Check for updates

ARTICLE   OPEN

Predictive modeling of adaptive behavior trajectories in autism: insights from a clinical cohort study

Annie Aitken [1,2✉], Maia C. Lazerwitz[1], Ally Eash[1], Neil Hattangadi[1], Pratik Mukherjee[3], Elysa J. Marco[1] and Kevin A. Shapiro[1,4]

© The Author(s) 2025

Research aimed at understanding how baseline clinical and demographic characteristics influence outcomes over time is critically important to inform individualized therapeutic programs for children with neurodevelopmental differences. This study characterizes adaptive behavior trajectories in children receiving medical and behavioral therapy within a network of care centers with a shared data-gathering mechanism for intake and longitudinal assessments. We then take the further step of utilizing intake data to develop machine-learning models which predict differences in those trajectories. Specifically, we evaluated data from 1225 autistic children, aged 20-90 months, using latent class growth mixture modeling (LCGMM) with scores on the Vineland Adaptive Behavior Scales, 3rd Edition, as the primary outcome measure. The LCGMM analysis revealed two distinct clusters of adaptive behavior trajectories. The "Less Impairment/Improving Trajectory" group (≥66% of the sample) exhibited greater developmental change in adaptive behavior, while the "Higher Impairment/Stable Trajectory" group (≤33% of the sample) showed little change over time relative to age-matched normative data. For a subset of 729 children, we used machine learning algorithms to forecast adaptive behavior trajectories using clinical and sociodemographic data collected at the initial assessment, comparing elastic net GLM, support vector machine, and random forest. The best-performing random forest model predicted adaptive behavior trajectory with an accuracy rate of 77%. The strongest predictors in our model were socioeconomic status, history of developmental regression, child temperament, paternal age at the time of the child's birth, baseline autism symptom severity, parent concerns about development, presence of ADHD symptoms, and parent concerns about mood. Notably, the inclusion of cumulative hours of applied behavioral analysis and developmental therapies in the machine learning models did not yield significant changes in performance metrics, indicating that increased therapy hours did not predict greater improvement. These findings extend our understanding of adaptive behavior development in autistic children and underscore the value of gathering comprehensive patient information at intake to tailor clinical care.

BACKGROUND

The diagnostic category of Autism Spectrum Disorder (ASD) encompasses a range of neurodevelopmental differences that result in atypical trajectories in the development of adaptive functions, particularly in the domains of socialization and communication [1, 2]. In clinical practice, the assessment of adaptive functioning skills, often referred to as "adaptive behaviors," is routinely used as a proxy for real-world ability and disability in autistic individuals over the course of the lifespan [3, 4].

Recently, researchers have made significant efforts to characterize patterns of adaptive behavior outcomes in autism. These efforts have revealed substantial variation, with some children exhibiting more favorable developmental changes compared to others. This information is highly relevant to clinicians who seek to develop targeted therapeutic and medical approaches based on an individual's unique psychosocial profile and neurobehavioral phenotype. Identifying how individuals may respond differently to clinical care is fundamental both to personalizing programs of intervention and to improving care delivery more broadly. Thus,

there is a pressing need to identify predictors of differential trajectories of adaptive behavior change within the context of clinical care from the point of intake.

The term "autism spectrum" itself underscores the inherent heterogeneity within the diagnostic construct. Individuals with an autism diagnosis differ in numerous aspects ranging from genetic risk factors to early-life experiences and sociodemographic variables, leading to diverse clinical and developmental outcomes [5]. One statistical technique that has been used to characterize the multifaceted nature of autism outcomes is latent class growth mixture modeling (LCGMM), which can capture variability in behavioral measures longitudinally from childhood to adulthood [6–9]. LCGMM offer several advantages over multilevel linear growth models, which are a traditional method for characterizing change and are particularly useful for modeling continuous change. In comparison, mixture models capture heterogeneity in patterns of change, enabling the identification of distinct trajectories that align more closely with real-world clinical observations. Prior studies have primarily analyzed standardized measures of adaptive behavior skills, such as the Vineland

[1]Cortica Healthcare, San Diego, CA, USA. [2]Department of Applied Psychology, New York University, New York, NY, USA. [3]University of California, San Francisco, CA, USA. [4]Boston College, Boston, MA, USA. ✉email: aitkenannie@gmail.com

Adaptive Behavior Scales (VABS) [2], in children drawn from community samples without regard to specific interventions or treatments.

Collectively, these studies have revealed that a significant proportion of autistic children follow trajectories characterized by little to no growth in adaptive behavior [6–10]. In other words, many children with a diagnosis of autism continue to lag significantly behind their peers in the development of adaptive behavior skills, despite presumably receiving some type of therapy within the community. However, the contribution of standard therapeutic interventions has not been comprehensively accounted for, as relevant clinical data are lacking in most research cohorts. It should also be noted that while research samples tend to emphasize psychometric characteristics, such as IQ, in predicting group membership, clinically relevant issues such as comorbid conditions and parent-reported concerns may not be routinely assessed.

On the other hand, research into the effectiveness of clinical interventions for autism often does not sufficiently account for heterogeneity in the population of individuals with that diagnostic label. This has resulted in a "one-size-fits-all" approach to clinical intervention, typically comprising numerous hours of intensive behavioral interventions such as Applied Behavior Analysis (ABA) therapy and ad hoc provision of other developmental therapies including speech-language therapy, occupational therapy, and physical therapy [11]. In theory, a more individually tailored approach to therapy could offer improvements in both cost efficiency and clinical efficacy, substantially reducing the lifelong socioeconomic burdens associated with an autism diagnosis.

Psychosocial, demographic, and longitudinal outcome data from various sources can be analyzed holistically using machine learning (ML) approaches, which have gained significant traction as a tool for predicting clinical outcomes in various mental health conditions [12]. In the realm of autism research, ML prediction models have been applied to identify subtypes of autism based on clinical data [7, 13], to develop precise diagnostic models [14, 15], and, more recently, to predict differential responses to medical treatments [16]. However, there is a paucity of studies employing ML to predict differential functional behavior outcomes among individuals with autism, and fewer still that systematically account for the impact of standard therapeutic interventions. From the perspective of autism care, therefore, there is a pressing need for studies that adopt a data-driven approach to explore predictors of variation in response to treatment. Such efforts should enhance the precision of intervention strategies and provide tailored support to individuals with autism, ultimately maximizing these individuals' overall quality of life [17].

Current study
This study aimed to characterize the variability in adaptive behavior outcomes among autistic children, and to identify key clinical predictors of differential outcomes. The research described here is unique in part because it utilizes data gathered within the course of clinical care, rather than in the isolated context of a research study. Thus, in addition to adaptive behavior outcomes data from the Vineland Adaptive Behavior Scales, 3rd Edition (VABS-3), we had access to in-depth clinical intake information, including baseline socio-demographic data as well as standardized parent ratings for a variety of neurodevelopmental concerns and medical comorbidities. We were also able to gather information about the type and duration of therapies children received.

The study had two primary objectives. First, we employed latent class growth mixture modeling (LCGMM) to uncover distinct trajectories of adaptive behavior change in a cohort of children (N = 1225) receiving clinical care within a network of centers with a uniform approach to autism treatment. We evaluated the Adaptive Behavior Composite (ABC) score and the standardized score from each subdomain on the VABS-3 to facilitate comparisons with previous latent class trajectory research [6, 8, 9, 11].

