




# Contributions of Artificial Intelligence to Analysis of Gut Microbiota in Autism Spectrum Disorder: A Systematic Review

Pau Climent-Pérez <sup>1</sup>, Agustín Ernesto Martínez-González <sup>2,\*</sup> and Pedro Andreo-Martínez <sup>3</sup>

<sup>1</sup> Department of Computing Technology, University of Alicante, 03690 San Vicente del Raspeig, Alicante, Spain; pau.climent@gcloud.ua.es

<sup>2</sup> Department of Developmental Psychology and Didactics, University of Alicante, 03690 San Vicente del Raspeig, Alicante, Spain

<sup>3</sup> Department of Agricultural Chemistry, Faculty of Chemistry, Regional Campus of International Excellence “Campus Mare Nostrum”, University of Murcia, Campus of Espinardo, 30100 Murcia, Spain; pam11@um.es

\* Correspondence: agustin.emartinez@ua.es

**Abstract:** Background: Autism spectrum disorder (ASD) is a highly heterogeneous neurodevelopmental disorder whose etiology is not known today, but everything indicates that it is multifactorial. For example, genetic and epigenetic factors seem to be involved in the etiology of ASD. In recent years, there has been an increase in studies on the implications of gut microbiota (GM) on the behavior of children with ASD given that dysbiosis in GM may trigger the onset, development and progression of ASD through the microbiota–gut–brain axis. At the same time, significant progress has occurred in the development of artificial intelligence (AI). Methods: The aim of the present study was to perform a systematic review of articles using AI to analyze GM in individuals with ASD. In line with the PRISMA model, 12 articles using AI to analyze GM in ASD were selected. Results: Outcomes reveal that the majority of relevant studies on this topic have been conducted in China (33.3%) and Italy (25%), followed by the Netherlands (16.6%), Mexico (16.6%) and South Korea (8.3%). Conclusions: The bacteria *Bifidobacterium* is the most relevant biomarker with regard to ASD. Although AI provides a very promising approach to data analysis, caution is needed to avoid the over-interpretation of preliminary findings. A first step must be taken to analyze GM in a representative general population and ASD samples in order to obtain a GM standard according to age, sex and country. Thus, more work is required to bridge the gap between AI in mental health research and clinical care in ASD.

**Keywords:** artificial intelligence; autism spectrum disorders; gut microbiota; machine learning



**Citation:** Climent-Pérez, P.; Martínez-González, A.E.; Andreo-Martínez, P. Contributions of Artificial Intelligence to Analysis of Gut Microbiota in Autism Spectrum Disorder: A Systematic Review. *Children* **2024**, *11*, 931. <https://doi.org/10.3390/children11080931>

Academic Editors: Francisco Alcantud-Marín, María Consuelo Sáiz-Manzanares and Yurena Alonso-Esteban

Received: 26 June 2024  
Revised: 26 July 2024  
Accepted: 27 July 2024  
Published: 31 July 2024



**Copyright:** © 2024 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/>).

## 1. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that is characterized by difficulties in social communication and interaction skills and the presence of restricted and repetitive patterns of behavior (RRBs) [1]. There has been an increase in the prevalence of ASD around the world [2–5]. This increase in prevalence may be influenced by different factors. One such factor is the validation of psychometric instruments, which has allowed progress in the detection and diagnosis of ASD, with the average age of global diagnosis being five years [6]. Furthermore, the recent development of cerebral organoids provides a powerful tool for studying both normal human embryonic brain development and, potentially, the origins of neurodevelopmental disorders, including ASD [7]. However, a series of epigenetic factors must also be considered that could influence the development of ASD [8]. In this sense, publications on gut microbiota (GM) and ASD have also seen a steady rise in the number [9,10].

Studies suggest that a series of symptoms exist that are associated with a possible dysbiosis in GM. Alterations to the gut microbiota–brain axis in ASD samples are often suspected following the appearance of pain and gastrointestinal symptoms (e.g., abdominal pain and constipation) [9,11,12]. Furthermore, sensory hyper-reactivity is highly related to

selective or restrictive eating patterns (e.g., being a very picky eater [13,14]) and such types of restricted behavior can contribute to an increase in gastrointestinal symptoms [15] and poor intestinal functioning [9,11,12]. In fact, one meta-analysis concluded that individuals with ASD have dysbiosis in the genus *Bifidobacterium* [9]. In addition, such bacteria are beneficial for inhibiting the growth of pathogens and modulating the immune system and they can be used as a probiotic [10,16].

The composition of GM is highly diverse and varies widely between populations as a function of dietary, cultural and biological factors [17–19]. Analysis of GM is fundamentally performed through pyrosequencing of the 16S rRNA bacterial gene (e.g., V2–V3 regions) in fecal samples [13]. 16s rRNA sequencing is a culture-free method that is used to identify and compare bacterial diversity. Application of this analytical technique must follow a highly rigorous methodology in order to avoid sample contamination.

With regard to artificial intelligence (AI), it is important to, first, introduce a number of concepts that will be employed throughout the text and must be explained to practitioners of non-engineering-related fields (i.e., practitioners outside of computer science and similar fields). Artificial intelligence is a broad term in computer science that was introduced in the 1960s and relates to any capability of a computer to replicate human intelligence, reasoning, decision-making or, even, perception and cognition [20,21]. It entails aspects such as computer vision (CV) and pattern recognition (PR), amongst other aspects. Within AI, a field exists that is dedicated to ‘learning’. For example, machine learning (ML) is capable of discerning patterns that differ from other repeated patterns and, based on similarities between said patterns, classify traits described by features or vectors of features as pertaining to a class or ‘label’. For instance, merging the fields of CV and ML gives rise to object recognition, which aims to discern different object types or ‘labels’ based on features such as shape, texture and color. From this, an internal ‘model’ of the observed reality is created, which provides an internal representation that mirrors a mathematical function for distinguishing between the different types of patterns observed during the ‘training’ process [22–24].

Furthermore, artificial neural networks (ANNs) are a type of ML model inspired by the synopsis process of neurons in the visual cortex of biological models, although they are simplified mathematical versions and have, since their inception, varied substantially from their biological counterparts [25]. By arranging neurons that perform very simple mathematical operations (a product of ‘weight’ with the addition of a ‘bias’) in layers, akin to their arrangement in the visual cortex, they are capable of learning increasingly complex mathematical functions by combining outcomes produced at the level of previous layers. As will be seen, several of the reviewed works use this type of neural network for classification tasks as part of their analysis of gut microbiota composition. When ANNs are involved, and specifically if using models with more than one hidden layer (a type of layer that represents neither the input nor the output layer of neurons), a method termed ‘deep learning’ (DL) [26] comes into play. These types of methods have bloomed since the mid-2010s and constitute most methodologies employed in the present day for a wide range of applications. They are popular because of their good outcomes regarding computer vision, mostly in part to the emergence of convolutional layers and convolutional neural networks (CNNs) [27] but also in specialized models for natural language processing (NLP), amongst other fields. However, a caveat associated with their use pertains to the amount of data required for training such models, since the number of neurons (weights, biases) that constitute a neural network amount to tens or even hundreds of millions. Adjustment of the neuron’s internal representations during the decision-making process entails ‘seeing’ many potential scenarios during training and, therefore, makes the training process a computationally expensive endeavor.

Recently, computer-based AI has facilitated analysis, detection and diagnosis in mental health work. As AI techniques continue to be refined and improved, it will be possible to help mental health practitioners redefine ASD more objectively than is currently possible using the DSM-5, whilst also enabling earlier identification of ASD. Consequently, inter-

ventions will be able to be put in place earlier and will be more personalized [28]. Thus, the present study aims to conduct an updated systematic review of AI-generated findings regarding GM in ASD.

## 2. Materials and Methods

### 2.1. Protocol and Registration

The PRISMA checklist was followed when designing this systematic review [29]. The PRISMA approach ensures a systematic approach to research conducted in the field of ASD, GM and AI by framing critical analysis of key parameters, such as diagnosis via AI or ML.

The protocol used for this systematic review was not registered in any registry for systematic reviews or meta-analyses.

