Application and research progress of machine learning in the diagnosis and treatment of neurodevelopmental disorders in children

Chao Song[1]*, Zhong-Quan Jiang[2], Dong Liu[3] and Ling-Ling Wu[1]

[1]Department of Developmental and Behavioral Pediatrics, The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Centre for Child Health, Hangzhou, China, [2]School of Public Health, Lanzhou University, Lanzhou, China, [3]Department of Neonatology, Shenzhen People's Hospital, Shenzhen, China

The prevalence of neurodevelopment disorders (NDDs) among children has been on the rise. This has affected the health and social life of children. This condition has also imposed a huge economic burden on families and health care systems. Currently, it is difficult to perform early diagnosis of NDDs, which results in delayed intervention. For this reason, patients with NDDs have a prognosis. In recent years, machine learning (ML) technology, which integrates artificial intelligence technology and medicine, has been applied in the early detection and prediction of diseases based on data mining. This paper reviews the progress made in the application of ML in the diagnosis and treatment of NDDs in children based on supervised and unsupervised learning tools. The data reviewed here provide new perspectives on early diagnosis and treatment of NDDs.

KEYWORDS

artificial intelligence, machine learning, child, neurodevelopmental disorder, diagnosis, treatment

Introduction

Neurodevelopmental disorders (NDDs) including autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), intellectual disability (ID), and learning disability (LD) are a class of diseases that affect brain development and function. These disorders occur during early development and affect the cognitive and emotional development of children (1–3). Evidence shows that burden of NDDs in children is becoming a global challenge, affecting about 3% of children worldwide (4). The incidence of NDDs has been on the rise globally. In ASD, the 2020 monitoring network report by the Centers for Disease Control and Prevention revealed that the prevalence of ASD among 8-year-old children was 1.68%, representing a 10% increase compared with 2018 (5). In 2021, a surveillance report showed that the prevalence of ASD had risen to 2.27% or 1 in every 44 children (6). Moreover, several

meta-analyses have reported varying global prevalence rates. For instance, the prevalence of ADHD in children was 7.2% (7), that of ID was 1–3% (8), whereas that of LD was 3–8% (9, 10). Of note, NDDs affect the health and social functioning of children, as well as imposes huge economic burden on families (11, 12).

Studies have shown that NDDs is mainly caused by genetic and environmental factors. However, the pathogenesis of NDDs, represented by ASD/ADHD, is unclear and there are no accurate biomarkers of this disorders (13). Currently, early diagnosis of NDDs is difficult due to the high heterogeneity of its phenotypes and etiological factors (14). This results in delayed intervention. Therefore, there is an urgent need to develop strategies for improving early detection and prediction of NDDs. In clinical practice, NDDs are mainly diagnosed based on behavioral symptoms of children and information provided by caregivers (2, 15). This calls for development of standardized diagnostic neuropsychological testing tools for this condition. Moreover, diagnosis based on behavioral symptoms is not accurate because it dependents on the pediatricians' experience and observation time. Currently, only about 8% of pediatric providers have the skills to diagnose NDDs (16). There are differences in the reliability and validity of standardized test tools for NDDs, but such tools cannot be easily obtained, due to geographical or cultural reasons (17). Currently, no testing tool or scale can directly diagnose NDDs. Even the available Autism Diagnostic Observation Scale and Autism Diagnostic Interview-Revised guidelines regarded as the "gold standard" for ASD diagnosis may lead to misdiagnosis (18).

Considering the inability of single scales, tools or indicators to accurately diagnose or predict NDDs, it has been proposed that objective index data (e.g., socio demographic information, EEG, skull imaging) should be combined to improve the diagnosis or prediction of NDDs. Machine learning (ML) has been found to offer good predictive performance on the occurrence of NDDs (19). Several ML methods such as, supervised, unsupervised, semi-supervised, and reinforcement learning, have been used in the diagnosis and treatment of NDDs (20–22). Semi supervised learning and reinforcement learning are rarely used in the field of NDDs. Semi-supervised learning and reinforcement learning are rarely used in the field of NDD with its unique data processing advantages, ML can facilitate the early identification and early diagnosis of NDD.