Second, utilizing predictive ML algorithms, we investigated the importance of socio-demographic variables and clinical phenotypes in predicting membership in one of the observed trajectories of adaptive behavior change identified through LCGMM analysis. We also evaluated whether including information about therapeutic intensity - operationalized as cumulative hours of ABA and developmental therapies - would substantially alter the prediction of outcomes of our models. What makes this sample unique is that for each autistic child receiving care, the parent or caregiver completed a detailed standardized intake form prior to the initial visit, providing a rich source of demographic, historical, and baseline clinical data. Caregivers also reported adaptive behavior skills using the VABS-3 [18] every six months as part of routine clinical care.

RESULTS
Dataset description
This study leveraged deidentified clinical data from a network of 13 clinical sites with a uniform operational and therapeutic approach to autism care (Cortica Healthcare). The network included 8 sites in California, 2 sites in Texas, 2 sites in Illinois, and 1 site in New Jersey. The study was conducted according to principles outlined in the Declaration of Helsinki, and was approved by the WIRB-Copernicus Group, Institutional review board (protocol # 20224562). We received a waiver of informed consent from the institutional review board given that all aspects of this study were based on deidentified retrospective clinical data.

We queried Cortica's clinical research database for children with an autism diagnosis between the ages of 19 and 90 months who had at least two assessments using the (VABS-3) from 2016 through 2023 (N = 1225). All participants met clinical criteria for Autism Spectrum Disorder based on the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition, as assessed by physicians and nurse practitioners with expertise in autism diagnosis. The clinical assessment was supported by one or more of three standardized diagnostic instruments: the Autism Diagnostic Interview – Revised; the Autism Diagnostic Observation Schedule, 2nd Edition; or the Childhood Autism Rating Scale, 2nd Edition.

The full sample of 1225 participants was included in the latent class growth mixture modeling (LCGMM) analysis. For a subset of these participants (N = 729), we had access to comprehensive socio-demographic, historical, and clinical intake information provided prior to their initial visit (See Table 1 for full variable list). We used this subset of patients for the ML analysis. Full sample demographics are presented in Supplemental Information Table 1.

Clusters of adaptive behavior growth trajectories
We used conditional LCGMM to identify distinct growth profiles of adaptive behavior change using repeated measures VABS-3 assessments in the Adaptive Behavior Composite, Socialization, Communication, and Daily Living Score (DLS) domains. We included initial VABS-3 symptom severity [19] as a covariate, as past literature has demonstrated a significant influence of baseline adaptive behavior levels on resultant outcomes [7]. For each domain, the 2-profile model was selected as the best fitting model, based on the highest entropy value, a low BIC, a significant LMR test, and each profile representing at least 5% of the entire sample (see Supplemental Information Table 2 for LCGMM iterative results). See Table 2 for details on the best fitting model for each VABS-3 outcome. The ABC, Socialization and Communication domains all showed a similar pattern of class distribution. Within these domains, class labels were determined based on the

Table 1. All variables included in machine learning analysis.

Time of Data Collection	Domain	Variable
Intake Prior to Patient Initial Visit	Sociodemographic	Gender
		Age at first Assessment
		Family Socioeconomic Status (SES)
	Historical	Prenatal Stress
		Prenatal/ Post-Partum Depression
		Pregnancy Problems
		Premature Labor
		Baby Temperament
		Maternal Age at Birth
		Paternal Age at Birth
		Developmental Regression
	Clinical Phenotyping	Sleep Problems
		Attention Deficit Hyperactivity (ADHD) Symptoms
		Autism Symptoms
		Sensory Processing Disorder Symptoms
		General Development Concerns
		Motor Development Concerns
		Sensory Over-Reactivity Concerns
		Communication-Language Concerns
		Impulsivity Concerns
		Attention Concerns
		Repetitive Behavior Concerns
		Mood Concerns
		Feeding Concerns
Over the Course of Care	Therapeutic Dosage	Cumulative Applied Behavioral Analysis (ABA) Hours
		Cumulative Developmental Therapy (DT) Hours

Variables listed were included as candidate predictors in the machine learning analyses. The table categorizes each variable by the time of data collection (either at intake or over the course of care), domain, and specific variable name.

directionality of the estimated slope and starting intercept (Less Impairment/Improving Trajectory; Higher Impairment/Stable Trajectory). The groups with negative slope coefficients were described as "stable" because the slope was not significantly different than zero, and scores on the VABS-3 reflect an individual's behavior relative to an age-matched normative sample (See Fig. 1). Thus, a negative slope reflects a deviation from typical change trajectories but does not imply regression of skills.

Within the DLS subdomain, entropy was high, but class distribution was highly unequal. While the high entropy value suggests clear separation, the second class comprised less than 1% of individuals, indicating that most individuals clustered within a single, stable class, while the second class likely represented outliers rather than a meaningful latent subgroup. This pattern suggests that variability in daily living scores over time is best captured within a single trajectory rather than distinct latent classes. For this reason, DLS scores were excluded from subsequent predictive learning group classification.

Definition of outcome groups
To improve balance between classification groups, we created a dummy variable that corresponded to a 1 (Stable/Limited Gains) if children fell into the stable LCGMM class for *any* of the VABS-3 outcomes (Adaptive Behavior Composite, Socialization, or Communication) and a 0 (Multi-Domain Improvement) if they fell into the improved LCGMM cluster for *all* the VABS-3 outcomes (see Fig. 2). Using this definition, in the full sample, 339 children (39%) were

classified as having Stable/Limited Gains, while 879 children (61%) demonstrated Multi-Domain Improvement. In the subset of children included in the predictive modeling analysis, 188 (35%) were classified as having Stable/Limited Gains, while 541 (65%) demonstrated Consistent Multi-Domain Improvement. For descriptive statistics of all predictor variables with the ML subgroup stratified by defined outcome group, see Supplemental Information Table 3.

Prediction models to differentiate stable and improved adaptive behavior trajectories
The goal of the machine learning analysis was to develop and evaluate prediction models capable of differentiating between Stable and Improved adaptive behavior trajectories with adequate accuracy. After data pre-processing and feature selection (see Methods), we tested the performance of six machine learning algorithms using 13 clinical intake variables with nested cross-validation (See Fig. 3). The three models employed and compared elastic net logistic regression, support vector machine, and random forest to predict group membership with the following sets of features: socio-demographic (age at first assessment, gender, socioeconomic status), historical (maternal age at birth, paternal age at birth, pre/perinatal mental health, infant temperament, history of developmental regression), and clinical phenotype (autism symptoms, sleep problems, attention deficit hyperactivity disorder (ADHD) symptoms, sensory processing disorder (SPD) symptoms, mood concerns, motor development concerns, communication/language concerns, seizures, tics or tremors, headaches, general developmental concerns, feeding problems, impulsivity, repetitive behaviors).

Table 2. Data-Driven clusters of adaptive behavior trajectories.

