Università di Catania

Dipartimento di Fisica e Astronomia "Ettore Majorana"

PHD IN COMPLEX SYSTEMS FOR PHYSICAL, SOCIO-ECONOMICAL AND LIFE SCIENCES
CYCLE XXXVIII
Coordinator: Prof. Andrea Rapisarda

Lara Cirnigliaro

Design of a Machine Learning-based classifier for enhancing the accuracy and applicability of DAES, a novel autism screening tool

---

PhD THESIS

Tutor:
Prof.ssa Rita Barone

Relators:
Prof.ssa Rita Barone
Prof. Antonio Di Maria

Academic year 2024/2025

CONTENTS

1. ABSTRACT.....1

2. INTRODUCTION.....2

2.1 Nosographic and epidemiological data of Autism Spectrum Disorder.....2

2.2 Diagnosis and outcome of ASD.....5

2.3 Screening tools for ASD: advances and limitations.....8

2.4 Developmental trajectories of ASD and assessment of the evolutive profile in ASD-risk children.....12

2.5 A novel screening tool for autism: DAES (Developmental Autism Early Screening).....14

2.5.a DAES development.....14

2.5.b DAES description and administration.....16

2.5.c DAES scoring.....16

2.5.d DAES validation analyses.....18

2.5.e DAES strengths and future perspectives.....20

2.6 Machine learning in ASD screening and early diagnosis.....21

3. AIMS.....25

4. MATERIAL AND METHODS.....26

4.1. Participants.....26

4.2 Data selection and pre-processing.....27

4.3 Cramer's V index computing.....29

4.4 Feature selection and filtering.....29

4.5 Machine learning classifier tuning.....30

4.6 Metric for ML performance.....32

4.7 Standardized measures.....35

5. RESULTS.....38

5.1 Data preparation.....38

5.2 Feature selection.....39

5.3 ML classification models.....41

6. DISCUSSION.....46

6.1 Study limitations and future perspectives.....51

7. CONCLUSIONS.....46

8. REFERENCES.....53

1. ABSTRACT

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterised by a wide range of clinical manifestations and developmental trajectories, which makes the diagnostic process complex and often challenging. Earlier identification of ASD enables earlier intervention and better outcomes. There is an increasing demand for reliable ASD screening tools that are easy and quick to administer. Recently, machine learning (ML) has been applied to improve the classification performance of first level screening tools, mainly parent-report questionnaires. In this study we used ML to improve the accuracy and applicability of a novel observational and interactive ASD screener, the Developmental Autism Early Screening (DAES). We previously developed this tool based on the Griffiths Scales of Child Development-third edition (Griffiths III) to intercept the early atypical developmental patterns in children with ASD risk, differentiating them from children with global developmental delay (DD) or neurotypical development (TD) in the first three years of life. In this study we explored and selected the potentially discriminative DAES items at two target age ranges (≤ 24 months and > 24 months) in a large sample of ASD, DD and TD children aged 12-48 months (n=610). We trained and tested five ML classifiers (random forest, RF, support vector classifier, SVC, decision tree, DT, logistic regression, LR, k-nearest neighbors, KNN) to classify ASD versus DD and TD children at the two different age ranges. RF and SVC were the two most effective algorithms for correctly detecting ASD children achieving very high accuracy (above 98%) from selections of 21 and 28 items (out of 36 DAES items) for children aged ≤ 24 and > 24 months, respectively. These findings confirm the validity of a shorter and faster version of the DAES using predictive items for specific age ranges. A widespread use of the tool will facilitate an earlier access to targeted intervention, allowing to redirect the atypical developmental trajectory towards a typical pathway in a greater number of children at risk of ASD.

2. INTRODUCTION

2.1 Nosographic and epidemiological data on Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental lifelong condition characterized by persistent social communication and interaction impairment, and restricted/repetitive interests and stereotypic patterns of behaviour [Lord et al., 2020]. According to the latest versions of the most frequently used classification systems in mental health (the Diagnostic and Statistical Manual of Mental Disorders – 5- Text Revision, DSM-5-TR, of the American Psychiatric Association, APA, and the International Classification of Diseases, ICD-11, of the World Health Organisation, WHO), ASD is classified within the broader category of "neurodevelopment disorders" with onset of symptoms usually during the early years of life [APA, 2022; WHO, 2021]. Both the DSM-5 and the ICD-11 aimed to make the diagnosis of ASD more straightforward in clinical practice, by referring to ASD as a unified set of core symptoms, including alterations in two domains: 1) social communication/social interactions and 2) restricted, repetitive or unusual sensory-motor behaviours (Table 1).

DSM-5 [APA, 2013] marks the shift from DSM-IV multi-categorial diagnostic system to a single diagnosis based on multiple dimensions, broadening the construct of autism as a "spectrum". This evolution reflects efforts to enhance the diagnostic specificity and sensitivity. The attribute of "Spectrum" indicates a wide range of clinical variability and the heterogeneity of the natural history of the disorder, with different levels of symptomatology severity based on the level of support needed for everyday life functioning [Rosen et al., 2021]. The two classification systems (DSM-5-TR and ICD-11) differ in their approaches to ASD conceptualization. The DSM-5-TR diagnostic criteria are more oriented toward a medical model of the disorder, specifying the number

of required observable behavioral symptoms and providing descriptions of severity levels. The ICD-11 gives more emphasis on a social model of disability, focusing on the inner experience of "diversity" and the discrepancy between individuals' characteristics and environmental demands [Cortese et al., 2025]. The conceptualization of autism in a dimensional view highlights the need for diagnostic approaches that intercept specific clinical features of each individual, in order to plan personalized management strategies [Kamp-Becker, 2024].

Table 1. Summary of DSM 5-TR and ICD-11 diagnostic criteria for Autism Spectrum Disorder.

DSM 5-TR	ICD-11
A. ***Persistent deficits in social communication and social interaction across multiple contexts, as manifested by all of the following:*** *1. Deficits in social-emotional reciprocity* *2. Deficits in nonverbal communicative behaviors used for social interaction* *3. Deficits in developing, maintaining, and understanding relationships*	- ***Persistent deficits in initiating and sustaining social communication and reciprocal social interactions*** *that are outside the expected range of typical functioning given the individual's age and level of intellectual development.* *Specific manifestations of these deficits vary according to chronological age, verbal and intellectual ability, and disorder severity.*
B. ***Restricted, repetitive patterns of behavior, interests, or activities, as manifested by at least two of the following:*** *1. Stereotyped or repetitive motor movements, use of objects, or speech* *2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns of verbal or nonverbal behaviour* *3. Highly restricted, fixated interests that are abnormal in intensity or focus* *4. Hyper- or hyporeactivity to sensory input or unusual interest in sensory aspects of the environment*	- ***Persistent restricted, repetitive, and inflexible patterns of behavior, interests, or activities*** *that are clearly atypical or excessive for the individual's age and sociocultural context.* - ***The onset of the disorder occurs during the developmental period, typically in early childhood****, but characteristic symptoms may not become fully manifest until later, when social demands exceed limited capacities.*
C. ***Symptoms must be present in the early developmental period***	- ***The symptoms result in significant impairment in personal, family, social, educational, occupational or other important areas of functioning****.* *Some individuals with Autism Spectrum Disorder are able to function adequately in many contexts through exceptional effort, such that their deficits may not be apparent to others. A diagnosis of Autism Spectrum Disorder is still appropriate in such cases.*
D. ***Symptoms cause clinically significant impairment in social, occupational, or other important areas of current functioning***	
E. ***These disturbances are not better explained by intellectual developmental disorder (intellectual disability) or global developmental delay***	

The significant health, educational and social needs of individuals with ASD and their families constitute a critical area for resources, research and professional education [Hyman et al., 2020].

Over the last decades, there has been a significant increase in ASD prevalence. The global prevalence of ASD consistently increased from 0.25% [95%CI=0.12-0.42] during the period from 1994 to 1999 to 0.99% [95%CI=0.73-1.28] in the 2015-2019 period [Talantseva et al., 2023]. The US Centers for Disease Control and Prevention (CDC) have most recently reported an ASD prevalence of 2.76% (1:36) in children aged 8 years [Maenner et al., 2023] compared to 2.3% (1:44) in 2018 [Maenner et al., 2021]. Prevalence of ASD among children aged 4 years increased 26%, from 1.7% in 2018 to 2.15% (1:46) in 2020. This finding suggests improved early identification of the disorder in recent years [Shaw et al., 2023].

The ASD project network in the European Union (ASDEU) reported an average prevalence of ASD between 1 and 1.5% [Bougeard et al., 2024].

A nationwide study promoted by the Ministry of Health in Italy has estimated an ASD prevalence of 1.3% (1.1-1.6%) in children aged 7-9 years [Scattoni et al., 2023].

The high inter- and intra-regional variability in global prevalence estimates, ranging from 0.42% to 3.13% in Europe, has been attributed to the heterogeneity of epidemiological studies in terms of design and methods, as well as the different characteristics of the populations studied (e.g., age, ethnicity, availability of services, socioeconomic status) [Bachmann et al., 2018; Bougeard et al., 2024].

The mechanism underlying the rise of ASD prevalence involves multiple factors, that can be divided in aetiologic (epigenetic and environmental risk factors) and non-aetiologic (i.e. expansion of diagnostic criteria and assessment tools for ASD, early diagnosis, inclusion of milder phenotypes of ASD, improved availability of diagnostic services, increased public and scientific awareness of the disorder) [Narzisi et al., 2018].

2.2 Diagnosis and outcome of ASD

The ASD clinical assessment and diagnostic process is complex and may be challenging, given the heterogeneity of ASD expression and its interplay with additional neurodevelopmental conditions, such as global developmental delay (DD) and intellectual disability, co-occurring in 48.5% and 37.9% of ASD children, respectively [Maenner, 2023; Shaw, 2023].

As reported in DSM-5-TR, symptoms may be masked during early development and fully manifest only when social demands exceed the individual's limited capacities.

Despite advances in understanding the neurobiology and genetics of ASD, the diagnosis of ASD continues to be based on identifying and reporting behaviourally defined clinical symptoms, as no biological markers for ASD are available [Hyman et al., 2020; Hodis et al., 2025].

The formal diagnosis is based on information gathered from (1) a detailed developmental, medical, and psychosocial history, obtained from parents/carers; (2) direct clinical observation of child behavior; (3) a comprehensive physical and neurological examination [Brian, 2019; Lord, 2018].

A combination of structured and semi-structured ASD-specific tools can improve diagnostic accuracy. The most widely used standardized diagnostic tools for ASD include a structured parent-interview, the Autism Diagnostic Interview, Revised (ADI-R) and a semi-structured observational evaluation tool, the Autism Diagnostic Observation Schedule, second edition (ADOS-2). Sensitivity and specificity reported for the ADI-R in clinical settings were 80% (95% CI, 79%-82%) and 72% (95% CI, 70%-74%), respectively. For the ADOS-2, the respective values were 91% (95% CI, 90%-92%) and 76% (95% CI, 74%-78%) [Lebersfeld et al., 2021].

Evidence indicates that diagnoses made with standardized evaluation are more reliable

across sites and more valid over time than single-clinician assessments [Yu et al., 2024]; however, relying solely on these tools for the diagnosis can lead to false positives and negatives [Lebersfeld et al., 2021]. Furthermore, the administration of these ASD-specific diagnostic tools is often time-consuming and requires a formal training and clinical expertise, as scores on these tools are highly dependent on how the tools are administered and interpreted [Thabtah et al., 2019].

ASD diagnosis conveys a great social and economic burden for families and community. Therefore, there is an imperative necessity to improve the early identification of children with ASD since early diagnosis is the only way to ameliorate the outcome of children with ASD through evidence-based early intervention practices during the period of maximum brain plasticity [Narzisi et al., 2014]. Indeed, early intervention (EI) before age of 2 years offers the best opportunity to modify developmental trajectories as the most sensitive period of neurodevelopment for EI is suggested to be between birth and age 2 [Lidstone, 2025] (Figure 1).

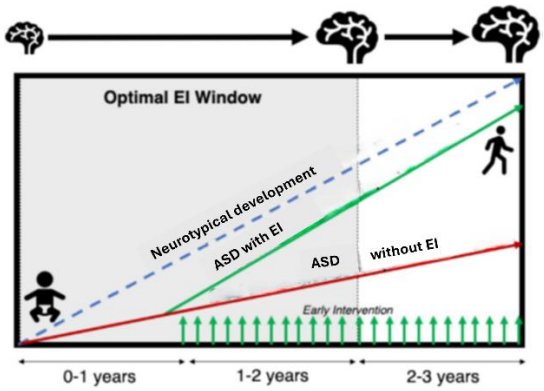


Figure 1. The neurotypical developmental trajectory (blue dashed line) compared to the slowed trajectory in children with ASD (red line) and the potential impact of intensive early intervention (EI) (green line) in modifying ASD-related delays [modified from Lidstone, 2025].

The National Survey of Children's Health (NSCH, 2021-2023) data analysis highlighted that only 6.7% of preschoolers with ASD were diagnosed before age 2, and only 15% of preschool children with ASD received EI before age 2 [Lidstone, 2025].

Later diagnosis of ASD (between 25 and 41 months) has been associated with greater impairment, revealing more severe delay and ASD symptoms [Miller et al., 2021].

According to a recent survey on ASD conducted in 14 European countries (ASDEU project network), the average age at diagnosis in Europe is approximately 36.4 months with most diagnoses occurring between 36 and 42 months. However, the average age at which concerns were first raised about the child subsequently being diagnosed with ASD was 18.3 months [Bejarano Martìn et al., 2020]. In the U.S. the average age at diagnosis of ASD is even later, ranging from 42 to 59 months [Maenner et al., 2023].

Many factors can influence the age of diagnosis, including the child's cognitive and language levels, as well as gender, ethnicity and socioeconomic status [Zwaigenbaum & Penner, 2018]. Children with milder ASD symptoms and/or average or above-average intelligence may not be identified until school-age, when differences in social language or personal rigidities affect their overall functioning [Hyman et al., 2020]. It has also been suggested that girls may have lesser intensity of symptoms and fewer externalizing behaviours or may manifest symptoms differently, according to "camouflage hypothesis" [Fombonne, 2020; Calderoni et al., 2023]. These differences may, in part, result in underdiagnosis in girls and in a lower prevalence of the disorder in females (ratio male to female 4:1) [Hodis et al., 2025]. Moreover, co-occurring neurodevelopmental and psychiatric disorders may prevent clinicians from recognizing symptoms of ASD in early childhood [Hirota et al., 2023]. Other factors that may impact diagnostic timing can be attributed to the family (e.g., caregiver understanding of typical development, caregiver denial of child delays): many caregivers are reluctant to pursue developmental concerns in early childhood and refuse clinical evaluations, particularly before the age of 18 months [Miller et al., 2021].

Overall, a diagnostic instability has been reported in early life more than any other age.

The variability and non-linearity of the ASD phenotype in early development defines a diagnostic instability over time ("lost or later diagnosis"), with some children meeting diagnosis at follow-up, and other children no longer meeting diagnostic criteria.

Diagnostic instability (e.g. "false positives" or "lost diagnoses") and missed cases ("false negatives" or "later diagnosed") demonstrate the variability and non-linearity of the ASD phenotype in early development [Landa et al., 2022].

