

Machine Learning Contribution to Autism Detection Based on Theory of Mind Tasks [†]

Efthymia Gkarneta ¹, Konstantinos-Filippos Kollias ^{1,*} , Alexandra Giola Genni ², Anna Triantafyllou ¹ ,
Panagiotis Sarigiannidis ¹  and George F. Fragulis ¹ 

¹ Department of Electrical & Computer Engineering, University of Western Macedonia, ZEP Campus, 50100 Kozani, Greece; ece01631@uowm.gr (E.G.); atriantafyllou@uowm.gr (A.T.); psarigiannidis@uowm.gr (P.S.); gfragulis@uowm.gr (G.F.F.)

² MetaMind Innovations P.C., Kila, 50100 Kozani, Greece; agenni@metamind.gr

* Correspondence: dece00063@uowm.gr

[†] Presented at the 8th International Global Conference Series on ICT Integration in Technical Education & Smart Society, Aizuwakamatsu City, Japan, 20–26 January 2026.

Abstract

Autism Spectrum Disorder (ASD) is a heterogeneous neurodevelopmental disorder marked by persistent social-communication difficulties and restricted behavioral patterns. Impairments in Theory of Mind (ToM), defined as the capacity to infer and interpret others' mental states, are widely recognized as central to the social challenges observed in ASD. Although ToM tasks have significantly advanced the theoretical understanding of socio-cognitive deficits, their application in early detection remains limited due to subjectivity, variability in behavioral responses, and time-intensive assessment procedures. This paper presents a theoretical examination of the contribution of Machine Learning (ML) to autism detection through data derived from ToM-based tasks. We argue that ML techniques can transform behavioral, eye-tracking, and neurophysiological responses into objective, multidimensional features, enabling the identification of discriminative patterns beyond conventional statistical analyses. By integrating socio-cognitive theory with computational modeling, we propose a conceptual framework in which ML-enhanced ToM paradigms support more accurate, scalable, and non-invasive approaches to ASD detection.

Keywords: machine learning; autism spectrum disorder; theory of mind; biomarker; classification

1. Introduction

Autism Spectrum Disorder (ASD) is a common and heterogeneous neurodevelopmental disorder characterized by deficits in cognitive, social, and communication functioning and restricted behaviors in children and adults [1]. These deficiencies have been reported to result in cascading consequences for the individual's social and professional integration [2]. Thus, there is a growing need for reliable and objective measures that can identify risk cases at an early stage.

In the DSM-5, core symptomatology is categorized into two domains: social communication and social interaction, and restrictive or repetitive patterns of behavior [3,4]. Despite the commonalities in symptoms, individuals on the spectrum may display a range of distinguishable behavioral and cognitive symptoms. That is to say, autistic people's abilities to learn, communicate, and reason range from highly proficient to significantly impaired [5]. For instance, certain individuals are well conversed, while others struggle



Academic Editors: Debopriyo Roy and Peter Ilic

Published: 8 July 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

with speech; some develop strong attachments, whereas others find it challenging to express emotions [6]; and some are independent in daily tasks, while others need substantial assistance for even basic activities [5]. In essence, it is generally accepted that there are behavioral, communicative, and linguistic manifestations that may vary in their degree of expression and development.

It has been postulated that the lack of Theory of Mind (ToM) may be a unique characteristic of autism, and it may further offer an explanation for difficulties in social interactions [2]. A widely accepted definition of ToM is the socio-cognitive ability to recognize and cognitively interpret postulated or observed mental states, such as opinions, intentions, and beliefs, independent of one's own [1]. The propositional attitudes involved in ToM distinguish its acquisition both conceptually and empirically from simpler social skills (e.g., general sociability). ToM tasks have been subject to several extensive studies and are potential targets for ASD identification due to the association with ToM deficits and core social dysfunction [7].

Although ToM tasks have substantially advanced our understanding of socio-cognitive differences in ASD, their role in classification aiming at detection remains limited and largely detached from computational approaches. At the same time, Machine Learning (ML) methods are increasingly employed in ASD detection, primarily relying on neuroimaging, behavioral screening scores, or physiological biomarkers. However, higher-order cognitive measures such as ToM are rarely integrated into these frameworks. This paper aims to theoretically address this gap by examining how ToM-derived features can contribute to more objective, scalable, reproducible, and theory-driven ML models for ASD detection.

2. Assessment of Autism Spectrum Disorder

Early ASD detection, through periodic developmental monitoring in infancy and toddlerhood, is preferable as it enables effective behavioral interventions in patients as young as 2 years old [8] and mitigates potential further developmental delays [9]. To expand on this, significant skill regression or plateauing has been noticed in the second year after the initial occurrence of autistic-like behaviors. In other words, timely assessment for children at risk is customary to assist in early access to intensive autism-specific services, which may result in significant gains in language, cognition, integration into the social, academic, and professional spheres, and overall improve the individual's quality of life [10].