ABC			
Entropy Value	0.83		
	Intercept (se)	Slope (se)	N
Group 1: Improved Behavior	72.3 (0.20)**	0.82 (0.21)**	1113 (91.3)
Group 2: Higher Impairment/Stable Trajectory	44.7 (1.90)**	−1.81 (0.55)*	106 (8.7%)
Socialization			
Entropy Value	0.74		
	Intercept	Slope	N
Group 1: Less Impairment/Improving Trajectory	73.5 (0.60)**	0.84 (0.34)**	1008 (82.7)
Group 2: Higher Impairment/Stable Trajectory	44.4 (1.50)**	−0.04 (0.68)^{ns}	211 (17.3%)
Communication			
Entropy Value	0.83		
	Intercept	Slope	N
Group 1: Less Impairment/Improving Trajectory	77.2 (0.52)**	1.87 (0.30)**	918 (75.3%)
Group 2: Higher Impairment/Stable Trajectory	40.9 (0.94)**	−1.21 (0.73)^{ns}	301 (24.7%)
Daily Living Score			
Entropy Value	0.97		
	Intercept	Slope	N
Group 1: Less Impairment/Stable Trajectory	71.7 (0.25)**	0.25 (0.22)^{ns}	1213 (99.5%)
Group 2: Outlier Group	102.4 (5.0)**	−22.4 (0.5)**	6 (0.05%)

Latent class growth mixture modeling (LCGMM) results for each Vineland-3 subdomain. Higher entropy values reflect greater class separation.
$p < 0.01 =$**, $p < 0.05 =$ *, $p < 0.10 =$ ^, $p > 0.10 =$ns.

Model validation and tuning

The first set of machine learning models evaluated the predictive value of socio-demographic, historical, and phenotypic intake variables for discriminating between individuals who fell into the Limited Functional Gains behavior group versus the Multi-Domain Improvement behavior group. Table 3 presents the accuracy and AUC-ROC for each algorithm during the 10-fold cross validation. The random forest model out preformed the others with a validation AUC-ROC of 0.73. The elastic net logistic regression and support vector machine reached a validation AUC-ROC of 0.72 and 0.71, respectively. The secondary metric used to evaluate ML model selection during development was accuracy. All algorithms achieved a 74% accuracy rate. See Supplementary Table 4a–c for ML model results on individual VABS-3 subdomain outcome.

Adding behavioral and developmental therapy hours as predictors

We conducted secondary machine learning analyses to evaluate whether differences in Applied Behavior Analysis (ABA) or developmental therapy (DT) dosage would predict Limited Functional Gains versus Multi-Domain Improvement group membership. Specifically, we included total hours of ABA therapy (direct services, i.e., including one-on-one therapy with the child and excluding hours of supervision among ABA providers) and developmental therapy. The model's performance metrics showed little to no change, resulting in a random forest ROC-AUC of 0.75 and an accuracy rate of 75%. Full details on the model's performance can be found in the Supplementary Information (SI Table 5). To mitigate model complexity and prevent overfitting, we evaluated our final validation model excluding hours of ABA or DT therapy in the final random forest model.

Final model external validation performance

Using the random forest algorithm, our final model demonstrated similar performance on the holdout validation set, yielding a ROC-AUC of 0.77 and an accuracy rate of 77.5%. Moreover, it exhibited a precision rate of 78.2%, a recall rate of

96.3%, and an F1 score of 0.863. Confusion matrices and predicted probability plots are depicted in Fig. 4a, b.

Predictive performance of variables

Next, we investigated the contributions of individual predictor variables to our model's accuracy to gain a better understanding of which factors most strongly influence adaptive behavior trajectories. Variable importance was identified using Shapley values, which reflect the additive importance of a particular variable in a model-agnostic fashion (see Fig. 5a). We considered variables with Shapley values above the grand average (0.016) to be the most significant predictors.

In the socio-demographic domain, we found that lower socioeconomic status (SES) was the most important predictor of group membership in the Stable/Limited Gains cohort. In the domain of prenatal and early-life experiences, a history of developmental regression, parent report of infant temperament, and father's age at birth were most predictive of group membership. Finally, among variables related to phenotype, severity of autism symptoms and ADHD symptoms, as well as parent-reported concerns about general development, motor development, communication/language, and mood were most predictive of adaptive function trajectory group membership (see Fig. 5b).

Directionally, children in the Stable/Limited Gains group had younger fathers, were more likely to have had a difficult temperament in infancy, and were more likely to have demonstrated a loss in previously acquired developmental skills. Children in the Stable/Limited Gains group also had more severe symptoms of autism and ADHD at the time of intake, with parents more commonly indicating concern for general development, motor development, communication/language, and mood.

DISCUSSION

In this study, we characterized trajectories of adaptive behavior change in a clinical cohort. Latent class models identified two distinct clusters of adaptive behavior change: a Less Impairment/
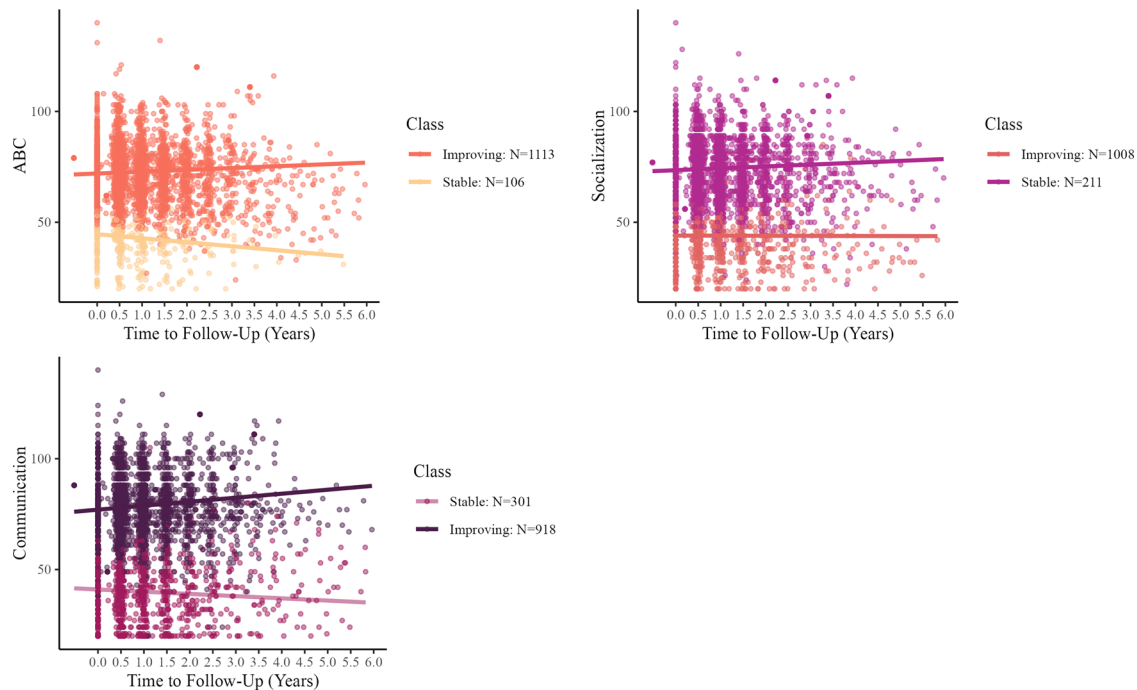
Fig. 1 LCGMM classes and trajectories. Plot depicting the estimated average slope identified from the latent class growth mixture models for each identified latent class trajectory cluster (Lower Impairment/Improving Trajectory and Higher Impairment/Stable Trajectory). Individual raw data points are overlayed.
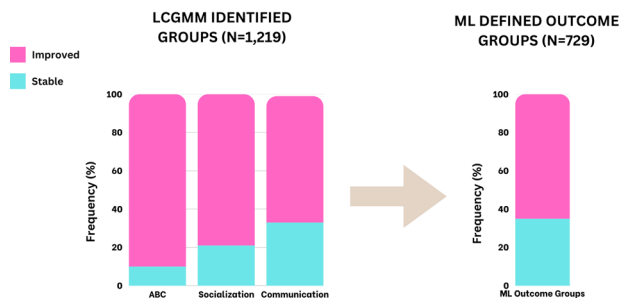

