5-12-2019

Identification and Analysis of Behavioral Phenotypes in Autism Spectrum Disorder via Unsupervised Machine Learning

Elizabeth Stevens
Chapman University

Dennis R. Dixon
Center for Autism and Related Disorders

Marlena N. Novack
Center for Autism and Related Disorders

Doreen Granpeesheh
Center for Autism and Related Disorders

Tristram Smith
University of Rochester

See next page for additional authors

Recommended Citation

E. Stevens, D. R. Dixon, M. N. Novack, et al., Identification and analysis of behavioral phenotypes in autism spectrum disorder via unsupervised machine learning, Int. J. Med. Inform. 129 (2019) 29-36.

Identification and Analysis of Behavioral Phenotypes in Autism Spectrum Disorder via Unsupervised Machine Learning

Authors
Elizabeth Stevens, Dennis R. Dixon, Marlena N. Novack, Doreen Granpeesheh, Tristram Smith, and Erik Linstead

International Journal of Medical Informatics 129 (2019) 29–36

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Identification and analysis of behavioral phenotypes in autism spectrum disorder via unsupervised machine learning

Elizabeth Stevens[a], Dennis R. Dixon[b], Marlena N. Novack[b], Doreen Granpeesheh[b], Tristram Smith[c], Erik Linstead[a,*]

[a] Chapman University, Schmid College of Science and Technology, Orange, CA, United States
[b] Center for Autism and Related Disorders, Woodland Hills, CA, United States
[c] University of Rochester Medical Center, Rochester, NY, United States

ARTICLE INFO

Keywords:
Machine learning
Autism spectrum disorder
Behavioral phenotypes
Cluster analysis
Treatment response

ABSTRACT

Background and objective: Autism spectrum disorder (ASD) is a heterogeneous disorder. Research has explored potential ASD subgroups with preliminary evidence supporting the existence of behaviorally and genetically distinct subgroups; however, research has yet to leverage machine learning to identify phenotypes on a scale large enough to robustly examine treatment response across such subgroups. The purpose of the present study was to apply Gaussian Mixture Models and Hierarchical Clustering to identify behavioral phenotypes of ASD and examine treatment response across the learned phenotypes.

Materials and methods: The present study included a sample of children with ASD ($N = 2400$), the largest of its kind to date. Unsupervised machine learning was applied to model ASD subgroups as well as their taxonomic relationships. Retrospective treatment data were available for a portion of the sample ($n = 1034$). Treatment response was examined within each subgroup via regression.

Results: The application of a Gaussian Mixture Model revealed 16 subgroups. Further examination of the subgroups through Hierarchical Agglomerative Clustering suggested 2 overlying behavioral phenotypes with unique deficit profiles each composed of subgroups that differed in severity of those deficits. Furthermore, differentiated response to treatment was found across subtypes, with a substantially higher amount of variance accounted for due to the homogenization effect of the clustering.

Discussion: The high amount of variance explained by the regression models indicates that clustering provides a basis for homogenization, and thus an opportunity to tailor treatment based on cluster memberships. These findings have significant implications on prognosis and targeted treatment of ASD, and pave the way for personalized intervention based on unsupervised machine learning.

1. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by persistent deficits in social communication and social interaction, in addition to restricted, repetitive patterns of behavior, interests, or activities [1]. Previously conceptualized as a group of related but distinct disorders (i.e., autistic disorder, Asperger's disorder, and childhood disintegrative disorder) in the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV) [2], ASD is currently classified as a single disorder with level of severity falling along a continuum in the latest edition of the DSM (i.e., DSM-5) [1]. The move from a categorical to dimensional approach was the result of research demonstrating that even well-trained clinicians using standardized diagnostic instruments did not reliably distinguish among the separate DSM-IV disorders [3]. The dimensional approach in DSM-5 involves rating symptom severity according to the level of support required for social communication and restricted, repetitive behavior. The DSM-5 also incorporates reports of co-occurring conditions and age of onset. The DSM-5 changes to ASD diagnostic criteria may facilitate efforts to classify potential ASD subtypes [4].

As demonstrated by current diagnostic practices, ASD is a heterogeneous disorder. ASD symptom severity, language abilities [5], and skills across other developmental domains (e.g., adaptive skills and executive functioning) [6,7] vary greatly across individuals diagnosed with ASD. Onset of symptoms also differ as some individuals experience delays or plateaus in development [8,9], and others experience

* Corresponding author.
E-mail address: linstead@chapman.edu (E. Linstead).

https://doi.org/10.1016/j.ijmedinf.2019.05.006
Received 4 August 2018; Received in revised form 25 February 2019; Accepted 9 May 2019

regression of previously acquired skills [10]. Many individuals with ASD also present with co-occurring disorders [11,12].

While there is no single known cause for ASD, genetics and environment play causal roles. It is estimated that genetic variants are present in approximately 10%–20% of ASD cases, [13,14] and more than 100 different genes have been linked to ASD [14]. Furthermore, in comparison to the general population, rates of ASD are higher among relatives of individuals with ASD [15] and among individuals with various genetic disorders [16]. Environmental factors also may be involved in the development of ASD. Prenatal exposures [17–24], increased parental age [25], duration between births [26], and birth complications [27] have been associated with an increased risk of ASD. While many potential environmental and genetic links to ASD have been identified, findings have not been well replicated, and no single factor has been found to be involved in all cases of ASD.

In addition to symptom presentation and etiology, individuals with ASD differ in their response to treatment. One of the most well-established treatment approaches for ASD is applied behavior analysis (ABA) [28–30]. ABA treatments involve applying principles and procedures of learning and motivation in order to alter behavior (e.g., promote skill acquisition, reduce challenging behaviors) [31]. ABA is typically initiated at a young age and delivered at a high intensity (e.g., 25–40 h per week) for several years [28,30]. While many studies report that ABA treatments are effective for ASD, research also indicates variation in individual response to treatment [32,33]. Several factors predict favorable response. While treatment specific factors, including treatment intensity [32,34–37], may account for a sizable amount of the variance observed in treatment response, child specific factors have also been identified. Younger age [36,38–44], lower severity of ASD symptoms [37–3943,45], higher IQ [32,37–3942,43,45,46], stronger adaptive skills [32,36,37,40,43,46,47], greater language skills [38,46,47], and stronger social skills [38,47] have been associated with superior outcomes. Despite some consistent findings, the ability to predict individual treatment response is limited [48].