At present, ASD detection relies on screening tools and practices described in classification systems, namely the DSM-5 [4] and the ICD-11 [11], since there are no universally established biophysiological tests for the disorder. Primary screening is conducted by child-care providers or pediatricians by means of informal counseling, access to screening tools, such as preliminary materials or structured interviews. General developmental screening instruments administered to toddlers identify delays in language, cognitive abilities, and motor skills, though they may not be sensitive enough to identify social symptoms associated with autism [3]. Secondary screening is performed in specialized clinics and involves systematic, professional observation and evaluation using specific screening tools and professional assessments.

2.1. Assessment Tools

Some standardized assessment instruments of high validity include the Screening Tool for Autism in Toddlers and Young Children (STAT), the Modified Checklist for Autism in Toddlers (M-CHAT), and the Autism Diagnostic Observation Schedule—Second Edition (ADOS-2).

The STAT is a 20-minute interactive observational assessment of toddlers aged 24 to 30 months. It involves 12 clinician-directed activities that evaluate abilities in areas such as

play, making requests, directing attention, and imitating adult actions [3,9,12]. Activities are scored as either “pass” with a score of 0, indicating that the behavior or skill is observed, or “fail” with a score of 0.25, 0.5, or 1, depending on the domain, meaning the behavior or skill is not observed.

The M-CHAT is a questionnaire-based assessment tool designed to identify risk cases for ASD [3]. It is administered to toddlers between 16 and 30 months, and it relies on parental observations of their behaviors. The checklist consists of 23 ‘yes/no’ items that cover a variety of developmental domains, as well as a parent interview to clarify responses and minimize the likelihood of false positives [13]. The Modified Checklist for Autism in Toddlers, Revised with Follow-Up (Questions) (M-CHAT-R/F) is an updated version that eliminates three items from the previous one [3] and incorporates a component for a professional review [13]. Children scoring eight or more are considered at high ASD risk and should be promptly evaluated further.

The ADOS-2 is a semi-structured screening tool that involves activity-based interactions and observations to evaluate ASD features [10,12]. It is frequently a component of both clinical research and actual assessments [3]. Clinicians use the data obtained from the test, as well as details about the individual’s peer interactions and further relevant history, to determine if the diagnostic criteria are met [3]. Comparison scores range on a scale from 1 to 10 and reflect different levels of impairment. The screening tool is suitable for individuals of all ages, as well as varying developmental levels and language abilities [3,10].

2.2. Challenges in Autism Assessment

Early ASD detection and intervention are challenging. First, there may be geographical and ethnic disparities in early ASD evaluation of children. It has been reported that children from rural or low-income backgrounds are more likely to disproportionately experience assessment delays. This issue arises from numerous factors. One explanation is found on the fact that screening services are often concentrated in metropolitan areas and require transportation or resources that may not be available to these families. The cost of care associated with ASD assessment, including developmental consultations and follow-up appointments, can exacerbate financial strain. Occasionally, the process may even extend over several months if the symptomology is not sufficiently evident, which contributes to the increased costs.

Cultural differences in behavioral norms, especially in areas such as eye contact or word preferences, can influence the emphasis that patients or healthcare professionals place on observed differences in these domains [14]. It has been noted that these variations have not yet been integrated into ASD assessment tools, thus resulting in a lack of cultural sensitivity among medical practitioners that can impact decision-making [14]. The cultural barriers may further discourage families from actively engaging in ASD research due to concerns regarding possible intrafamilial conflict and stigmatization.

In the context of gender, there are reports that autism is underdiagnosed in girls and women [15,16]. One hypothesis for this observation is that their expression of ASD does not meet the current diagnostic criteria that are derived mostly from male samples [16]. Namely, autistic girls are less likely to display overt restricted interests and more likely to present their difficulties as passive manifestations, which would reduce the probability of them receiving an accurate ASD evaluation [16]. Furthermore, research has shown that women are more prone to use conscious or unconscious masking strategies to hide their autistic difficulties during a social setting by mimicking facial expressions or forcing themselves to maintain eye contact [15]. This process is colloquially known as “camouflaging” or simply “masking”.

Beyond the issues of health equity, ASD detection is complicated by its dependence on standardized screening instruments within medical and behavioral disciplines. Since there is no quantifiable technology available yet to specifically address autism, the screening tools require clinical judgment from an ASD expert to secure an accurate assessment [10], which can introduce variability based on the evaluator's experience and expertise. In other words, assessment relies heavily on empirical judgment. Even when valid and reliable tools are employed, the sensitivity of assessment methods is rarely, if ever, perfect, leading to some positive cases being unidentified.

Overall, full confidence in ASD assessment occurs when symptoms are stable, and the overt syndrome is clearly displayed; however, delaying evaluation risks missing the optimal time for intervention and can lead to further financial burdens on the family.

3. Theory of Mind

The fundamental socio-cognitive skills associated with ToM underline a person's competence to build social relationships from an early age and foster prosocial behaviors. For instance, parents often infer their child's feelings of unfairness to provide alternative points of view and encourage understanding; friends recognize each other's emotional states, such as distress through tone or body language, to provide appropriate support; teachers assess a student's confusion during lessons to offer clarifications or assistance and guide their learning. Even in the arts, such as literature, cinema, or theater, mentalizing abilities implicitly motivate people to attribute characters' mental states and understand motivations, streams of consciousness, irony or deception, and different points of view in the narrative. Engagement with fiction prompts individuals to immerse themselves in characters' experiences and reflect on diverse perspectives, sometimes even gaining new insights into behavior and social dynamics, which are closely related to developed ToM.

