CrossRef](#)]
34. Baron-Cohen, S.; Wheelwright, S.; Robinson, J.; Woodbury-Smith, M. The Adult Asperger Assessment (AAA): A diagnostic method. *J. Autism Dev. Disord.* **2005**, *35*, 807–819. [[CrossRef](#)]
35. Sizoo, B.B.; Horwitz, E.H.; Teunisse, J.P.; Kan, C.C.; Visser, C.T.W.M.; Forceville, E.J.M.; Van Voorst, A.J.P.; Van Voorst, H.M. Predictive validity of self-report questionnaires in the assessment of autism spectrum disorders in adults. *Autism* **2015**, *19*, 842–849. [[CrossRef](#)]
36. Wakabayashi, A.; Baron-Cohen, S.; Wheelwright, S.; Tojo, Y. The Autism-Spectrum Quotient (AQ) in Japan: A cross-cultural comparison. *J. Autism Dev. Disord.* **2006**, *36*, 263–270. [[CrossRef](#)]
37. Kenny, H.; Alison, J.S. How useful are the Adult Asperger Assessment and AQ-10 within an adult clinical population of all intellectual abilities? *Adv. Autism* **2016**, *2*, 118–130. [[CrossRef](#)]
38. Fusar-Poli, L.; Ciancio, A.; Gabbiadini, A.; Meo, V.; Patania, F.; Rodolico, A.; Saitta, G.; Voza, L.; Petralia, A.; Signorelli, M.S.; et al. Self-reported autistic traits using the AQ: A comparison between individuals with ASD, psychosis, and non-clinical controls. *Brain Sci.* **2020**, *10*, 291. [[CrossRef](#)] [[PubMed](#)]
39. Adamou, M.; Jones, S.L.; Wetherhill, S. *AAA Screening in Adults with ASD: A Retrospective Cohort Study*; Emerald Publishing Limited: Bingley, UK, 2021.
40. Ashwood, K.L.; Gillan, N.; Horder, J.; Hayward, H.; Woodhouse, E.; McEwen, F.S.; Findon, J.; Eklund, H.; Spain, D.; Wilson, C.E.; et al. Predicting the diagnosis of autism in adults using the Autism-Spectrum Quotient (AQ) questionnaire. *Psychol. Med.* **2016**, *46*, 2595–2604. [[CrossRef](#)] [[PubMed](#)]
41. Bishop, S.L.; Seltzer, M.M. Self-reported autism symptoms in adults with autism spectrum disorders. *J. Autism Dev. Disord.* **2012**, *42*, 2354–2363. [[CrossRef](#)] [[PubMed](#)]

42. Booth, T.; Murray, A.L.; McKenzie, K.; Kuenssberg, R.; O'Donnell, M.; Burnett, H. Brief report: An evaluation of the AQ-10 as a brief screening instrument for ASD in adults. *J. Autism Dev. Disord.* **2013**, *43*, 2997–3000. [[CrossRef](#)]
43. Allison, C.; Auyeung, B.; Baron-Cohen, S. Toward brief “red flags” for autism screening: The short autism spectrum quotient and the short quantitative checklist in 1000 cases and 3000 controls. *J. Am. Acad. Child Adolesc. Psychiatry* **2012**, *51*, 202–212. [[CrossRef](#)]
44. Kurita, H.; Koyama, T.; Osada, H. Autism-Spectrum Quotient-Japanese version and its short forms for screening normally intelligent persons with pervasive developmental disorders. *Psychiatry Clin. Neurosci.* **2005**, *59*, 490–496. [[CrossRef](#)]
45. Lepage, J.-F.; Lortie, M.; Taschereau-Dumouchel, V.; Théoret, H. Validation of French-Canadian versions of the Empathy Quotient and Autism Spectrum Quotient. *Can. J. Behav. Sci.* **2009**, *41*, 272–276. [[CrossRef](#)]
46. Ketelaars, C.; Horwitz, E.; Sytema, S.; Bos, J.; Wiersma, D.; Minderaa, R.; Hartman, C.A. Brief report: Adults with mild autism spectrum disorders (ASD): Scores on the autism spectrum quotient (AQ) and comorbid psychopathology. *J. Autism Dev. Disord.* **2008**, *38*, 176–180. [[CrossRef](#)]
47. Hirota, T.; So, R.; Kim, Y.S.; Leventhal, B.; Epstein, R.A. A systematic review of screening tools in non-young children and adults for autism spectrum disorder. *Res. Dev. Disabil.* **2018**, *80*, 1–12. [[CrossRef](#)]
48. Williams, J.; Brayne, C. Screening for autism spectrum disorders: What is the evidence? *Autism* **2006**, *10*, 11–35. [[CrossRef](#)] [[PubMed](#)]