### 2.2. Eligibility Criteria

Eligibility criteria defining the scope of the present systematic review were agreed upon at a meeting of all authors and were applied to all papers retrieved from the database search. The following inclusion criteria were applied: (1) studies that examine diagnoses of ASD and GM made using AI or ML; (2) articles published prior to 25 April 2024 and (3) articles reporting comprehensive results and/or information. The following exclusion criteria were applied: (1) unsystematic narrative reviews; (2) studies published in a language other than English; (3) dissertations and conference proceedings; (4) books or book chapters; (5) editorial material; (6) articles examining ASD diagnosed according to any technique other than that already described as the main interest of the present study.

### 2.3. Information Sources

In order to minimize potential bias, literature searches were conducted in four different comprehensive databases. Namely, the comprehensive databases Web of Science (n = 23), Scopus (n = 41), PubMed (n = 22) and Science Database (n = 41) were searched. Works published prior to 25 April 2024 were included in these searches.

### 2.4. Search

The following search terms with relevant Boolean operators (including wildcards) were used: (gut\* OR intestine\* OR bowel\* OR gastrointestinal\*) AND (microbiota\* OR microflora\* OR bacteria\* OR microbiome\* OR flora\* OR bacterial\* OR bacteria\* OR microorganism\* OR feces\* OR stool\*) AND ('Autism' OR 'ASD' OR 'autism spectrum disorder') AND ('Artificial Intelligence' OR 'Machine Learning'). Furthermore, Web of Science was searched according to 'theme', whilst Scopus was searched according to 'title, abstract and keywords', PubMed according to 'all fields' and finally, Science according to 'all fields except full text (NOFT)'. No language restriction was applied to any of the searches.

### 2.5. Study Selection

Following the completion of database searches, a three-stage process was followed to review all records in accordance with the previously established eligibility criteria. First, titles were reviewed for eligibility, followed by abstracts and, finally, full texts. Articles gathered from all four databases were screened using 'EndNote 20' software with the aim of identifying duplicates and classifying papers according to inclusion/exclusion criteria. The three authors formed a 'review team', which took steps to minimize bias and possible random error at all stages of the review by independently verifying article selection according to title, abstract and full text. All three authors are experts in the fields of GM, AI and ASD.

### 2.6. Data Collection and Data Items

In order to ensure accuracy and impartiality at this step, data extraction was carried out by the review team. Using a data extraction form developed specifically for the present study, qualitative and quantitative data were extracted from the 12 articles included in this systematic review. The data items included for data extraction are given in Table 1.

**Table 1.** Characteristics of included studies.

Ref	Participant Characteristic		Country	Data Source	Study Type	Main Finding	ML/DL Method Used	Model Predicting Microbiome
	Experimental Group	Control Group						
[30]	ASD (n = 249)	NT siblings (n = 106) NT control (n = 101)	South Korea	Original	PCR, 16S rRNA (V3–V4 regions)	Negative association between <i>Bifidobacterium longum</i> and Childhood Autism Rating Scale outcomes, as well as a negative association between <i>Streptococcus salivarius</i> and Social Responsiveness Scale (SRS) outcomes in ASD.	ML: XGB regression of high/low SRS/VABS values Prediction of microbial age	<i>Bacteroides vulgatus</i> , <i>Roseburia cecicola</i> group, <i>Lachnospiraceae</i> and <i>Agathobaculum butyriproducience</i> showed significantly different abundances between high and low SRS groups in the E1 model ( $p \leq 0.01$ ), but not in the E2-SRS classification model. <i>Streptococcus salivarius</i> significantly differed between high and low SRS groups in model E2 ( $p \leq 0.01$ ), but not in E1.
[31]	ASD (n = 48) (from 2 to 7 years old, average BMI = 17.4, 10 females and 38 males) ASD (n = 77)	NT (n = 48) (all at 48 months, no allergies, 24 females and 24 males, average BMI = 16.3). NT (n = 50)	Mexico	[32,33]	16S rRNA (V3–V4 regions) (V4 region)	See [32,33]	ML: SVM, RF DL: ANNs Classification ASD v. HC	<i>Lachnospira</i> (primary predictor in the RF- and ANN-based models, ranking second in the SVM) Of the five main predictors in SVM and ANN models: <i>Bacteroides</i> ( $p = 2.4 \times 10^{-3}$ ), <i>Escherichia–Shigella</i> ( $p = 2.39 \times 10^{-2}$ ), <i>Akkermansia</i> ( $p = 2.51 \times 10^{-2}$ ) and <i>Dialister</i> ( $p = 3.67 \times 10^{-2}$ ) are statistically different. SVM [32]: <i>Bacteroides</i> , <i>Lachnospira</i> , <i>Blautia</i> , <i>Lachnospiraceae</i> and <i>Subdoligranulum</i> ANN [32]: <i>Lachnospira</i> , <i>Bacteroides</i> , <i>Lachnospiraceae</i> , <i>Blautia</i> and <i>Subdoligranulum</i> RF [32]: <i>Lachnospira</i> , <i>Escherichia–Shigella</i> , <i>Bacteroides</i> , <i>Blautia</i> and <i>Roseburia</i> ANN performed better than SVM on training and validation partitions, with 97.01% for training and 82.21% for validation. SVM [33]: <i>Ruminococcus torques</i> , <i>Anaerobutyricum Dorea</i> , <i>Subdoligranulum</i> and <i>Bacteroides</i> ANN [33]: <i>Anaerobutyricum</i> , <i>Bacteroides</i> , <i>Ruminococcus torques</i> , <i>Dorea</i> and <i>Subdoligranulum</i> RF [33]: <i>Anaerobutyricum</i> , <i>Faecalibacterium</i> , <i>Clostridium sensu stricto</i> , <i>Ruminococcus torques</i> and <i>Agathobacter</i>
[34]	ASD (n = 111)	NT (n = 143)	Mexico	[35]	16S rRNA (V4 region)	See [35]	ML: RF, SVM, kNN, NB DL: ANNs Classification ASD v. HC	Main predictor: <i>Prevotella_2</i> . Other significant predictors: <i>Ruminiclostridium_6</i> and the <i>Alloprevotella</i> . The ANN model demonstrates a 6% increase in sensitivity compared to kNN and RF models
[36]	ASD (n = 60) ASD (n = 77) ASD (n = 48)	Siblings (n = 57) HC (n = 50) HC (n = 48)	The Netherlands	Original [32,33]	16S rRNA (V3–V4 regions) (V4 region)	See [32,33]	ML: REFS Feature selection through REFS	ASVs: 26 ASVs for differential abundances. ↓ <i>Actinobacteria phylum</i> , <i>Bifidobacterium</i> and <i>Collinsella</i> in ASD ↑ <i>Bacteroidota phylum</i> , <i>Prevotellaceae</i> and <i>Parabacteroides</i> in ASD ↑bacterial taxa in ASD phenotype: <i>Clostridia</i> , <i>Sarcina</i> and <i>Parabacteroides</i>

Table 1. Cont.