Fig. 2 Definition of outcome groups for machine learning. Stacked bar chart representing the adaptive behavior outcome group defined from the latent class growth mixture model.

Improving Trajectory group, with adaptive behavior scores that approximated age-matched norms over time, and a Higher Impairment/Stable Trajectory group, with slower change in adaptive skills as compared to peers. We then developed a machine learning model based on clinical intake information and data regarding interventions to predict which children might show more or less growth in adaptive skills over time. Using a wide range of socio-demographic, historical, and clinical phenotyping predictor variables, random forest machine learning models were able to predict course trajectory during care with an AUC-ROC of 0.77 and an accuracy of 77.5% in the external validation set. This study was motivated by a pressing need to characterize trajectories of adaptive behavior among autistic children in a clinically relevant context.

Past literature using latent class growth mixture model (LCGMM) approaches to identify clusters of adaptive behavior change over time has relied on research cohorts of autistic children recruited from the community, with little information about the nature of the interventions and therapies those children were receiving [6, 8, 9, 11]. Interestingly, while these study populations did encompass children of diverse ages and varying

levels of autism severity, a consistent finding has been that most children (70–100%) exhibit stable or regressing trajectories.

Our study fills an important gap in the literature by examining outcomes in a large cohort receiving care delivered according to relatively standardized clinical criteria. In this cohort, approximately 75–91% of individuals show an improving growth trajectory across different VABS-3 domain scores.

One possible explanation for this disparity is the nature of care received by our cohort. Unlike many prior studies, which assessed either naturalistic trajectories in the absence of detailed information about therapeutic interventions or specific standardized intervention programs, our study examined outcomes in children receiving care within a comprehensive clinical framework. All children in our cohort had access to a care model that integrated medical treatment alongside behavioral intervention and developmental therapies. In theory, such a comprehensive approach may support broader adaptive improvements. On the other hand, the model is inherently personalized, and therefore the heterogeneity of interventions received across children in the cohort makes it difficult to determine how differences in the observed growth trajectories might be related to treatment. Apart from any treatment-related differences, differences in population characteristics (including socioeconomic and demographic variables and baseline levels of function) and methodological approaches may have contributed to the finding of overall better outcomes in our cohort [20–22]. It is also important to note that slope estimates for our improving growth trajectory were relatively modest and may not necessarily reflect clinically significant change. Nevertheless, these findings highlight the potential impact of individualized, multi-modal care models and underscore the need for future research to clarify factors driving these discrepancies.

Moreover, we were able develop machine learning models to predict which patients were more likely to show improvements in adaptive behavior. Cross-fold validation for our final model had an observed AUC-ROC of 0.73 and an accuracy of 74.0%, and performance in the external validation set showed an AUC-ROC of 0.77 and an accuracy of 77.5%. Our models appear to improve on

the accuracy of previous machine learning models designed to predict autism developmental courses. For example, Préfontaine et al. [23] used machine learning to predict adaptive behavior change during behavioral intervention based on demographic information, IQ, and baseline autism severity features. Their best fitting model showed an accuracy of 68% [23]. Thus, our study extends the literature by applying a comprehensive workflow of machine learning development to predict adaptive behavior outcomes as defined from data-driven latent class approaches.

Although very few studies have employed machine learning, other researchers have used inferential statistics to identify predictors of adaptive behavior development. This more traditional statistical approach has demonstrated that cognitive and language skills, presence of epilepsy, and severity of autism are risk factors for low growth trajectories [24]. These results overlap with our findings that autism symptom severity and parent concerns about development are the most robust predictors of outcome trajectory.

Interestingly, when therapy intensity - operationalized as cumulative hours of ABA or DT - was added to our model, predictive accuracy did not improve. This would seem to offer counter evidence to research proposing a linear relation between duration and intensity of therapy with behavioral improvement [25, 26]. On the other hand, these findings add to growing literature suggesting the high total treatment duration of ABA therapy does not lead to improvement in treatment gains above average dosage levels [27]. It may in fact be the case, as demonstrated here, that increasing therapeutic hours is not necessarily effective or cost-efficient in improving real-world outcomes for all children [28].

Our analysis of variable importance identified variables across several domains contributing significantly to the model. Importantly, SES was one of the most important indicating variables in our models. Individuals from lower SES backgrounds may lack some of the key resources required to support their autistic child; in many cases, accessing therapeutic services can be associated with a severe economic burden [29–31]. There is a dearth of research evaluating health disparities in autism diagnoses and

outcomes, warranting a need for more intensive research within this domain.

In addition, our data suggest that older paternal age at the time of the child's birth predicted a more favorable adaptive behavior outcome. While this may seem counterintuitive given the literature suggesting that autism is more prevalent in children of older parents, the explanation may relate to the findings about SES that we have just discussed: it may be that older parents are more likely to be financially secure and stable.

It is also notable that infant temperament was identified as a robust predictive variable: infants with fussier temperaments were more likely to fall into the Stable/Limited Gains adaptive behavior trajectory. This finding is in line with a large body of developmental literature describing infant temperament as an early marker for autism risk [32, 33]. Another top predictor within the domain of early life experiences was the presence of a history of developmental regression. Again, a great deal of literature has demonstrated the children with who have experienced a loss in previously attained abilities show less favorable developmental outcomes, perhaps attributable to underlying biological or developmental conditions that lead to autistic regression. This highlights the importance of a thorough etiologic evaluation not only to understand the contributors to a child's neurodevelopmental condition, but also to inform prognosis and develop a plan of care [34, 35].

Within the domain of phenotypes, several variables predicted adaptive behavior trajectory. These include two variables that are directly involved with the core features of autism: namely, global autism symptom severity (as measured by the Autism Quotient [36]) and parent communication and language concerns (identified using the ESSENCE-Q-REV [34]). This finding is not surprising given historical evidence showing that baseline autism symptoms are associated with response to treatment and developmental outcomes [7, 37, 38]. The item about general developmental concerns on the ESSENCE-Q-REV captures a range of parental observations about child development; the globally inclusive nature of this item is undoubtedly inherent to its predictive power [39]. Interestingly, motor developmental concerns were also identified as a top predictive variable. It is increasingly recognized that heterogeneity in the acquisition of motor skills is an important early predictor of neurodevelopmental outcomes, and specifically that motor developmental delays are a risk factor for later autism prognosis and severity.