2.3 Screening tools for ASD: advances and limitations

Adequate autism-specific screening programmes may help to improve the early identification of children at risk, reducing the diagnostic delays and facilitating the access to early targeted intervention [McCarty et al., 2020].

The American Academy of Pediatrics (AAP) recommends screening all children for symptoms of ASD through a combination of developmental surveillance and standardized autism-specific screening tests at 18 and 24 months of age in their primary care visits [Johnson et al., 2007; Hyman et al., 2020].

The large majority of ASD screening tools are designed to help caregivers identify and report symptoms observed in children at high risk for ASD [Hyman et al., 2020]. All of these instruments are based on a conceptual analysis of early communication development and "red flags", commonly described in the scientific literature as early ASD indicators, including difficulties with joint attention, social interaction, visual contact and playing skills [Magan-Maganto et al., 2017].

Screening measures for the early detection of the ASD risk are classified as either Level 1 instruments for the general population, or Level 2 tools for children under observation for development concerns, siblings of children with ASD, or children who fail Level 1 screening [Petrocchi et al., 2020].

The most commonly used screening tests are parent-completed questionnaires. However, parent reports are not always as reliable as direct observations [Towle et al., 2016], so an interactive system is recommended in order to elicit early signs of ASD in young children aged between 18 and 36 months [Choueri et al., 2023].

Emerging evidence suggested that a multi-step approach with the administration of different screening tools may be the best method to detect early signs of autism during the first years of life and to maximize the predictive value of screening [Levante et al., 2020; McCarty et al., 2020].

Main limitations of current screening tools are related to under- or over-estimation of the early risk sign by parent reports, low performances in discriminating ASD from different neurodevelopmental disorders (e.g. DD), consistency and reliability issues, requirements for training in test administration, lengthy test administration time [Brewer et al., 2020].

The most studied and used Level 1 ASD screening tool worldwide is the Modified Checklist for Autism in Toddlers, Revised with Follow-Up (M-CHAT-R/F), a two-stage parent-report screener for children aged 16 to 30 months with a structured follow-up interview administered to caregivers to obtain additional information [Robins et al., 2014]. Recent meta-analyses of the specific performances of M-CHAT-R/F showed a pooled sensitivity of 0.83 (95% CI, 0.77-0.88) and a pooled specificity of 0.94 (95% CI, 0.89-0.97) [Wieckowski et al., 2023]. The probability of ASD diagnosis following M-CHAT-R/F positivity (pooled positive predictive values, PPV) was estimated at 57.7% (95% CI 48.6–66.8). This implies that a positive screening with the M-CHAT-R/F is predictive of an ASD diagnosis in approximately 58% of children. However, the pooled PPV of the M-CHAT-R/F for the presence of any developmental disorder rises to 89% (95% CI 82.9-94.6) [Aishworiya et al., 2023]. It has been pointed out that several factors, such as false responses due to lack of parental awareness of expected socio-

communicative milestones, are associated with lower PPV regardless of screening sensitivity [McCarty & Frye, 2020].

The wide variability in psychometric properties of M-CHAT(-R/F) across studies highlights differences in screener use that should be considered in research and practice [Wieckowski et al., 2023]. Repeated screening of the same child at multiple timepoints, such as at 18 and 24 months as recommended by the American Academy of Pediatrics, is critical to improve the accuracy of ASD screening [Aishworiya et al., 2023].

The Quantitative CHecklist for Autism in Toddlers (Q-CHAT) is another Level 1 screener that differs from the M-CHAT-R/F in its approach and scoring system. It is a dimensional and quantitative measure of threshold and sub-threshold autistic features, consisting of a 25-item parent-reported questionnaire that assesses behavioural, language and non-verbal communication patterns of autistic traits. It was tested in both case–control studies and primary-care settings, displaying fair-to-good psychometric properties and predictive validity [Allison et al., 2008; Ruta et al., 2019]. The Autism Spectrum Quotient, in the version for children, (AQ-Child) is another parent-administered questionnaire for children aged 4-11 years. It contains 50 items covering the area of social skills, attention switching, imagination, communication and attention to detail [Thabtah et al., 2019]. A shortened version of the Q-CHAT and the AQ, including the ten most predictive items (Q-CHAT-10 and AQ-10, respectively), was also developed, aiming to create a quick screening tool suitable for the time constraints of paediatric check-ups and to help reduce delays in potential referrals [Allison et al., 2012].

The American Academy of Pediatrics has recommended the use of a Level 2 screener after failing a Level 1 screener, before referring children on for a full comprehensive evaluation for ASD [McCarty, 2020].

Among Level 2 screeners, some have undergone a greater number of validation studies and, therefore, warrant consideration for clinical application. The main characteristics and

psychometric properties of these tools are shown in Table 2. In particular, the STAT and the RITA-T represent the two secondary interactive and observational screening instruments with the most evidence, valid for children aged 36 months or less. All these tests require specific training for administration and scoring.

Table 2. Level 2 ASD Screening tools which have undergone a greater number of validation studies

Autism screening tests	Administration	Age range (months)	Average No. items	Administration time (minute)	Psychometric properties			References
					Sensitivity	*Specificity*	*PPV/NPV*	
STAT ***(Screening Tool for Autism in Two-Year-Olds)***	Trained clinician-directed, interactive/ observational	24-35	12	20-30	0.83	0.86	0.77 / 0.90	Stone WL, et al. *JADD.* 2000; Stone WL, et al. *JADD.* 2004; Attar SM, et al. *Autism Res.* 2022
ADEC ***(Autism Detection in Early Childhood)***	Trained clinician-directed, Observational Checklist	12-36	16	10	0.93	64%	0.83 / 0.81	Nah YH, et al. *Psychol Assess.* 2014; Hedley D, et al. *JADD.* 2015
BISCUIT ***(Baby and Infants Screen for Children with aUtIsm Traits)***	Parent-interview	17-37	62	20-30	0.93	0.76	-	Matson JL, et al. *Dev. Neurorehabil.* 2010; Horovitz M, et al. *J. Dev. Phys. Disabil.* 2014
SORF ***(Systematic Observation of Red Flags)***	Observational, trained video-recorded administration	16-24	22	30	0.77	0.72	-	Dow D, et al. *Autism.* 2017; Dow D, et al. *Autism Res.* 2020; Pileggi et al. *Am J Speech Lang Pathol*. 2021
RITA-T ***(Rapid Interactive Test for Autism in Toddlers)***	Trained clinician-directed, interactive/ observational	18-36	9	10	0.97	0.83	0.95 / 0.79	Choueiri et al. 2015; Lemay JF, et al. *J Dev Behav Pediatr.* 2020; Choueiri R, et al.*JADD.* 2021

Further longitudinal validation studies are needed for all measures to examine their psychometric properties, to compare the results of different measures and to increase their criterion validity through comparison with gold standard measures [Brewer et al., 2020; Petrocchi et al., 2020]. Suggested priorities for future research include the use of large

samples to enable reliable evaluations of Level 2 screening items in the 12- to 36-month age range, the exploration of potentially discriminative items within the targeted age ranges, and an attempt to unravel the complexities of developmental trajectories in autistic children [Brewer et al., 2020].

2.4 Developmental trajectories of ASD and assessment of the evolutive profile in ASD-risk children

The variability and non-linearity of the ASD phenotype in early development may be explained by different patterns of developmental trajectories identified in children with ASD in the first three years of life [Landa et al., 2013]. After a prodromal period, a developmental divergence from typical development becomes evident between 14 and 24 months and particularly affects language and social development; the ASD symptom consolidation usually occurs in the second and third years of life [Landa et al., 2022].

Any sign of atypical developmental trajectory should result in referral for further developmental assessment and an ASD-specific screener should be administered to all children at regular intervals. This practice may increase early detection of atypical development signs that may be associated with preclinical ASD or that are associated with non-ASD delays. Defining patterns of developmental trajectory within and across different developmental profiles in ASD is important for developing screening tools and detecting ASD early for access to early intervention [Landa et al., 2013].

Systematic evaluation of early developmental profiles illustrated some weakness in language, social and communication skills in children with ASD when compared to peers with developmental and/or language delays [Delehanty et al., 2018; Torrens & Ruiz, 2021]. Developmental assessment of ASD children is instrumental for understanding the child's early developmental profiles, revealing strengths and weakness in the different

domains, in order to plan appropriate therapeutic interventions [Jansen et al., 2020]. Moreover, a standardised developmental assessment may help discriminate ASD from global developmental delay at certain ages, intercepting early atypical developmental patterns in children at autism risk in the first years of life [Wang et al., 2021].

The Griffiths Scales of Child Development 3rd Edition (Griffiths III) is a validated tool for assessing the children's psychomotor developmental profile. It was standardized in 2015 on a representative sample from the UK and Ireland, thereafter it was published and adapted for other population samples, with a normative age range between 1 and 72 months [Green et al., 2016; Stroud et al., 2016].

A recent study compared the Griffiths III developmental profiles of children with co-occurring ASD and DD with those of a group of children with DD. Both groups exhibited lower age-equivalent scores than their chronological age in all developmental domains. However, while the DD group showed a uniform decrease in expected performance across all domains, the ASD + DD group exhibited an uneven profile, with relative failures in the Language and Communication and Personal-Social–Emotional subscales [Taddei et al., 2023].

Based on these insights, we hypothesised that a screening tool derived from the Griffiths III might timely intercept predictors of atypical developmental trajectories in children at risk of ASD, recognise specific risk signs and enable early, tailored therapeutic intervention to achieve a more favourable outcome.

2.5 A novel screening tool for autism: DAES (Developmental Autism Early Screening)

We developed DAES (Developmental Autism Early Screening), an observational and interactive Level 2 screening test for ASD, at the Child Neuropsychiatry of the University Hospital of Catania, through a systematic analysis of all Griffiths III items in the first 3 years of age, in young children at risk of ASD, DD and with typical development (TD). The DAES, based on Griffiths III, was designed to intercept significant differences in the early atypical developmental patterns of children at ASD-risk in the first 3 years of life, compared to DD or typically developing (TD) children [Cirnigliaro et al., 2023].

2.5.a DAES Development

For test development, scores obtained on each Griffiths III subscale were compared in children at risk of ASD, DD and TD children (CTRL group). Participants were matched according to their developmental age (DA), as measured by the Griffith III A-subscale, expression of the learning base, in order to compare children with the same non-verbal cognitive level.

The one-way ANOVA statistical test was applied for each Griffiths III subscale (A-E) in order to find out possible significant differences among groups. A significant difference was found for the B ($F = 13$; $p = 0.00001$) and D subscale DA means ($F = 12.3$; $p = 0.00002$). No differences were found in the remaining Griffith III subscales (Table 3).

Table 3. Demographic features and developmental profiles (developmental ages) on the Griffiths III of participants recruited for test development [adapted from Cirnigliaro et al., 2023].

Participants	ASD-risk	DD-risk	CTRL	F2	p-Value
N. 78 (Sex)	26 (M: 20)	26 (M: 22)	26 (M: 18)		
CA (months) (Mean ± SD)	39,46±12,48	34,07±7,88	26,38±7,48	12.288	**0.00002**
A - Subscale DA (months) (Mean ± SD)	24.38±5.80	24.5±6.28	25.23±5.87	0.1524	0.8588
B - Subscale DA (months) (Mean ± SD)	14.61±7.35	18.27±7.47	25.46±8.52	13.00	**0.00001**
C - Subscale DA (months) (Mean ± SD)	22.92±7.50	23.19±6.89	25.23±7.03	0.8097	0.4488
D - Subscale DA (months) (Mean ± SD)	17.0±6.21	22.15±7.20	26.53±6.64	12.345	**0.00002**
E - Subscale DA (months) (Mean ± SD)	25.73±6.60	26.07±7.07	26±7.08	0.0179	0.9822

CA: chronological age; DA: developmental age; M: males; ASD: autism spectrum disorder; DD: global developmental delay; CTRL: neurotypical children

According to previous studies [Jansen et al., 2020; Li et al., 2020; Pino et al., 2022; Taddei et al., 2023], we proved that B- (Language and Communication) and D- (Personal-Social-Emotional) subscales are more sensitive in intercepting differences between the groups. We therefore focused on each individual item of the B- and D-subscales that differed significantly in the comparisons between ASD and the remaining two groups.

All items of B- and D-subscales were grouped by year of age (first 3 years of life) and by constructs. The scores obtained for each single item in the three different groups were analysed by the Pearson's Chi-square independence test, which was applied to assess which items showed significant differences over the three groups and between each pair of groups: ASD/CTRL, ASD/DD and DD/CTRL group. The items significantly different between groups, representing the most predictive ones (N=36) for ASD risk, were used

to figure the novel screening tool based on differences in early developmental profiles measured on the Griffith III: Developmental Autism Early Screening (DAES).

2.5.b DAES description and administration

DAES items (N=36) are organized according to the Griffiths III specific constructs of B- and D-subscales, and probe the following developmental areas: listening, attention and communicative intent, expressive communication, receptive language development, social-emotional reciprocity, social communication and interaction, play, social cognition, joint attention, self-awareness, emotional understanding and expression and empathy. The tool is administered starting with the first items (starting points) regardless of the child's age. According to Griffith III, the discontinue administration rule is established after six consecutive items not passed within each subscale. No specific training is required for users trained in the use of the Griffith III.

2.5.c DAES scoring

Based on statistical analyses, we assigned a score of + 2 to the items that were significantly different in the ASD children compared to both the DD and CTRL groups, and a score of + 1 to the items that were significantly different in the ASD children compared to the CTRL group.

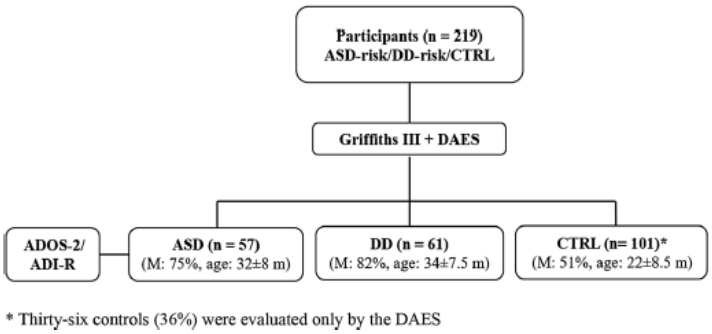
For each item, 0 represents the skill being expressed. Therefore, if the child fails the item (skill not expressed), a score of + 1/ + 2 is assigned during the observation, so that a higher total score indicates a higher risk of ASD (Table 4).