Ref	Participant Characteristic		Country	Data Source	Study Type	Main Finding	ML/DL Method Used	Model Predicting Microbiome
	Experimental Group	Control Group						
[37]	ASD (n = 540)	HC (n = 419)	Italy	[35,38–42]	16S rRNA (Different regions)	See cited articles	ML: RF, SVM, GB Classification ASD v. HC	<p>Main bacterial genera for all three algorithms: <i>Alloprevotella</i>, <i>Sutterella</i>, <i>Haemophilus</i>, <i>Faecalibacterium</i> and an unclassified <i>Clostridia</i> 'UCG 014'.</p> <p>RF and the GBM algorithms: [<i>Eubacterium</i>] <i>siraecum_group</i>, <i>Tyzzereella</i>, <i>Negativibacillus</i>, <i>Muribaculaceae</i>, <i>Gastranaerophilales</i>, <i>Megamonas</i> and <i>Rombustia</i>.</p> <p>GBM and SVM algorithms: <i>Bacteroides</i> and <i>Subdoligranulum</i> identified as important.</p> <p>↓ <i>Alloprevotella</i> genus in ASD sample (abundance in ASD samples = <math>0.34 \pm 0.20</math>, abundance in HC samples = <math>0.12 \pm 0.14</math>).</p> <p>↑ <i>Parasutterella</i> (ASD samples = <math>0.57 \pm 0.16</math>, HC samples = <math>0.38 \pm 0.17</math>), <i>Haemophilus</i> (ASD samples = <math>0.57 \pm 0.19</math>, HC samples = <math>0.33 \pm 0.17</math>), <i>Faecalibacterium</i> (ASD samples = <math>0.86 \pm 0.14</math>, HC samples = <math>0.70 \pm 0.21</math>) and <i>Clostridiales</i> UCG 14 (ASD samples = <math>0.60 \pm 0.21</math>, HC samples = <math>0.34 \pm 0.17</math>) in ASD.</p>
[43]	ASD (n = 60) ASD (n = 77) ASD (n = 48)	HC (n = 57) HC (n = 50) HC (n = 48)	The Netherlands	[32,33,44]	16S rRNA (V3–V4 regions) (V4 region)	See cited articles	ML: REFS Feature selection: biomarker identification	<p>Better performance in AUC and MCC compared to K-Best and 10-time random selection methods.</p> <p>↓ <i>Bifidobacterium</i>, <i>Enterobacteriaceae</i>, <i>Lachnospira</i>, <i>Lachnospiraceae</i> and <i>Clostridium</i> in ASD</p>
[45]	ASD (n = 41)	NT (n = 35)	Italy	Original	PCR 16S rRNA (V3–V4 regions)	<i>Bifidobacterium</i> was negatively correlated with indole and skatole Positive correlations between <i>Carnobacteriaceae</i> , <i>Actinobacillus</i> , <i>Pepetostreptococcaceae</i> , pentanoic acid, 2,6-dimethyl-pyrazine, nonadecane and 3-methyl-butanoic acid.	ML: PCA, PLS-DA Feature selection: biomarker identification	<p>Hist Gradient Boosting Classifier was the best performing model with 89% accuracy.</p> <p>VOCs associated with ASDs: methyl isobutyl ketone, benzeneacetaldehyde, phenyl ethyl alcohol, ethanol, butanoic acid, octadecane, acetic acid, skatole and tetradecanal (myristyl aldehyde)</p> <p>Positive correlations with OTUs-VOCs couples: -<i>Bifidobacteriaceae</i>/2-dodecanol -<i>Serratia</i>/benzyl alcohol -<i>Roseburia</i>/1-butanol -<i>Firmicutes</i>/butanoic acid -<i>Pasteurellaceae</i>/3-methyl 1-butanol.</p>

Table 1. Cont.

Ref	Participant Characteristic		Country	Data Source	Study Type	Main Finding	ML/DL Method Used	Model Predicting Microbiome
	Experimental Group	Control Group						
[46]	ASD (n = 41)	NT (n = 35)	Italy	Original	PCR 16S rRNA (V3–V4 regions)	<p>30 ASD with GI symptoms: 93% with a high level of severity</p> <p>Phylum level:                      ↓<i>Actinobacteria</i>, <i>Cyanobacteria</i> and TM7 in ASD                      ↑<i>Proteobacteria</i> and <i>Bacteroidetes</i> in ASD                      (<i>Bacteroidetes</i> was observed in the ASD without GI symptoms group)</p> <p>Family level:                      ↓<i>Coriobacteriaceae</i>,  <i>Bifidobacteriaceae</i>, <i>Actinomycetaceae</i> and                      (<i>Tissierellaceae</i>) in ASD.                      ↑<i>Alcaligenaceae</i>, <i>Lactobacillaceae</i>,  <i>Prevotellaceae</i> and <i>Bacteroidaceae</i> in ASD.                      ↑<i>Bacteroidaceae</i> and                      Lactobacillaceae ASD without GI symptoms.</p> <p>Genus level:                      ↑<i>Bacteroides</i> and <i>Klebsiella</i>                      in ASD. <i>Klebsiella</i> and <i>Lactobacillus</i> were higher                      in ASD without GI symptoms.                      ↓<i>Bifidobacterium</i> and <i>Actinomyces</i> in ASD.</p> <p>ASD-related microbial                      biomarkers (p-value &lt; 0.05):                      -<i>Bacteroidetes/Proteobacteria</i>.                      -<i>Bacteroidaceae/Rikenellaceae/</i>  <i>Lactobacillaceae/Prevotellaceae/Pasteurellaceae/</i>  <i>Alcaligenaceae</i>.                      -<i>Bacteroides/Lactobacillus/Prevotella/</i>  <i>Klebsiella/Roseburia/Haemophilus/Sutterella</i>.</p>	ML: LR, SGD, RF, ET, GB, XGB, etc. DL: MLP Classification ASD v. HC	Contextually, model classification analysis based on ML identified both KOs and ko pathways able to classify 73% of patients with ASD versus CTRLs (p-value < 0.05). Specific selected OTUs for ASD and CTRLs revealed the main bacteria: ↓ <i>Bacteroides</i> , <i>Lactobacillus</i> , <i>Prevotella</i> , <i>Staphylococcus</i> and <i>Sutterella</i> in ASD ↓ <i>Ruminococcus</i> , <i>Blautia</i> , <i>Coprococcus</i> , <i>Bifodobacterium</i> and <i>Streptococcus</i> in ASD.

Table 1. Cont.

Ref	Participant Characteristic		Country	Data Source	Study Type	Main Finding	ML/DL Method Used	Model Predicting Microbiome
	Experimental Group	Control Group						
[47]	ASD (n = 43)	TD (n = 31)	China	[48,49]	IgA detection via ELISA [49]  StoolGen fecal DNA extraction kit (CWBiotech Co., Beijing, China) and NanoDrop 2000 (Thermo Scientific, Foster City, CA, USA). A total of 5 µg (or more) of DNA was required for library construction using the TruSeq DNA sample preparation kit (Illumina, San Diego, CA, USA) [48].	VFGM genes related to ASD: cpsH, cpsJ and cpsO genes related to high levels of <i>Streptococcus agalactiae</i> 2603 V/R in the gut of ASD children with/without GI symptoms	ML: RF Classification ASD v. HC	The main genes involved according to machine learning via the random forest method were mtrE, kfiC, pvdM and hasA.
[50]	ASD (n = 73)	TD (n = 71)	China	[48,51]	Illumina NovaSeq 6000 Illumina HiSeq 4000, Illumina Inc. San Diego, CA, USA	See cited articles	ML: RF Classification ASD v. HC	Predicted performance was evaluated according to AUROC.  In the China cohort, a high AUROC value of 0.984 and 97% accuracy were achieved with only one round of a 100-iteration run.  The Moscow cohort produced a poor average AUROC outcome of 0.81 and only 67% accuracy following six rounds of the 100-iteration run.  Overall, average values for AUROC and accuracy were 0.86 and 80%, respectively, with an average feature set of 67 species.  <i>Eubacterium_sp_CAG_248</i> and <i>Prevotella copri</i> were the most likely biomarkers involved in ASD.
[52]	ASD (n = 169)	NT (n = 128)	China	[39,42,53–55]	PCR 16S rRNA (Different regions)	See cited articles	ML: LDA (LefSe)+ RF, kSVM + RBF, DT DL: MLP Feature selection + Classification	Dominant major genera: ↓ <i>Prevotella</i> , ↓ <i>Ruminococcus</i> and <i>Roseburia</i> as potential biomarkers of ASD.  <i>Prevotella</i> , <i>Roseburia</i> , <i>Ruminococcus</i> , <i>Megasphaera</i> and <i>Catenibacterium</i> as potential biomarkers in ASD patients. However, only <i>Prevotella</i> significantly differed between the two groups.

Table 1. Cont.

Ref	Participant Characteristic		Country	Data Source	Study Type	Main Finding	ML/DL Method Used	Model Predicting Microbiome
	Experimental Group	Control Group						
[56]	ASD (n = 569)	HC (n = 450)	China	[32,33,35,38,40,41,46,57–61]	PCR 16S rRNA (V3–V4, V4, V4–V5 regions)  Illumina MiSeq	See cited articles	ML: RF Classification of ASD v. HC (after feature selection)	AUC of the training set and verification set was 0.688 and 0.706.  Dominant genera of the ASD group included <i>Lachnospiraceae_incertae_sedis</i> , <i>Clostridium_XVIII</i> , <i>Eubacterium</i> , <i>Anaerostipes</i> , <i>Clostridium_sensu_stricto</i> , <i>Coprococcus</i> , <i>Dorea</i> and <i>Faecalibacterium</i> .  Subgroup analysis followed different sequencing platforms to examine dominant genera in ASD. Dominant genera in the ASD group included <i>Eubacterium</i> , <i>Bifidobacterium</i> , <i>Blautia</i> , <i>Dialister</i> , <i>Coprococcus</i> and <i>Lachnospiraceae_incertae_sedis</i> .