Finally, ADHD symptom severity (identified from the SNAP-IV [40]) and mood concerns (identified from the ESSENCE-Q-REV) demonstrated significant predictive ability in the machine learning model. It is likely that psychiatric comorbidities of autism – including ADHD, anxiety, and depression —may influence a child's ability to show progress in adaptive behavior [41–43].

Limitations and future directions
While we believe this study offers a major contribution to the literature on adaptive behavior outcomes in autism, several
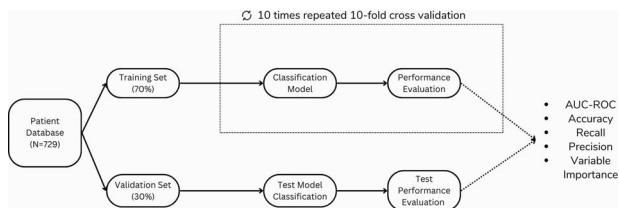
Fig. 3 Schematic of machine learning model development and validation. Flowchart of machine learning models used for the prediction of adaptive behavior outcomes.

Table 3. Machine learning models.

		Elastic Net Logistic Regression	Support Vector Machine	Random Forest
Development	Accuracy	0.74	0.74	**0.75**
	AUC-ROC	0.72	0.71	**0.76**
Tuning Parameters (Best Model)	Parameter 1	Penalty = 0.008	Cost = 0.003	**Trees = 675**
	Parameter 2	Mixture = 0.430	Margin = 0.15	**Mtry = 2**
	Parameter 3	–	–	**Min_n = 8**

Model performance results across 10-fold cross validation. Penalty=reflects the balance between feature selection and coefficient shrinkage offered from ridge and lasso regression, Mixture=value that controls the strength of regularization. Cost=number reflecting the hardness of the margin, Margin=value reflecting the size of the margin. Trees=number of decision tress in the random forest, mtry= the number of features to consider at each decision tree, min_n=minimum number of data points in a node required for another split.
The **Development** summarizes the model's average performance across a 10-fold cross-validation procedure applied to the training dataset.
The **Tuning Parameters** section displays the final hyperparameter values that yielded the best performance during cross-validation for each algorithm.
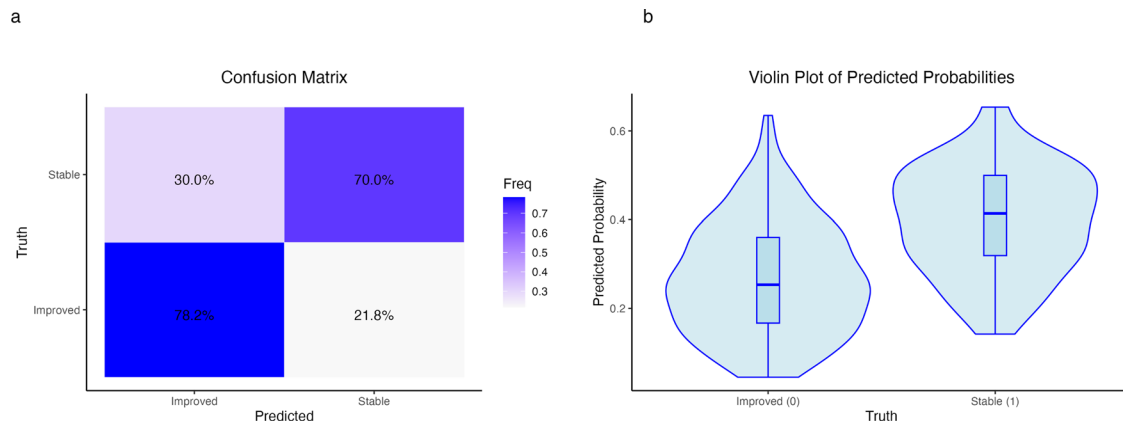
Fig. 4 Model predictions in the external holdout set. a Confusion matrix for the real vs. predicted binary classification of the adaptive behavior trajectory group. Values correspond to proportions of predictions for each row, respectively. **b** Violin plot depicting the spread of the predicted probabilities for the binary classification of growth trajectories. Lower values in the Multi-Domain Improvement category correspond to better accuracy (closer to 0) and higher values in the Stable/Limited Gains category reflect better accuracy (closer to 1).
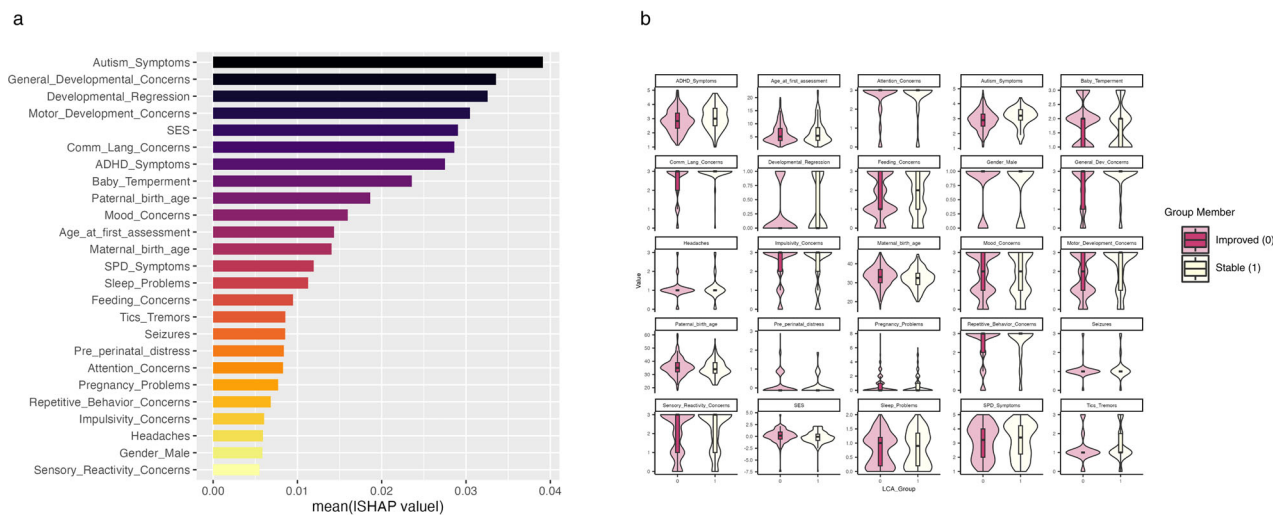
Fig. 5 Variable importance in predicting adaptive behavior outcomes. Variables included in the final random forest machine learning model for prediction of adaptive behavior change. **a** A stacked bar plot of Shapley importance. Absolute average values of Shapley importance scores "(mean|SHAP value|)"are indicated on the x-axis. **b** Violin plots of the distribution raw values for each feature variable between outcome groups with boxplots representing the median and interquartile range of each variable.

limitations require further discussion. Because the design of the study was based on chart review within an opportunistic clinical sample, there was no control group consisting of children not receiving autism care, or children receiving different types or models of therapy. In comparison to the general population of autistic individuals, our population likely drew from families with more resources (i.e., access to commercial medical insurance), or who were more likely to seek out a specific type of clinical environment. The geographical distribution of the sample was also constrained by the location of centers in the Cortica Healthcare network. Nonetheless, even given these constraints, we were able to identify robust demographic predictors of outcome, including socioeconomic status.