Table 4. Developmental Autism Early Screening (DAES)

Chronological age: 18-48 months (developmental age: 12-36 months)		
	Subscale B – Language and Communication	
FIRST YEAR		SCORE
B1.1	***Makes eye contact with speaker***	+2
B1.6	***Responds when called - get the child's attention in some way***	+2
B1.10	Knows own name	+1
B1.11	*Babbled phrases (non-reduplicated or variegated babbling)*	+1
B1.13	***Short babbled sentences used to interact with others***	+2
B1.14	*Looks at speaker when they are talking*	+1
B1.16	*Identifies one objects from four by looking/pointing*	+1
		TOT:
SECOND YEAR		SCORE
B2.2	*Maintains attention for two minutes*	+1
B2.3	*Identifies one picture from The Animals' Day Picture Book*	+1
B2.9	*Names two objects from six*	+1
B2.10	***Has a vocabulary of 10 words***	+2
B2.13	***Communicates with two-word phrase***	+2
B2.14	*Follows a two-step instruction with or without gestural cues*	+1
		TOT:
	Subscale D – Personal – Social - Emotional	
FIRST YEAR		SCORE
D1.3	*Social smile*	+1
D1.4	***Responds positively to strangers***	+2
D1.13	*Mirror task (2) – looks at reflection and smiles, tries to touch, kiss, vocalise to it*	+1
D1.14	***Eye Gaze (1) – follows head turn and eye gaze***	+2
D1.15	*Plays simple interactive games*	+1
D1.16	*Gives affection to other familiar people*	+1
D1.18	***Eye Gaze (2) – directs adult's attention***	+2
		TOT:
SECOND YEAR		SCORE
D2.2	***Claps hands in imitation***	+2
D2.3	***Points to share interest***	+2
D2.4	*Holds open cup for drinking*	+1
D2.7	*Plays alongside other children*	+1
D2.8	***Identifies Body Parts (1) – one feature***	+2
D2.9	*Funny act*	+1
D2.10	*Dressing (1) – helps actively to dress and undress*	+1
D2.13	*Shows Concern (1) – shows concern through expression or actions*	+1
D2.14	*Identifies Body Parts (2) – four features*	+1
D2.15	*Eats independently – some mess*	+1
		TOT:
THIRD YEAR		SCORE
D3.1	**Mirror Task (4) – names self**	+1
D3.2	**Gives name on request (1) – first name**	+2
D3.4	*Knows own gender*	+1
D3.5	*Facial Expressions (1) – identifies two of four*	+1
D3.7	***Knows age***	+2
D3.8	***Participates in group play***	+1
		TOT:

2.5.d DAES validation analyses

For preliminary validation analyses, the DAES was administered to 219 children recruited at the Child Neuropsychiatry of the University Hospital of Catania, over a 12-month period (ASD-risk, DD and CTRL) [Cirnigliaro et al., 2023]. Participants had a chronological age ranging from 18 to 48 months with a Griffith III A-Subscale DA equal to 12–36 months. They were assessed by clinical evaluation and by the Griffiths III. The DAES was administered at the time of clinical evaluation by two research assistants. Two senior board-certified researchers blind to the DAES scores generated final diagnoses independently of the DAES based on a full assessment (history, observation, and all testing measures) (Figure 2).



ASD: Autism Spectrum Disorder, *DD:* Global Developmetal Delay, *CTRL:* control group, *ADOS-2:* Autism Diagnostic Observation Schedule, *ADI-R:* Autism Diagnostic Interview-Revised, *M:* males, *m:* months

Figure 2. Participant flowchart for DAES preliminary validation [adapted from Cirnigliaro et al., 2023].

Receiver operating characteristic (ROC) curve analysis was performed to validate the predictive performance of DAES scores by determining its optimal cut-off score in discriminating ASD group from non-ASD participants (DD/CTRL groups). We found that the area under the curve (AUC) of DAES Total Score was 0.994 (95% CI 0.98–1), thus supporting the DAES capacity in classifying the ASD group from the non-ASD group. The DAES Total Score cut-off of 12.5 showed higher optimal sensitivity (93%)

and specificity (98.4%), with a PPV of 96.3% and NPV of 96.9%, determined by using Youden's J index and Euclidean distance [Akobeng, 2007] (Figure 3).

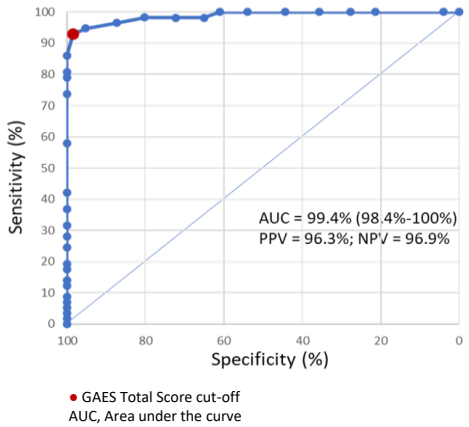


Figure 3. Receiver operating characteristic (ROC) curve analyses in ASD group (n=57) vs non-ASD (DD+CTRL, n=126) group [adapted from Cirnigliaro et al., 2023].

Linear regression was used to assess correlations between DAES score and ADOS-2. To allow comparisons between different modules, ADOS-2 CSS (Calibrated Severity Score) were used for correlation analyses (Spearman's rank correlation). The DAES total score was found to be significantly correlated with the ADOS-2 total, SA and RRB CSS, which provide a measure of ASD symptom severity (Figure 4).

Because of its significant correlation with ADOS-2 scores, the DAES has its value in stratifying those children at risk for ASD. In fact, following the overall ADOS-CSS risk stratification, we further identified an ASD risk score divided into three risk bands, with a total DAES score ≥ 15 indicating moderate to high risk.

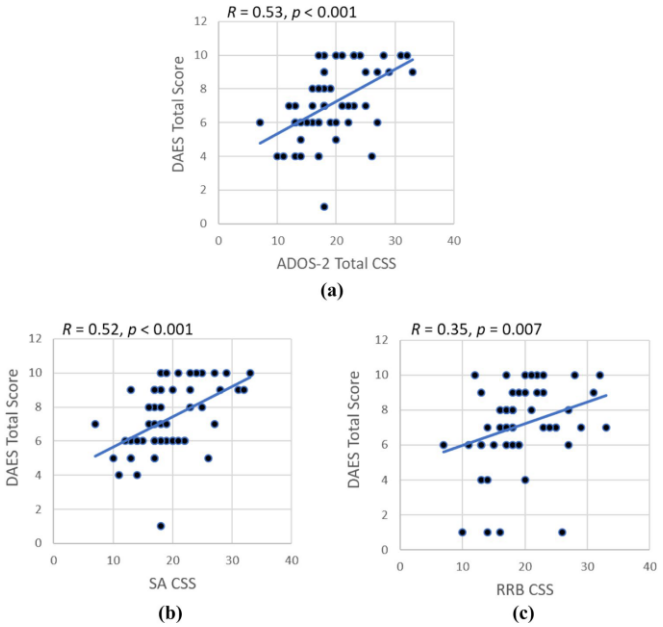


Figure 4. Linear correlations between DAES total score and **a** ADOS-2 total CSS (R = 0.53, p < 0.001), **b** SA CSS (R = 0.52, p < 0.001), and **c** RRB CSS (R = 0.35, p = 0.007), calculated using the Spearman's rank correlation [adapted from Cirnigliaro et al., 2023].

2.5.e DAES strengths and future perspectives

Overall, the DAES, based on differences in early developmental profiles on the Griffiths III, may assist in the identification of young children at risk of ASD differentiating them from DD and TD children.

A strength of the DAES is that it can be promptly used in the clinical practice with "clinically referred" children who have already been assessed with the Griffiths III, to measure ASD risk at 12–36 months of age. The DAES could complement current screeners and have the potential for widespread use due to its ease of administration, scoring and interpretation.

Further studies are needed to extend the assessment to ASD-risk children younger than 18 months of age. Moreover, replication in larger independent referral samples is necessary. Future priorities for the DAES include the exploration of potentially discriminating items at the target age range, while attempting to unravel the complexity of developmental trajectories in children with ASD.

2.6 Machine Learning in ASD screening and early diagnosis

Much progress has been made in identifying biomarkers of early brain development (e.g. using MRI, fMRI, EEG) and early behavioural features of infants later diagnosed with autism (e.g. atypical attention, prelinguistic communication, social engagement, sensorimotor processing). However, the challenge ahead is to translate these scientific findings into validated screening tools that can be used in clinical practice [Dawson et al., 2023; Cortese et al., 2025].

To overcome the limitations of the current screening and diagnostic procedures, in recent years there has been a significant increase in research dedicated to the development of machine learning (ML) models to support the clinical diagnosis or screening of ASD [Siddiqui et al., 2018; Cortese et al., 2025; Yuwattana et al., 2025]. For example, ML approaches have been applied to predict a formal diagnosis of ASD at 3 years from parent-rated learning and adaptive functioning measures [Bussu et al., 2018] or to identify specific behavioural patterns of autistic children from home-video recording [Jin et al., 2024].

ML and computer vision approaches have also been used to quantify early signs in children with ASD, such as atypical visual attention or social visual engagement (SVE) by analysis of eye-tracking data [Minissi et al., 2022; Klin, 2023; Wei et al., 2024].

Furthermore, computer vision analysis has been applied to quantify autism-related behaviours (e.g. social attention, facial expressions and motor behaviors) and integrate them into a digital application (app) using ML [Perochon et al., 2023]. Integrating this digital phenotyping tool with traditional screener (M-CHAT-R/F) resulted in an improved predictive screening validity [Krishnappa Babu et al., 2024].

ML algorithms are able to improve the psychometric properties of screening and diagnostic tools, by identifying the least number of items needed to maintain satisfying predictive and classification accuracy.

Sparsifying ML models have been applied to gold-standard tools for ASD diagnosis, such as ADOS and ADI-R, showing that a lower number of items could be sufficient to diagnose ASD and be accurate as the full and time-consuming assessments [Levy et al., 2017], while also combining clinical and molecular data [Yuwattana et al., 2025]. Furthermore, ML strategies have been used to process items from multiple instruments, such as the Social Responsiveness Scale (SRS) and the ADI-R [Bone et al., 2016], or combine structured parent-reported questionnaires and behavioural data from semi-structured home videos of children, resulting into a single assessment of higher accuracy [Abbas et al., 2018].

Research on ML/AI methods for early diagnosis of ASD has also led to the development of software and devices that are currently implemented in some clinical contexts, such as the Food and Drug Administration (FDA)-approved Canvas Dx (https://cognoa.com/) which applies ML algorithms to data received by parents/caregivers (e.g., questionnaires and home videos), video analysts, and healthcare professionals to facilitate the ASD diagnosis [Megerian et al., 2022; Wall et al., 2023].

To the best of our knowledge, ML approaches have mostly been applied as screening tools to caregiver questionnaires, such as the M-CHAT-R/F, Q-CHAT, AQ-10 [Achenie et al., 2019; Thabtah et al., 2019; Tartarisco et al., 2021]

Machine learning has been proven useful for implementing efficient automatic scoring of the M-CHAT-R/F, thereby avoiding the need for laborious follow-up and reducing the risk of human error. In terms of accuracy, it was comparable to the M-CHAT-R/F, while using fewer items [Achenie et al., 2019]. The application of ML algorithms to the Q-CHAT improved the tool's classification accuracy by selecting the subset of items that

most effectively discriminated between young autistic and typically developing children [Tartarisco et al., 2021].

Furthermore, ML algorithms were applied to datasets collected using a mobile application (ASDTests), consisting of four modules based on autism screening tools, the short form of the Autism Spectrum Quotient (AQ-10 age-appropriate versions for children and teens) and the Q-CHAT (Q-CHAT-10 for toddlers) [Thabtah et al., 2019]. These datasets have been used over the years to classify and predict ASD, with high accuracy [Shahamiri et al., 2022; Thabtah et al., 2022; Shrivastava et al., 2024]. The studies that applied ML to screening tools for ASD in children are summarized in Table 5.

Overall, the use of machine learning for autism screening is still a developing field, but it has shown some promising results.

We foresee that applying ML to an observational and interactive Level 2 screening tool, such as the DAES, could be useful to create a shorter and faster version of the tool, ensuring its predictive accuracy, even for age groups.

Table 5. Summary of studies using machine learning applied to screening tools for ASD children

Reference	Sample size	Age	Screening tool	ML Method	Key findings
Achenie et al., 2019	14995 toddlers (well-visit archival data)	16-30 months	M-CHAT-R/F	fANN model	ML method accuracy of ASD classification (99.72%) was comparable to the original version of M-CHAT-R/F while using fewer items
Erkan and Thanh, 2019	249 children (126 ASD) (UCI datasets)	4-11 years	AQ-10-Child	KNN, SVM, RF	RF achieved 100% accuracy in ASD classification
Thabtah, 2019	290 children (150 ASD-risk)	6.3 years (mean age)	AQ-10-Child	LR, NB	Proposal for a new mobile-based ASD screening tool (ASDTests). Best accuracy with LR (97.94%) in detecting autistic traits
Akter et al., 2019	1054 toddlers, 248 children (ASDTests app data)	18-36 months, 4-11 years	Q-CHAT-10, AQ-10-Child	Log, Z-score, and Sine FT; Adaboost, FDA, C5.0, Glmboost, LDA, MDA, PDA, SVM, CART	SVM and Adaboost showed the best performance for the toddler and child datasets (accuracy: 98.77% and 97.20%). The FT resulting in the best classifications was sine function for toddler and Z-score for child datasets.
Baadel et al., 2020	509 children (ASDTests app data)	4-11 years	AQ-10-Child	CATC; OMCOKE, RIPPER, PART, RF, RT, ANN	Proposal of a method using both clustering and classification in ASD screening: employing a new semi-supervised ML approach (CATC) significantly improved the performance of the classifiers, especially RF (accuracy: 97%).
Tartarisco et al., 2021	259 children (137 ASD / 122 TD)	22-43 months	Q-CHAT	RF, NB, SVM, KNN, LR	SVM showed the best discriminant validity (accuracy = 91%, sensitivity = 87%, specificity = 96% from a selection of 14 out of 25 items)
Thabtah et al., 2022	401 toddlers (142 ASD), 278 children (~50 ASD) (ASDTests app data)	18-36 months, 4-11 years	Q-CHAT-10, AQ-10-Child	New model combining SOM and NB or RF	The model built combining RF and SOM achieved the higher accuracy (96%) than model built with NB.
Shahamiri et al., 2022	257 ASD-risk children out of 509 (ASDTests app data)	4-11 years	AQ-10-Child	Bayes Net, C4.5, PART, RIPPER, ANN	Superior performance of ANN with higher specificity and sensitivity (100%) than other algorithms
Shrivastava et al., 2024	1054 toddler, 292 child datasets (ASDTests app data)	18-36 months, 4-11 years	Q-CHAT-10, AQ-10-Child	SVM, KNN, RF, ANN; kNN imputer	RF, combined with kNN imputer at the preprocessing stage, achieved accuracy of 100% with negligible overfitting to classify ASD/non-ASD
Lu et al., 2024	371 children (48 ASD / 323 non-ASD)	18-24 months	CHAT-23	mRMR feature selection; SVM, DT, GBDT, LGBM, RF, XGB, AdaBoost	The best-performing LGBM model achieved a sensitivity of 0.909 and a specificity of 0.922 with 9 features out of 23 to identify ASD children

fANN: feed-forward artificial neural network; kNN: k-nearest neighbors; SVM: support vector machine; RF: random forests; UCI: University of California, Irvine; LR: Logistic Regression, NB: Naïve Bayes; FT: feature transformation; FDA: Flexible Discriminant Analysis; Glmboost: Boosted Generalized Linear Model; LDA: Linear Discriminant Analysis; C5.0: Decision Tree; MDA: Mixture Discriminant Analysis; PDA: Penalized Discriminant Analysis; CART: Classification and Regression Trees; CATC: Clustering based Autistic Trait Classification; RIPPER: Repeated Incremental Pruning to Produce Error Reduction;; OMCOKE: Multi-Cluster Overlapping K-Means Extension; RT: Random Trees; PART: partial decision tree algorithm; SOM: Self-Organizing Map (clustering approach using ANN to transform data onto a two-dimensional map); C4.5: Decision Tree; kNN imputer: a distance-based imputation technique (it uses the euclidean distance between K neighbors values present in a particular attribute and replaces the missing value by mean value of K nearest neighbors); DT: Decision Tree; GBDT: Gradient Boosting Decision Tree; LGBM: Light Gradient Boosting Machine; XGB: eXtreme Gradient Boosting; AdaBoost: Adaptive Boosting; mRMR: Max-Relevance and Min-Redundancy

3. AIMS

The present study was undertaken for validating the DAES, a novel observational and interactive Level 2 screener, based on Griffiths III, applying machine learning (ML) on a larger sample of children. The DAES has been designed to intercept early atypical developmental patterns in ASD-risk children, differentiating them from children with DD or typical development.