3.1. Theory of Mind Development in Childhood

It has been assumed that ToM's development begins to occur around the age of 4, involving deliberate reasoning processes and cognitive functions [2]. However, some researchers argue that even infants possess fundamental ToM abilities, and they sequentially develop with age and experience in social interactions throughout one's lifespan [17]. They propose that at 14 months, neurotypical infants exhibit shared attention by noticing objects and reactions. By age 2, they engage in symbolic play and imitation. At age 3, neurotypical children establish basic emotion comprehension, passing the first ToM tests and indicating ToM growth. By age 4, they begin to grasp the concept of incorrect belief content, suggesting an evolutionary cognitive development. From ages 5 to 8, children reason about deception and lies, thus refining their understanding through experience. By age 9, they exhibit good performance on ToM tasks related to recognizing social cues and their effects on others, which demonstrates advanced ToM development [1] (Figure 1).

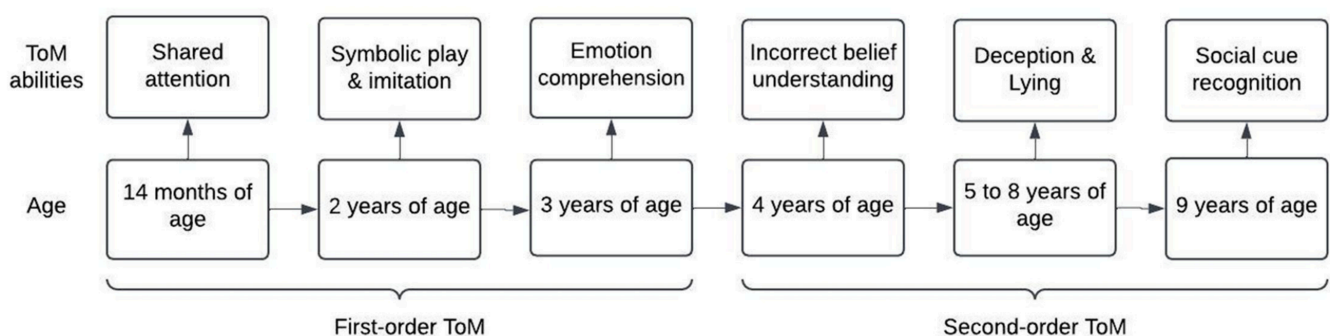


Figure 1. Chronological milestones in development of ToM abilities.

Nevertheless, some academics argue that the assumption that adults fully acquire ToM remains debatable, as they often make mistakes when attempting to understand others' mental states [18], which indicates that, in general, ToM may be more context-dependent than previously assumed.

3.2. Theory of Mind in Autism Spectrum Disorder

The capacity to understand what others think and feel remains one of the most elusive mental faculties. As suggested by previous research, ToM development in neurodivergent children has been linked to difficulties in imputing mental and emotional states [1]. Autistic children, in particular, may have impaired or even delayed ToM development compared to their neurotypical counterparts [2,19].

Consistent with this, it has been proposed that autistic children find it challenging to interpret the mental states of others. To expand on this, children with ASD can recognize and express basic emotions, such as happiness and sadness; however, they have trouble comprehending and communicating more cognitive emotions, such as embarrassment and surprise [1].

It is also commonly believed that children on the spectrum are less prone to initiate joint attention and may exhibit atypical gaze-leading behaviors [20]. In addition, they often demonstrate significantly poorer attention spans and cognitive sensitivities, struggle with interpreting social cues, and exhibit lower levels of cognitive empathy compared to the general population. This is hypothesized to be largely due to their underdeveloped social cognition and emotional awareness [1].

3.3. Theory of Mind Tasks

There is a variety of ToM tools and methods used to evaluate cognitive or affective abilities. The most commonly used are the Faux Pas Recognition Test (FPRT), the Sally-Anne False Belief Test (FBT), the Theory of Mind Task Battery (ToM-TB), and the Theory of Mind Inventory-2 (ToMI-2), which are presented in Table 1.

Table 1. Overview of Theory of Mind tools.

Tool	Description	Components Assessed
FPRT	Verbally examined test that assesses comprehension of faux pas in daily social scenarios	Cognitive and affective mental states involved in recognizing faux pas
FBT	Measures understanding of false beliefs through the Sally-Anne story	First- and second-order false beliefs, understanding of others' mental states
ToM-TB	List of test questions that evaluate ToM through a series of scenarios in a story-book format	Emotional recognition, perspective-taking, action, and comprehension of false beliefs
ToMI-2	Caregiver-report questionnaire that measures basic to advanced ToM functioning and competence	Thought and emotion recognition, false beliefs, sarcasm, and pragmatics

First, the FPRT is widely used to evaluate ToM abilities in children with ASD, with a focus on their comprehension of social nuances, as measured through faux pas detection [21]. A faux pas is an unintentional act or remark that proves to be incorrect or socially inappropriate within the context of the interaction [22]. In the FPRT, individuals examine verbally presented stories involving a faux pas event, which describes interpersonal interactions in daily life situations and then answer questions regarding the thoughts and emotions of the person affected in each narrative. The test comprises 20 stories in total, with 10 being control stories and 10 being faux pas stories [21]. The participants collect one point for each correct answer to the questions asked, and they obtain a collective score named the Faux Pas Score (FPS) [21,22].