Note: ASD = autism spectrum disorder; NT = neurotypical; HC = healthy control; GS = gastrointestinal symptoms; w = with; w/o = without; AUC = area under the curve; AUROC = area under the receiver operating characteristic curve; VFGM = virulence factor-related gut microbiota; MCC = the Matthews correlation coefficient = MCC; ASVs = amplicon sequence variants; ML = machine learning; REFS = recursive ensemble feature selection; PCA = principal component analysis; PLS-DA = partial least squares-discriminant analysis; LDA = linear discriminant analysis; LefSe = LDA effect size; SVM = support vector machines = SVM; RF = random forest classifier; kNN = k-nearest neighbor classifier; NB = naïve Bayes classifier; LR = logistic regression; SGD = Stochastic gradient descent classifier; ET = extra trees classifier; DT = decision trees classifier; GB = gradient boosting; XGB = extreme GB; kSVM + RBF = kernelized SVMs (RBF kernel); ANNs = artificial neural networks; MLP = multi-layer perceptron; DL = deep learning. Feature selection: REFS = recursive ensemble feature selection; PCA = principal component analysis; PLS-DA = partial least squares-discriminant analysis; LDA = linear discriminant analysis; LefSe = LDA effect size. ML = machine learning; SVM = support vector machines; RF = random forest classifier; kNN = k-nearest neighbor classifier; NB = naïve Bayes classifier; LR = logistic regression; SGD = stochastic gradient descent classifier; ET = extra trees classifier; DT = decision trees classifier; GB = gradient boosting; XGB = extreme GB; kSVM + RBF = kernelized SVMs (RBF kernel). DL = deep learning; although very basic DL methodologies, the following have been identified: ANNs = artificial neural networks; MLP = multi-layer perceptron.



### 3. Results

#### 3.1. Study Selection

As can be seen in Figure 1, 127 articles corresponding to the study aim or research question were selected. Following the elimination of duplicates using EndNote 20 software, the number of included studies was reduced to 88. A total of 34 articles were eliminated following the application of eligibility criteria. Despite five articles being deemed doubtful for final inclusion, 16 articles were included in the initial screening by the review team. After discussion between all authors, this number was reduced to 12 articles. Of the doubtful papers that were ultimately excluded, the paper of Shi et al. [62] was discarded due to its unsuitable format (letter to the editor), whilst of Liu et al. [63] was excluded for not focusing on ASD and being an unsuitable format (a proposal).

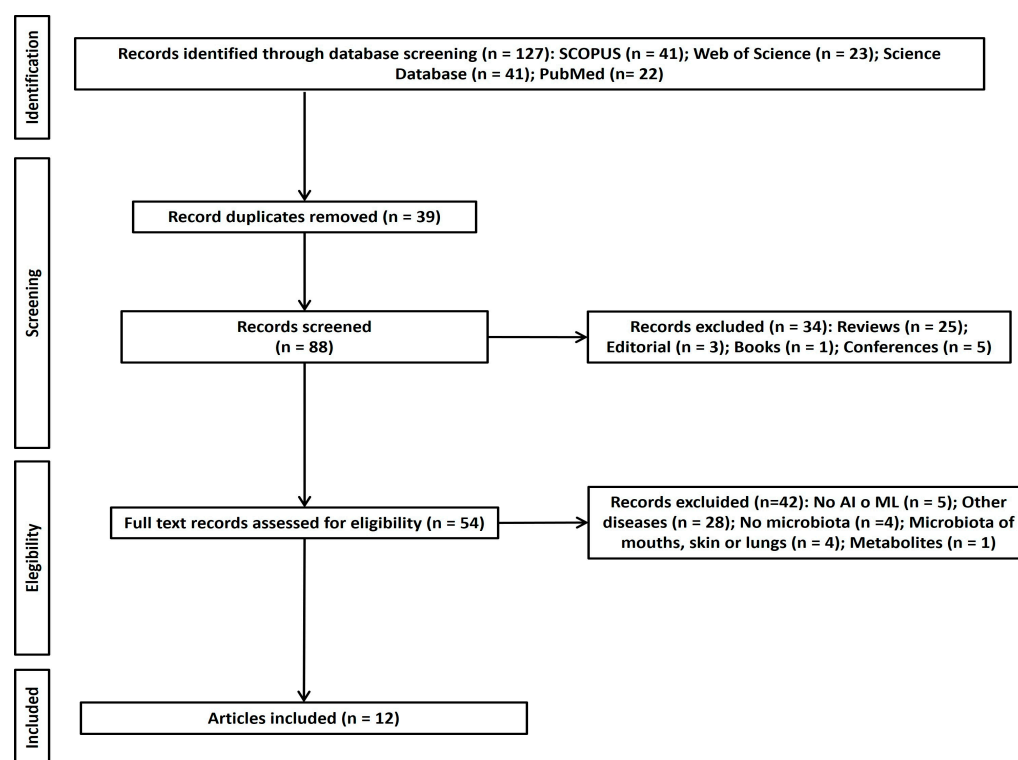


Figure 1. PRISMA flow chart of study selection.

#### 3.2. Individual Study Characteristics and Outcomes

The 12 included articles are summarized in Table 1 according to reference, participant characteristics, country, data source, study type, main findings, ML/DL method used and model used to predict microbiome.

The 12 included articles were published between 2020 and 2024 in five different countries (four [33.3%] in China, three [25.0%] in Italy, two [16.6%] in the Netherlands, two [16.6%] in Mexico and one [8.3%] in South Korea). Included articles were published in 10 different journals with 2 (16.6%) appearing in *Biomedicines*, 2 (16.6%) in *Frontiers in Microbiology* and 1 (8.3%) each in *Psychiatry Research*, *Neural Computing and Applications*, *Scientific Reports*, *BMC Bioinformatics*, *Computational and Structural Biotechnology Journal*, *Microorganisms*, *Microbial Pathogenesis* and *Microbiology Spectrum*.

Three (25%) articles gathered their own data to perform study analyses [30,45,46], whilst one (8.3%) article compared data they collected with data from two publicly available studies [36], one (8.3%) article exclusively used data from a publicly available study [34], three (8.3%) articles used data from two publicly available studies [31,47,50], one (8.3%) article used data from three publicly available studies [43], one (8.3%) article used data from

five publicly available studies [52], one (8.3%) article used data from six publicly available studies [37] and one (8.3%) article used data from ten publicly available studies [56].

Some of the bacteria found to be implicated in ASD found by studies included in the present systematic review include *Bacteroides* [30,31,36,46], *Bifidobacterium* [36,43,46,56], *Prevotella* [36,46,50,52], *Faecalibacterium* [37,56], *Ruminococcus* [31,46,52] and *Clostridium* [31,36,37,43]. Outcomes indicate low levels of abundance pertaining to the bacteria *Bifidobacterium*, *Bacteroides*, *Prevotella*, *Ruminococcus*, *Lachnospira* and *Clostridium*. These findings suggest that these bacteria may be biomarkers of ASD.

### 3.3. Risk of Bias

The Methods Guide for Comparative Effectiveness Reviews was used to evaluate the risk of bias for each included article in the present systematic review [64]. This risk assessment method has also been used by the authors in other publications [13,65,66].

The strength of publications is expressed in terms of risk of bias: low (10–12); M = medium (6–9); H = high (1–5).

Five articles included in the present systematic review exhibited a low risk of bias, according to their risk of bias score, whilst seven studies exhibited a moderate risk of bias. Thus, a moderate risk of bias was found in 41.6% of included studies. In general, the greatest bias emerged in relation to sample size, identification of the sample and controlling for comorbidity factors such as intellectual disability, severity of ASD and diet. Overall and individual item scores for included articles are presented in Table 2.