We aimed to train and select ML classifiers in order to create a shorter and faster version of the tool with high overall performance, improving its applicability in the clinical practice and enabling a widespread use in various settings. In particular, considering chronological age rather than equivalent age before administering the DAES, could enable the tool to be used independently of the Griffiths III assessment. This would facilitate the applicability of the tool by general paediatricians and rehabilitation therapists, reaching a larger proportion of the population.

Additionally, analysing and identifying more sensitive and specific items by age group (under 24 months and over 24 months) in the first three years of life could help predict an autism condition with higher accuracy, ensuring timely access to early diagnosis and facilitating a tailored intervention during the period of maximum brain plasticity for a more favourable outcome.

4. Materials and Methods

4.1. Participants

At study entry, a sample of n= 520 young children were recruited in a multicenter study of two Italian Centres (the Child Neurology and Psychiatry Unit, Policlinico Hospital, University of Catania, and IRCCS "Bonino-Pulejo", Messina) between January 2023 and June 2024. A new sample of n= 128 children, recruited between July 2024 and May 2025 at the Child Neurology and Psychiatry Unit, Policlinico Hospital, University of Catania, was subsequently included in the study. Therefore, clinical data were obtained for a total sample of 648 children, included in the three different datasets.

The study subjects were 'clinically referred' children with parental or professional concerns about ASD (Autism Spectrum Disorder) and/or DD (Global Developmental Delay), and children without developmental concerns (TD). All participants were assessed by clinical evaluation using DSM-5 criteria and underwent to a developmental assessment using the Griffiths Scales of Child Development 3rd Edition (Griffiths III).

The inclusion criteria were: chronological age ranging from 12 to 48 months, Griffith III A-subscale developmental age (DA) of 10–36 months; clinical referral for ASD-risk or DD-risk (first and second group, respectively). TD children were defined as having General Development Quotient ≥ 90 at Griffiths III. They were healthy children prospectively screened for transient neonatal jaundice, suspected maternal infection, or preterm birth (gestational age > 34 weeks) without signs of neonatal distress (Apgar scores: 9/10). Exclusion criteria included bilingualism; syndromes or genetic abnormalities; diagnosis of other neurological disorders (i.e. epilepsy, hearing and visual defects).

The DAES was administered at the time of clinical evaluation by two research assistants, blind to the final diagnoses, at the referral University Hospital for Child Neuropsychiatry

of Catania and IRCCS of Messina. The Cohen κ statistic was calculated for each rater and varied between 0.7 and 1, indicating good to excellent agreement. Two senior board-certified researchers blind to the DAES scores generated final diagnoses independently of the DAES, using all available clinical information and all testing measures, including the Autism Diagnostic Interview-Revised (ADI-R) and the Autism Diagnostic Observation Schedule, 2nd edition (ADOS-2) when clinically indicated, at the Centres of Catania and Messina. The datasets were then divided into three classes according to diagnostic category (ASD, DD, TD).

Written informed consent from both parents was acquired before the beginning of the study. The current study was part of an overall larger study aimed at identifying markers, predictors and developmental trajectories of ASD. The larger overall study was approved by the Institutional Review Board at the University Hospital Referral Centre for ASD (n° 759). All procedures performed in the present study were in accordance with the 1964 Declaration of Helsinki and its later amendments (2013).

4.2 Data selection and pre-processing

For the purposes of the analysis, the participants were divided into two groups, according to chronological age, ≤ 24 months and > 24 months.

To enable comparison of children from the three classes (ASD, DD and TD) within the same age range, any children younger than 12 months in terms of equivalent age, who had missing DAES scores for the second year, were excluded. This ensured that all remaining children had completed all DAES items for the first and second years.

Consequently, the total sample comprised 610 children instead of 648. Demographic characteristics of the entire sample are reported in Table 6.

Table 6. Demographic data of participants

Participants	***ASD***	***DD***	***TD***	***Total sample***
N. Tot.	200	214	196	610
Gender M:F (M%)	141:59 (70.5%)	156:58 (72.9%)	127:69 (64.8%)	424:186 (69.5%)
Age in months (mean, SD)	32 ± 8.5	33.4 ± 8.8	26.8 ± 7.7	30.8 ± 8.8
DA	21.8 ± 5.7	24.3 ± 6.3	26.1 ± 6.9	24.1 ± 6.6
N. ≤ 24 months	59	56	78	193
N. > 24 months	141	158	118	417

The dataset, including 482 children instead of the initial 520 from the clinical samples of Catania and Messina, has been used to: i) analyse the correlation between the equivalent age and the other features (chronological age, DAES's items) by using the V index of Cramer; ii) select the features according to chronological age (≤24 months and >24 months); iii) train the classification models.

The dataset including 128 children, recruited at a later stage of the study at the Child Neuropsychiatry of the University Hospital of Catania, has been used to form the Test set of the generated classifier models.

As the DAES is administered considering the child's equivalent age assessed by Griffiths III, some children with ASD or DD, who exhibited discrepancies between their chronological and equivalent ages, did not complete the third-year items. For example, if a child had a chronological age of 35 months and an equivalent age of 22 months, only the items corresponding to the first and second years were administered, not those relating to the third year. Consequently, children with ASD or DD who were older than 24 months in terms of chronological age, but younger than 24 months in terms of equivalent age did not complete the items related to the third year, unlike typically developing children of the same chronological age. Therefore, the third-year DAES item scores were not included in the analysis to mitigate potential bias during classifier model training for the

24- to 48-month age range and to develop a classifier capable of distinguishing between ASD, DD and TD. However, the inability to analyse DAES data for each item in the third year could result in reduced accuracy. To address this issue, we aim to demonstrate that the DAES can be administered based on chronological age rather than equivalent age, thanks to the hypothesized correlation between certain DAES items and equivalent age. This approach would eliminate the previously described limitation, enabling a more comprehensive assessment of children (including all DAES items for children over 24 months old) regardless of discrepancies between chronological and equivalent age.

4.3 Cramer's V index computing

The Cramer's V index computing was performed in order to investigate whether it is necessary to establish the child's equivalent age prior to administering the DAES, or if this step can be omitted. To evaluate this, the correlation between DAES items and equivalent age was computed. We hypothesized that some DAES items would be significantly correlated with equivalent age, thus reflecting the child's development level. This would eliminate the need to determine the equivalent age prior to DAES assessment and allow the administration of all DAES items to all children, including those older than 24 months, according to their chronological age. Cramér's statistic (V C) facilitates the interpretation of nominal-variable association estimates, as this index ranges from 0 to +1. A higher V C indicates a stronger association [Kearney, 2017].

3.4 Feature selection and filtering

To improve the classification performance, a selection of DAES items has been carried out. The aim was to remove items which are not able to discriminate among the three groups of children (ASD, DD and TD). The Chi-Square technique was used for this

purpose. It helps to determine whether the observed differences between the expected and actual data are statistically significant or due to chance [McHugh, 2013].

Initially, the three cohorts of children were grouped into pairs, defined as (child type 1 (ct1), child type 2 (ct2)), where ct1 ∈ [ASD, DD] and ct2 ∈ [ASD, TD]. Then, for each DAES item (itx), the following tuple was computed: ((ct1, ct2), itx, vx1, vx2), where vx1 and vx2 are two integer vectors containing the scores obtained by the ct1 and ct2 groups for the itx item, respectively. The Chi-Square test was applied to each tuple, resulting in the addition of two parameters: the Chi-Square statistic and the corresponding p-value. Tuples with a p-value greater than 0.05 were filtered out, as they indicate an absence of statistical significance. The remaining tuples were sorted in descending order based on their Chi-Square values. These retained tuples correspond to the DAES items that are most informative for building the classifier.

4.5 Machine-learning classifier tuning

Once the features were selected, two classification strategies were developed:

i) The first one involves the implementation of a multiclass classifier based on three different classes: ASD, DD, and TD. While this approach will create models that perform well in discriminating between ASD vs DD and ASD vs TD, it is less effective in distinguishing between DD and TD. This limitation primarily stems from the DAES methodology. The DAES items have been designed to identify children at risk of ASD, and distinguish them from children with DD or TD, rather than differentiate between the DD and TD groups. Figures 5 and 6 show violin plots of DAES total scores for children with ASD, DD, and TD aged 24 months or younger and for children with ASD, DD and TD aged over 24 months, respectively. There is minimal overlap between the DAES

scores of the ASD group and those of the DD and TD groups. In contrast, the plots for the DD and TD groups substantially overlap.

ii) The second strategy, instead, involves a binary classification, comparing the ASD group to the non-ASD group (including DD and TD groups). Thanks to the almost total lack of overlap between ASD and the other classes (DD and TD), the classifier will be able to distinguish children with ASD from those with DD or TD.

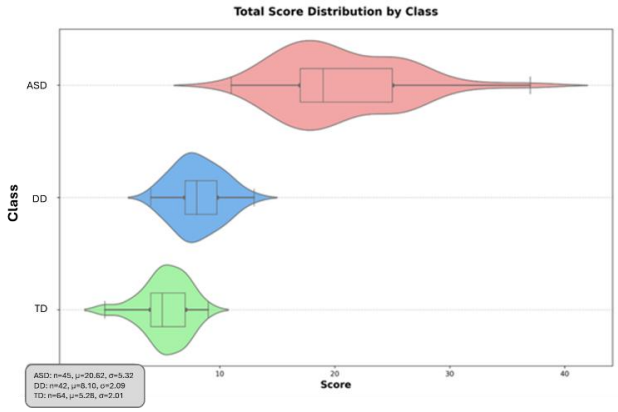


Figure 5. Violin plot: DAES total score distribution in ASD, DD and TD children aged 24 months or younger

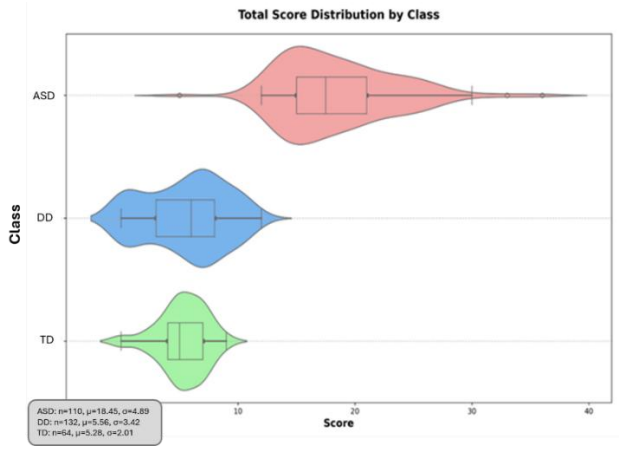


Figure 6. Violin plot: DAES total score distribution in ASD, DD and TD children aged over 24 months.

In addition to the features selected from the first and second years, the total scores from the B and D subscales were used to train the models. These scores were computed for each child by adding together the scores of the selected items from the first and second

years, as well as the third-year DAES scores. Consequently, the total D-subscale scores also reflect the DAES items from the third year, which were not included in the analysis due to discrepancies between chronological and equivalent ages of ASD and DD children.

We tested five supervised classifiers to understand the relationship between DAES items and the diagnostic label: random forest (RF) [Breiman, 2001], decision tree (DT), support vector machine (SVM), K-nearest neighbors (KNN) [Peterson, 2009] and logistic-regression (LR) [Kleinbaum et al., 2002] algorithms.

We conducted the selection of the best classifier model parameters using the GridSearchCV, related to the Scikit-learn library in Python. This will tune the hyperparameters on the data, that has already been preprocessed by StandardScaler. In addition, a tenfold cross-validation has been employed to ensure model robustness, testing the accuracy of each classifier.

Data analysis, model training and evaluation were performed using the Python 3.12 version and the Pandas, NumPy and Scikit-learn libraries.

4.6 Metric for ML performance

The Test set was used for final model validation to test each model's performance on previously unseen data.

The performance of each classifier was examined using the following metrics: confusion matrix, accuracy, sensitivity/recall, precision, F1 score and AUC-ROC curve.

These metrics were used for the two classification strategies: multiclass (ASD, DD and TD) and binary (ASD, non-ASD).

Accuracy represents the overall success rate of each classifier and is computed as: accuracy = (TP + TN)/(TP + FP + FN + TN). TP (true positive) and TN (true negative) represent the number of instances that are correctly classified; FP (false positive) and FN

(false negative) represent the number of instances that are misclassified. Sensitivity/recall is defined as the percentage of correctly classified instances and is computed as: sensitivity/recall = TP/(TP + FN); precision is defined as percentage of incorrectly classified instances and computed as: precision = TP/(TP + FP). F1 score computed as F1 = (2TP)/(2TP + FP + FN) is the measure of a test's accuracy and it is based on the harmonic mean of specificity and sensitivity. It reaches its best value at 1 and worst at 0. Our comprehensive workflow is illustrated in Figure 7.