Given the diversity in ASD symptom presentation, etiology, and treatment response, researchers have looked towards subgroups within ASD to account for these differences [49]. There are substantial implications for research on ASD subgroups, including revealing different etiologies [49,50,51], aiding diagnosis [50], predicting treatment response [49,50], and advancing targeted treatments [49,50,51],

Numerous studies have identified diverse behavioral phenotypes within ASD. Beglinger and Smith [49] conducted a review of research efforts to subtype ASD, finding support for ASD subtypes based on level of functioning and social abilities. Since Beglinger and Smith's review, cluster analyses by Shen and colleagues [51] and Hu and Steinberg [52] identified four ASD subtypes that differed based on severity of ASD symptoms. Ingram, Takahashi, and Miles [53] used a taxometric methodology to assess ASD subgroups. Findings supported subgrouping based on social interaction and communication, intelligence, and morphological abnormalities, while other domains (i.e., insistence on sameness, repetitive sensory motor actions, language acquisition, and adaptive functioning) were found to be dimensional. Most recently, Lombardo and colleagues [54] used a clustering method based on task performance of reading emotions and mental states. Five ASD subgroups were found with performance ranging from severe deficits to minimal impairments. While a number of studies have identified behavioral subgroups, no subgroups have been well replicated [55].

While it is useful to identify subgroups of ASD that differ behaviorally, validation of those subgroups relies on further associations (e.g., biologic or genetic factors, treatment response, etc.) [53]. Preliminary research has revealed that behavioral subgroups of ASD may exhibit increased rates of epilepsy [56], and may also display differences in gene expression [57–59].

Another novel way to identify subgroups is through an evaluation of developmental trajectories. While research is limited, this approach has yielded interesting results. For example, Fein and colleagues [60] conducted a cluster analysis in a sample of 194 preschool-aged participants with pervasive developmental disorder-not otherwise specified (PDD-NOS) and identified two subgroups, a high-functioning subgroup and a low-functioning subgroup. Using a large battery of standardized assessments, subgroups were found to differ in areas of cognitive and social development. Participants were reevaluated once they were school aged and, overall, subgroup membership was found to be stable. The subgroups were found to have different developmental trajectories with the high-functioning group showing greater overall improvement while the low-functioning group demonstrated little to no improvement. Likewise, Stevens and colleagues [61] further examined subgroups in a sample of 138 school-age children diagnosed with autistic disorder. A hierarchical cluster analysis once again identified two subgroups, a high-functioning subgroup and a low-functioning subgroup, based on measurements of social, language, and nonverbal ability. Overall, preschool-age and school-age subgroup membership was found to be stable for the low-functioning subgroup while school-aged outcomes for the high-functioning preschool-age subgroup were divided. More recently, Pickles and colleagues [62] evaluated the developmental trajectories displayed by 192 children referred for evaluation of ASD and subsequently were evaluated on language development at six points over the course of 17 years. Latent class growth analyses identified seven classes based upon language development. Development patterns varied between classes from ages 2–6 years, after which developmental trajectories were found to be parallel from ages 6–19 years. Lord, Bishop, and Anderson [63] further examined developmental trajectories and identified three groups based upon outcome (ASD IQ $<$ 70, ASD IQ $>$ 70, Very Positive Outcome). These groups showed distinct developmental trajectories on measures of social communication deficits, social adaptive functioning, and repetitive behaviors. While longitudinal studies reveal different developmental trajectories across ASD subgroups, no studies have investigated treatment response between subgroups [49].

Subtypes of ASD may contribute to the variation in treatment response observed among individuals with ASD. As such, assessments able to classify subgroups of ASD will aid the development of targeted treatment [64]. Similar to the targeted treatments of subtypes of cancer, for instance leukemia [65] and breast cancer [66], targeted treatments of subtypes of ASD may improve outcomes. The purpose of the present study was to leverage unsupervised machine learning in the form of Gaussian Mixture Models and Hierarchical Agglomerative Clustering to identify behavioral phenotypes of ASD from a sample of 2400 children using data from a detailed assessment of skills across developmental domains, and evaluate treatment response between the identified clusters. The largest and most extensive study of its type to date, results indicate that machine learning not only successfully extracted behavioral subtypes and the relationships among subtypes, but also yielded novel information that has the potential to guide clinical practice. Indeed, the clusters provide for a more accurate assessment of treatment response across the spectrum. These finding have direct and significant implications for tailoring treatment curriculum with the ultimate goal of delivering personalized and precise intervention in the context of applied behavior analysis.

2. Materials and methods

2.1. Dataset

Deidentified, retrospective data from a large archival database was used in the present study. The Skills system was used to assess deficits and track ongoing treatment progress. The Skills Assessment is a comprehensive measurement of over 3000 skills across eight developmental domains, including social, language, adaptive, academic, cognitive, executive functioning, play, and motor skills. The Skills Language Index has been found to have strong reliability across subscales with internal consistency ranging from 0.73 to 0.997, inter-rater reliability ranging

from 0.866 to 0.941, and test-retest reliability ranging from 0.815 to 0.968. [67] The Skills assessment has also been shown to have strong concurrent validity with moderate to high correlations found between parent report and direct observation ranging from 0.646 to 0.954 across domains [68].

Participants were children receiving ABA treatment services from a community-based provider, operating centers in numerous states, during a 36-month period (January 1, 2014 through December 31, 2016). A pool of 7822 clinical records were reviewed and were subject to the following inclusion criteria: (a) between the ages 18 months and 12 years old; (b) had a DSM-5 diagnosis of ASD [1], or a DSM-IV diagnosis of autistic disorder [2], PDD-NOS [2], or Asperger's disorder [2] by an independent licensed clinician (e.g., psychologist, pediatrician, neurologists, etc.); (c) received at least 20 h of treatment per month; and (d) had at least one month of continuous services. After applying these criteria, a sample of 2400 clinical records remained. Of the participants, 1948 were male and 452 were female. The average age of participants was 7.8 years old, ranging from 2.66 years to 12 years. Participants resided in the states of Arizona, California, Colorado, Illinois, Louisiana, New York, Texas, and Virginia.