Second, the FBT is a measure designed to assess an individual's understanding of false beliefs in others [19]. Typically, in FBT, a child is presented with the story of Sally and Anne, in which Sally hides a marble in a basket before leaving and in her absence, Anne moves that marble to a box nearby. The child is then questioned about where Sally will look for her marble when she returns. Recent ToM studies suggest that children instinctively attribute incorrect beliefs to others even before they pass the explicit false belief test. This test assesses not only the first-order belief, which is understanding another person's mental state, but also the second-order belief, which involves knowing why a third party holds a specific belief. In other words, the false belief test evaluates both the comprehension of another's psychological state and the reasoning behind a third party's belief [1].

Third, the ToM-TB is a reference tool that directly evaluates a child's comprehension of a series of ToM scenarios. It is a list of 15 test questions arranged in ascending order of difficulty in the three subscales (i.e., early, basic, and advanced) [19,23], embedded within nine stories [7]. The task is presented as brief vignettes within a story-book format, with colorful illustrations and a corresponding test on each page [23], so that children with limited receptive language abilities can be assessed as well [7]. The tasks test emotion recognition, visual perspective-taking, desire-based emotion, perception-based belief, action, and comprehension of first- and second-order false beliefs [19]. Each question is scored as either 1 (pass) or 0 (fail) and has four available answers, which reduces the probability of finding the correct one by random chance to 25%. The ToM-TB includes 11 additional memory control questions, which are unrelated to ToM, to ensure that children who struggle with the ToM questions have adequate comprehension of the story presented to them [7,19].

Last, the ToMI-2 is a caregiver-report tool designed to measure a child's ToM functioning and applied competence [23]. Presented in a questionnaire format [19], the ToMI-2 includes 60 items that evaluate particular ToM aspects, ranging from basic to advanced skills, such as basic recognition of others' thoughts to more complex comprehension of hidden emotions, false beliefs, irony and sarcasm. Each item is rated on a standard scoring system, where the respondents (e.g., a parent) mark their answer with a vertical line on a continuous scale with endpoints labeled "Definitely Not" and "Definitely" [19,23]. Scores for individual items and subscales lie between 0 and 20, with higher values reflecting greater parental confidence in their child's ToM abilities [23].

3.4. Limitations of Theory of Mind Tasks

The aforementioned methods are widely used to assist in ASD assessment. However, the psychometric properties of ToM assessment tools remain a subject of considerable discussion [7]. In recent years, numerous researchers have drawn attention to certain limitations that should be addressed.

First, one concern regards ecological validity, since ToM tasks are often presented as simplified scenarios designed for monitoring in controlled environments. In particular, several ToM tasks do not focus on social interactions [18,19] and often rely simply on passive observation rather than active engagement. This approach can potentially compromise the assessment process due to observational bias.

Second, it has been noted that ToM performance in older people can be influenced by aspects that emerge from diverse life experiences [18] and individual differences in mentalizing abilities across adult populations. However, current research is primarily child-centric. These points point to a gap in tasks designed for older participants that can explore and evaluate ToM in practical contexts [18]. As a result, existing tasks often demonstrate insufficient range in behavioral performance since they may not adequately take experiential differences into consideration. To look deeper into the issue of adult applicability, social cognition research for autistic adults has primarily adapted methods

originally designed for minors. This approach often results in tasks that are insufficiently sensitive and limited in their effectiveness.

4. Machine Learning

In recent years, medicine and behavioral sciences disciplines have witnessed the emergence of advanced Artificial Intelligence (AI) tools and algorithms with high potential to assist healthcare professionals in decision-making. ML algorithms are increasingly recognized as effective tools in healthcare practice due to their ability to rapidly extract patterns from large-scale datasets and support precision interventions. In this context, emerging Digital Twin frameworks are being explored as promising approaches for personalized autism intervention [24]. Further, in the context of ASD prediction and classification research, there have been several studies that have included the application of intelligent ML algorithms using biomarkers from various modalities.

4.1. Machine Learning Models

First, using extracted features derived from functional connectivity metrics, Support Vector Machine (SVM) has been used to classify ASD with high accuracy [23], in addition to demonstrating an accuracy of 98.77% in early detection scenarios for toddlers [25].

Second, Decision Tree (DT) models have shown low error rates in their predictions and high positive rates, thus effective performance. To expand on this, recent multi-classifier experiments have demonstrated that Random Forest (RF) and DT methods specifically outperform other models in accuracy [26].

Third, researchers have explored the K-Nearest Neighbors (KNN) algorithm for ASD classification, which demonstrated notable performance, producing the highest accuracy of 87.143% among the evaluated models [27].