Our decision to use VABS-3 standard scores as dependent variables within our LCGMM framework was driven primarily by considerations of comparability with previous research in this domain, which have also generally focused on standard scores [6, 8, 9, 11]. However, standard scores have limitations in their interpretation as indices of change. Further research should capitalize on metrics that may be more favorable for analyzing change, such as growth scale values or age-equivalent scores.

It is also important to note that our phenotypic data consisted almost entirely of parent-reported questionnaires, as opposed to psychometric assessments that are typically conducted by a clinician. While these questionnaires (such as the Autism Quotient, SNAP-IV, and ESSENCE-Q-REV) have been validated in a research context, it is arguably the case that clinician-administered assessments would provide more consistent and accurate phenotypic data. On the other hand, such assessments are time- and labor-intensive and are often not practical in clinical environments, especially given the severity of current bottlenecks in access to diagnosis and therapy for autism [44, 45]. The intake measures used in this study are practical and easy to gather in the context of routine clinical care.

Finally, we note that while we tested multiple machine learning algorithms, the possibilities for tuning of all hyper-parameters were not exhausted. It is possible that different hyper-parameter tuning may have led to altered predictions. Our analytic approach was selected to prioritize a straightforward model and reduce over-fitting and model complexity. Future research validating our machine learning models in an independent sample is warranted.

8

CONCLUSION

The current study showed that more than 60% of children receiving care at a network of community neurodevelopment centers with a comprehensive approach to autism care demonstrated improvement across multiple domains of adaptive behavior. We also described a set of clinical intake variables, including information about patients' socio-demographic backgrounds and behavioral phenotypes, that can predict adaptive behavior trajectories with moderate accuracy. Of note, cumulative hours of behavioral and developmental therapy did *not* seem to contribute to the prediction of outcome trajectories. Our results add to a growing body of research that uses latent growth curve modelling to characterize adaptive behavior change, and machine learning models to predict treatment outcomes in neurodevelopmental populations.

METHODS

Outcome variables

Adaptive behavior outcomes. The primary outcome measure used was the Vineland Adaptive Behavior Scales, Third Edition (VABS-3) Parent/Caregiver Form [18]. The parent report VABS-3 evaluates adaptive behaviors in five different domains: Social, Communication, Daily Living Skills, and Motor. Based on the summarized scores in each domain, an Adaptive Behavior Composite (ABC) is also calculated. Items are scored on a scale of 0 (never), 1 (sometimes), or 2 (usually) for the level at which the child can independently perform a skill. Each domain reflects a standard score based on a population mean of 100 and a standard deviation of 15. The ABC, Social, and Communication domains were included in the current analysis.

Clinical intake predictor variables

Sociodemographic variables. Child gender and age at initial assessment were gathered directly from Cortica's medical record database. Income and level of education for the primary caregiver(s) were together used as a proxy for socioeconomic status (SES). Specifically, total years of education and total income were scaled and averaged to create a composite SES variable.

Historical variables. Caregivers were asked to report on whether they experienced prenatal stress, prenatal depression, and/or post-partum depression. Responses were summed into a pre/perinatal mental health variable. We collected information about the age of each parent at the time of the child's birth. Parents rated their child's temperament as a baby on a three-point categorical scale (1 = "Easy baby/little crying", 2 = "Average baby/average crying", 3 = "Difficult baby/increased crying"). Finally, parents reported whether their child had lost previously attained developmental abilities (i.e., history of developmental regression).

Clinical phenotyping variables. At the time of patient intake, several clinical phenotyping variables were gathered. To assess sleep problems, the Bedtime Issues, Excessive Daytime Sleepiness, Night Awakenings, Regularity and Duration of Sleep, Snoring (BEARS) questionnaire [46] was administered. Attention deficit and hyperactivity disorder (ADHD) symptoms were evaluated using questions adapted from the Swanson, Nolan, and Pelham (SNAP-IV) questionnaire [40]. To measure autism symptoms, the Autism Quotient (AQ) [36] was administered. Sensory processing disorder (SPD) symptoms were measured using questions adapted from the Sensory Processing Scales (SPS) [47]. Finally, we administered the early symptomatic syndromes eliciting neurodevelopmental clinical examinations-questionnaire (ESSENCE-Q-REV) [48] which includes the following subset of concerns: general development, motor development, sensory reactivity, communication and language, impulsivity, attention, repetitive behavior, mood, and feeding. Parents rate their concerns on a scale of 1 (No Concern), 2 (Maybe/A Little Concern), 3 (Yes Concern).

Conditional latent growth curve analysis

All latent class growth mixture model (LCGMM) analyses were conducted in Mplus Version 8.1 [49]. Full Information Maximum Likelihood (FIML) estimator was used to account for missing data in all LCGMM analyses yield unbiased estimates when data are missing at random (Muthen & Muthen 2012). Each LCGMM was initialized 200 times with 50 integrations for final stage of optimization. The best-fitting model was selected based on Bayesian information criterion (BIC), statistically significant Lo–Mendell–Rubin likelihood ratio test (LMR), and entropy values [50]. Lower BIC reflects better model fit while higher entropy values (closest to 1.0) indicate greater class separation and lower classification error [51]. Data were structured in a wide panel format, with each assessment timepoint (scheduled in 6-month increments) represented as a separate column. Time was parameterized in 1-year increments, with baseline set at 0 and subsequent assessments spaced at intervals of 0.5 (i.e., 6 months). This means that the slope estimates reflect the rate of change in the outcome variable per year rather than per assessment wave. Change was modeled with respect to clinical time-to-follow-up rather than chronological age, as the Vineland-3 standard scores are already age-normed. This approach ensures that the identified trajectories reflect variations in adaptive behavior over the course of clinical care rather than differences in age at assessment.

Machine learning predictive analysis

We compared three machine learning algorithms to predict between the two LCGA growth trajectory groups. All models were run using the tidymodels package in R [52]. First we used penalized elastic-net logistic regression from the R package glmnet [53], followed by support vector machine (SVM) classifiers from the R package kernlab [52, 54], and lastly, random forest classification from the R package ranger [55]. Elastic net logistic regression supports sparse model development through feature selection. SVM employs a technique where data is projected into a higher-dimensional space, enabling the classes to be distinguished by a hyperplane. By moving the data into this augmented space, the classifier effectively addresses issues of class separation by selecting the class-separating hyperplane that maximizes margins between features. Finally, random forest uses bootstrapped aggregation (or "bagging") to combine predictions from many tress where each tree is fit with a bootstrapped dataset.

Feature engineering. Data was preprocessed to optimize model performance. First, variables with missing data were imputed with k-nearest neighbor (kNN). This process is applied using the training set to impute missing data by considering the similarity from its k-nearest neighbor which contained non-missing values. Gower's distance function is used as the primary metric for calculating distances between data points. Next, all variables were normalized and variables with near-zero variance and highly correlated variables were removed. Variables with a high degree of correlation (>0.90) were removed to avoid redundant information. Finally, we applied Synthetic Minority Over-sampling Technique (SMOTE) [56] from the *themis* R package [57] using an over-ratio of 0.5 to mitigate the impact of class imbalance. SMOTE counteracts bias stemming from under-represented samples by generating synthetic instances of the minority class.