2.2. Treatment

Participants in the present study were receiving comprehensive, individualized ABA treatment targeting all domains in which deficits were present, including language, social, adaptive, cognitive, executive function, academic, play, and motor skills. Depending on funding agency requirements and other variables, treatment services were provided in the participant's home, clinic, school, community, or a variety of settings. Participants' treatment services were delivered according to the CARD model [69].

2.3. Data analysis

To extract latent clusters from the data, a data matrix, D, was constructed. D is represented as a collection of vectors, $D = \{X_1, X_2, ..., X_m\}$. Each vector, X_i, represents a unique data instance (patient), and each vector element, $X_{i,j}$, represents a specific measurement (attribute) for that point corresponding to proficiency in one of the eight treatment domains (language, social, adaptive, cognitive, executive function, academic, play, and motor skills). This proficiency was determined by summing the number of Skills assessment questions answered affirmatively, and then normalizing by the number of questions applicable for the domain given the subject's age, resulting in a 2400×8 dimensional matrix with each element in [0,1]. The matrix, D, serves as the input to a Gaussian Mixture Model (GMM). Gaussian Mixture Models provide a statistical framework for learning latent cluster memberships from data that allows for probabilistic interpretation of assignments. Additionally, GMMs are better able to adapt to differences in cluster geometry compared to partitioning-based methods like K-Means.

In GMMs, it is assumed that each X_i is generated from a probability density function, $p(X)$. The data points are assumed to be independent and identically distributed (iid). In this model, it is also assumed that there are K latent clusters in the data, and that each data instance belonged to all K clusters but with different probability. This results in a Gaussian Mixture Model for $p(X)$ defined as:

$$p(X|\theta) = \sum_{k=1}^{K} \alpha_k p_k(X|\theta_k)$$

The parameters of the model, θ, consists of the following: $\theta = \{\theta_1, \theta_2, ..., \theta_K; \alpha_1, \alpha_2, ..., \alpha_K\}$. Each cluster is represented by a multivariate Gaussian distribution parameterized by θ_k, where $\theta_k = \mu_k, \sigma_k$, the mean and covariance of the distribution, respectively. Thus, substituting these parameters into the density function for a multivariate Gaussian, the probability of observation i belonging to cluster k can be computed as follows:

$$p_k(X_i|\theta_k) = \frac{1}{\sqrt{(2\pi)^n|\sigma_k|}} \exp(-\frac{1}{2}(X_i - \mu_k)^T \sigma_k^{-1}(X_i - \mu_k)$$

The mixing parameters, $\alpha_1 ... \alpha_k$, are subject to the following constraint and represent the significance of each cluster in the overall model:

$$\sum_{k=1}^{K} \alpha_k = 1$$

The probability of a data instance, X_i, belonging to cluster k can be computed directly from the posterior probability distribution, $b_{i,k} = p(Cluster = k|X_i, \theta)$. By definition, this implies:

$$\sum_{k=1}^{K} b_{i,k} = 1$$

Algorithmically, the goal of the EM algorithm is to learn the parameters, θ, from the training data. It achieves this in two steps, the expectation (E) step and the maximization (M) step. The algorithm alternates between these steps until convergence is achieved.

The expectation (E) step

To randomly initialize the parameters, θ, before training, care was taken to ensure that the covariance matrices were not singular, which was achieved by adding a small positive constant to the diagonals. Once the parameters are initialized at the start of modeling, the expectation step simply consists of computing and updating all belief weights. That is, for all data instances, i, and clusters, k, compute $b_{i,k}$ as described above.

The maximization (M) step

Once the belief weights are calculated in the E step, they are used to reevaluate the remaining parameters of the model as follows:

$$\alpha_k^{next} = \frac{1}{M} \sum_{i=1}^{M} b_{i,k}, \ 1 \le k \le K$$

$$\mu_k^{next} = \frac{1}{\sum_{i=1}^{M} b_{i,k}} \sum_{i=1}^{M} b_{i,k} X_i, \ 1 \le k \le K$$

$$\sigma_k^{next} = \frac{1}{\sum_{i=1}^{M} b_{i,k}} \sum_{i=1}^{M} b_{i,k} (X_i - \mu_k^{next})(X_i - \mu_k^{next})^T$$

While EM provides an algorithmically straight-forward way for identifying grouping, a substantial challenge is determining the number of clusters, K, to be modeled. In the absence of domain information to suggest the correct parameter setting, several mechanisms exist for statistically determining the most likely K, including nonparametric statistical techniques. Here, K is determined by using the Bayesian Information Criterion (BIC), an information-theoretic approach for model selection [70] that also prevents overfitting by penalizing growth in the number of parameters,

Once the clusters were fit to the data, the impact of treatment intensity on learning outcomes was examined for each cluster. This was achieved by summing the total number of treatment hours for each participant, as well as the total number of mastered exemplars that occurred during the treatment period, subject to the inclusion criteria described above. A linear regression model was then fit to the data using treatment intensity as the independent variable and mastered exemplars as the dependent variable.

3. Results

The EM model described above was run on the dataset, and the number of clusters determined by minimizing BIC. This resulted in the identification of 16 latent clusters, each with a distinct signature described by measured ability across the eight skill domains. From

Fig. 1. A radar graph depicting the skill profiles of each cluster and subgroup.

inspection of Fig. 1, which visually represents each cluster in the feature space, it is apparent that there is a hierarchical relationship between many of the individual clusters. That is, the overall "shape" of the skill signature is similar, but the magnitude of each skill present varies. To simplify the model, and take these hierarchical relationships into account, clusters were merged by employing hierarchical agglomerative clustering (HAC). HAC starts with individual clusters. At each step, the pair of clusters that are most similar are grouped together. This continues until only a single large cluster remains. The steps taken by the algorithm to combine clusters can be visualized in a dendrogram, represented in Fig. 2. The leaves of the dendrogram represent the 16 clusters detected by the EM model, with every layer in the tree representing a merging of cluster groups.