Fourth, experiments have demonstrated that Logistic Regression (LR) is highly effective for the classification of ASD in children and adolescents. For instance, LR was found to achieve a high average F1 score of 0.97, indicating great model performance and effectively balancing precision and sensitivity [28].

Fifth, Deep Learning (DL) models, such as the Auto-ASD-Network [29], have achieved more than 70% accuracy on four different datasets and increased performance up to 26%, compared to traditional methods, with a maximum accuracy of 80%.

Sixth, Federated Learning (FL) is a recent focus in the healthcare field, with a growing number of researchers starting to evaluate its efficiency for timely, early-stage ASD detection [30]. To be precise, FL combined with locally trained ML models, such as SVM and LR, used on multiple datasets reached 81% accuracy in detecting the disorder in adults with the SVM model and 98% accuracy in children with the LR model [30].

4.2. Biomarkers

Predicting outcomes and forecasting treatment responses in the clinical field often requires a reflection of the underlying biological processes or mental states. Hence, researchers have proposed the integration of scalable biomarkers, i.e., objectively measured indicators that represent those biological processes [3], with traditional evaluation techniques to overcome weaknesses in assessments. In this context, instruments with the potential to study behavioral biomarkers related to ASD have been the target of several extensive studies.

First, Electroencephalography (EEG) is a non-invasive technique used to examine brain function with high time resolution. In relation to ASD, electrophysiologic studies suggest differences in auditory and visual processing, recognition memory, and neural connectivity in ASD [3]. Research shows that the EEG signals of autistic children differ significantly

from those of their neurotypical peers, with a substantial decline in EEG complexity [6]. In fact, it has been proposed that EEG-based assessments of early brain development can correctly identify signs of ASD even before behavioral symptoms emerge [31].

Moreover, Functional Magnetic Resonance Imaging (fMRI) is a neuroimaging method that observes changes in morphology and activation patterns in the brain. Studies report abnormalities in brain volume and cortical thickness in people with ASD [23], while resting-state fMRI has further shown distinct anterior–posterior patterns consistent with reduced connectivity in the autistic brain [5].

Furthermore, abnormalities in gaze patterns have been consistently linked with autism [32–34]. Individuals with ASD show difficulties prioritizing biological motion and further impairments in global visual processing, with reduced focus on the eye region in early infancy [3], and deviations from the typical “left visual field bias” observed in neurotypical face perception.

Virtual Reality (VR) is an additional emerging technology capable of detecting and classifying potential biomarkers, such as body movements [35], with VR-derived features assessing cognitive impairment through behavioral data. The disorder is typically characterized by exaggerated or repetitive behaviors, such as head spinning, arm flapping, body rocking, feet stamping, and finger wiggling [35], and studies show that autistic children exhibit more pronounced and varied movements during imitation tasks, with ML models confirming these differences.

5. Integration of ToM Measures in Machine Learning Frameworks

While numerous ML approaches have been proposed for ASD classification, the majority of them rely on neuroimaging, behavioral screening scores, or physiological biomarkers, with limited attention to higher-order socio-cognitive constructs. In particular, the integration of ToM-derived measures into ML-based detection frameworks remains underexplored. This section examines the existing studies, to the best of current knowledge, that have incorporated such features and demonstrates the conceptual and methodological implications of introducing ToM features into ML models.

For example, a study investigated alterations in brain connectivity associated with autism by analyzing causal influences of effective connectivity [36]. The participants included 15 individuals with ASD (mean age: 21.14 years) and 15 age-and-IQ-matched controls (mean age: 22.18 years) who interpreted black-and-white comic strip vignettes involving physical and intentional causality while undergoing fMRI. Using the regions of interest (ROIs) that were identified, such as the temporal parietal junction, the inferior frontal gyrus, and the middle temporal gyrus, the study applied a multivariate autoregressive model to capture Granger causal relationships. The study leveraged connectivity metrics, assessment scores, functional connectivity values, and fractional anisotropy in a Recursive Cluster Elimination SVM framework, progressively eliminating the low-scoring clusters. The model achieved 95.9% accuracy (specificity: 94.8%, sensitivity: 96.9%), with effective connectivity path weights emerging as the most discriminative features. These findings underscore that brain regions associated with mentalizing processes can serve as highly discriminative features in ML classification, which indirectly supports the relevance of ToM-related neural mechanisms in ASD detection.

Another exception is presented in [23], which included behavioral features from the ToMI-2 and ToM-TB assessments into its ML framework. The dataset consisted of behavioral and neuroimaging measurements from 28 children aged 7–14 years. The study used a novel Conjunctive Clause Evolutionary Algorithm to produce a parsimonious model by searching for combinations of features and their corresponding value ranges, therefore, detecting feature interactions even in the absence of strong individual predictors. From 2438

conjunctive clauses, the eight best second-order models were selected, which further identified neuroanatomical features as potential biomarker candidates, including brain volume, area, cortical thickness, and mean curvature in specific brain regions. The analysis extended to third-order models, which incorporated ToM-related behavioral features and validation with a separate KNN classifier, showing accuracies of 89.29% for second-order features, 78.57% for third-order features, and 85.71% when combining both. Additional testing using the same features demonstrated that the second-order model achieved 87.5% accuracy, the third-order 81.25%, and the combined model 93.75%. Notably, these performances indicate that the inclusion of ToM behavioral measures alongside neuroanatomical features improved classification performance, which suggests that higher-order socio-cognitive variables may enhance model sensitivity beyond purely structural biomarkers.