**Table 2.** Quality assessment outcomes for included studies.

Item	[31]	[34]	[36]	[37]	[43]	[45]	[46]	[47]	[50]	[52]	[56]	[30]
1. Clear stated aim	2	2	2	2	2	2	2	2	1	2	2	2
2. Appropriate study size	1	2	2	2	2	1	1	1	2	2	2	2
3. Identified and assessed	1	1	2	2	1	1	2	1	1	1	2	2
4. Blinding of participants and personnel	1	1	2	1	1	1	2	1	1	1	2	2
5. Other bias (controls for dietary intake, reports on comorbidity and ASD severity)	1	1	1	1	1	1	2	1	1	1	1	2
6. AI or ML in GM	2	2	2	2	2	2	2	1	1	2	2	2
<b>TOTAL</b>	<b>8</b>	<b>9</b>	<b>11</b>	<b>10</b>	<b>9</b>	<b>8</b>	<b>11</b>	<b>8</b>	<b>6</b>	<b>9</b>	<b>11</b>	<b>12</b>
Risk of bias	4	3	1	2	3	4	1	4	6	3	1	0
Risk of bias classification	M	M	L	L	M	M	L	M	M	M	L	L

Note: AI = artificial intelligence; ML = machine learning; GM = gut microbiota; 0 = not reported; 1 = not adequately assessed; 2 = adequately assessed; L = low (10–12); M = medium (6–9); H = high (1–5).

Most studies examine samples from countries other than the original source of the publication. Data registries on GM come mainly from studies in China, Italy, the USA and Russia. Only four studies include an original sample [30,36,45,46], with such samples being highly limited and not representative of the ASD population [43,45,46].

### 3.4. Limitations

As previously discussed, the present systematic review was, by definition, limited by the databases used, the search terms chosen and the inclusion/exclusion criteria established [67].

#### 4. Discussion

Present findings point to a relationship between low levels of abundance in the bacteria *Bifidobacterium*, *Bacteroides*, *Prevotella*, *Ruminococcus*, *Lachnospira* and *Clostridium* in ASD. This coincides with a previous meta-analysis study that found a dysbiosis or alteration of *Bifidobacterium* in ASD [9]. Thus, these outcomes produced using classic human methodology appear to coincide with those produced using ML. *Bifidobacterium* is possibly the bacterial biomarker most clearly implicated in the neurodevelopment of ASD. *Bifidobacterium* is one of the bacteria that appears most in studies with [30,36,43,45,46,56] and without AI in autism [9]. This is one of the first bacteria to colonize the intestine of neonates and is, therefore, highly important in neurodevelopment. Furthermore, it is associated with low levels of anxiety and gastrointestinal symptoms [10].

It can be observed that most of the machine learning and artificial intelligence (ML/AI) methodologies presented in the reviewed literature were employed for either (1) feature selection, i.e., to reduce the dimensionality of data and extract the most discriminative characteristics of data and, subsequently, use this subset of features for prediction, or as a support for the training of machine learning models; or (2) classification of ASD vs. HC (healthy controls), oftentimes also referred to as 'logistic regression', to be able to discern two types of gut bacteria profiles and, resultantly, determine the presence or absence of ASD-related dysbiosis in a new unseen sample.

One included study stands out [30] for performing regression for the prediction of microbial age instead of classification, whilst also notably employing extreme gradient boosting (XGB) to regress high/low SRS/VABS values.

However, a larger body of analyzed research uses very similar approaches, i.e., uses machine learning techniques, for sample classification as either ASD or HC participants. The techniques used vary but mostly correspond to 'classical' machine learning techniques, such as support vector machines (SVMs), whether the vanilla version [31,34,37] or kernelized with RBF (radial basis function kernel) [52], random forests (RF), k-nearest neighbors (k-NN) [34], gradient boosting (GB) [37,46], stochastic gradient descent (SGD) classifier [46], extra trees (ET) [46] and decision trees (DT) [52]. Random forest classifiers (RF) were the most commonly used method in the literature analyzed, as they tend to outperform other methods, and the complexity of training required is not very high. RF classifiers are employed in a number of included articles [31,34,37,46,47,50,56].

Another body of work focuses on the selection of features, with the two main aims of this being either direct biomarker discovery [36,43,45] or as a preliminary step towards simplifying model training within a machine learning classifier model [52,56] similar to those presented in the previous paragraph. For this, included studies present several methods, namely, principal component analysis (PCA), a classical method [45], recursive ensemble feature selection (REFS) [36,43], partial least squares discriminant analysis (PLS-DA) [45] and linear discriminant analysis (LDA) with effect size estimation (LEfSe) [52].

Finally, neural network-based models are seldom used in the studies analyzed. Indeed, only four following studies used classical models such as artificial neural networks (ANNs) [31,34] or multi-layer perceptron models (MLPs) [46,52]. From the descriptions available, it was not possible to determine whether 'deep' models were employed in these cases, although it seems implausible due to the nature of the methods used, which seem to refer to the prepackaged implementations available in off-the-shelf machine learning and data mining software libraries, or statistical analysis tools such as MATLAB, R, SPSS, Weka, Scikit-Learn, etc.

This leads to the first limitation inherent to included studies, which emerged as a lack of more advanced methods, although it was also the case that some deeper models would have encountered problems when dealing with low-dimensional data for which there are also too few samples. Data-driven learning, which is employed for deep neural network training, is 'data-hungry' in this sense and can lead to overfitting fairly easily. It would be interesting, however, to see more research conducted on this issue.

When using relatively small datasets, one of the most common techniques in ML/AI model training is to use techniques such as cross-validation (CV) or n-fold CV, which split data into one large and one small subset with the former being used for the main analysis and the latter ('fold') being used to determine model accuracy. Following the initial analysis, the second subset is then shifted, the system retrained, and average outcomes are calculated from the two resulting models. This has traditionally yielded good results; however, it does not provide additional information from other potentially existing broader datasets. For instance, it is now common in ML model training to perform transfer learning (TL) and, specifically, use fine-tuning processes during training. Within this training approach, the ML model is first trained on a larger domain (wider problem, e.g., general image classification) and then harnessed to address a restricted-domain problem (e.g., melanoma image classification). A similar approach, namely, domain adaptation (DA), could have been conducted here. When using DA, the aim is to learn from source data distribution, applying the model to different (yet related) target data distributions. This enriches the model, as it handles more data overall. In contrast, when using CV techniques, a very limited set of data is used; therefore, model overfitting becomes a common issue. In ASD GM studies, DA could be used to learn from pooled data from datasets originating from small-to-medium studies and applied to an unseen set of data from another small(er) study. Finally, it would also be interesting to introduce continual learning (CL) approaches, in which models are trained on 'shifting' datasets, i.e., data distributions that change over time, as more data are incorporated into the model, e.g., from growing corpora of collected ASD and HC GM-related data.

A second limitation of the studies included in the present review is that the majority are based on very small samples of 50 to 70 participants, which is far too low to support the conclusions drawn by the authors of these original studies. The risk of bias in each individual study must also be considered, not only because of the sample size, which is not very representative of the ASD population but also because of the poor representation of countries and cultures. Furthermore, it can be observed that the bacterial profiles used in the presented studies come from very specific populations and do not consider cultural differences in dietary intake, which is a key factor in the GM profiles of the general population in any given geographical area. It would be of interest for future work to conduct pooled analyses of these populations, whilst also completing the geographical profiles with additional data from less observed countries and regions. This could then be used to train a larger model, or even a deeper neural model, as the risk of bias and overfitting would be diminished. A more global examination of GM is necessary as a function of evolutionary periods within both neurotypical populations and those with ASD. In addition, ASD severity must be considered, alongside the existence of gastrointestinal symptoms. A huge challenge to research is posed by the standardization of global GM, which should also correspond to a preliminary step in the development of AI. At this time, more basic research is needed to analyze the diversity of GM in the general population.