From Fig. 2, it can be seen that the 16 individual clusters detected by HAC are merged into 5 high level clusters with the following memberships and demographics:

Cluster A (n = 154, mean age 7.98 years (SD = 2.03), 79.2% male): {13}

Cluster B (n = 280, mean age 7.67 years (SD = 2.53), 80.7% male): {7,10}

Cluster C (n = 780, mean age 7.34 years (SD = 2.34), 82.3% male): {4,5,9,12,16}

Cluster D (n = 345, mean age 6.93 years (SD = 2.39), 82.6% male): {2,3}

Cluster E (n = 841, mean age 8.64 years (SD = 2.23), 80.0% male): {1,6,8,11,14,15}

Fig. 1 visually depict clusters A through E using a radar graph. Each dimension of the radar graph shows measured ability on a specific skill

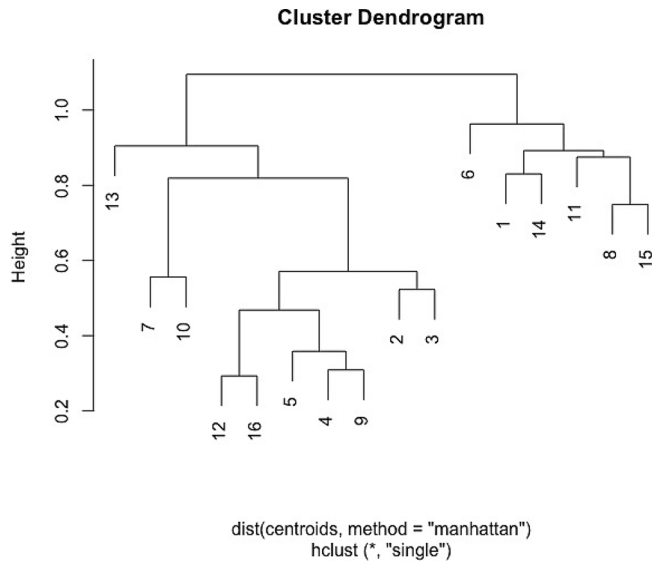

Fig. 2. Dendrogram of clusters depicting their hierarchical composition.

domain. Looking at Cluster A, for example, a group of children (n = 154) is shown with average domain mastery scores of 50% for academic, 30% for adaptive, 22% for cognitive, 20% for executive function, 65% for language, 65% for motor, 50% for play, and 25% for social skills. Overall, this cluster appears to represent what would have historically been referred to as atypical autism or high functioning autism.

The remaining clusters can be interpreted in a similar fashion using the radar graphs and dendrogram. If Cluster A represents high functioning autism, then Cluster C appears to represent the other side of the spectrum, "low functioning autism," with all domains except motor skill showing extreme impairment (less than 20% mastery). Clusters B and D represent the spectrum between Clusters A and C. In comparison to Cluster C, Clusters B and D show an increased presence of skills in domains such as language and play. Referring to the dendrogram, Cluster E falls to the right of the tree, and therefore was deemed to be most different from the other clusters as part of the HAC process. Thus, the cluster profiles in this group would be expected to be substantially different from the others. Indeed, inspecting this group in Fig. 1, a distinct skill profile is shown that, unlike previous groups, exhibits high skill development across multiple domains, including play, language, academic, and motor, but variable skill development in the remaining domains (i.e., social, adaptive, cognitive, and executive). Overall, Cluster E appears to represent what was previously described as Asperger's disorder [2]. However, we also observe that the average age of members of cluster E is slightly higher than that of the other clusters, which could contribute to overall higher skill profile demonstrated by this group.

After identifying the shape of the latent clusters with unsupervised machine learning, a linear regression model was fit to understand how each of the groups responded to varying treatment intensity. Specifically, a univariate regression model was fit with treatment intensity (hours) as the independent variable and learning outcomes (mastered exemplars) as the dependent variable for a subset of data points (n = 1034) for which learning outcome data was available. To account for the fact that these variables can naturally span over orders of magnitude, a logarithmic transform was applied. Fig. 3 depicts the fit of these models for each cluster. Table 1 presents the respective variance accounted for and the slope of the linear regression equations.

4. Discussion

The present study examined behavioral phenotypes in a large sample (N = 2,400) of children with ASD. A total of 16 unique clusters were identified. Further computational analysis found a hierarchy of 5 distinct subgroups (i.e., Clusters A through E). Within each subgroup, clusters appeared to represent different degrees of severity across developmental domains (i.e., language, social, adaptive, cognitive, executive function, academic, play, and motor skills). While it is interesting from a methodological perspective that machine learning can identify subgroups of individuals with ASD, the value of this lies in identifying ways to improve treatment response.

The subgroups did indeed display unique treatment response profiles when modeled with linear regression as described above, in particular the r^2 values. The linear relationships between treatment hours and mastery of learning objectives were strong within subgroups, with the regression line accounting for between 64% and 75% of the variance observed in treatment response. In comparison, treatment hours have accounted for 35% of the variance observed in treatment response in a large sample of children with ASD [35]. That is, children within the current clusters respond more similarly to treatment than children with ASD as a whole. Overall, these r^2 values were substantially higher than previous studies [35,71], on average explaining 20% more variance. This suggests that not only do the cluster memberships assigned by the machine learning algorithm have predictive validity, but also provide a method for homogenizing treatment groups. In this way, the clusters can form a computationally and mathematically robust basis for research on tailoring behavioral intervention for different subgroups of children with ASD.

The subgroups displayed different rates of skill mastery in response to treatment intensity. The differences in treatment response may appear to be minimal (i.e., slopes ranging from 0.91 to 1.02); however, these different slopes have the potential to lead to substantial differences in outcomes when treatment is applied over multiple years and at a high intensity. Overall, participants included in the clusters within the higher-functioning subgroups mastered skills more rapidly than participants included in the clusters in the lower-functioning subgroups. The current findings expand on research demonstrating variation in individual response to ABA treatment, in that those with less severe ASD symptoms [37–3943,45] and who are higher functioning in various developmental domains, including adaptive [32,36,37,40,43,46,47], cognitive [32,37–3942,43,45,46], language [38,46,47], and social [38,47], have greater success in treatment. The current findings are in line with such research showing that individuals who are higher functioning make greater gains during treatment, with implications that those individuals may make up a distinct subtype of ASD.