Collectively, these findings suggest that although ToM-related features are rarely incorporated into ML-based ASD classification models, preliminary evidence indicates that their inclusion may enhance discriminative performance. The shortage of such studies highlights a significant gap in the literature and underscores the need for frameworks that explicitly integrate socio-cognitive theory with computational biomarker development. Addressing this gap could overall contribute to more theoretically grounded and clinically informative ML models.

6. Challenges and Limitations

With respect to the attributes of the samples, age and gender disparities [37] reflect potential biases that can affect the outcomes of ML models. Such imbalances may lead to underrepresented data patterns or incorrect results. It is important to further note that datasets for classification models based on ToM tasks are small and include limited feature representation. Most of the studies are single-case designs in pilot stages and lack large-scale research. As a result, their effect size is difficult to evaluate [38].

Data acquisition and integration are especially challenging as healthcare institutions are hesitant to share patient records due to internal policies and data protection legislations [30]. Furthermore, the process of collecting classification data can be disrupted by sensory sensitivities and behavioral characteristics commonly seen in people with ASD. To expand on this, gathering accurate data from tools, such as neuroimaging scanners, is challenging, given the distracting noises and the potential difficulty for the autistic individuals to remain motionless during the scans [23].

7. Future Directions

Future research directions in ML are promising with the integration of additional diverse data modalities. However, to foster effective innovation in mental health through AI, it is crucial to conduct extensive analyses of the moral implications to pinpoint areas of concern [38]. AI's ethical consequences are shaped by the social practices and norms surrounding its deployment, rather than just the computational algorithms themselves. Thus, achieving ethical AI requires developing a strategy that includes improving both the technology and the organizational practices that govern its use.

In the context of autism classification, transparency, fairness, patient autonomy, and accountability are foundational imperatives to ensure quality care. The early identification of ethical concerns helps researchers to integrate them into the design and production of future intelligent software for medical applications [38]. Therefore, it is essential for practitioners to consider all four principles of medical ethics: autonomy, beneficence, nonmaleficence, and justice, throughout all aspects of care [39].

8. Conclusions

ML demonstrates significant potential in enhancing ASD detection when integrated with ToM-based measures. While ToM tasks have long provided valuable theoretical insights into socio-cognitive differences in ASD, their application in detection remains constrained by subjectivity, limited ecological validity, and variability across age and gender. The reviewed evidence suggests that classification performance was consistently higher when ToM features were combined with established neurobiological biomarkers such as EEG, fMRI, and eye-tracking data.

Although initial results are promising, research at the intersection of ToM and ML-driven ASD classification remains limited and methodologically heterogeneous. Small sample sizes, dataset imbalances, and a lack of standardized protocols and feature extraction strategies restrict generalizability and cross-study comparisons. Future large-scale, multimodal, and ethically and culturally grounded investigations are necessary to validate the clinical applicability of ML-enhanced ToM frameworks. Ultimately, integrating socio-cognitive theory with computational modeling may contribute to more objective, scalable, reliable, and early ASD detection strategies.

Author Contributions: Conceptualization, E.G., K.-F.K., A.G.G., A.T., P.S. and G.F.F.; methodology, E.G.; validation, E.G., K.-F.K., A.G.G., A.T., P.S. and G.F.F.; formal analysis, E.G.; investigation, E.G. and K.-F.K.; resources, E.G. and K.-F.K.; data curation, E.G. and K.-F.K.; writing—original draft preparation, E.G., K.-F.K. and G.F.F.; writing—review and editing, E.G., K.-F.K. and G.F.F.; visualization, E.G.; supervision, K.-F.K. and G.F.F.; project administration, K.-F.K. and G.F.F.; funding acquisition, A.G.G., A.T. and P.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No. 101178789 (EVOSST). Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created.

Conflicts of Interest: Author Alexandra Giola Genni was employed by the company MetaMind Innovations P.C. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Zong, Z. Analyze social behavior in autistic children by using theory of mind. *Lect. Notes Educ. Psychol. Public Media* **2023**, *6*, 564–575. [[CrossRef](#)]
2. Czajeczny, D.; Jaroszkiewicz, A.; Daroszewski, P.; Kopczyński, P.; Warchoł-Biedermann, K.; Pigłowska, A.; Samborski, W.; Wójciak, R.W.; Mojs, E. Application of Eye-tracking in research on the theory of mind in ASD. *Eur. Rev. Med. Pharmacol. Sci.* **2022**, *26*, 1364–1373. [[CrossRef](#)] [[PubMed](#)]
3. Hyman, S.L.; Levy, S.E.; Myers, S.M.; Council on Children with Disabilities, Section on Developmental and Behavioral Pediatrics; Kuo, D.Z.; Apkon, S.; Davidson, L.F.; Ellerbeck, K.A.; Foster, J.E.A.; Noritz, G.H.; et al. Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics* **2020**, *145*, e20193447. [[CrossRef](#)] [[PubMed](#)]
4. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR)*, 5th ed.; American Psychiatric Publishing: Arlington, VA, USA, 2022.
5. Chola Raja, S.K.K. Conditional Generative Adversarial Network Approach for Autism Prediction. *Comput. Syst. Sci. Eng.* **2023**, *44*, 741–755. [[CrossRef](#)]