Abbreviations: ADHD, Attention deficit hyperactivity disorder; ASD, Autism spectrum disorder; ANN, Artificial neural network; BP, Back Propagation; CART, Classification and regression tree; CNN, Convolutional neural network; DRA, Dopamine receptor antagonist; ID, Intellectual disability; ID3, Iterative dichotomiser 3; LD, Learning disability; LSTM, Long and short-term memory model; ML, Machine learning; MLP, Multilayer perceptron; NDD, Neurodevelopment disorder; PCA, Principal component analysis; Rs-fMRI, Resting-state functional Magnetic Resonance Imaging; SVM, Support vector machines; TS, Tourette syndrome.

Reviewing the progress of ML in the field of NDD is a reflection of the cross-fertilization of medicine and engineering, which helps to expand the boundaries of ML applications and deepen the understanding of NDD among medical professionals. Therefore, this paper focuses on the application of supervised and unsupervised learning in NDD to provide a scientific basis for improving the quality of life of NDD patients.

Supervised learning

Supervised learning can be applied in early detection, prediction of NDDs, and identification of risk factors. Regression analysis, decision tree, support vector machine, and artificial neural network are the commonly used supervised ML methods.

Regression analysis

Regression analysis is the most basic and widely utilized ML model. Linear regression, logistic regression, and regularized regression are interpretable and are extensively. For instance, Wang et al. adopted multivariate binary logistic regression analysis to identify factors associated with ASD. They found that gender, living area, age, and education level are contributing factors contributing to ASD occurrence (23). Tourette syndrome (TS) is the most common neurodevelopmental movement disorder (2). Elsewhere, Burd et al. used binary logistic regression analysis to develop a regression model for evaluating factors contributing to TS. They found that being male, without a family history of TS, and high number of comorbidities influence the occurrence of TS (24). Bertoncelli et al. established a binary logistic regression analysis model comprising 91 adolescents with cerebral palsy for predicting cerebral palsy in children and the associated risk factors. The average accuracy, specificity and sensitivity of the model were 78%. It also suggested that poor motor skills, epilepsy and cerebral palsy were related risk factors. This implies that a prediction model based on binary logistics can effectively identify children with cerebral palsy (25).

There are a lot of influential factors in NDDs, which inevitably leads to collinearity problems. If these factors are not controlled and filtered, they affect the model performance and even lead to production of misleading results. To address this problem, regularization technology has been proposed. In the European multicenter children's TS study (EMTICS), 187 first-degree relatives of TS children aged between 3 and 10 were followed up for 7 years. Subsequently, a lasso logistic regression prediction model for Tourette was established. The interpretation of this method were relatively simple and its prediction accuracy was good (26), indicating the extensive use of regression analysis in the field of NDDs.

Decision tree

The decision tree was first proposed in 1986 (27). It possesses tree classifier classification properties and can produce interpretable and accurate results without parameter assumptions. Iterative dichotomiser 3 (ID3), classification and regression tree (CART) are the most widely used to generate medical decision rules for NDDs. Mohamma et al. used features such as, child behavior, neuropsychology, and electrophysiological markers to build models. They then constructed an early childhood predictive model for ADHD using the classic ID3 algorithm. They reported that the decision tree model yielded excellent classification accuracy (100%). Also, subtypes of ADHD can be distinguished by key nodes in decision-making rules such as behavioral, neuropsychiatric and electrophysiological parameters (28). New algorithms based on classical decision tree algorithms, including the ones using alternate decision trees, multi-class alternate decision trees, have been used to construct models based on genomic and magnetic resonance data. It has been found that the decision tree outperforms other ML models. Consequently, rs878960 in GABRB3 (gamma-aminobutyric acid A receptor, beta 3) has been selected by all tree-based models (29). In practical application, the decision tree is prone to overfitting. Effective sampling methods and pruning methods should be developed to solve the problem of overfitting. CART, which is extensively used, utilizes a cost complexity pruning algorithm. Previously, the predictive significance of birth weight, term infants, and Apgar score in ADHD was explored. A total of 132 boys diagnosed with ADHD and 146 typical developmental boys in the control group. The decision tree model constructed using the CART algorithm revealed that the Apgar score used to reflect the degree of neonatal asphyxia had the highest predictive value, whereas a low Apgar score was among the most critical risk factors in the perinatal period of ADHD children, suggesting that perinatal asphyxia may be related to later occurrence of NDDs symptoms. Therefore, application of complexity pruning algorithm for post pruning improves the prediction accuracy of the decision tree (30).