Hyperparameter tuning. Using the transformed and quality-controlled feature variables, we then preformed hyperparameter tuning on each machine learning algorithm. Elastic net is tuned by optimizing the balance between feature selection and coefficient shrinkage from ridge and lasso regression and strength of regularization. SVN's tuning parameters include optimizing margin hardness and size. Random forest tuning parameters include the number of features to consider at each split, the number of decision tress included in the forest, and the minimum number of data points in a node required for another split.

A 30% holdout test dataset was created using a random split stratified by outcome group. Hyperparameters were optimized for each algorithm via grid search within stratified 10-fold cross-validation repeated 10 times [58]. The grid search was preformed iteratively using the stratified cross-validation folds. For each repetition, the test dataset was randomly split into 10 equally sized subsets, of which two were used as a training set to create the model and the hold out set was sued as the test set. To evaluate training performance of the model, area under the receiver operating curve (AUROC) and accuracy were aggregated across the cross-validation folds. The study selected the hyperparameter combination that yielded the best model performance as the final model.

Model validation. We deployed the final model on the previously unseen holdout test data for model validation. To evaluate classification performance of the final model, ROCAUC, accuracy, precision, recall, and F1 score metrics were calculated. A confusion matrix and prediction plot were generated to compare the predicted values to true values of trajectory group membership.

Variable importance. To assess the significance of individual variables within the final model, we employed Shapley values. Originating from cooperative game theory, the Shapley method provides a model-agnostic approach to quantifying variable importance, aiming to estimate the contribution of each feature to the predictions generated by a machine learning model [59]. Shapley values were calculated using the 'fastshap' package in R. We conducted 100 Monte Carlo simulations to obtain robust estimates of the Shapley values. These values capture the weighted differences in model outcomes as features are systematically included or excluded for all possible permutations.

DATA AVAILABILITY
The datasets used in the current study are available from the corresponding author on reasonable request.

REFERENCES

1. Kenworthy L, Case L, Harms MB, Martin A, Wallace GL. Adaptive behavior ratings correlate with symptomatology and IQ among individuals with high-functioning autism spectrum disorders. J Autism Dev Disord. 2010;40:416–23.
2. Volkmar FR, Sparrow SS, Goudreau D, Cicchetti DV, Paul R, Cohen DJ. Social deficits in autism: an operational approach using the Vineland Adaptive Behavior Scales. J Am Acad Child Adolesc Psychiatry. 1987;26:156–61.
3. Kanne SM, Gerber AJ, Quirmbach LM, Sparrow SS, Cicchetti DV, Saulnier CA. The role of adaptive behavior in autism spectrum disorders: implications for functional outcome. J Autism Dev Disord. 2011;41:1007–18.
4. Zander E, Bölte S. The new DSM-5 impairment criterion: a challenge to early autism spectrum disorder diagnosis?. J Autism Dev Disord. 2015;45:3634–43.
5. Lombardo MV, Lai M-C, Baron-Cohen S. Big data approaches to decomposing heterogeneity across the autism spectrum. Mol Psychiatry. 2019;24:1435–50.
6. Farmer C, Swineford L, Swedo SE, Thurm A. Classifying and characterizing the development of adaptive behavior in a naturalistic longitudinal study of young children with autism. J Neurodev Disord. 2018;10:1.
7. Mandelli V, Landi I, Busuoli EM, Courchesne E, Pierce K, Lombardo MV. Prognostic early snapshot stratification of autism based on adaptive functioning. Nature Ment Health. 2023;1:327–36.
8. Szatmari P, Georgiades S, Duku E, Bennett TA, Bryson S, Fombonne E, et al. Developmental trajectories of symptom severity and adaptive functioning in an inception cohort of preschool children with autism spectrum disorder. JAMA Psychiatry. 2015;72:276–83.
9. Baghdadli A, Michelon C, Pernon E, Picot MC, Miot S, Sonié S, et al. Adaptive trajectories and early risk factors in the autism spectrum: a 15-year prospective study. Autism Res. 2018;11:1455–67.
10. Yerys BE. The promise and peril of predicting adaptive behavior in toddlers with autism. Nature Ment Health. 2023;1:302–3.
11. Pierce K, Gazestani VH, Bacon E, Barnes CC, Cha D, Nalabolu S, et al. Evaluation of the diagnostic stability of the early autism spectrum disorder phenotype in the general population starting at 12 months. JAMA Pediatr. 2019;173:578–87.
12. Reichow B, Hume K, Barton EE, Boyd BA. Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD). Cochrane Database Syst Rev. 2018;5:CD009260.
13. Dinga R, Marquand AF, Veltman DJ, Beekman ATF, Schoevers RA, van Hemert AM, et al. Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. Transl Psychiatry. 2018;8:241.
14. Washington P, Wall DP. A review of and roadmap for data science and machine learning for the neuropsychiatric phenotype of autism. Annu Rev Biomed Data Sci. 2023;6:211–28.
15. Farooq MS, Tehseen R, Sabir M, Atal Z. Detection of autism spectrum disorder (ASD) in children and adults using machine learning. Sci Rep. 2023;13:9605.
16. Küpper C, Stroth S, Wolff N, Hauck F, Kliewer N, Schad-Hansjosten T, et al. Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. Sci Rep. 2020;10:4805.
17. Quillet JC, Siani-Rose M, McKee R, Goldstein B, Taylor M, Kurek I. A machine learning approach for understanding the metabolomics response of children with autism spectrum disorder to medical cannabis treatment. Sci Rep. 2023;13:13022.
18. Sparrow SS, Cicchetti DV, Saulnier CA. *Vineland Adaptive Behavior Scales*. (2016).
19. de Bildt A, Kraijer D, Sytema S, Minderaa R. The psychometric properties of the Vineland Adaptive Behavior Scales in children and adolescents with mental retardation. J Autism Dev Disord. 2005;35:53–62.
20. Perry A, Flanagan HE, Dunn Geier J, Freeman NL. Brief report: the Vineland Adaptive Behavior Scales in young children with autism spectrum disorders at different cognitive levels. J Autism Dev Disord. 2009;39:1066–78.
21. Gentles S, Duku E, Kerns C, McVey AJ, Hunsche MC, Ng Cordell EC, et al. Trajectory research in children on the autism spectrum: a scoping review protocol. BMJ Open. 2021;11:e053443.
22. Bhasin TK, Schendel D. Sociodemographic risk factors for autism in a US metropolitan area. J Autism Dev Disord. 2007;37:667–77.
23. Préfontaine I, Lanovaz MJ, Rivard M. Brief report: machine learning for estimating prognosis of children with autism receiving early behavioral intervention-a proof of concept. J Autism Dev Disord. 2022;54:1605–10. https://doi.org/10.1007/s10803-022-05641-9.
24. Baghdadli A, Assouline B, Sonié S, Pernon E, Darrou C, Michelon C, et al. Developmental trajectories of adaptive behaviors from early childhood to adolescence in a cohort of 152 children with autism spectrum disorders. J Autism Dev Disord. 2012;42:1314–25.
25. Virués-Ortega J. Applied behavior analytic intervention for autism in early childhood: meta-analysis, meta-regression and dose-response meta-analysis of multiple outcomes. Clin Psychol Rev. 2010;30:387–99.
26. Reed P, Osborne LA, Corness M. Brief report: relative effectiveness of different home-based behavioral approaches to early teaching intervention. J Autism Dev Disord. 2007;37:1815–21.
27. Vietze P, Lax LE. Early intervention ABA for toddlers with ASD: effect of age and amount. Curr Psychol. 2020;39:1234–44.
28. Ostrovsky A, Willa M, Cho T, Strandberg M, Howard S, Davitian C. Data-driven, client-centric applied behavior analysis treatment-dose optimization improves functional outcomes. World J Pediatr. 2023;19:753–60.
29. Bishop-Fitzpatrick L, Kind AJH. A scoping review of health disparities in autism spectrum disorder. J Autism Dev Disord. 2017;47:3380–91.
30. Eilenberg JS, Paff M, Harrison AJ, Long KA. Disparities based on race, ethnicity, and socioeconomic status over the transition to adulthood among adolescents and young adults on the autism spectrum: a systematic review. Curr Psychiatry Rep. 2019;21:32.
31. Rogge N, Janssen J. The economic costs of autism spectrum disorder: a literature review. J Autism Dev Disord. 2019;49:2873–2900.
32. Clifford SM, Hudry K, Elsabbagh M, Charman T, Johnson MH. Temperament in the first 2 years of life in infants at high-risk for autism spectrum disorders. J Autism Dev Disord. 2013;43:673–86.
33. Pijl MKJ, Bussu G, Charman T, Johnson MH, Jones EJH, Pasco G, et al. Temperament as an early risk marker for autism spectrum disorders? a longitudinal study of high-risk and low-risk infants. J Autism Dev Disord. 2019;49:1825–36.
34. Rogers SJ. Developmental regression in autism spectrum disorders. Ment Retard Dev Disabil Res Rev. 2004;10:139–43.
35. Goin-Kochel RP, Esler AN, Kanne SM, Hus V. Developmental regression among children with autism spectrum disorder: onset, duration, and effects on functional outcomes. Res Autism Spectr Disord. 2014;8:890–8.
36. Allison C, Auyeung B, Baron-Cohen S. Toward brief "Red Flags" for autism screening: The Short Autism Spectrum Quotient and the Short Quantitative Checklist for Autism in toddlers in 1000 cases and 3000 controls [corrected]. J Am Acad Child Adolesc Psychiatry. 2012;51:202–212.e7.
37. Rogers SJ, Yoder P, Estes A, Warren Z, McEachin J, Munson J, et al. A multisite randomized controlled trial comparing the effects of intervention intensity and intervention style on outcomes for young children with autism. J Am Acad Child Adolesc Psychiatry. 2021;60:710–22.
38. Hudry K, McConachie H, Le Couteur A, Howlin P, Barrett B, Slonims V. Predictors of reliable symptom change: Secondary analysis of the Preschool Autism Communication Trial. Autism Developmental Lang Impairments. 2018;3:2396941518764760.
39. Kattimani S, Joseph R, Mondal N, Johnson A. Predictive validity of ESSENCE Q screening tool for early detection of neurodevelopmental disorder in children. Asian J Psychiatr. 2022;73:103162.
40. Swanson JM, Schuck S, Porter MM, Carlson C, Hartman CA, Sergeant JA, et al. Categorical and dimensional definitions and evaluations of symptoms of ADHD: history of the SNAP and the SWAN rating scales. Int J Educ Psychol Assess. 2012;10:51–70.
41. Leffa DT, Caye A, Rohde LA. ADHD in Children and Adults: Diagnosis and Prognosis. In New Discoveries in the Behavioral Neuroscience of Attention-Deficit Hyperactivity Disorder (eds. Stanford, SC & Sciberras, E) 1–18 (Springer International Publishing, Cham, 2022).
42. Mannion A, Leader G. Comorbidity in autism spectrum disorder: a literature review. Res Autism Spectr Disord. 2013;7:1595–616.
43. Kirsch AC, Huebner ARS, Mehta SQ, Howie FR, Weaver AL, Myers SM, et al. Association of comorbid mood and anxiety disorders with autism spectrum disorder. JAMA Pediatr. 2020;174:63–70.
44. Ozonoff S, Goodlin-Jones BL, Solomon M. Evidence-based assessment of autism spectrum disorders in children and adolescents. J Clin Child Adolesc Psychol. 2005;34:523–40.
45. Youngstrom EA, Choukas-Bradley S, Calhoun CD, Jensen-Doss A. Clinical guide to the evidence-based assessment approach to diagnosis and treatment. Cogn Behav Pr. 2015;22:20–35.