The findings of the current study suggest that there are two distinct behavioral subgroups within ASD (Clusters A through D and E). In contrast to the current classification of ASD [1], that is a single disorder with a spectrum of severity, findings of the present study suggest there may be two distinct autism spectrums, each with different levels of severity. These finding are in line with Fein and colleagues [60] and Stevens and colleagues [61], whose cluster analyses revealed two distinct ASD subgroups, a high-functioning subgroup and a low-functioning subgroup. The current study expands on these findings by breaking down the specific developmental deficits that are unique to each subgroup. However, while individuals in Cluster E tended to have higher levels of skills than Clusters A through D, the distinction is more than simply a low and high functioning group. Individuals in Clusters A through D showed a distinct pattern of skills. Furthermore, the current findings are in line with the recent studies exploring developmental trajectories of ASD subgroups [62,63], in that within the two larger ASD subgroups, there were multiple clusters that appear to describe development along a common continuum. Additionally, the unique treatment response profiles observed across clusters in the present study may influence developmental trajectories and outcomes.

There may be an immediate application of these clustering techniques within a clinical setting, such that a tool developed to prospectively identify cluster membership for each new patient could

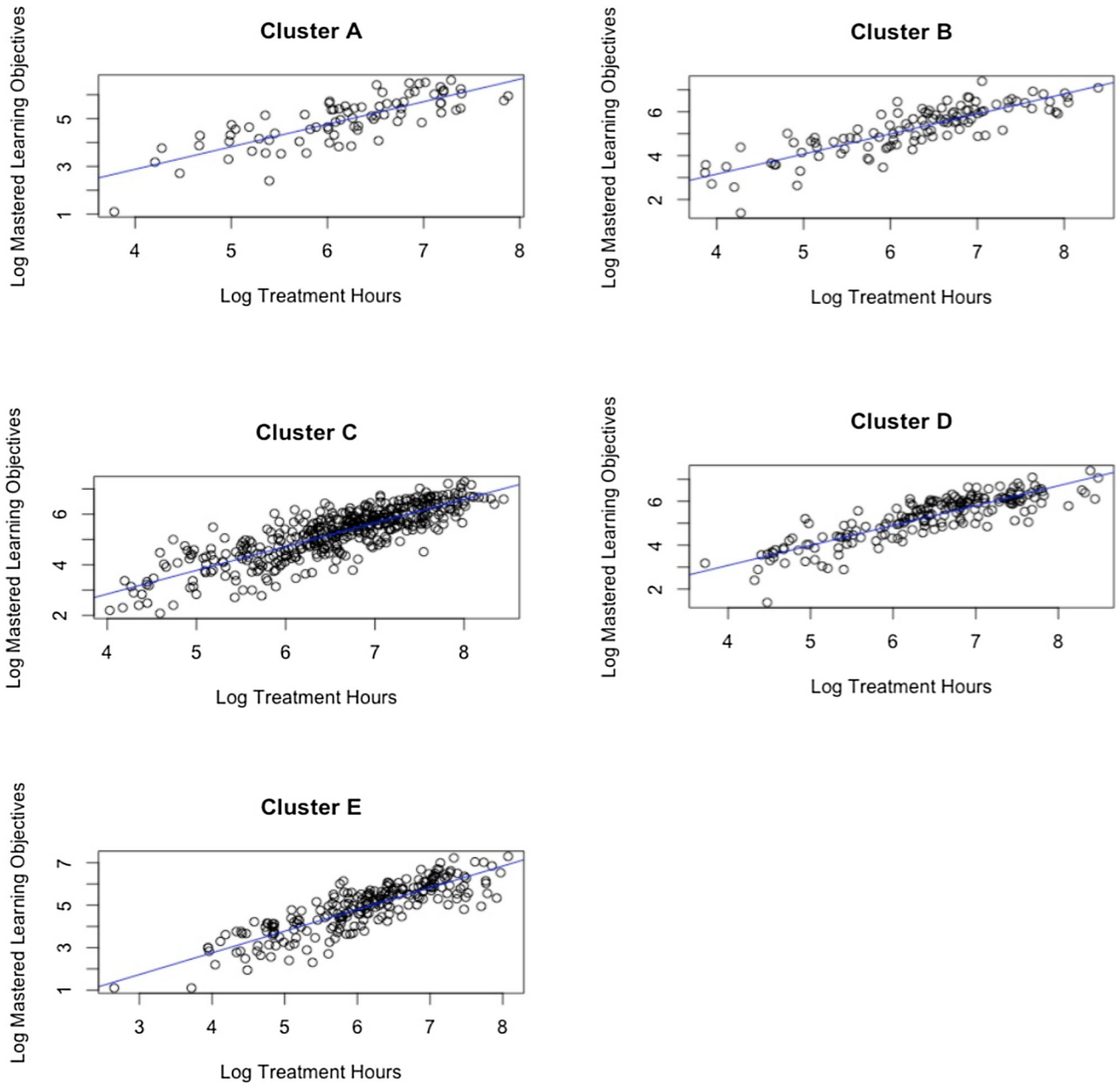

Fig. 3. Scatter plots depicting treatment intensity and learning outcomes for each cluster.

Table 1
Linear regression parameters by cluster.

Cluster	Intercept	Intensity	R^2	p
A	−0.89344	0.94295	0.6355	< 0.001
B	−0.4933	0.91304	0.7107	< 0.001
C	−0.90675	0.9373	0.698	< 0.001
D	−0.5483	0.9065	0.7499	< 0.001
E	−1.33326	1.02234	0.7257	< 0.001

improve initial treatment planning and setting treatment expectations. In light of previous research [71], there may be an optimal treatment approach for each cluster wherein the particular skill deficits lead to different recommendations in the initial dose and targets of therapy. Clearly, such a tool would require significant further development and validation. We are currently in the initial stages of packaging the cluster models described here in order to integrate them into the Skills assessment system. A future study comparing the treatment paths of these new patients to patients used to train the clusters models is planned once a suitable amount of data has been collected.