Support vector machines

Previously, Cortes et al. proposed a linear classifier model which had the largest spacing in feature space and a support vector machine (SVM). The model can solve a separation hyperplane that correctly divides the training dataset with the largest geometric intervals (31). SVM has good performance on small sample implementations. Notably, linear kernel functions, polynomial kernel functions, sigmoid, radial basis function kernels are frequently utilized kernel functions. For instance, Conti et al. used retrospective cohort data from 68 children aged 34–74 months from the head of MRI to construct an early differential diagnostic model of ASD and Childhood Apraxia of Speech (CAS) of linear nuclear function SVM. It was found that the linear kernel function SVM model effectively achieved early differential diagnosis and individualized intervention of ASD and CAS (32). Similarly, Agastinose Ronicko et al. used Gaussian kernel SVM, random forest, and convolutional neural network to construct a predictive model based on Resting-state functional Magnetic Resonance Imaging (Rs-fMRI) data for early diagnosis and treatment of ASD. They found that compared with other machine learning mentioned above, Gaussian kernel SVM has stronger performance in early diagnosis and treatment of ASD (33). To improve the performance of individual SVM classifiers, Bi et al. constructed an ensemble SVM model by integrating Rs-fMRI data from 46 normal children and 61 children with ASD. The proposed ensemble SVM model showed good classification performance based on all features, implying that the ensemble SVM method can be used as an auxiliary diagnosis of ASD (34). Objective imaging data obtained by Rs-fMRI technology is more effective for the diagnosis of ASD compared with behavioral observation. SVM has excellent performance in the above imaging data and small samples.

Artificial neural network

An artificial neural network (ANN) is a complex network structure formed by interconnection of numerous processing units. It is a form of abstraction, simplification, and simulation of the structure and operation mechanism of the human brain. ANN can perform simulations, image recognition, and prediction functions. In an investigation aimed at evaluating the relationship between athletic capacity and other clinical features of ASD, Fulceri et al. performed exploratory analysis via ANN. Poor motor performance is a common clinical feature in preschoolers with ASD, associated with repetitive stereotyped behaviors and weak language skills (35). Single-layer neural networks cannot solve the XOR problem in the context of artificial neural networks. In contrast, two-layer neural networks can resolve this problem. At the same time, it demonstrates a strong non-linear classification effect. Rumelhar et al. proposed the Back Propagation (BP) algorithm in 1986 (36). BP solves the complex computational quantity problem required by two-layer neural networks and the computational problem of multilayer perceptron (MLP). The concept of implicit layer was introduced to act as a kernel function of an SVM that maps sample spaces to high-dimensional linear separable spaces. Moreover, Hossain et al. analyzed demographic data, clinical indicators, and imaging data to identify ASD features and construct the MLP classifier model to improve the accuracy of automated diagnosis of children with ASD. It was observed that the MLP outperformed all other benchmark classification models, achieving a 100% accuracy with the lowest number of attributes in the toddler, child, adolescent, and adult datasets (37).