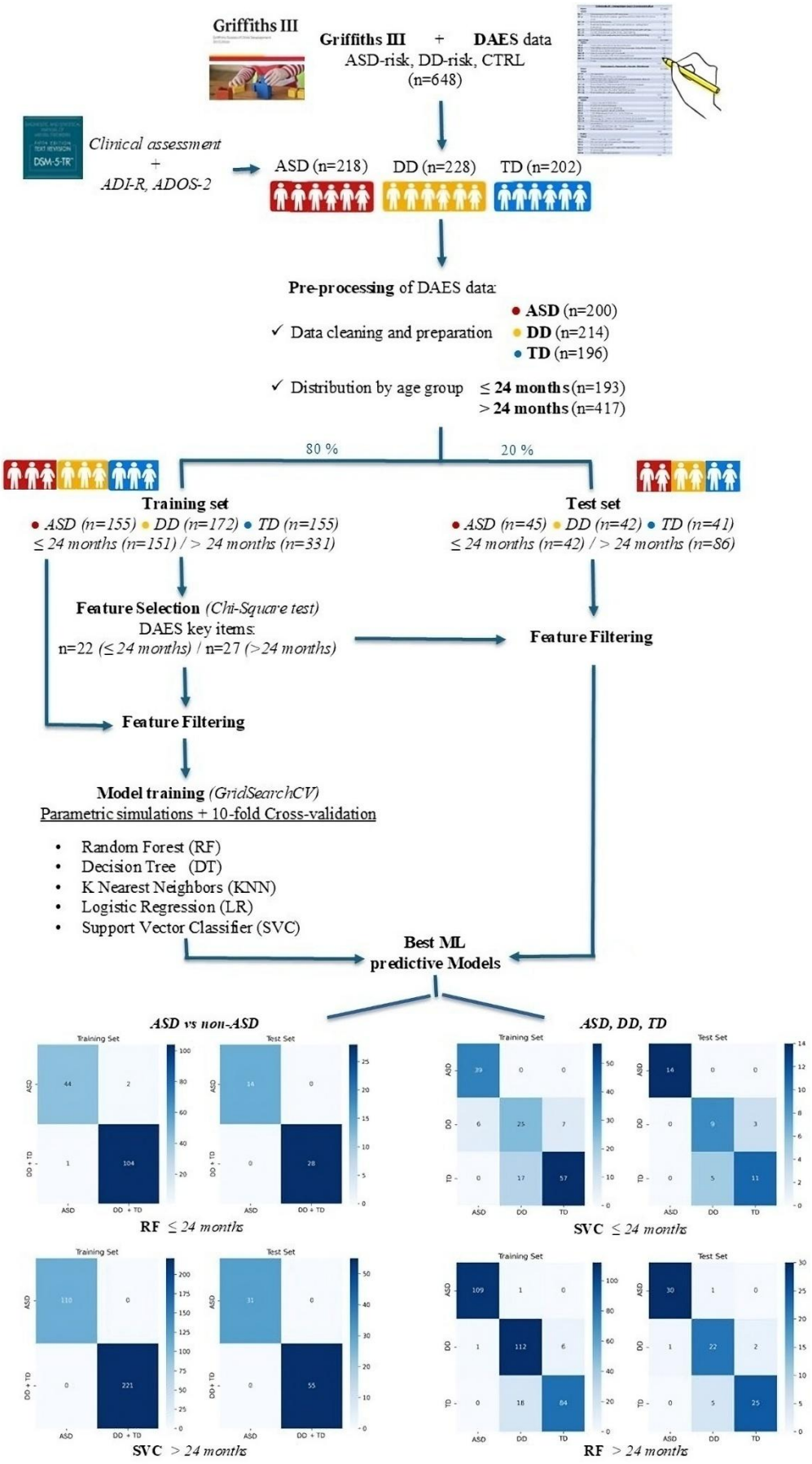


Figure 7. Workflow of the entire analytical process: ML framework for ASD screening using DAES

4.7 Standardized measures

All the participants were assessed using the **Griffiths III** that provides an overall measure of child psychomotor development (from birth to 72 months). The Griffiths III displays good reliability and validity for early detection of developmental delay, as well as an individual profile of strengths and needs across five subscales: (A) Foundations of Learning, (B) Language and Communication, (C) Eye and Hand Coordination, (D) Personal-Social-Emotional, (E) Gross Motor domains [Green et al., 2016].

Subscale A assesses a child's skills for learning (including attention, problem-solving abilities, sequential reasoning, processing speed, visuospatial skills), ways of thinking, memory and capacity to engage with real objects in imaginative ways. Subscale B refers to the ability to understand the meaning of words and to use language effectively. It evaluates the development of both receptive and expressive language and the child's use of language to communicate socially with others. Subscale C assesses the child's early development of visual perception, visually directed reaching and object manipulation and more complex fine motor skills. Subscale D evaluates self-help and self-care skills, as well as the child's development of self-awareness and concept of the self. It assesses early social and emotional development through items measuring imitation, joint attention and perspective taking, emotional expression and understanding in the self and the others, moral development and theory of mind. The subscale also explores the child's interest in and interactions with others, humour and friendships. Subscale E refers to the child's early development of postural control gross body coordination, balance and visual-spatial coordination.

The assessment starts with the administration of subscale A. The child's chronological age gives a rough position to start. It is best to have established a rough developmental age for the child before starting on the activities in order to avoid lengthening

administration time by using inappropriate activities with the child. The appropriate items for the child can be completed until the basal and the ceiling for the subscale are obtained. The basal is the last of the six consecutive items passed; the ceiling is the last passed item (the child must fail six consecutive items). The basal at subscale A gives an indication of where to start in the next subscale. Each item is scored as a pass or a fail: +1 or 0. Subscale raw scores and general development raw scores are calculated to then determine the Developmental Age (months), Scaled Score and Development Quotient, according to the norm tables.

DSM-5 criteria defined by the American Psychiatric Association were used for ASD and DD diagnosis (American Psychiatric Association, 2013). Symptoms of ASD were established using the gold-standard tools for ASD diagnosis: Autism Diagnostic Interview-Revised (ADI-R) [Lord et al., 1994] and the Autism Diagnostic Observation Schedule, 2nd edition (ADOS-2) [Lord et al., 2012]. The **ADI-R** is a structured interview to the parents of children referred for evaluation of possible autism spectrum disorder. The **ADOS-2** is a semi-structured, standardized assessment of social interaction, language and communication, restricted and repetitive interests and behaviours and play and imagination [Lord et al., 2012]. It contains five modules that are differentiated by children's developmental and language levels (module Toddler, 1, 2, 3, 4).

In the current study participants completed the ADOS-2 Toddler Module or the Module 1, administered by trained professionals. The ADOS Toddler Module was designed specifically for children 12-30 months old with limited language to measure symptoms of ASD. Module 1 is used for children aged from 31 months who do not consistently use phrase speech. The measure provides symptom domain scores (Social Affect, AS, and Restricted and Repetitive Behaviour, RRB) and a total score. The Toddler Module provides scores which indicate level of ASD risk: "little-to-no risk", "mild-to-moderate risk" and "moderate-to-severe risk". In the Module 1 an algorithm with cut-off scores

defines autistic disorder (AD), autism spectrum (ASD) and non-ASD. Moreover, the ADOS Comparison Score allow to obtain a mean of severity/ASD-symptom level.

5. RESULTS

5.1 Data preparation

The DAES was administered to a total sample of 648 children (ASD/DD risk and TD) recruited over a 28-month period at two different times. Initially, a sample of 520 children was recruited at the Catania and Messina Centres; a novel sample of 128 children was subsequently recruited at the Catania Centre. According to DSM-5 criteria, Griffiths III, ADI-R and ADOS-2 scores, 218 children were diagnosed with ASD and 228 children were diagnosed with DD. 202 children were in the TD group.

Following the exclusion of children with missing DAES scores for the second year (data cleaning), the final sample comprised 610 children, including 200 ASD children and 410 non-ASD (DD and TD) children. According to the data preparation, the sample recruited at the Catania and Messina centres comprised 482 children for model training and validation. The novel sample of 128 children, recruited at a later stage, was used to form the test set for final model validation (Figure 7).

Mean total DAES scores were higher in the ASD group (19.3 ± 5.2) than in the DD (6.1 ± 3.2) and TD (4 ± 2.7) groups, as expected.

TD children were significantly younger than both ASD and DD participants ($p < 0.001$, ANOVA test). They also had a significantly higher DA compared to both ASD and DD groups ($p < 0.001$ and $p = 0.002$, respectively). DA was significantly lower in children with ASD than in children with DD ($p < 0.001$), indicating that, on average, participants with ASD were developmentally delayed. In fact, a minority of the ASD group in the study (16%) had no difference in DA/CA.

Applying Cramer's V index, we found that the equivalent age was significantly correlated with some DAES items, particularly within the ASD and DD groups (Table 7). Thanks to these correlations, the equivalent age does not need to be determined before

administering the DAES, as some items intrinsically reflect it. Therefore, the DAES can be administered according to chronological age rather than equivalent age, enabling a more comprehensive assessment of children regardless of discrepancies between their chronological and equivalent age.

Table 7. Correlation between the equivalent age and DAES items for ASD, DD and TD children

ASD				
First feature	*Second feature*	*Cramér V*	*p-value*	*Significance*
Equivalent age	*Chronological age*	0,664	0,010	*
	B1.1	0,692	0,048	*
	B1.10	0,711	0,0324	*
	B1.13	0,702	0,0398	*
	B1.16	0,722	0,0245	*
	B2.13	0,706	0,0363	*
	D1.4	0,782	0,0050	**
	D1.15	0,772	0,0067	**
	D2.10	0,753	0,0111	*
DD				
Equivalent age	*Chronological age*	0,660	0,011	*
	B1.13	0,715	0,017	*
	B1.16	0,696	0,028	*
	B2.10	0,752	0,0003	***
	D2.2	0,755	0,0060	**
	D2.3	0,858	0,00019	***
	D2.8	0,701	0,0255	*
	D2.13	0,7023	0,0243	*
	D2.14	0,6960	0,0287	*
	D2.15	0,7867	0,0023	**
TD				
Equivalent age	*Chronological age*	0,660	0,005	**
	B1.13	0,835	0,005	**
	D2.14	0,835	0,005	**

5.2 Feature selection

The training and test set samples were divided into two age groups (≤ 24 months and > 24 months) to identify the most predictive ASD-risk items in relation to developmental level, and to design classifiers that can better distinguish between children with ASD and those with DD or TD in each age group.

The DAES items that best discriminated between the three groups (ASD, DD, and TD) across the two age groups were selected applying the Chi-square technique to the training set sample. Twenty-one out of 36 DAES items were found to be the most predictive for

the ≤ 24-month age range, and 28 items for the >24-month age group. In addition, the total scores of B and D subscales were computed for each child within the groups. These scores, along with the selected items, were used to train the classifiers and improve classification performance (Table 8).

Table 8. Selected DAES features for each age range

AGE RANGE	**≤ 24 months**		**> 24 months**	
DAES FEATURES	**Selected DAES items**	**Developmental domains**	**Selected DAES items**	**Developmental domains**
	B1.1	*Listening, attention, communicative intent*	B1.1	*Listening, attention, communicative intent*
	B1.6	*Listening, attention*	B1.6	*Listening, attention*
	B1.10	*Preverbal Receptive communication*	B1.10	*Preverbal Receptive communication*
	B1.11	*Preverbal Expressive communication*	B1.11	*Preverbal Expressive communication*
	B1.13	*Communicative intent*	B1.13	*Communicative intent*
	B1.14		B1.14	
	B1.16	*Preverbal Receptive communication*	B1.16	*Preverbal Receptive communication*
	B2.2	*Listening, attention*	B2.2	*Listening, attention*
	B2.3	*Preverbal Receptive communication*	B2.3	*Preverbal Receptive communication*
	D1.3	*Social interaction*	B2.9	*Expressive language*
	D1.4		B2.10	
	D1.13	*Self-concept*	B2.13	
	D1.14	*Social-emotional cognition, joint attention*	B2.14	*Receptive language*
	D1.15	*Social interaction*	D1.3	*Social interaction*
	D1.16	*Emotional expression*	D1.4	
	D1.18	*Social-emotional cognition, joint attention*	D1.13	*Self-concept*
	D2.2	*Social interaction, emotional expression*	D1.14	*Social-emotional cognition, joint attention*
	D2.3	*Joint attention*	D1.15	*Social interaction*
	D2.7	*Social interaction*	D1.16	*Emotional expression*
	D2.8	*Self-concept*	D1.18	*Social-emotional cognition, joint attention*
	D2.9	*Social cognition*	D2.2	*Social interaction, emotional expression*
			D2.3	*Joint attention*
			D2.7	*Social interaction*
			D2.8	*Self-concept*
			D2.9	*Social cognition*
			D2.10	*Self-help*
			D2.13	*Emotional expression, empathy*
			D2.14	*Self-concept*
	DAES B subscale Total scores		DAES B subscale Total scores	
	DAES D subscale Total scores		DAES D subscale Total scores	

5.3 ML classification models

We trained and selected the best parameters for five machine learning algorithms (RF, SVC, DT, KNN and LR) using the training data, according to two classification strategies: binary (ASD/non-ASD) and multiclass (ASD, DD and TD) in the two age groups (≤ 24 months and > 24 months). After applying GridSearchCV with 10-fold cross-validation the best parameters were found for each classifier (Table 9).

Table 9. Best parameters for each trained classifier

	Parameter	ASD / non-ASD ≤ 24 months	ASD / non-ASD > 24 months	ASD, DD, TD ≤ 24 months	ASD, DD, TD > 24 months
SUPPORT VECTOR (SVC)	C	1.0	10.0		
	class_weight	balanced			
	degree	2			
	gamma	0.001	auto	0.001	
	kernel	rbf			
RANDOM FOREST (RF)	bootstrap	True			
	ccp_alpha	0.0			0.01
	class_weight	balanced			
	criterion	gini			entropy
	max_depth	None	5	2	8
	max_features	None		log2	None
	min_samples_split	5		20	10
	n_estimators	100		200	100
DECISION TREE (DT)	ccp_alpha	0.05	0.0	0.001	0.01
	class_weight	balanced			
	criterion	gini	entropy	gini	entropy
	max_depth	5	none	5	none
	min_impurity	0.001	0.0	0.0001	0.0
	min_samples_split	10	5	20	10
	min_weight	0.0			
	splitter	random	best	random	best
LOGISTIC REGRESSION (LR)	C	0.1	0.01	0.1	
	max_iter	10000			
	penalty	11	12	11	12
	solver	saga	lbfgs	saga	lbfgs
K-NEAREST NEIGHBORS (KNN)	algorithm	auto	ball_tree	kd_tree	auto
	leaf_size	25		35	25
	metric	euclidean			
	n_neighbors	7	13	17	7
	weights	uniform			

The effectiveness of each classification model was assessed using a confusion matrix on both the training and the test data, divided by age group (Figures 8 and 9).

The performance metrics of each classifier (train and test accuracy, test precision, test recall, test F1 score, ROC AUC) were computed for both the binary (ASD/non-ASD) and multiclass (ASD, DD and TD) classifications, considering the DAES selected items for each age group. The detailed performance metrics are reported in Tables 10 and 11.