A limitation of the current study is the lack of data from standardized assessments. The use of standardized assessments could have enabled better comparison to previous cluster research, including the findings of Fein and colleagues [60] and Stevens and colleagues [61], as well as to other ABA outcome studies, and it could provide information on the practical relevance of the skills that were mastered. A criticism of standardized assessments, particularly for evaluations of treatment response, is that they measure broad behaviors [34,[72]. The use of sensitive measures that detect response to treatment has advantages in the study of subgroups [50]. The Skills assessment, used in the current study, measures over 3000 skills across eight areas of child development and gives a fine-grained representation of a child's developmental deficits. Furthermore, the use of proximal outcomes, such as the rate of skill acquisition, allows for greater comparison of individualized treatment progress than more distal outcomes such as standardized assessments [72],[73]. Rate of skill acquisition does have its own

limitations. For instance, skills can vary greatly in level of difficulty (i.e., some skills are more challenging and therefore take longer to acquire than others). It should be mentioned that restricted repetitive behaviors, a core symptom of ASD, and challenging behaviors were not included in this analysis, which could offer further behavioral insights into the identified clusters.

While this study was able to benefit from a large volume of data of known provenance by leveraging the Skills system, it is important to emphasize that large sample sizes and the existence of clean data is a common hurdle to the application of machine learning in general. As with many computational models, the quality of the results produced by machine learning is closely tied to the quality of the data that serves as input to the algorithms. This problem is referred to colloquially by computer scientists as "garbage in, garbage out." Thus, successful applications of machine learning in this domain require, from the outset, an interdisciplinary team of clinicians and data scientists to ensure that the quality of data is suitable for the planned machine learning tasks. Similarly, building machine learning models on data sets with large numbers of attributes (high dimensionality) also requires a suitable volume of data to avoid overfitting so that the models will generalize to unseen data. This can prove challenging in some clinical settings due to the time and cost associated with data collection. In these situations, practitioners can benefit from adopting a hybrid model whereby machine learning algorithms are augmented with human expertise in-the-loop. This allows machine learning pipelines, which are typically closed, to benefit from human oracles which can provide interactive guidance to improve the models learned by the algorithms when data is sparse [74]. Similarly, semi-supervised learning [75], in which a small amount of labeled training data is used in conjunction with large volumes of unlabeled training data, can provide a practical alternative to the fully unsupervised machine learning models leveraged in this study.

The findings presented in this paper are preliminary and require replication across different samples of individuals with ASD. Furthermore, longitudinal research is needed to evaluate if group membership across behavioral subgroups is stable or if individuals may move between these groups (e.g., as they make progress in treatment). Other research initiatives may evaluate biological and genetic differences between the behavioral subgroups. Such study may lead to clearer understanding of etiological differences in ASD. Furthermore, if biological and/or genetic markers are identified, such markers may support detection of ASD subtypes for targeted treatment. As more evidence emerges supporting distinct ASD subtypes, research across fields, whether etiologic, biologic, genetic, treatment, etc., should consider potential segmentation within the disorder. While segmentation has proven difficult, the current study has demonstrated that unsupervised machine learning provides a computationally efficient method to approach these research questions.

Competing interests

The authors declare no competing interests.

Funding

No external funding was obtained for this research.

Contributors

All authors made significant contributions to the manuscript. DD and EL conceived the research design. ES and EL conducted the analysis. All authors contributed to drafting and revising the manuscript.

All persons listed as authors qualify as such based on their contributions.

All authors made significant contributions to the manuscript. DD and EL conceived the research design. ES and EL conducted the analysis. All authors contributed to drafting and revising the manuscript.

Furthermore, the authors declare that this contribution is original, and not under consideration for publication in any other venues.

Summary table

Previously Known

- ASD is a complex disorder with diverse behaviors across the spectrum
- Presence of behavioral subtypes supported by previous high-level analyses

Contributions of this work

- Low-level analysis of ASD behavioral phenotypes based on 8 dimensions of skills
- Use of unsupervised machine learning to detect behavioral profiles, and relationship between them, on large (N = 2400) sample size
- Analysis of response to treatment across behavioral clusters.

References

[1] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 5th ed., American Psychiatric Publishing, Arlington, VA, 2013.
[2] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 4th ed., American Psychiatric Association, Washington, DC, 2000 text rev.
[3] C. Lord, E. Petkova, V. Hus, et al., A multisite study of the clinical diagnosis of different autism spectrum disorders, Arch. Gen. Psychiatry 69 (2012) 306–313.
[4] R. Grzadzinski, M. Huerta, C. Lord, DSM-5 and autism spectrum disorders (ASDs): an opportunity for identifying ASD subtypes, Mol. Autism 4 (2013) 1–6.
[5] C. Lord, S. Risi, A. Pickles, Trajectory of language development in autistic spectrum disorders, in: M.L. Rice, S.F. Warren (Eds.), Developmental Language Disorders: From Phenotypes to Etiologies, Lawrence Erlbaum, Mahwah, NJ, 2004, pp. 7–29.
[6] S.M. Kanne, A.J. Gerber, L.M. Quirmbach, et al., The role of adaptive behavior in autism spectrum disorders: implications for functional outcome, J. Autism Dev. Disord. 41 (2011) 1007–1018.
[7] D.R. Dajani, M.M. Llabre, M.B. Nebel, et al., Heterogeneity of executive functions among comorbid neurodevelopmental disorders, Sci. Rep. 6 (2016) 1–10.
[8] D.R. Dixon, D. Granpeesheh, J. Tarbox, et al., Early detection of autism spectrum disorders, in: J.L. Matson, P. Sturmey (Eds.), InTernational Handbook of Autism and Pervasive Developmental Disorders, Springer, New York, NY, 2011, pp. 197–214.
[9] L. Zwaigenbaum, S. Bryson, T. Rogers, et al., Behavioral manifestations of autism in the first year of life, Int. J. Dev. Neurosci. 23 (2005) 143–152.
[10] B.D. Barger, J.M. Campbell, J.D. McDonough, Prevalence and onset of regression within autism spectrum disorders: a meta-analytic review, J. Autism Dev. Disord. 65 (2013) 1–23.
[11] S.E. Levy, E. Giarelli, L.C. Lee, et al., Autism spectrum disorder and co-occurring developmental, psychiatric, and medical conditions among children in multiple populations of the United States, J. Dev. Behav. Pediatr. 31 (2010) 267–275.
[12] J.L. Matson, P.E. Cervantes, Commonly studied comorbid psychopathologies among persons with autism spectrum disorder, Res. Dev. Disabil. 35 (2014) 952–962.
[13] B.S. Abrahams, D.H. Geschwind, Advances in autism genetics: on the threshold of a new neurobiology, Nat. Rev. Genet. 9 (2008) 341–355.
[14] C. Betancur, Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting, Brain Res. 22 (2011) 42–77.
[15] E. Pisula, K. Ziegart-Sadowska, Broader autism phenotype in siblings of children with ASD – a review, Int. J. Mol. Sci. 16 (2015) 13217–13258.
[16] C. Richards, C. Jones, L. Groves, et al., Prevalence of autism spectrum disorder phenomenology in genetic disorders: a systematic review and meta-analysis, Lancet Psychiatry 2 (2015) 909–916.
[17] R. Raz, A.L. Roberts, K. Lyall, et al., Autism spectrum disorder and particulate matter air pollution before, during, and after pregnancy: a nested case-control analysis within the Nurses' Health Study II Cohort, Environ. Health Perspect. 123 (2015) 264–270.
[18] H.E. Volk, F. Lurmann, B. Penfold, et al., Traffic-related air pollution, particulate matter, and autism, JAMA Psychiatry 70 (2013) 71–77.
[19] S. St-Hilaire, V.O. Ezike, H. Stryhn, et al., An ecological study on childhood autism, Int. J. Health Geogr. 11 (2012) 1–8.
[20] L.A. Croen, J.K. Grether, C.K. Yoshida, et al., Antidepressant use during pregnancy and childhood autism spectrum disorders, Arch. Gen. Psychiatry 68 (2011) 1104–1112.
[21] A.D. Rasalam, H. Hailey, J.H. Williams, et al., Characteristics of fetal anticonvulsant syndrome associated autistic disorder, Dev. Med. Child Neurol. 47 (2005) 551–555.
[22] H.O. Atladóttir, P. Thorsen, L. Østergaard, et al., Maternal infection requiring hospitalization during pregnancy and autism spectrum disorders, J. Autism Dev. Disord. 40 (2010) 1423–1430.