In summary, the following strengths and weaknesses are seen in studies of AI in the study of GM in ASD: (1) acceptable samples for validation studies of psychometric instruments, which tend to imply few variables or factors with 250–500 individuals with ASD. However, variability in GM is greater because each family of bacteria acts as a higher dimension or factor that, in turn, includes thousands of genera and species of bacteria. Thus, samples should comprise at least 1000 individuals with ASD representing the three levels of autism severity according to the DSM-5. (2) ASD study samples should be representative of the rural and urban areas in the countries, in which the research is conducted and avoid extremes with regard to socioeconomic resources given that socioeconomic context influences diet [12,68] and autism severity [69]. (3) Additionally, GM studies should be performed with individuals with ASD who have comorbidities. (4) Due to the small study samples, correlations according to age and sex have not been performed with GM individuals with ASD. Future studies should consider these factors. (5) The use of AI should have a rigorous prior methodological basis based on sample size, age and evolutionary

period. In this way, predictions about biomarkers could be more adjusted to the reality of GM in autism. In conclusion, more global and rigorous analyses regarding AI would be possible if all these methodological factors were considered. In any case, the incorporation of AI is an important advance that saves time and cuts economic costs in ASD research.

## 5. Conclusions

Current methods for the prediction of ASD in children based on GM data are still hampered by bias due to the small sample sizes used by relevant research. This limits the viability of such methods for use in a clinical setting, as they are still in need of further in-depth research. Further data are required to be able to construct large predictive models that can be generalized to global populations and consider diverse cultural and ethnic backgrounds. Generally speaking, larger samples are needed, comprising thousands of data points or more, in order to avoid the bias observed in the methods employed in the present review. Future studies should consider and strive to control for environmental factors such as diet and country of origin, as well as analytical factors such as sequencing platform and hypervariable region when comparing GM in patients with ASD with neurotypical patients. Finally, further research is required on this topic in order to facilitate early diagnosis and provide tools for a better prognosis of ASD in children and teenagers.

**Author Contributions:** Conceptualization, A.E.M.-G. and P.C.-P.; methodology, P.A.-M.; formal analysis, P.A.-M. and A.E.M.-G.; investigation, A.E.M.-G.; data curation, P.A.-M.; writing—original draft preparation, A.E.M.-G., P.C.-P. and P.A.-M.; writing—review and editing, A.E.M.-G. and P.A.-M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study due to the fact that it presents a systematic review.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. APA (American Psychiatric Association). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition Text Revision, DSM-5-TR*; American Psychiatric Association: Washington, DC, USA, 2022.
2. Maenner, M.J.; Shaw, K.A.; Baio, J.; Washington, A.; Patrick, M.; DiRienzo, M.; Christensen, D.L.; Wiggins, L.D.; Pettygrove, S.; Andrews, J.G.; et al. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. *MMWR Surveill. Summ.* **2020**, *69*, 1–12. [[CrossRef](#)] [[PubMed](#)]
3. Myers, S.M.; Voigt, R.G.; Colligan, R.C.; Weaver, A.L.; Storlie, C.B.; Stoeckel, R.E.; Port, J.D.; Katusic, S.K. Autism Spectrum Disorder: Incidence and Time Trends Over Two Decades in a Population-Based Birth Cohort. *J. Autism Dev. Disord.* **2019**, *49*, 1455–1474. [[CrossRef](#)] [[PubMed](#)]
4. Russell, G.; Stapley, S.; Newlove-Delgado, T.; Salmon, A.; White, R.; Warren, F.; Pearson, A.; Ford, T. Time trends in autism diagnosis over 20 years: A UK population-based cohort study. *J. Child Psychol. Psychiatry* **2022**, *63*, 674–682. [[CrossRef](#)] [[PubMed](#)]
5. Qiu, S.; Lu, Y.; Li, Y.; Shi, J.; Cui, H.; Gu, Y.; Li, Y.; Zhong, W.; Zhu, X.; Liu, Y.; et al. Prevalence of autism spectrum disorder in Asia: A systematic review and meta-analysis. *Psychiatry Res.* **2020**, *284*, 112679. [[CrossRef](#)] [[PubMed](#)]
6. van 't Hof, M.; Tisseur, C.; van Berckeleer-Onnes, I.; van Nieuwenhuyzen, A.; Daniels, A.M.; Deen, M.; Hoek, H.W.; Ester, W.A. Age at autism spectrum disorder diagnosis: A systematic review and meta-analysis from 2012 to 2019. *Autism* **2021**, *25*, 862–873. [[CrossRef](#)] [[PubMed](#)]
7. Chan, W.K.; Griffiths, R.; Price, D.J.; Mason, J.O. Cerebral organoids as tools to identify the developmental roots of autism. *Mol. Autism* **2020**, *11*, 58. [[CrossRef](#)] [[PubMed](#)]
8. LaSalle, J.M. Epigenomic signatures reveal mechanistic clues and predictive markers for autism spectrum disorder. *Mol. Psychiatry* **2023**, *28*, 1890–1901. [[CrossRef](#)] [[PubMed](#)]
9. Andreo-Martínez, P.; Rubio-Aparicio, M.; Sánchez-Meca, J.; Veas, A.; Martínez-González, A.E. A Meta-analysis of Gut Microbiota in Children with Autism. *J. Autism Dev. Disord.* **2022**, *52*, 1374–1387. [[CrossRef](#)]
10. Martínez-González, A.E.; Andreo-Martínez, P. Implications of Gut Microbiota and Gastrointestinal Symptoms in Autism. In *Advances in Health and Disease*; Duncan, L.T., Ed.; NOVA Science Publishers: New York, NY, USA, 2020; Volume 29, pp. 16–21.