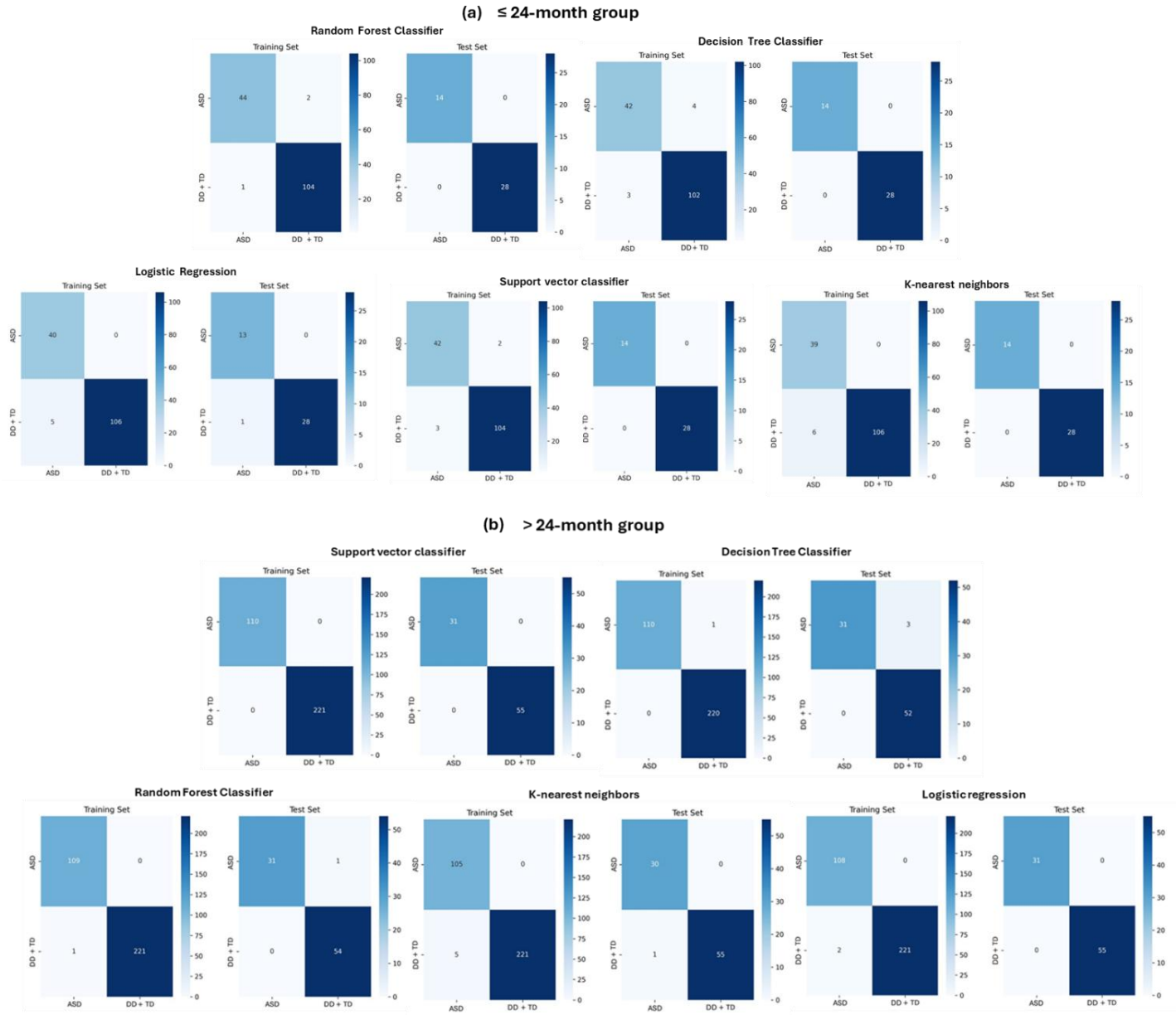
The RF algorithm showed the best train and test accuracy (98% and 100% respectively) in discriminating ASD vs non-ASD children aged ≤ 24 months, considering the 21 selected DAES items. According to the confusion matrix results for the binary classification (ASD/non-ASD), the train false negative rate (FNR) for the RF was 2.2 % and the test FNR was 0. The train false positive rate (FPR) for the RF was 1.8% and the test FNR was 0. These results confirmed that the RF achieved a higher train and test sensitivity (97% and 100% respectively) and train and test specificity (98% and 100% respectively) than the remaining algorithms in children aged ≤ 24 months.

The SVC algorithm achieved higher train and test accuracy values than the other algorithms (100%) in discriminating between ASD and non-ASD children aged >24 months, considering the 28 selected DAES items. According to the confusion matrix results for the binary classification (ASD/non-ASD), the train and test FNR and FPR for the SVC was 0, therefore the SVC achieved a higher train and test sensitivity and specificity (100%) than the remaining algorithms in children aged > 24 months.

In addition, the SVC algorithm demonstrated a higher train and test accuracy (80%) than the other algorithms in discriminating among the three classes (ASD, DD and TD) within the ≤ 24-month age range, while RF was the most effective model for the >24-month age group, achieving train and test accuracies of 92% and 89%, respectively. The lower accuracy of the classifiers in discriminating among the three classes (ASD, DD and TD) may be explained by the specific design of the DAES for identifying children at risk of

ASD and distinguishing them from those with DD or TD. Consequently, it is to be expected that the DAES is less accurate in discriminating DD from TD children.

Figure 8. Confusion matrix of different classifiers on ASD / non-ASD classes in children (a) aged ≤ 24 months and (b) > 24 months



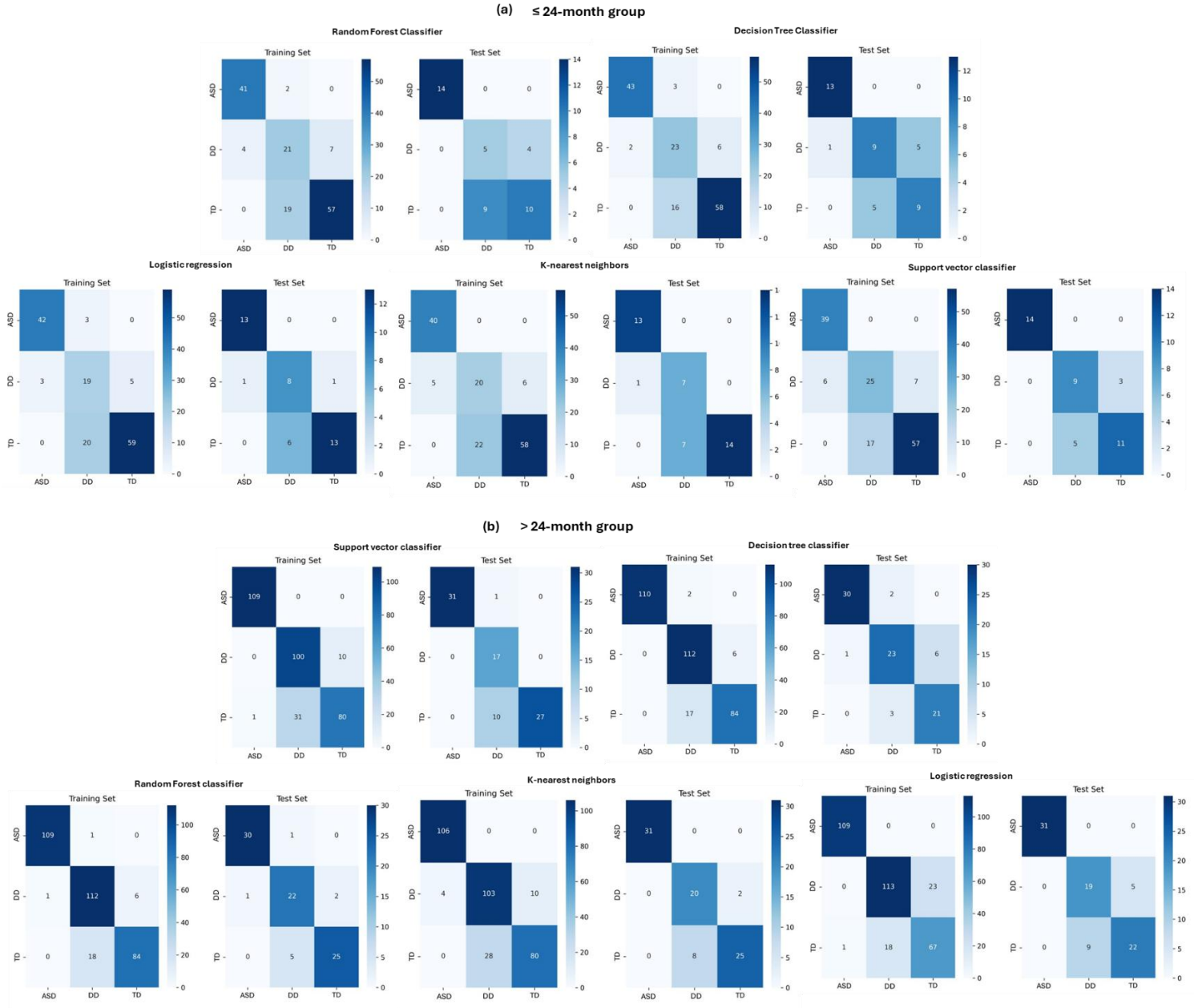


Figure 9. Confusion matrix of different classifiers on ASD, DD and TD classes in children (a) aged ≤ 24 months and (b) > 24 months

Table 10. Detailed performance metrics of ML classifiers in ASD vs non-ASD classes for age groups.

Classification Model	Age group	Train accuracy	Test accuracy	Precision	Recall	F1 score	ROC AUC
RFC	≤ 24	0.98	1.00	1.00	1.00	1.00	1.00
	> 24	0.99	0.98	1.00	0.98	0.99	1.00
SVC	≤ 24	0.96	1.00	1.00	1.00	1.00	1.00
	> 24	1.00	1.00	1.00	1.00	1.00	1.00
DTC	≤ 24	0.95	1.00	1.00	1.00	1.00	1.00
	> 24	0.99	0.96	1.00	0.94	0.97	0.97
KNN	≤ 24	0.96	1.00	1.00	1.00	1.00	1.00
	> 24	0.98	0.98	0.98	1.00	0.99	0.99
LR	≤ 24	0.96	0.97	0.96	1.00	0.98	1.00
	> 24	0.99	1.00	1.00	1.00	1.00	1.00

Table 11. Detailed performance metrics of ML classifiers in ASD, DD and TD classes for age groups.

Classification Model	Age group	Train accuracy	Test accuracy	Precision	Recall	F1 score
RFC	≤ 24	0.78	0.69	0.69	0.69	ASD: 1.00 DD: 0.43 TD: 0.60
	> 24	0.92	0.89	0.89	0.89	ASD: 0.96 DD: 0.83 TD: 0.87
SVC	≤ 24	0.80	0.80	0.81	0.80	ASD: 1.00 DD: 0.69 TD: 0.73
	> 24	0.87	0.87	0.90	0.87	ASD: 0.98 DD: 0.75 TD: 0.84
DTC	≤ 24	0.82	0.73	0.74	0.73	ASD: 0.96 DD: 0.62 TD: 0.64
	> 24	0.92	0.86	0.86	0.86	ASD: 0.95 DD: 0.79 TD: 0.82
KNN	≤ 24	0.78	0.80	0.84	0.80	ASD: 0.96 DD: 0.63 TD: 0.80
	> 24	0.87	0.88	0.89	0.88	ASD: 1.00 DD: 0.80 TD: 0.83
LR	≤ 24	0.79	0.80	0.82	0.80	ASD: 0.96 DD: 0.66 TD: 0.78
	> 24	0.87	0.83	0.84	0.83	ASD: 1.00 DD: 0.73 TD: 0.75

6. DISCUSSION

In this study, we applied machine learning (ML) to enhance the accuracy and applicability of the Developmental Autism Early Screening (DAES). This novel screening tool for autism, based on the Griffiths III, has previously been shown to effectively distinguish children at risk of ASD from those with DD and TD in the first three years of life [Cirnigliaro et al., 2023].

In the present study, applying ML analyses in a larger sample, we improved the DAES validity in discriminating children with ASD-risk from DD or TD children, selecting the most predictive subset of items in two age ranges (≤ 24 months and > 24 months).

A priority of ASD screening tools is exploring potentially discriminative items within targeted age ranges (12-36 months) in an attempt to unravel the complexities of developmental trajectories in ASD children [Brewer et al., 2020]. Previously, an assessment of the developmental pathway of ASD at different ages was performed by integrating information from multiple functional domains and time points: a supervised classifier analysis was applied on Mullen Scales of Early Learning (MSEL), Vineland Adaptive Behavior Scales (VABS) and the Autism Observational Scale for Infants (AOSI) scores at 8 and 14 months, in order to predict the clinical outcome at 36 months. The study results, which showed moderate classifier performance (AUC=71%), may be explained by the high inter-individual variability in clinical symptoms and developmental trajectories of ASD [Bussu et al., 2018].

We tested five machine learning (ML) algorithms (random forest, RF, support vector classification, SVC, decision tree, DT, k-nearest neighbours, KNN, and logistic regression, LR) and selected the best performing classifiers (RF and SVC) for each age range, using a reduced subset of specific DAES items. We foresee that a shorter and faster version of the tool, derived from Griffiths III, might promptly intercept predictors of

atypical developmental trajectories in children at risk of ASD, when administered across different age ranges.

In addition, we have optimised the applicability of the DAES by demonstrating that it is possible to administer this tool based on chronological age rather than equivalent age obtained by Griffiths III. Applying Cramer's V index, we found a significant correlation between the equivalent age and some DAES items. Thanks to these correlations, the equivalent age does not need to be determined before administering the DAES, as some items intrinsically reflect it. This approach will make the DAES easier to administer regardless of discrepancies between equivalent and chronological age in children with ASD, enabling its wider use in various settings and thus reaching a larger proportion of the population.

To the best of our knowledge, no studies have yet investigated the application of ML to second-level interactive screening tools. However, many studies have applied ML to parent questionnaires (first-level screening tools) [Achenie et al., 2019; Tartarisco et al., 2021; Shahamiri et al., 2022] or to gold-standard diagnostic tools for autism, such as the ADOS and ADI-R [Levy et al., 2017; Yuwattana et al., 2025]. In addition, some studies have used home video recordings to train ML algorithms in order to identify behavioural patterns that discriminate between ASD and non-ASD children [Jin et al., 2024].

Previous studies demonstrated that ML algorithms can classify autistic vs non-autistic children with a small subset of items, selected from parent-report questionnaires, such as M-CHAT-R/F and QCHAT, maintaining high classification accuracy in order to create a quick, easy and high-performance tool in primary care setting [Achenie et al, 2019; Lu et al., 2024]. In particular, in a multicentre study, a supervised approach of binary classification (ASD/TD children) was applied to the Italian Q-CHAT data of 259 children with an age range between 22 and 43 months. Five algorithms (RF, Naïve Bayes, KNN, LR, SVM) were tested to understand the relationship between Q-CHAT items and the

diagnostic label. Among these, SVM showed the best discriminant validity, reaching an accuracy of about 93% using 14 out of the 25 items. The study demonstrated that a reduced number of items was sufficient to accurately predict an autism condition with implications for facilitating more effective and quick screening procedures [Tartarisco et al., 2021].

In a series of studies, ML was applied to datasets of different age groups, including Q-CHAT-10 data in toddlers (12-36 months) and AQ-10 data in children (4-11 years), adolescents (12-17 years) and adults (18 years) to classify ASD from TD cases, achieving high accuracy, especially in the child and toddler datasets (approximately 98%) [Akter et al., 2019; Thabtah et al., 2020; Vakadkar et al., 2021; Shrivastava et al., 2024]. Among these studies, the more recent one used various performance metrics, including the confusion matrix, the area under the ROC curve (AUC-ROC) and the F1 score, to validate the performance of the SVM, RF, KNN and artificial neural network (ANN) classifiers. RF was found to be the most effective classifier when using different training and testing data, achieving 100% accuracy for all four datasets [Shrivastava et al. 2024].

Following this line of evidence, we trained and selected the optimal parameters for each of the five machine learning algorithms (RF, SVC, DT, KNN and LR), using GridSearchCV with 10-fold cross-validation. Focusing on two age ranges (≤24 months and >24 months), we employed two classification strategies: binary (ASD/non-ASD) and multiclass (ASD, DD, and TD) within each age group. The effectiveness of each classification model was assessed using a confusion matrix on both the training and the test data (previously unseen by the system), divided by age group. For the binary classification (ASD/non-ASD), RF resulted the best model, achieving high train and test sensitivity (97% and 100% respectively) and train and test specificity (98% and 100% respectively) with a train FNR of 2.2 % and a test FNR of 0 in children aged ≤ 24 months. According to the confusion matrix results, SVC was the best predictive model in children

aged over 24 months, with an FNR of 0 in both training and testing, achieving higher sensitivity and specificity (100%) than the other algorithms.

The low FNR for both classifiers confirmed the robustness of our model for identifying children at risk of ASD at an early stage, facilitating timely access to targeted intervention prior to formal diagnosis, with a more favourable outcome.