[23] B.K. Lee, C. Magnusson, R.M. Gardner, et al., Maternal hospitalization with infection during pregnancy and risk of autism spectrum disorders, Brain Behav. Immun. 44 (2015) 100–105.
[24] O. Zerbo, Y. Qian, C. Yoshida, et al., Maternal infection during pregnancy and autism spectrum disorders, J. Autism Dev. Disord. 45 (2015) 4015–4025.
[25] L.A. Croen, D.V. Najjar, B. Fireman, et al., Maternal and paternal age and risk of autism spectrum disorders, Arch. Pediatr. Adolesc. Med. 161 (2007) 334–340.
[26] A. Conde-Agudelo, A. Rosas-Bermudez, M.H. Norton, Birth spacing and risk of autism and other neurodevelopmental disabilities: a systematic review, Pediatrics 137 (2016) 1–13.
[27] H. Gardener, D. Spiegelman, S.L. Buka, Perinatal and neonatal risk factors for autism: a comprehensive meta-analysis, Pediatrics 128 (2011) 344–355.
[28] S. Eldevik, R.P. Hastings, J.C. Hughes, et al., Meta-analysis of early intensive behavioral intervention for children with autism, J. Clin. Child Adolesc. Psychol. 38 (2009) 439–450.
[29] B. Reichow, Overview of meta-analyses on early intensive behavioral intervention for young children with autism spectrum disorders, J. Autism Dev. Disord. 42 (2012) 512–520.
[30] B. Reichow, E.E. Barton, B.A. Boyd, et al., Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD), Cochrane Database Syst. Rev. (2012) 10.
[31] D. Granpeesheh, J. Tarbox, D.R. Dixon, Applied behavior analytic interventions for children with autism: a description and review of treatment research, Ann. Clin. Psychiatry 21 (2009) 162–173.
[32] S. Eldevik, R.P. Hastings, J.C. Hughes, et al., Using participant data to extend the evidence base for intensive behavioral intervention for children with autism, Am. J. Intellect. Dev. Disabil. 115 (2010) 381–405.
[33] P. Howlin, I. Magiati, T. Charman, Systematic review of early intensive behavioral interventions for children with autism, Am. J. Intellect. Dev. Disabil. 114 (2009) 23–41.
[34] D. Granpeesheh, D.R. Dixon, J. Tarbox, et al., The effects of age and treatment intensity on behavioral intervention outcomes for children with autism spectrum disorders, Res. Autism Spectr. Disord. 3 (2009) 1014–1022.
[35] E. Linstead, D.R. Dixon, R. French, et al., Intensity and learning outcomes in the treatment of children with autism spectrum disorder, Behav. Modif. 41 (2017) 229–252.
[36] M.K. Makrygianni, P. Reed, A meta-analytic review of the effectiveness of behavioural early intervention programs for children with autistic spectrum disorders, Res. Autism Spectr. Disord. 4 (2010) 577–593.
[37] B. Remington, R.P. Hastings, H. Kovshoff, et al., Early intensive behavioral intervention: outcomes for children with autism and their parents after two years, Am. J. Ment. Retard. 112 (2007) 418–438.
[38] E. Ben Itzchak, D.A. Zachor, Who benefits from early intervention in autism spectrum disorders? Res. Autism Spectr. Disord. 5 (2011) 345–350.
[39] S. Eldevik, R.P. Hastings, E. Jahr, et al., Outcomes of behavioral intervention for children with autism in mainstream pre-school settings, J. Autism Dev. Disord. 42 (2012) 210–220.
[40] H.E. Flanagan, A. Perry, N.L. Freeman, Effectiveness of large-scale community-based intensive behavioral intervention: a waitlist comparison study exploring outcomes and predictors, Res. Autism Spectr. Disord. 6 (2012) 673–682.
[41] D. Granpeesheh, D.R. Dixon, J. Tarbox, et al., The effects of age and treatment intensity on behavioral intervention outcomes for children with autism spectrum disorders, Res. Autism Spectr. Disord. 3 (2009) 1014–1022.
[42] S.L. Harris, J.S. Handleman, Age and IQ at intake as predictors of placement for young children with autism: a four- to six-year follow-up, J. Autism Dev. Disord. 30 (2000) 137–142.
[43] A. Perry, A. Cummings, J. Dunn Geier, et al., Predictors of outcome for children receiving intensive behavioral intervention in a large, community-based program, Res. Autism Spectr. Disord. 5 (2011) 592–603.
[44] J. Virués-Ortega, V. Rodríguez, C.T. Yu, Prediction of treatment outcomes and longitudinal analysis in children with autism undergoing intensive behavioral intervention, Int. J. Clin. Health Psychol. 13 (2013) 91–100.
[45] T. Smith, A.D. Groen, J.W. Wynn, Randomized trial of intensive early intervention for children with pervasive developmental disorder, Am. J. Ment. Retard. 105 (2000) 269–285.
[46] I. Magiati, J. Moss, T. Charman, et al., Patterns of change in children with autism spectrum disorders who received community based comprehensive interventions in their pre-school years: a seven year follow-up study, Res. Autism Spectr. Disord. 5 (2011) 1016–1027.
[47] G.O. Sallows, T.D. Graupner, Intensive behavioral treatment for children with autism: four-year outcome and predictors, Am. J. Ment. Retard. 110 (2005) 417–438.