46. Owens JA, Dalzell V. Use of the "BEARS" sleep screening tool in a pediatric residents' continuity clinic: a pilot study. Sleep Med. 2005;6:63–69.

47. Schoen SA, Miller LJ, Sullivan JC. Measurement in sensory modulation: the sensory processing scale assessment. Am J Occup Ther. 2014;68:522–30.

48. Gillberg C. The ESSENCE in child psychiatry: early symptomatic syndromes eliciting neurodevelopmental clinical examinations. Res Dev Disabil. 2010;31:1543–51.

49. Muthén, B, Muthén, L Mplus. Handbook of item response theory (2017).

50. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a monte carlo simulation study. Struct Equ Modeling. 2007;14:535–69.

51. Jedidi K, Ramaswamy V, Desarbo WS. A maximum likelihood method for latent class regression involving a censored dependent variable. Psychometrika. 1993;58:375–94.

52. Kuhn M, Wickham H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. Preprint at https://www.tidymodels.org (2020).

53. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33:1–22.

54. Karatzoglou A, Smola A, Hornik K kernlab: Kernel-Based Machine Learning Lab. Preprint at https://CRAN.R-project.org/package=kernlab (2023).

55. Wright MN, Ziegler A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J. Stat. Softw. 2017;77:1–17.

56. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. jair. 2002;16:321–57.

57. Hvitfeldt E. themis: Extra Recipes Steps for Dealing with Unbalanced Data. R Package Version 0.1. 0 (2022).

58. Yarkoni T, Westfall J. Choosing prediction over explanation in psychology: lessons from machine learning. Perspect Psychol Sci. 2017;12:1100–22.

59. Roth AE Introduction to the Shapley value. The Shapley value 1–27 (1988).

AUTHOR CONTRIBUTIONS

AA proposed the hypothesis and approach, preformed analyses and wrote the manuscript; ML supported manuscript writing and editing; AE contributed to data acquisition and processing for the analysis; NH, PM, EM, and KS assisted with conceptual study development and manuscript writing. All authors contributed to manuscript revisions.

COMPETING INTERESTS

ML, AE, NH, EM and KS are employed by Cortica, AA received consultation fees from Cortica.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41398-025-03592-0.

Correspondence and requests for materials should be addressed to Annie Aitken.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.