The performance metrics of each classifier (train and test accuracy, test precision, test recall, test F1 score, ROC-AUC) were computed for both binary (ASD/non-ASD) and multiclass (ASD, DD and TD) classifications, considering the DAES selected items for each age group. The RF algorithm reached the best train (98%) and test accuracy (100%) in discriminating ASD from DD and TD children (non-ASD class) aged ≤ 24 months, considering the 21 selected DAES items (out of 36 initial items) in this age range. The SVC algorithm achieved higher accuracy values (100%) than the other algorithms in differentiating ASD/non-ASD children aged >24 months, considering the 28 selected DAES items (out of 36 initial items) in this age range.

Regarding the multiclass classification (ASD, DD and TD), SVC demonstrated the best accuracy in discriminating the three classes in ≤ 24-month age group (80%), whereas RF showed the best train and test accuracy in children aged over 24 months old (92% and 89%, respectively). The relatively lower accuracy of the classifiers in discriminating between the three groups (ASD, DD and TD) is related to the specific design of the DAES for identifying children at risk of ASD and distinguishing them from those with DD or TD. Consequently, it is to be expected that the DAES is less accurate in discriminating DD from TD children. Indeed, mean total DAES scores in the DD group were slightly higher but not significant different from TD children, while were significantly lower than the ASD group.

Notably, we found that the difference between training and testing accuracy was minimal. This ensures that there was no overfitting issue, considering that we used ten-fold cross-validation for model training and selection.

Further validation studies involving a larger sample of children are needed to confirm these findings. The performance of our models may be improved by testing other sophisticated ML algorithms that require a larger amount of data.

To date, only a few studies have investigated the ability of screening tools to differentiate between children at risk of ASD and those with DD or TD across different age ranges using ML. One such study focused on training two algorithms to identify children at-risk for autism, one based on structured parent-report questionnaires (M-CHAT and Child Behavior Checklist, CBCL) and the other based on semi-structured home video recordings. A combination algorithm was then used to merge the results from the classifiers into a single screening assessment. Specializing the ML model on a dichotomy of age groups (< 48 months and ≥ 48 months) revealed that the screener for younger children relied on non-verbal behavioral features, such as eye contact, gestures, and facial expressions, whereas the screener for older children focused more on verbal communication and interactions with other children [Abbas et al., 2018]. In the present study, we found similar results regarding the items selected for each age group, despite considering different lower age ranges. In particular, the 28 DAES items selected for children aged over 24 months mainly explore receptive and expressive verbal communication, self-concept, social cognition and empathy, whereas the 21 items for the ≤ 24-month age range focus on preverbal communication, emotional expression and joint attention. Selecting DAES items by age group enabled the screening tool to be refined using machine learning, which played a very important role in improving its discriminatory effectiveness.

6.1 Study limitations and future perspective

The present study has certain limitations. The main limitation concerns the DAES administration according to equivalent age, obtained from Griffiths III, since the DAES items were selected during the test development phase based on equivalent age. This limitation can be overcome by administering the DAES according to chronological age, considering that we found some items to be significantly correlated with equivalent age. Thanks to these correlations, the equivalent age does not need to be determined before administering the DAES, as some items intrinsically reflect it.

Furthermore, the third-year DAES item scores could not be included in the analysis, due to discrepancies between chronological and equivalent age in some children with ASD or DD, who did not complete the items related to the third year, unlike typically developing children of the same chronological age.

Further replication on large-scale independent samples with different ML models and settings is needed to translate this screening tool into clinical practice, including primary care.

We foresee that these limitations may be overcome in the next step of the workflow.

New data is currently being collected from a new sample of children at-risk of ASD by administering the DAES according to the chronological age.

In addition, an app based on the ML classifiers validated in the present study for specific age ranges is being developed. This app will enable the quick collection of a large amount of data by completing the DAES based on the child's chronological age. In this way, the model can be updated taking into account the features of the third year, which will be enabled on the app based on the child's age. The model can be adapted to the new data for continuous learning, in order to refine the classifiers, enhancing their accuracy and effectiveness. Furthermore, continuous data collection over an extended period will

enable the real-time monitoring of the child development trajectory, detecting trends and changes in behavior, which may indicate an ASD-risk.

The app will facilitate the integration of this screening tool into existing clinical workflows and healthcare systems, which is crucial for follow-up assessments and timely interventions.

7. CONCLUSIONS

In this study we applied ML to validate a novel ASD screening tool, the DAES, improving its ability to differentiate children at ASD-risk from DD or TD children. We explored and selected the potentially discriminative DAES items at two target age ranges (≤ 24 months and > 24 months), attempting to intercept specific atypical developmental patterns which might unravel predictors of the complex developmental trajectories in children with ASD. We trained five ML classifiers (RF, SVC, DT, LR, KNN) and selected the best algorithms to classify ASD versus DD and TD children at the two different age ranges. Our results showed that two ML classifiers (RF and SVC) were able to correctly detect ASD children with very high accuracy (above 98%) from a selection of 21 DAES items for children ≤ 24 months and 28 items for children > 24 months.

These findings confirm the validity of the DAES as an early screening tool for autism, as well as the potential of ML to improve its accuracy, using a reduced number of items that are predictive of ASD-risk for specific age ranges.

A shorter and easier-to-administer version of the tool will enable more effective and quick screening procedures and a widespread use of the tool, with important clinical implications for facilitating an earlier access to targeted intervention. This could enable atypical developmental trajectories to be promptly redirected towards a typical pathway in a larger number of children at risk of ASD.

8. REFERENCES

1. Abbas H, Garberson F, Glover E, Wall DP. Machine learning approach for early detection of autism by combining questionnaire and home video screening. *J Am Med Inform Assoc*. 2018;25(8):1000-1007. doi:10.1093/jamia/ocy039
2. Achenie LEK, Scarpa A, Factor RS, Wang T, Robins DL, McCrickard DS. A Machine Learning Strategy for Autism Screening in Toddlers. *J Dev Behav Pediatr*. 2019;40(5):369-376. doi:10.1097/DBP.0000000000000668
3. Aishworiya R, Ma VK, Stewart S, Hagerman R, Feldman HM. Meta-analysis of the Modified Checklist for Autism in Toddlers, Revised/Follow-up for Screening. *Pediatrics*. 2023;151(6):e2022059393. doi:10.1542/peds.2022-059393
4. Akobeng A.K. Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatr.* 2007;96(5):644-7. doi: 10.1111/j.1651-2227.2006.00178.x.
5. Akter T., Satu S., Khan I., et al. Machine learning-based models for early stage detection of autism spectrum disorders. *IEEE Access*. 2019; 7:8895818 166509-166527. https://doi.org/10.1109/ACCESS.2019.2952609
6. Allison, C.; Baron-Cohen, S.; Wheelwright, S.; Charman, T.; Richler, J.; Pasco, G.; Brayne, C. The Q-CHAT (Quantitative Checklist for Autism in Toddlers): A Normally Distributed Quantitative Measure of Autistic Traits at 18–24 Months of Age: Preliminary Report. *J. Autism Dev. Disord*. 2008, 38, 1414–1425.
7. Allison, C.; Auyeung, B.; Baron-Cohen, S. Toward Brief "Red Flags" for Autism Screening: The Short Autism Spectrum Quotient and the Short Quantitative CHecklist in 1000 Cases and 3000 Controls. *J. Am. Acad. Child Adolesc. Psychiatry* 2012, 51, 202–212.
8. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5), 5th ed. Washington, DC: American Psychiatric Association; 2013.
9. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th ed. (TR). 2022.
10. Attar SM, Ibanez LV, Stone WL. Separate scoring algorithms for specific identification priorities optimize the screening properties of the Screening Tool for Autism in Toddlers (STAT). *Autism Res*. 2022;15(11):2069-2080. doi:10.1002/aur.2799

11. Baadel S, Thabtah F, Lu J. A clustering approach for autistic trait classification. *Inform Health Soc Care*. 2020;45(3):309-326. doi:10.1080/17538157.2019.1687482
12. Bejarano-Martín Á, Canal-Bedia R, Magán-Maganto M, et al. Early Detection, Diagnosis and Intervention Services for Young Children with Autism Spectrum Disorder in the European Union (ASDEU): Family and Professional Perspectives. *J Autism Dev Disord.* 2020;50(9):3380-3394. doi:10.1007/s10803-019-04253-0
13. Bone, D., Bishop, S. L., Black, M. P., Goodwin, M. S., Lord, C., & Narayanan, S. S. (2016). Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion. *Journal of child psychology and psychiatry, and allied disciplines*, *57*(8), 927–937. https://doi.org/10.1111/jcpp.12559
14. Bougeard C, Picarel-Blanchot F, Schmid R, Campbell R, Buitelaar J. Prevalence of Autism Spectrum Disorder and Co-Morbidities in Children and Adolescents: A Systematic Literature Review. *Focus (Am Psychiatr Publ)*. 2024;22(2):212-228. doi:10.1176/appi.focus.24022005
15. Breiman L. Random Forests. *Machine Learning*. 2001;45:5–32. https://doi.org/10.1023/A:1010933404324.
16. Brewer N, Young RL, Lucas CA. Autism Screening in Early Childhood: Discriminating Autism From Other Developmental Concerns. *Front Neurol*. 2020;11:594381. doi:10.3389/fneur.2020.594381
17. Brian JA, Zwaigenbaum L, Ip A. Standards of diagnostic assessment for autism spectrum disorder. *Paediatr Child Health*. 2019;24(7):444-460. doi:10.1093/pch/pxz117
18. Bussu G, Jones EJH, Charman T, Johnson MH, Buitelaar JK; BASIS Team. Prediction of Autism at 3 Years from Behavioural and Developmental Measures in High-Risk Infants: A Longitudinal Cross-Domain Classifier Analysis. *J Autism Dev Disord*. 2018;48(7):2418-2433. doi:10.1007/s10803-018-3509-x
19. Calderoni S. Sex/gender differences in children with autism spectrum disorder: A brief overview on epidemiology, symptom profile, and neuroanatomy. *J Neurosci Res*. 2023;101(5):739-750. doi:10.1002/jnr.25000
20. Choueiri R, Wagner S. A New Interactive Screening Test for Autism Spectrum Disorders in Toddlers. *J Pediatr*. 2015;167(2):460-466. doi:10.1016/j.jpeds.2015.05.029

21. Choueiri R, Lindenbaum A, Ravi M, Robsky W, Flahive J, Garrison W. Improving Early Identification and Access to Diagnosis of Autism Spectrum Disorder in Toddlers in a Culturally Diverse Community with the Rapid Interactive screening Test for Autism in Toddlers. *J Autism Dev Disord*. 2021;51(11):3937-3945. doi:10.1007/s10803-020-04851-3
22. Choueiri R, Garrison WT, Tokatli V. Early Identification of Autism Spectrum Disorder (ASD): Strategies for Use in Local Communities. *Indian J Pediatr*. 2023;90(4):377-386. doi:10.1007/s12098-022-04172-6
23. Cirnigliaro L, Valle MS, Casabona A, et al. The Developmental Autism Early Screening (DAES): A Novel Test for Screening Autism Spectrum Disorder. *J Autism Dev Disord*. 2025;55(1):221-236. Epub 2023 Dec 18. doi:10.1007/s10803-023-06184-3
24. Cortese S, Bellato A, Gabellone A, et al. Latest clinical frontiers related to autism diagnostic strategies. *Cell Rep Med*. 2025;6(2):101916. doi:10.1016/j.xcrm.2024.101916
25. Dawson G, Rieder AD, Johnson MH. Prediction of autism in infants: progress and challenges. *Lancet Neurol*. 2023;22(3):244-254. doi:10.1016/S1474-4422(22)00407-0
26. Delehanty AD., Stronach S, Guthrie W, Slate E, Wetherby AM. Verbal and nonverbal outcomes of toddlers with and without autism spectrum disorder, language delay, and global developmental delay. *Autism & Developmental Language Impairments*. 2018; 3. doi:10.1177/2396941518764764
27. Dow D, Guthrie W, Stronach ST, Wetherby AM. Psychometric analysis of the Systematic Observation of Red Flags for autism spectrum disorder in toddlers. *Autism*. 2017;21(3):301-309. doi:10.1177/1362361316636760
28. Dow D, Day TN, Kutta TJ, Nottke C, Wetherby AM. Screening for autism spectrum disorder in a naturalistic home setting using the systematic observation of red flags (SORF) at 18-24 months. *Autism Res*. 2020;13(1):122-133. doi:10.1002/aur.2226
29. Erkan, U. & Thanh, D. N. H. Autism Spectrum Disorder Detection with Machine Learning Methods. *Curr. Psychiatry Res. Rev.* 2019;15(4): 297–308. doi:10.2174/2666082215666191111121115
30. Fombonne E. Camouflage and autism. *J Child Psychol Psychiatry*. 2020;61(7):735-738. doi:10.1111/jcpp.13296

31. Green E., Stroud L., Bloomfield S., Cronje J., Foxcroft C., Hurter K., et al. The Griffiths Scales of Childhood Development (3rd ed.). Oxford, UK: Hogrefe Ltd, 2016.
32. Hedley D, Nevill RE, Monroy-Moreno Y, et al. Efficacy of the ADEC in Identifying Autism Spectrum Disorder in Clinically Referred Toddlers in the US. *J Autism Dev Disord*. 2015;45(8):2337-2348. doi:10.1007/s10803-015-2398-5
33. Hirota T, King BH. Autism Spectrum Disorder: A Review. *JAMA*. 2023;329(2):157-168. doi:10.1001/jama.2022.23661
34. Hodis B, Mughal S, Saadabadi A. Autism Spectrum Disorder. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; January 17, 2025.
35. Horovitz M, Matson JL. The Baby and Infant Screen for Children with aUtIsm Traits-Part 1: Age-based scoring procedures. *J Dev Phys Dis.* 2014; *26*(1):1-22. https://doi.org/10.1007/s10882-013-9340-6
36. Hyman S.L., Levy S.E., Myers S.M.; Council on Children with Disabilities, Section on Developmental and Behavioral Pediatrics. Identification, Evaluation, and Management of Children With Autism Spectrum Disorder. *Pediatrics*, 2020; 145(1): e20193447. doi: 10.1542/peds.2019-3447.
37. Jansen J.M., Green E., Stroud L.A. and Watson M.B. Using the Griffiths III and Quartile Charts in assessing Autism Spectrum Disorder: a case study. *Journal of Education and Learning*, 2020; 9(1): 30-40. doi: 10.5539/jel.v9n1p30.
38. Jin L, Cui H, Zhang P, Cai C. Early diagnostic value of home video-based machine learning in autism spectrum disorder: a meta-analysis. *Eur J Pediatr*. 2024;184(1):37. doi:10.1007/s00431-024-05837-4
39. Johnson C.P., Myers S.M.; American Academy of Pediatrics Council on Children With Disabilities. Identification and evaluation of children with autism spectrum disorders. *Pediatrics*. 2007;120(5):1183–1215. doi: 10.1542/peds.2007-2361.
40. Kamp-Becker, I. Autism spectrum disorder in ICD-11-a critical reflection of its possible impact on clinical practice and research. *Mol. Psychiatry.* 2024; 29(3):633–638. https://doi.org/10.1038/s41380-023-02354-y.
41. Kearney MW. Cramér's V. In book: Sage Encyclopedia of Communication Research Methods. Editor: M. R. Allen. December 2017. doi:10.4135/9781483381411.n107
42. Kleinbaum D.G., Dietz K., Gail M., Klein M. Logistic Regression; Springer: Berlin, Germany, 2002.