49. Posserud, M.-B.; Lundervold, A.J.; Gillberg, C. Validation of the autism spectrum screening questionnaire in a total population sample. *J. Autism Dev. Disord.* **2009**, *39*, 126–134. [[CrossRef](#)] [[PubMed](#)]
50. Ritvo, R.A.; Ritvo, E.R.; Guthrie, D.; Yuwiler, A.; Ritvo, M.J.; Weisbender, L. A scale to assist the diagnosis of autism and Asperger’s disorder in adults (RAADS): A pilot study. *J. Autism Dev. Disord.* **2008**, *38*, 213–223. [[CrossRef](#)] [[PubMed](#)]
51. Jones, S.L.; Johnson, M.; Alty, B.; Adamou, M. The effectiveness of RAADS-R as a screening tool for adult ASD populations. *Autism Res. Treat.* **2021**, *2021*, 9974791. [[CrossRef](#)] [[PubMed](#)]
52. Kember, S.M.; Williams, M.N. Autism in Aotearoa: Is the RAADS-14 a valid tool for a New Zealand population? *Eur. J. Psychol. Assess.* **2021**, *37*, 247. [[CrossRef](#)]
53. Conner, C.M.; Cramer, R.D.; McGonigle, J.J. Examining the diagnostic validity of autism measures among adults in an outpatient clinic sample. *Autism Adulthood* **2019**, *1*, 60–68. [[CrossRef](#)]
54. Picot, M.-C.; Michelon, C.; Bertet, H.; Pernon, E.; Fiard, D.; Coutelle, R.; Abbar, M.; Attal, J.; Amestoy, A.; Duverger, P.; et al. The French version of the revised ritvo autism and Asperger diagnostic scale: A psychometric validation and diagnostic accuracy study. *J. Autism Dev. Disord.* **2021**, *51*, 30–44. [[CrossRef](#)]
55. Brugha, T.; Tyrer, F.; Leaver, A.; Lewis, S.; Seaton, S.; Morgan, Z.; Tromans, S.; van Rensburg, K. Testing adults by questionnaire for social and communication disorders, including autism spectrum disorders, in an adult mental health service population. *Int. J. Methods Psychiatr. Res.* **2020**, *29*, e1814. [[CrossRef](#)]
56. Horwitz, E.H.; Schoevers, R.A.; Ketelaars, C.E.J.; Kan, C.C.; van Lammeren, A.M.D.; Meesters, Y.; Spek, A.A.; Wouters, S.; Teunisse, J.P.; Cuppen, L.; et al. Clinical assessment of ASD in adults using self- and other-report: Psychometric properties and validity of the Adult Social Behavior Questionnaire (ASBQ). *Res. Autism Spectr. Disord.* **2016**, *24*, 17–28. [[CrossRef](#)]
57. Andersen, L.M.J.; Näswall, K.; Manouilenko, I.; Nylander, L.; Edgar, J.; Ritvo, R.A.; Ritvo, E.; Bejerot, S. The Swedish version of the ritvo autism and Asperger diagnostic scale: Revised (RAADS-R). A Validation study of a rating scale for adults. *J. Autism Dev. Disord.* **2011**, *41*, 1635–1645. [[CrossRef](#)]
58. Westwood, H.; Eisler, I.; Mandy, W.; Leppanen, J.; Treasure, J.; Tchanturia, K. Using the autism-spectrum quotient to measure autistic traits in anorexia nervosa: A systematic review and meta-analysis. *J. Autism Dev. Disord.* **2016**, *46*, 964–977. [[CrossRef](#)] [[PubMed](#)]
59. Romero, M.; Aguilar, J.M.; Del-Rey-Mejías, Á.; Mayoral, F.; Rapado, M.; Peciña, M.; Barbancho, M.Á.; Ruiz-Veguilla, M.; Lara, J.P. Psychiatric comorbidities in autism spectrum disorder: A comparative study between DSM-IV-TR and DSM-5 diagnosis. *Int. J. Clin. Health Psychol.* **2016**, *16*, 266–275. [[CrossRef](#)] [[PubMed](#)]
60. Mannion, A.; Leader, G. Comorbidity in autism spectrum disorder: A literature review. *Res. Autism Spectr. Disord.* **2013**, *7*, 1595–1616. [[CrossRef](#)]
61. Lai, M.-C.; Kasseh, C.; Besney, R.; Bonato, S.; Hull, L.; Mandy, W.; Szatmari, P.; Ameis, S.H. Prevalence of co-occurring mental health diagnoses in the autism population: A systematic review and meta-analysis. *Lancet Psychiatry* **2019**, *6*, 819–829. [[CrossRef](#)]
62. Lugnegård, T.; Hallerbäck, M.U.; Gillberg, C. Asperger syndrome and schizophrenia: Overlap of self-reported autistic traits using the Autism-spectrum Quotient (AQ). *Nord. J. Psychiatry* **2015**, *69*, 268–274. [[CrossRef](#)]