6. Ali, N.A.; Syafeeza, A.R.; Jaafar, A.S.; Mohd Fitri Alif, M.K. Autism spectrum disorder classification on electroencephalogram signal using deep learning algorithm. *IAES Int. J. Artif. Intell. (IJ-AI)* **2020**, *9*, 91–99. [\[CrossRef\]](#)
7. Gosling, C.J.; Cartigny, A.; Stevanovic, D.; Moutier, S.; Delorme, R.; Attwood, T. Known-groups and convergent validity of the theory of mind task battery in children with autism spectrum disorder. *Br. J. Clin. Psychol.* **2023**, *62*, 525–535. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Singh, U.; Shukla, S.; Gore, M.M. Detection of autism spectrum disorder using multi-scale enhanced graph Convolutional Network. *Cogn. Comput. Syst.* **2024**, *6*, 12–25. [\[CrossRef\]](#)
9. Tagavi, D.M.; Dick, C.C.; Attar, S.M.; Ibanez, L.V.; Stone, W.L. The implementation of the screening tool for autism in toddlers in Part C early intervention programs: An 18-month follow-up. *Autism* **2022**, *27*, 173–187. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Briguglio, M.; Turriziani, L.; Currò, A.; Gagliano, A.; Di Rosa, G.; Caccamo, D.; Tonacci, A.; Gangemi, S. A machine learning approach to the diagnosis of autism spectrum disorder and multi-systemic developmental disorder based on retrospective data and Ados-2 score. *Brain Sci.* **2023**, *13*, 883. [\[CrossRef\]](#) [\[PubMed\]](#)
11. World Health Organization. *International Classification of Diseases, Eleventh Revision (ICD-11)*, 11th ed.; World Health Organization: Geneva, Switzerland, 2022.
12. Corona, L.L.; Wagner, L.; Hooper, M.; Weitlauf, A.; Foster, T.E.; Hine, J.; Miceli, A.; Nicholson, A.; Stone, C.; Vehorn, A.; et al. A Randomized Trial of the Accuracy of Novel Telehealth Instruments for the Assessment of Autism in Toddlers. *J. Autism Dev. Disord.* **2024**, *54*, 2069–2080. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Lordan, R.; Storni, C.; De Benedictis, C.A. Autism spectrum disorders: Diagnosis and Treatment. In *Autism Spectrum Disorders*; Exon Publications: Brisbane, Australia, 2021; pp. 17–32. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Tromans, S.J.; Chester, V.; Gemegah, E.; Roberts, K.; Morgan, Z.; Yao, G.; Brugha, T. Autism Identification across Ethnic Groups: A Narrative Review. *Adv. Autism* **2020**, *7*, 241–255. [\[CrossRef\]](#)
15. Hull, L.; Petrides, K.V.; Mandy, W. The female autism phenotype and camouflaging: A narrative review. *Rev. J. Autism Dev. Disord.* **2020**, *7*, 306–317. [\[CrossRef\]](#)
16. Leedham, A.; Thompson, A.R.; Smith, R.; Freeth, M. 'I was exhausted trying to figure it out': The experiences of females receiving an autism diagnosis in middle to late adulthood. *Autism* **2020**, *24*, 135–146. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Kulke, L.; Johannsen, J.; Rakoczy, H. Why can some implicit theory of mind tasks be replicated and others cannot? A test of mentalizing versus submentalizing accounts. *PLoS ONE* **2019**, *14*, e0213772. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Feng, S.C.; Luk, G. Assessing Theory of Mind in Bilinguals: A Scoping Review on Tasks and Study Designs. *Biling. Lang. Cogn.* **2023**, *27*, 531–545. [\[CrossRef\]](#)
19. Chen, K.-L.; Jiang, D.-R.; Yu, Y.-T.; Lee, Y.-C. Development and psychometric evidence of the Chinese Version of the Theory of Mind Inventory-2 (ToMI-2) in children with autism spectrum disorder. *Res. Autism Spectr. Disord.* **2023**, *103*, 102132. [\[CrossRef\]](#)
20. Stephenson, L.J.; Edwards, S.G.; Bayliss, A.P. From Gaze Perception to Social Cognition: The Shared-Attention System. *Perspect. Psychol. Sci. J. Assoc. Psychol. Sci.* **2021**, *16*, 553–576. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Şandor, S.; İşcen, P. Faux-pas recognition test: A Turkish adaptation study and a proposal of a standardized short version. *Appl. Neuropsychol. Adult* **2021**, *30*, 34–42. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Garcia-Molina, I.; Clemente-Estevan, R.A. Autism and faux pas. influences of presentation modality and working memory. *Span. J. Psychol.* **2019**, *22*, E13. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Han, Y.; Rizzo, D.M.; Hanley, J.P.; Coderre, E.L.; Prelock, P.A. Identifying neuroanatomical and behavioral features for autism spectrum disorder diagnosis in children using machine learning. *PLoS ONE* **2022**, *17*, e0269773. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Kollias, K.-F.; Moysis, L.; Siniosoglou, I.; Argyriou, V.; Koumboulis, F.N.; Sarigiannidis, P.; Fragulis, G.F. A Digital Twin Design for Autism Intervention. In *Proceedings of the Horizons of AI: Ethical Considerations and Interdisciplinary Engagements*; Farmanbar, M., Tzamtzi, M., Schoeffmann, K., Kouvakas, N., Verma, A.K., Eds.; Springer Nature: Singapore, 2025; pp. 583–595.
25. Akter, T.; Satu, S.; Khan, I.; Ali, M.H.; Uddin, S.; Lio, P.; Quinn, J.M.W.; Moni, M.A. Machine learning-based models for early stage detection of autism spectrum disorders. *IEEE Access* **2019**, *7*, 166509–166527. [\[CrossRef\]](#)
26. Shinde, A.V.; Patil, D.D. A Multi-Classifer-Based Recommender System for Early Autism Spectrum Disorder Detection using Machine Learning. *Healthc. Anal.* **2023**, *4*, 100211. [\[CrossRef\]](#)
27. Bousidrah, N.; Khamees, Z.; Ali, S. Detection of Autism Spectrum Disorder by a Case Study Model Using Machine Learning Techniques An Experimental Analysis on Child, Adolescent and Datasets. *Sci. J. Univ. Benghazi* **2024**, *37*. [\[CrossRef\]](#)
28. Zheng, Y.; Deng, T.; Wang, Y. Autism classification based on logistic regression model. In *Proceedings of the 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, Nanchang, China, 26–28 March 2021.
29. Eslami, T.; Saeed, F. Auto-ASD-Network: A Technique Based on Deep Learning and Support Vector Machines for Diagnosing Autism Spectrum Disorder using fMRI Data. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, Niagara Falls, NY, USA, 7–10 September 2019; pp. 646–651.
30. Farooq, M.S.; Tahseen, R.; Sabir, M.; Atal, Z. Detection of autism spectrum disorder (ASD) in children and adults using machine learning. *Sci. Rep.* **2023**, *13*, 9605. [\[CrossRef\]](#) [\[PubMed\]](#)