11. Andreo-Martínez, P.; García-Martínez, N.; Sánchez-Samper, E.P.; Quesada-Medina, J.; MacFabe, D. Metabolites of the gut microbiota involved in the autism spectrum disorder. *Rev. Dis. Clin. Neuro.* **2018**, *5*, 39–48. [[CrossRef](#)]
12. Andreo-Martínez, P.; García-Martínez, N.; Sánchez-Samper, E.P.; Martínez-González, A.E. An approach to gut microbiota profile in children with autism spectrum disorder. *Environ. Microbiol. Rep.* **2019**, *12*, 115–135. [[CrossRef](#)]
13. Martínez-González, A.E.; Andreo-Martínez, P. The Role of Gut Microbiota in Gastrointestinal Symptoms of Children with ASD. *Medicina* **2019**, *55*, 408. [[CrossRef](#)] [[PubMed](#)]
14. Berding, K.; Donovan, S.M. Diet Can Impact Microbiota Composition in Children With Autism Spectrum Disorder. *Front. Neurosci.* **2018**, *12*, 515. [[CrossRef](#)]
15. Martínez-González, A.E.; Cervin, M.; Pérez-Sánchez, S. Assessing gastrointestinal symptoms in people with autism: Applying a new measure based on the Rome IV criteria. *Dig. Liver Dis.* **2024**, *in press*. [[CrossRef](#)]
16. Martínez-González, A.E.; Andreo-Martínez, P. Prebiotics, probiotics and fecal microbiota transplantation in autism: A systematic review. *Rev. Psiquiatr. Salud Ment.* **2020**, *13*, 150–164. [[CrossRef](#)] [[PubMed](#)]
17. Lindström, E.S.; Langenheder, S. Local and regional factors influencing bacterial community assembly. *Environ. Microbiol. Rep.* **2012**, *4*, 1–9. [[CrossRef](#)] [[PubMed](#)]
18. Suzuki, T.A.; Worobey, M. Geographical variation of human gut microbial composition. *Biol. Lett.* **2014**, *10*, 20131037. [[CrossRef](#)] [[PubMed](#)]
19. Gupta, V.K.; Paul, S.; Dutta, C. Geography, Ethnicity or Subsistence-Specific Variations in Human Microbiome Composition and Diversity. *Front. Microbiol.* **2017**, *8*, 1162. [[CrossRef](#)] [[PubMed](#)]
20. Minsky, M. Steps toward Artificial Intelligence. *Proc. IRE* **1961**, *49*, 8–30. [[CrossRef](#)]
21. Feigenbaum, E. Artificial intelligence research. *IEEE Trans. Inf. Theory* **1963**, *9*, 248–253. [[CrossRef](#)]
22. Duan, Y.; Edwards, J.S.; Dwivedi, Y.K. Artificial intelligence for decision making in the era of Big Data—Evolution, challenges and research agenda. *Int. J. Inf. Manag.* **2019**, *48*, 63–71. [[CrossRef](#)]
23. Stahl, D.; Pickles, A.; Elsabbagh, M.; Johnson, M.H.; The, B.T. Novel Machine Learning Methods for ERP Analysis: A Validation From Research on Infants at Risk for Autism. *Dev. Neuropsychol.* **2012**, *37*, 274–298. [[CrossRef](#)] [[PubMed](#)]
24. Maenner, M.J.; Yeargin-Allsopp, M.; Van Naarden Braun, K.; Christensen, D.L.; Schieve, L.A. Development of a Machine Learning Algorithm for the Surveillance of Autism Spectrum Disorder. *PLoS ONE* **2016**, *11*, e0168224. [[CrossRef](#)] [[PubMed](#)]
25. Rosenblatt, F. Principles of neurodynamics. perceptrons and the theory of brain mechanisms. *Am. J. Psychol.* **1963**, *76*, 705.
26. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
27. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. *Neural Inf. Process. Syst.* **2012**, *25*, 84–90. [[CrossRef](#)]
28. Graham, S.; Depp, C.; Lee, E.E.; Nebeker, C.; Tu, X.; Kim, H.C.; Jeste, D.V. Artificial Intelligence for Mental Health and Mental Illnesses: An Overview. *Curr. Psychiatry Rep.* **2019**, *21*, 116. [[CrossRef](#)] [[PubMed](#)]
29. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [[CrossRef](#)] [[PubMed](#)]
30. Jung, Y.; Lee, T.; Oh, H.S.; Hyun, Y.; Song, S.; Chun, J.; Kim, H.W. Gut microbial and clinical characteristics of individuals with autism spectrum disorder differ depending on the ecological structure of the gut microbiome. *Psychiatry Res.* **2024**, *335*, 115775. [[CrossRef](#)] [[PubMed](#)]
31. Olaguez-Gonzalez, J.M.; Chairez, I.; Breton-Deval, L.; Alfaro-Ponce, M. Machine Learning Algorithms Applied to Predict Autism Spectrum Disorder Based on Gut Microbiome Composition. *Biomedicines* **2023**, *11*, 2633. [[CrossRef](#)]
32. Zou, R.; Xu, F.; Wang, Y.; Duan, M.; Guo, M.; Zhang, Q.; Zhao, H.; Zheng, H. Changes in the Gut Microbiota of Children with Autism Spectrum Disorder. *Autism Res.* **2020**, *13*, 1614–1625. [[CrossRef](#)]
33. Ding, X.; Xu, Y.; Zhang, X.; Zhang, L.; Duan, G.; Song, C.; Li, Z.; Yang, Y.; Wang, Y.; Wang, X.; et al. Gut microbiota changes in patients with autism spectrum disorders. *J. Psychiatr. Res.* **2020**, *129*, 149–159. [[CrossRef](#)] [[PubMed](#)]
34. Olaguez-Gonzalez, J.M.; Schaeffer, S.E.; Breton-Deval, L.; Alfaro-Ponce, M.; Chairez, I. Assessment of machine learning strategies for simplified detection of autism spectrum disorder based on the gut microbiome composition. *Neural Comput. Appl.* **2024**, *36*, 8163–8180. [[CrossRef](#)]
35. Dan, Z.; Mao, X.; Liu, Q.; Guo, M.; Zhuang, Y.; Liu, Z.; Chen, K.; Chen, J.; Xu, R.; Tang, J.; et al. Altered gut microbial profile is associated with abnormal metabolism activity of Autism Spectrum Disorder. *Gut Microbes* **2020**, *11*, 1246–1267. [[CrossRef](#)] [[PubMed](#)]
36. Peralta-Marzal, L.N.; Rojas-Velazquez, D.; Rigters, D.; Prince, N.; Garssen, J.; Kraneveld, A.D.; Perez-Pardo, P.; Lopez-Rincon, A. A robust microbiome signature for autism spectrum disorder across different studies using machine learning. *Sci. Rep. (Nat. Publ. Group)* **2024**, *14*, 814. [[CrossRef](#)]
37. Pietrucci, D.; Teofani, A.; Milanese, M.; Fosso, B.; Putignani, L.; Messina, F.; Pesole, G.; Desideri, A.; Chillemi, G. Machine Learning Data Analysis Highlights the Role of Parasutterella and Alloprevotella in Autism Spectrum Disorders. *Biomedicines* **2022**, *10*, 2028. [[CrossRef](#)]
38. Averina, O.V.; Kovtun, A.S.; Polyakova, S.I.; Savilova, A.M.; Rebrikov, D.V.; Danilenko, V.N. The bacterial neurometabolic signature of the gut microbiota of young children with autism spectrum disorders. *J. Med. Microbiol.* **2020**, *69*, 558–571. [[CrossRef](#)]