43. Klin A. Translating advances in developmental social neuroscience into greater access to early diagnosis in autism spectrum disorder. *Medicina (B Aires)*. 2023;83 Suppl 2(Suppl 2):32-36.
44. Krishnappa Babu PR, Di Martino JM, Aiello R, et al. Validation of a Mobile App for Remote Autism Screening in Toddlers. *NEJM AI*. 2024;1(10):10.1056/AIcs2400510. doi:10.1056/AIcs2400510
45. Landa RJ, Gross AL, Stuart EA, Faherty A. Developmental trajectories in children with and without autism spectrum disorders: the first 3 years. *Child Dev*. 2013;84(2):429-442. doi:10.1111/j.1467-8624.2012.01870.x
46. Landa RJ, Reetzke R, Holingue CB, Herman D, Hess CR. Diagnostic Stability and Phenotypic Differences Among School-Age Children Diagnosed With ASD Before Age 2. *Front Psychiatry*. 2022;13:805686. doi:10.3389/fpsyt.2022.805686
47. Lebersfeld JB, Swanson M, Clesi CD, O'Kelley SE. Systematic Review and Meta-Analysis of the Clinical Utility of the ADOS-2 and the ADI-R in Diagnosing Autism Spectrum Disorders in Children. *J Autism Dev Disord*. 2021;51(11):4101-4114. doi:10.1007/s10803-020-04839-z
48. Lemay JF, Amin P, Langenberger S, McLeod S. Experience with the Rapid Interactive Test for Autism in Toddlers in an Autism Spectrum Disorder Diagnostic Clinic. *J Dev Behav Pediatr*. 2020;41(2):95-103. doi:10.1097/DBP.0000000000000730
49. Levy, S., Duda, M., Haber, N., and Wall, D.P. (2017). Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. *Mol. Autism.* 2017;8:65. doi:10.1186/s13229-017-0180-6.
50. Li HH, Wang CX, Feng JY, Wang B, Li CL, Jia FY. A Developmental Profile of Children With Autism Spectrum Disorder in China Using the Griffiths Mental Development Scales. *Front Psychol*. 2020;11:570923. doi:10.3389/fpsyg.2020.570923
51. Lidstone DE. Missed Early Intervention Opportunities for Children With Autism Spectrum Disorder. *Autism Res*. 2025;18(5):1097-1103. doi:10.1002/aur.70043
52. Lord C., Rutter M., Le Couteur A. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord.* 1994; 24:659–85. doi: 10.1007/BF02172145.

53. Lord C., Rutter M., Di Lavore P.C., Risi S., Gotham K., Bishop S. Autism Diagnostic Observation Schedule, Second Edition. Torrance: Western Psychological Services, 2012.
54. Lord C, Brugha TS, Charman T, et al. Autism spectrum disorder. Nat Rev Dis Primers. 2020;6(1):5. doi:10.1038/s41572-019-0138-4
55. Lu H, Zhang H, Zhong Y, Meng XY, Zhang MF, Qiu T. A machine learning model based on CHAT-23 for early screening of autism in Chinese children. *Front Pediatr*. 2024;12:1400110. doi:10.3389/fped.2024.1400110
56. Maenner, M.J., Warren, Z., Williams, A.R., et al. Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2020. *MMWR. Surveill. Summ.* 2023;72(2):1–14. https://doi.org/10.15585/mmwr.ss7202a1.
57. Maenner M.J., Shaw K.A., Bakian A.V., Bilder D.A., Durkin M.S., Esler A., et al. Prevalence and characteristics of Autism Spectrum Disorder among children aged 8 years - Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2018. *MMWR Surveill. Summ.* 2021;70(11):1-16. https://doi.org/10.15585/mmwr.ss7011a1.
58. Magán-Maganto M., Bejarano-Martín A., Fernández-Alvarez C., Narzisi A., García-Primo P., Kawa R., et al. Early Detection and Intervention of ASD: A European Overview. *Brain Sci.*, 2017; 7(12): 159. doi:10.3390/brainsci7120159.
59. Matson JL, Boisjoli JA, Hess JA, Wilkins J. Factor structure and diagnostic fidelity of the Baby and Infant Screen for Children with aUtIsm Traits-Part 1 (BISCUIT-part 1). *Dev Neurorehabil*. 2010;13(2):72-79. doi:10.3109/17518420903213576
60. McCarty P, Frye RE. Early Detection and Diagnosis of Autism Spectrum Disorder: Why Is It So Difficult?. *Semin Pediatr Neurol*. 2020;35:100831. doi:10.1016/j.spen.2020.100831
61. McHugh ML. The chi-square test of independence. *Biochem Med (Zagreb)*. 2013;23(2):143-149. doi:10.11613/bm.2013.018
62. Megerian, J. T., Dey, S., Melmed, R. D., Coury, D. L., Lerner, M., Nicholls, C. J., Sohl, K., Rouhbakhsh, R., Narasimhan, A., Romain, J., Golla, S., Shareef, S., Ostrovsky, A., Shannon, J., Kraft, C., Liu-Mayo, S., Abbas, H., Gal-Szabo, D. E., Wall, D. P., & Taraman, S. (2022). Evaluation of an artificial intelligence-based medical device for diagnosis of autism spectrum disorder. *NPJ digital medicine*, *5*(1), 57. https://doi.org/10.1038/s41746-022-00598-6

63. Miller LE, Dai YG, Fein DA, Robins DL. Characteristics of toddlers with early versus later diagnosis of autism spectrum disorder. *Autism*. 2021;25(2):416-428. doi:10.1177/1362361320959507
64. Minissi ME, Chicchi Giglioli IA, Mantovani F, Alcañiz Raya M. Assessment of the Autism Spectrum Disorder Based on Machine Learning and Social Visual Attention: A Systematic Review. *J Autism Dev Disord*. 2022;52(5):2187-2202. doi:10.1007/s10803-021-05106-5
65. Nah YH, Young RL, Brewer N, Berlingeri G. Autism detection in early childhood (ADEC): reliability and validity data for a Level 2 screening tool for autistic disorder. *Psychol Assess*. 2014;26(1):215-226. doi:10.1037/a0034472
66. Narzisi A., Costanza C., Ballottin U., Muratori F. Non-pharmacological treatments in autism spectrum disorders: an overview on early interventions for pre-schoolers. *Curr. Clin. Pharmacol.,* 2014; 9:17–26. doi: 10.2174/15748847113086660071.
67. Narzisi A, Posada M, Barbieri F, et al. Prevalence of Autism Spectrum Disorder in a large Italian catchment area: a school-based population study within the ASDEU project. *Epidemiol Psychiatr Sci*. 2018;29:e5. doi:10.1017/S2045796018000483
68. Perochon, S., Di Martino, J. M., Carpenter, K. L. H., Compton, S., Davis, N., Eichner, B., Espinosa, S., Franz, L., Krishnappa Babu, P. R., Sapiro, G., & Dawson, G. (2023). Early detection of autism using digital behavioral phenotyping. *Nature medicine*, *29*(10), 2489–2497. https://doi.org/10.1038/s41591-023-02574-3
69. Peterson LE. K-Nearest Neighbor. *Scholarpedia* 2009;4: 1883.
70. Petrocchi, S., Levante, A., & Lecciso, F. Systematic Review of Level 1 and Level 2 Screening Tools for Autism Spectrum Disorders in Toddlers. *Brain Sci.* 2020; 10(3), 180. doi.org/10.3390/brainsci10030180
71. Pileggi ML, Brane N, Bradshaw J, et al. Early Observation of Red Flags in 12-Month-Old Infant Siblings Later Diagnosed With Autism Spectrum Disorder. *Am J Speech Lang Pathol*. 2021;30(4):1846-1855. doi:10.1044/2020_AJSLP-20-00165
72. Pino MC, Donne IL, Vagnetti R, Tiberti S, Valenti M, Mazza M. Using the Griffiths Mental Development Scales to Evaluate a Developmental Profile of Children with Autism Spectrum Disorder and Their Symptomatologic Severity. *Child Psychiatry Hum Dev*. 2024;55(1):117-126. doi:10.1007/s10578-022-01390-z
73. Robins DL, Casagrande K, Barton M, Chen CM, Dumont-Mathieu T, Fein D. Validation of the modified checklist for Autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics*. 2014;133(1):37-45. doi:10.1542/peds.2013-1813

74. Rosen NE, Lord C, Volkmar FR. The Diagnosis of Autism: From Kanner to DSM-III to DSM-5 and Beyond. *J Autism Dev Disord.* 2021;51(12):4253-4270. doi:10.1007/s10803-021-04904-1
75. Ruta, L.; Arduino, G.M.; Gagliano, A.; Apicella, F.; Leonardi, E.; Famà, F.I.; Chericoni, N.; Costanzo, V.; Turco, N.; Tartarisco, G. Psychometric Properties, Factor Structure and Cross-Cultural Validity of the Quantitative CHecklist for Autism in Toddlers (Q-CHAT) in an Italian Community Setting. *Res. Autism Spectr. Disord.* 2019, 64, 39–48.
76. Shaw KA, Bilder DA, McArthur D, et al. Early Identification of Autism Spectrum Disorder Among Children Aged 4 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2020. *MMWR Surveill Summ*. 2023;72(1):1-15. doi:10.15585/mmwr.ss7201a1
77. Scattoni ML, Fatta LM, Micai M, et al. Autism spectrum disorder prevalence in Italy: a nationwide study promoted by the Ministry of Health. *Child Adolesc Psychiatry Ment Health*. 2023;17(1):125. Published 2023 Oct 28. doi:10.1186/s13034-023-00673-0
78. Shahamiri SR, Thabtah F, Abdelhamid N. A new classification system for autism based on machine learning of artificial intelligence. *Technol Health Care*. 2022;30(3):605-622. doi:10.3233/THC-213032
79. Shrivastava T, Singh V, Agrawal A. Autism spectrum disorder detection with kNN imputer and machine learning classifiers via questionnaire mode of screening. *Health Inf Sci Syst*. 2024;12(1):18. doi:10.1007/s13755-024-00277-8
80. Siddiqui S, Gunaseelan L, Shaikh R, Khan A, Mankad D, Hamid MA. Food for Thought: Machine Learning in Autism Spectrum Disorder Screening of Infants. *Cureus*. 2021;13(10):e18721. doi:10.7759/cureus.18721
81. Stone WL, Coonrod EE, Ousley OY. Brief report: screening tool for autism in two-year-olds (STAT): development and preliminary data. *J Autism Dev Disord*. 2000;30(6):607-612. doi:10.1023/a:1005647629002
82. Stone WL, Coonrod EE, Turner LM, Pozdol SL. Psychometric properties of the STAT for early autism screening. *J Autism Dev Disord*. 2004;34(6):691-701. doi:10.1007/s10803-004-5289-8
83. Stroud, L., Foxcroft, C., Green, E., Bloomfield, S., Cronje, J., Hurter, K., Lane, H., Marais, R., Marx, C., McAlinden, P., O'Connell, R., Paradice, R., & Venter, D.

Griffiths Scales of Child Development (3rd Edition) Part I: Overview, development, and psychometric properties. Oxford, United Kingdom: Hogrefe Ltd, 2016.

84. Taddei M, Bulgheroni S, Toffalini E, Pantaleoni C, Lanfranchi S. Developmental profiles of young children with autism spectrum disorder and global developmental delay: A study with the Griffiths III scales. *Autism Res*. 2023;16(7):1344-1359. doi:10.1002/aur.2953
85. Tartarisco G, Cicceri G, Di Pietro D, et al. Use of Machine Learning to Investigate the Quantitative Checklist for Autism in Toddlers (Q-CHAT) towards Early Autism Screening. *Diagnostics (Basel)*. 2021;11(3):574. Published 2021 Mar 22. doi:10.3390/diagnostics11030574
86. Thabtah F., Peebles D. Early Autism Screening: A Comprehensive Review. *Int. J. Environ. Res. Public Health*. 2019; 16:3502. doi:10.3390/ijerph16183502.
87. Thabtah F. An accessible and efficient autism screening method for behavioural data and predictive analyses. *Health Informatics J*. 2019;25(4):1739-1755. doi:10.1177/1460458218796636
88. Thabtah F, Peebles D. A new machine learning model based on induction of rules for autism detection. *Health Inform J.* 2020;26(1):264–86. doi:10.1177/14604 58218 824711.
89. Thabtah F, Spencer R, Abdelhamid N, et al. Autism screening: an unsupervised machine learning approach. *Health Inf Sci Syst*. 2022;10(1):26. doi:10.1007/s13755-022-00191-x
90. Torrens V, Ruiz C. Language and Communication in Preschool Children with Autism and Other Developmental Disorders. *Children (Basel)*. 2021;8(3):192. doi:10.3390/children8030192
91. Towle PO, Patrick PA. Autism Spectrum Disorder Screening Instruments for Very Young Children: A Systematic Review. *Autism Res Treat*. 2016;2016:4624829. doi:10.1155/2016/4624829
92. Talantseva OI, Romanova RS, Shurdova EM, et al. The global prevalence of autism spectrum disorder: A three-level meta-analysis. *Front Psychiatry* 2023;14:1071181. doi:10.3389/fpsyt.2023.1071181
93. Vakadkar K, Purkayastha D, Krishnan D. Detection of autism spectrum disorder in children using machine learning techniques. *SN Comput Sci.* 2021;2(5):1–9. doi:10. 1007/s42979-021-00776-5.

94. Wall, D. P., Liu-Mayo, S., Salomon, C., Shannon, J. & Taraman, S. Optimizing a de novo artificial intelligence-based medical device under a predetermined change control plan: Improved ability to detect or rule out pediatric autism. *Intell.-Based Med.* 2023;8:100102.
95. Wang H, Du Y, Mao Z, et al. Use of the Griffiths mental development scale-Chinese in the assessment of children with autism spectrum disorder and global developmental delay/intellectual disability. *Medicine (Baltimore)*. 2021;100(13):e25407. doi:10.1097/MD.0000000000025407
96. Wei, Q., Dong, W., Yu, D., Wang, K., Yang, T., Xiao, Y., Long, D., Xiong, H., Chen, J., Xu, X., & Li, T. (2024). Early identification of autism spectrum disorder based on machine learning with eye-tracking data. *Journal of affective disorders*, *358*, 326–334. https://doi.org/10.1016/j.jad.2024.04.049
97. World Health Organization (WHO). International Statistical Classification of Diseases and Related Health Problems (11th ed). 2021.
98. Wieckowski AT, Williams LN, Rando J, Lyall K, Robins DL. Sensitivity and Specificity of the Modified Checklist for Autism in Toddlers (Original and Revised): A Systematic Review and Meta-analysis. *JAMA Pediatr*. 2023;177(4):373-383. doi:10.1001/jamapediatrics.2022.5975
99. Yu Y, Ozonoff S, Miller M. Assessment of Autism Spectrum Disorder. *Assessment*. 2024;31(1):24-41. doi:10.1177/10731911231173089
100. Yuwattana W, Saeliw T, van Erp ML, et al. Machine learning of clinical phenotypes facilitates autism screening and identifies novel subgroups with distinct transcriptomic profiles. *Sci Rep*. 2025;15(1):11712. doi:10.1038/s41598-025-95291-5
101. Zwaigenbaum L, Penner M. Autism spectrum disorder: advances in diagnosis and evaluation. *BMJ*. 2018;361:k1674. doi:10.1136/bmj.k1674