63. Tebartz Van Elst, L.; Pick, M.; Biscaldi, M.; Fangmeier, T.; Riedel, A. High-functioning autism spectrum disorder as a basic disorder in adult psychiatry and psychotherapy: Psychopathological presentation, clinical relevance and therapeutic concepts. *Eur. Arch. Psychiatry Clin. Neurosci.* **2013**, *263* (Suppl. S2), S189–S196. [[CrossRef](#)]
64. Wigham, S.; Rodgers, J.; Berney, T.; Couteur, A.L.; Ingham, B.; Parr, J.R. Psychometric properties of questionnaires and diagnostic measures for autism spectrum disorders in adults: A systematic review. *Autism* **2018**, *23*, 287–305. [[CrossRef](#)]
65. Happé, F.G.; Mansour, H.; Barrett, P.; Brown, T.; Abbott, P.; Charlton, R.A. Demographic and cognitive profile of individuals seeking a diagnosis of autism spectrum disorder in adulthood. *J. Autism Dev. Disord.* **2016**, *46*, 3469–3480. [[CrossRef](#)]
66. Sewani, H.; Kashef, R. An autoencoder-based deep learning classifier for efficient diagnosis of autism. *Children* **2020**, *7*, 182. [[CrossRef](#)]

67. Kashef, R. ECNN: Enhanced convolutional neural network for efficient diagnosis of autism spectrum disorder. *Cogn. Syst. Res.* **2022**, *71*, 41–49. [[CrossRef](#)]
68. Kanimozhiselvi, C.S.; Jayaprakash, D. Machine learning based autism grading for clinical decision making. *Int. J. Recet. Technol. Eng.* **2019**, *8*, 7443–7446.
69. Eslami, T.; Almuqhim, F.; Raiker, J.S.; Saeed, F. Machine learning methods for diagnosing autism spectrum disorder and attention-deficit/hyperactivity disorder using functional and structural MRI: A survey. *Front. Neuroinform.* **2021**, *14*, 62. [[CrossRef](#)] [[PubMed](#)]
70. Hyde, K.K.; Novack, M.N.; LaHaye, N.; Parlett-Pelleriti, C.; Anden, R.; Dixon, D.R.; Linstead, E. Applications of supervised machine learning in autism spectrum disorder research: A review. *Review J. Autism Dev. Disord.* **2019**, *6*, 128–146. [[CrossRef](#)]
71. Batsakis, S.; Adamou, M.; Tachmazidis, I.; Antoniou, G.; Kehagias, T. Data-driven decision support for autism diagnosis using machine learning. In Proceedings of the 13th International Conference on Management of Digital EcoSystems, Hammamet, Tunisia, 1–3 November 2021; pp. 30–34.
72. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: An update. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 10–18. [[CrossRef](#)]
73. Batsakis, S.; Tachmazidis, I.; Baryannis, G.; Antoniou, G. Semantic artificial neural networks. In *European Semantic Web Conference*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 39–44.
74. Došilović, F.K.; Brčić, M.; Hlupić, N. Explainable artificial intelligence: A survey. In Proceedings of the 2018 41st International Convention On Information And Communication Technology, Electronics And Microelectronics (MIPRO), Opatija, Croatia, 21–25 May 2018; pp. 210–215.
75. Shavlik, J.W.; Towell, G.G. An approach to combining explanation-based and neural learning algorithms. In *Applications of Learning and Planning Methods*; World Scientific: Singapore, 1991; pp. 71–98.
76. Towell, G.G.; Shavlik, J.W. Knowledge-based artificial neural networks. *Artif. Intell.* **1994**, *70*, 119–165. [[CrossRef](#)]
77. Tsamardinos, I.; Charonyktakis, P.; Lakiotaki, K.; Borboudakis, G.; Zenklusen, J.C.; Juhl, H.; Chatzaki, E.; Lagani, V. *Just Add Data: Automated Predictive Modeling and Biosignature Discovery*; Cold Spring Harbor Laboratory: Laurel Hollow, NY, USA, 2020.
78. Tsamardinos, I.; Greasidou, E.; Borboudakis, G. Boot-strapping the out-of-sample predictions for efficient and accurate cross-validation. *Mach. Learn.* **2018**, *107*, 1895–1922. [[CrossRef](#)]