39. Pulikkan, J.; Maji, A.; Dhakan, D.B.; Saxena, R.; Mohan, B.; Anto, M.M.; Agarwal, N.; Grace, T.; Sharma, V.K. Gut Microbial Dysbiosis in Indian Children with Autism Spectrum Disorders. *Microb. Ecol.* **2018**, *76*, 1102–1114. [[CrossRef](#)] [[PubMed](#)]
40. Zurita, M.F.; Cárdenas, P.A.; Sandoval, M.E.; Peña, M.C.; Fornasini, M.; Flores, N.; Monaco, M.H.; Berding, K.; Donovan, S.M.; Kuntz, T.; et al. Analysis of gut microbiome, nutrition and immune status in autism spectrum disorder: A case-control study in Ecuador. *Gut Microbes* **2019**, *11*, 453–464. [[CrossRef](#)]
41. Coretti, L.; Paparo, L.; Riccio, M.P.; Amato, F.; Cuomo, M.; Natale, A.; Borrelli, L.; Corrado, G.; Comegna, M.; Buommino, E.; et al. Gut Microbiota Features in Young Children With Autism Spectrum Disorders. *Front. Microbiol.* **2018**, *9*, 3146. [[CrossRef](#)]
42. Son, J.S.; Zheng, L.J.; Rowehl, L.M.; Tian, X.; Zhang, Y.; Zhu, W.; Litcher-Kelly, L.; Gadow, K.D.; Gathungu, G.; Robertson, C.E.; et al. Comparison of Fecal Microbiota in Children with Autism Spectrum Disorders and Neurotypical Siblings in the Simons Simplex Collection. *PLoS ONE* **2015**, *10*, e0137725. [[CrossRef](#)]
43. Rojas-Velazquez, D.; Kidwai, S.; Kraneveld, A.D.; Tonda, A.; Oberski, D.; Garssen, J.; Lopez-Rincon, A. Methodology for biomarker discovery with reproducibility in microbiome data using machine learning. *BMC Bioinform.* **2024**, *25*, 26. [[CrossRef](#)] [[PubMed](#)]
44. David, M.M.; Tataru, C.; Daniels, J.; Schwartz, J.; Keating, J.; Hampton-Marcell, J.; Gottel, N.; Gilbert, J.A.; Wall, D.P. Children with Autism and Their Typically Developing Siblings Differ in Amplicon Sequence Variants and Predicted Functions of Stool-Associated Microbes. *mSystems* **2021**, *6*, e00193-20. [[CrossRef](#)] [[PubMed](#)]
45. Vernocchi, P.; Marangelo, C.; Guerrera, S.; Del Chierico, F.; Guarrasi, V.; Gardini, S.; Conte, F.; Paci, P.; Ianiro, G.; Gasbarrini, A.; et al. Gut microbiota functional profiling in autism spectrum disorders: Bacterial VOCs and related metabolic pathways acting as disease biomarkers and predictors. *Front. Microbiol.* **2023**, *14*, 1287350. [[CrossRef](#)] [[PubMed](#)]
46. Vernocchi, P.; Ristori, M.V.; Guerrera, S.; Guarrasi, V.; Conte, F.; Russo, A.; Lupi, E.; Albitar-Nehme, S.; Gardini, S.; Paci, P.; et al. Gut Microbiota Ecology and Inferred Functions in Children With ASD Compared to Neurotypical Subjects. *Front. Microbiol.* **2022**, *13*, 871086. [[CrossRef](#)] [[PubMed](#)]
47. Wang, M.B.; Doenyas, C.; Wan, J.; Zeng, S.J.; Cai, C.Q.; Zhou, J.X.; Liu, Y.Q.; Yin, Z.Q.; Zhou, W.H. Virulence factor-related gut microbiota genes and immunoglobulin A levels as novel markers for machine learning-based classification of autism spectrum disorder. *Comp. Struct. Biotechnol. J.* **2021**, *19*, 545–554. [[CrossRef](#)] [[PubMed](#)]
48. Wang, M.; Wan, J.; Rong, H.; He, F.; Wang, H.; Zhou, J.; Cai, C.; Wang, Y.; Xu, R.; Yin, Z.; et al. Alterations in Gut Glutamate Metabolism Associated with Changes in Gut Microbiota Composition in Children with Autism Spectrum Disorder. *mSystems* **2019**, *4*, e00321-18. [[CrossRef](#)] [[PubMed](#)]
49. Zhou, J.; He, F.; Yang, F.; Yang, Z.; Xie, Y.; Zhou, S.; Liang, J.; Xu, R.; Wang, Y.; Guo, H.; et al. Increased stool immunoglobulin A level in children with autism spectrum disorders. *Res. Dev. Disabil.* **2018**, *82*, 90–94. [[CrossRef](#)] [[PubMed](#)]
50. Wang, W.J.; Fu, P.C. Gut Microbiota Analysis and In Silico Biomarker Detection of Children with Autism Spectrum Disorder across Cohorts. *Microorganisms* **2023**, *11*, 19. [[CrossRef](#)]
51. Kovtun, A.S.; Averina, O.V.; Alekseeva, M.G.; Danilenko, V.N. Antibiotic Resistance Genes in the Gut Microbiota of Children with Autistic Spectrum Disorder as Possible Predictors of the Disease. *Microb. Drug Resist.* **2020**, *26*, 1307–1320. [[CrossRef](#)]
52. Wu, T.; Wang, H.C.; Lu, W.W.; Zhai, Q.X.; Zhang, Q.X.; Yuan, W.W.; Gu, Z.N.; Zhao, J.X.; Zhang, H.; Chen, W. Potential of gut microbiome for detection of autism spectrum disorder. *Microb. Pathog.* **2020**, *149*, 10. [[CrossRef](#)]
53. Strati, F.; Cavalieri, D.; Albanese, D.; De Felice, C.; Donati, C.; Hayek, J.; Jousson, O.; Leoncini, S.; Renzi, D.; Calabro, A.; et al. New evidences on the altered gut microbiota in autism spectrum disorders. *Microbiome* **2017**, *5*, 24. [[CrossRef](#)] [[PubMed](#)]
54. Kang, D.W.; Park, J.G.; Ilhan, Z.E.; Wallstrom, G.; Labaer, J.; Adams, J.B.; Krajmalnik-Brown, R. Reduced incidence of Prevotella and other fermenters in intestinal microflora of autistic children. *PLoS ONE* **2013**, *8*, e68322. [[CrossRef](#)] [[PubMed](#)]
55. Liu, J.; Liu, X.; Xiong, X.-Q.; Yang, T.; Cui, T.; Hou, N.-L.; Lai, X.; Liu, S.; Guo, M.; Liang, X.-H.; et al. Effect of vitamin A supplementation on gut microbiota in children with autism spectrum disorders—A pilot study. *BMC Microbiol.* **2017**, *17*, 204. [[CrossRef](#)] [[PubMed](#)]
56. Xu, Y.S.; Wang, Y.H.; Xu, J.S.; Song, Y.; Liu, B.Q.; Xiong, Z.F. Leveraging Existing 16SrRNA Microbial Data to Define a Composite Biomarker for Autism Spectrum Disorder. *Microbiol. Spectr.* **2022**, *10*, 12. [[CrossRef](#)] [[PubMed](#)]
57. Chiappori, F.; Cupaioli, F.A.; Consiglio, A.; Di Nanni, N.; Mosca, E.; Licciulli, V.F.; Mezzelani, A. Analysis of Faecal Microbiota and Small ncRNAs in Autism: Detection of miRNAs and piRNAs with Possible Implications in Host-Gut Microbiota Cross-Talk. *Nutrients* **2022**, *14*, 1340. [[CrossRef](#)] [[PubMed](#)]
58. Chen, Z.; Shi, K.; Liu, X.; Dai, Y.; Liu, Y.; Zhang, L.; Du, X.; Zhu, T.; Yu, J.; Fang, S.; et al. Gut microbial profile is associated with the severity of social impairment and IQ performance in children with autism spectrum disorder. *Front. Psychiatry* **2021**, *12*, 789864. [[CrossRef](#)] [[PubMed](#)]
59. Ha, S.; Oh, D.; Lee, S.; Park, J.; Ahn, J.; Choi, S.; Cheon, K.A. Altered Gut Microbiota in Korean Children with Autism Spectrum Disorders. *Nutrients* **2021**, *13*, 3300. [[CrossRef](#)]
60. Huang, M.; Liu, K.; Wei, Z.; Feng, Z.; Chen, J.; Yang, J.; Zhong, Q.; Wan, G.; Kong, X.J. Serum Oxytocin Level Correlates With Gut Microbiome Dysbiosis in Children With Autism Spectrum Disorder. *Front. Neurosci.* **2021**, *15*, 721884. [[CrossRef](#)] [[PubMed](#)]
61. Zhao, R.H.; Zheng, P.Y.; Liu, S.M.; Tang, Y.C.; Li, E.Y.; Sun, Z.Y.; Jiang, M.M. Correlation between gut microbiota and behavior symptoms in children with autism spectrum disorder. *Zhongguo Dang Dai Er Ke Za Zhi* **2019**, *21*, 663–669. [[CrossRef](#)]

62. Shi, K.; Zhang, L.; Yu, J.; Chen, Z.; Lai, S.; Zhao, X.; Li, W.G.; Luo, Q.; Lin, W.; Feng, J.; et al. A 12-genus bacterial signature identifies a group of severe autistic children with differential sensory behavior and brain structures. *Clin. Transl. Med.* **2021**, *11*, e314. [[CrossRef](#)]
63. Liu, T.; Pan, X.; Wang, X.; Feenstra, K.A.; Heringa, J.; Huang, Z. Predicting the relationships between gut microbiota and mental disorders with knowledge graphs. *Health Inf. Sci. Syst.* **2021**, *9*, 3. [[CrossRef](#)] [[PubMed](#)]
64. Viswanathan, M.; Patnode, C.D.; Berkman, N.D.; Bass, E.B.; Chang, S.; Hartling, L.; Murad, M.H.; Treadwell, J.R.; Kane, R.L. Assessing the risk of bias in systematic reviews of health care interventions. In *Methods Guide for Effectiveness and Comparative Effectiveness Reviews [Internet]*; Agency for Healthcare Research and Quality (US): Rockville, MD, USA, 2017.
65. Andreo-Martínez, P.; Ortiz-Martínez, V.M.; Salar-García, M.J.; Veiga-del-Baño, J.M.; Chica, A.; Quesada-Medina, J. Waste animal fats as feedstock for biodiesel production using non-catalytic supercritical alcohol transesterification: A perspective by the PRISMA methodology. *Energy Sustain. Dev.* **2022**, *69*, 150–163. [[CrossRef](#)]
66. Guillamón, E.; Andreo-Martínez, P.; Mut-Salud, N.; Fonollá, J.; Baños, A. Beneficial Effects of Organosulfur Compounds from *Allium cepa* on Gut Health: A Systematic Review. *Foods* **2021**, *10*, 1680. [[CrossRef](#)] [[PubMed](#)]
67. Ortiz-Martínez, V.M.; Andreo-Martínez, P.; García-Martínez, N.; Pérez de los Ríos, A.; Hernández-Fernández, F.J.; Quesada-Medina, J. Approach to biodiesel production from microalgae under supercritical conditions by the PRISMA method. *Fuel Process. Technol.* **2019**, *191*, 211–222. [[CrossRef](#)]
68. Castro, K.; Faccioli, L.S.; Baronio, D.; Gottfried, C.; Perry, I.S.; dos Santos Riesgo, R. Effect of a ketogenic diet on autism spectrum disorder: A systematic review. *Res. Autism. Spectr. Disord.* **2015**, *20*, 31–38. [[CrossRef](#)]
69. Martínez-González, A.E.; Piqueras, J.A. Differences in the severity of Autistic Spectrum Disorder symptoms according to the educational context. *Eur. J. Educ. Psychol.* **2019**, *12*, 153–164. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.