31. Alhassan, S.; Soudani, A.; Almusallam, M. Energy-efficient EEG-based scheme for Autism Spectrum Disorder Detection Using wearable sensors. *Sensors* **2023**, *23*, 2228. [[CrossRef](#)] [[PubMed](#)]
32. Kollias, K.-F.; Syriopoulou-Delli, C.K.; Sarigiannidis, P.; Fragulis, G.F. Autism detection in High-Functioning Adults with the application of Eye-Tracking technology and Machine Learning. In Proceedings of the 2022 11th International Conference on Modern Circuits and Systems Technologies (MOCAST), Bremen, Germany, 8–10 June 2022; pp. 1–4.
33. Kollias, K.-F.; Maraslidis, G.S.; Sarigiannidis, P.; Fragulis, G.F. Application of machine learning on eye-tracking data for autism detection: The case of high-functioning adults. In *Proceedings of the AIP Conference Proceedings*; AIP Publishing LLC: Melville, NY, USA, 2024; Volume 3220, p. 050012.
34. Kaloforidis, N.; Kollias, K.-F.; Radoglou-Grammatikis, P.; Sarigiannidis, P.; Fragulis, G.F. Autism Spectrum Disorder Classification in Children Using Eye-Tracking Data and Machine Learning. *Eng. Proc.* **2025**, *107*, 12. [[CrossRef](#)]
35. Alcañiz Raya, M.; Marín-Morales, J.; Minissi, M.E.; Garcia, G.T.; Abad, L.; Giglioli, I.A.C. Machine Learning and virtual reality on body movements' behaviors to classify children with autism spectrum disorder. *J. Clin. Med.* **2020**, *9*, 1260. [[CrossRef](#)] [[PubMed](#)]
36. Deshpande, G.; Libero, L.; Sreenivasan, K.R.; Deshpande, H.; Kana, R.K. Identification of neural connectivity signatures of autism using machine learning. *Front. Hum. Neurosci.* **2013**, *7*, 670. [[CrossRef](#)] [[PubMed](#)]
37. Gao, S.; Wang, X.; Su, Y. Examining whether adults with autism spectrum disorder encounter multiple problems in theory of mind: A study based on meta-analysis. *Psychon. Bull. Rev.* **2023**, *30*, 1740–1758. [[CrossRef](#)] [[PubMed](#)]
38. Fiske, A.; Henningsen, P.; Buyx, A. Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. *J. Med. Internet Res.* **2019**, *21*, e13216. [[CrossRef](#)] [[PubMed](#)]
39. Farhud, D.D.; Zokaei, S. Ethical issues of Artificial Intelligence in medicine and Healthcare. *Iran. J. Public Health* **2021**, *50*, 1–5. [[CrossRef](#)] [[PubMed](#)]

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.