[48] Z. Warren, M.L. McPheeters, N. Sathe, et al., A systematic review of early intensive intervention for autism spectrum disorders, Pediatrics 127 (2011) e1303-11.
[49] L.J. Beglinger, T.H. Smith, A review of subtyping in autism and proposed dimensional classification model, J. Autism Dev. Disord. 31 (2001) 411–422.
[50] S. Georgiades, P. Szatmari, M. Boyle, Importance of studying heterogeneity in autism, Neuropsychiatry 3 (2013) 123–125.
[51] J.J. Shen, P.H. Lee, J.J. Holden, et al., Using cluster ensemble and validation to identify subtypes of pervasive developmental disorders, AMIA Annu. Symp. Proc. (2007) 666–670.
[52] V.W. Hu, M.E. Steinberg, Novel clustering of items from the Autism Diagnostic Interview-Revised to define phenotypes within autism spectrum disorders, Autism Res. 2 (2009) 67–77.
[53] D.G. Ingram, T.N. Takahashi, J.H. Miles, Defining autism subgroups: a taxometric solution, J. Autism Dev. Disord. 38 (2008) 950–960.
[54] M.V. Lombardo, M.C. Lai, B. Auyeung, et al., Unsupervised data-driven stratification of mentalizing heterogeneity in autism, Sci. Rep. 6 (2016) 1–15.
[55] V. Eapen, R. Crnčec, A. Walter, Exploring links between genotypes, phenotypes, and clinical predictors of response to early intensive behavioral intervention in autism spectrum disorder, Front. Hum. Neurosci. 7 (2013) 1–10.
[56] M.L. Cuccaro, R.F. Tuchman, K.L. Hamilton, et al., Exploring the relationship between autism spectrum disorder and epilepsy using latent class cluster analysis, J. Autism Dev. Disord. 42 (2012) 1630–1641.
[57] O.J. Veatch, J. Veenstra-Vanderweele, M. Potter, et al., Genetically meaningful phenotypic subgroups in autism spectrum disorders, Genes Brain Behav. 13 (2014) 276–285.
[58] V.W. Hu, Y. Lai, Developing a predictive gene classifier for autism spectrum disorders based upon differential gene expression profiles of phenotypic subgroups, N. Am. J. Med. Sci. 6 (2013) 1–18.
[59] H. Bruining, L. de Sonneville, H. Swaab, et al., Dissecting the clinical heterogeneity of autism spectrum disorders through defined genotypes, PLoS One 5 (2010) 1–7.
[60] D. Fein, M. Stevens, M. Dunn, et al., Subtypes of pervasive developmental disorder: clinical characteristics, Child Neuropsychol. 5 (1999) 1–23.
[61] M. Stevens, D. Fein, M. Dunn, et al., Subgroups of children with autism by cluster analysis: a longitudinal examination, J. Am. Acad. Child Adolesc. Psychiatry 39 (2000) 346–352.
[62] A. Pickles, D.K. Anderson, C. Lord, Heterogeneity and plasticity in the development of language: a 17-year follow-up of children referred early for possible autism, J. Child Psychol. Psychiatry 55 (2014) 1354–1362.
[63] C. Lord, S. Bishop, D. Anderson, Developmental trajectories as autism phenotypes, Am. J. Med. Genet. C Semin. Med. Genet. 169 (2015) 198–208.
[64] O. Ousley, T. Cermak, Autism spectrum disorder: defining dimensions and subgroups, Curr. Dev. Disord. Rep. 1 (2014) 20–28.
[65] Z. Gu, M. Churchman, K. Roberts, et al., Genomic analyses identify recurrent MEF2D fusions in acute lymphoblastic leukaemia, Nat. Commun. 7 (2016) 1–10.
[66] B.D. Lehmann, J.A. Bauer, X. Chen, et al., Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies, J. Clin. Invest. 121 (2011) 2750–2767.
[67] D.R. Dixon, J. Tarbox, A.C. Najdowski, et al., A comprehensive evaluation of language for early behavioral intervention programs: the reliability of the SKILLS Language Index, Res. Autism Spectr. Disord. 5 (2011) 506–511.
[68] A. Persicke, M.R. Bishop, C.M. Coffman, et al., Evaluation of the concurrent validity of a skills assessment for autism treatment, Res. Autism Spectr. Disord. 8 (2014) 281–285.
[69] D. Granpeesheh, J. Tarbox, A.C. Najdowski, et al., Evidence-Based Treatment for Children with Autism: the CARD Model, Elsevier, Waltham, MA, 2014.
[70] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464.
[71] E. Linstead, D.R. Dixon, E. Hong, et al., An evaluation of the effects of intensity and duration on outcomes across treatment domains for children with autism spectrum disorder, Transl. Psychiatry 19 (2017) 1–6.
[72] J.L. Matson, R.L. Goldin, Early intensive behavioral interventions: selecting behaviors for treatment and assessing treatment effectiveness, Res. Autism Spectr. Disord. 8 (2014) 13–142.
[73] P.J. Yoder, K. Bottema-Beutel, T. Woynaroski, et al., Social communication intervention effects vary by dependent variable type in preschoolers with autism spectrum disorders, Evid. Commun. Assess. Interv. 7 (2013) 150–174.
[74] Holzinger, Andreas. Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Inform. 3 (2016) 119–131.
[75] X. Zhu, A.B. Goldberg, Introduction to semi-supervised learning, Synth. Lect. Artif. Intell. Mach. Learn. 3 (2009) 1–130.