With the development of computer technology, the number of layers of neural network is increasing, and the problem of local optimal solution is becoming more and more prominent. The "convolutional kernel" is an intermediary, model which ensures that the original position relationship is preserved after an image is convoluted, thereby limiting the risk of falling into a locally optimal solution. Therefore, several convolutional neural networks (CNNs) have been proposed. Thomas et al. trained 3D-CNNs on an open ASD dataset to distinguish ASD using Rs-fMRI images and constructed a CNN-based ASD recognition model. Results showed that 3D-CNN had better distinguishing effect. Moreover, its performance exceeded that of the SVM model. However, valuable information cannot be extracted from time series in 3D-CNNs (38). Scientists have developed a long and short-term memory model (LSTM) to solve the disappearance of gradients in time. This model fulfills the time memory function by switching the gate and preventing the gradient from disappearing. Vikas et al. developed CNN, LSTM, and MLP (based on DSM-V) models for accurate diagnosis and assessment of severity of individuals with ASD. Comparative analysis revealed that LSTM functions better in the diagnosis of ASD unlike other neural network algorithms (e.g., CNN, MLP). This suggests that AI algorithms can improve the diagnosis of ASD (39). DSM-V is the most widely used diagnostic criteria for NDDs worldwide. The combination of DSM-V and ML not only enriches the connotation of DSM-V, but also proves that ML is suitable for the diagnosis and treatment of NDDs.

Ensemble learning

Ensemble learning accomplishes learning tasks by constructing and integrating multiple weak learners. Common ensemble learning methods include boosting, bagging, and stacking (40–42). AdaBoost is an efficient boosting algorithm that allows weak learning algorithms with approximate random accuracy to be strong learning algorithms (43). PU Putra et al. explored responses and gaze performance of children during Go/No-Go missions. Based on the AdaBoost algorithm, the eye tracker was used to track the gaze data of children and construct a distinguishing model for ASD. As a result, the accuracy rate of AdaBoost's algorithm predicting ASD reached 88.60%, which has an application value (44). The collected the gaze data was huge and complex, and it was difficult to analyze such data with traditional statistical methods, and can only be processed by ML.

Of note, the Bagging algorithm is a parallel integration strategy that differs from Boosting. Bagging insights are applied to decision trees to obtain random forest models, further improving the predictive performance of the decision tree model (45). Feczko E et al. utilized Rs-fMRI brain connection data from 47 children with ASD and 58 healthy children to construct a random forest model to distinguish ASD. The findings showed

a prediction accuracy of the random forest model of 72.71%, a specificity of 80.74%, and a sensitivity of 63.15%. Besides, unique behavioral characteristics of 3 ASD and 4 subsets of normal children were simultaneously revealed, showing that the random forest model performs effectively with extremely high value in the interpretation of features (46). In an exploratory analysis, random forests are extensively used for favorable robustness. Gao et al. sampled feces from 49 tic children and 50 healthy children for intestinal microbiome analysis to investigate the intestinal microbial features in tic patients and the effects of dopamine receptor antagonist (DRA) drugs on the composition and metabolic function of the intestinal microbiota. A random forest model was constructed to predict tic. The results showed that the model had an AUC of 0.884. Moreover, a significant correlation was noted between the severity of tic symptoms and abundance of multiple bacteria as well as the metabolic function of the gut microbiota (47).

Based on boosting and bagging, a stacking technique using different models for integration has emerged (48), however, literature related to NDDs is few; therefore, the application value warrants further investigations.

Unsupervised learning

Unsupervised learning aims to train a model to learn the data structure, then provide valuable information about a new sample. The most significant distinction between unsupervised and supervised learning is whether the data contains learning labels or not. The most common scenarios for unsupervised learning include association rules, clustering, and dimensionality reduction.

Association rule

Association rule use metrics to differentiate between strong rules existing in a database. The most common algorithm that uses this rule is the Apriori algorithm (49). Kim et al. applied the Apriori algorithm to extract ADHD comorbidities in Korean national health insurance data. Mood/affective disorders were the most common comorbidities of ADHD. Based on the outcomes of the association rules, 9 association rules were generated, providing a reference for subsequent research on ADHD (50). Many comorbidities are among the characteristics of NDDs. Such comorbidities can be used in the differential diagnosis of NDDs. ML provides a new path for early identification of comorbidities in NDDs, and it can also help to formulate more comprehensive intervention plans to improve outcomes in children with NDDs. Tai et al. also used the Apriori algorithm to evaluate the comorbid network of children with ADHD. Consequently, the risk of comorbidity between ADHD and psychosis was significantly higher than that

with other physical diseases (51). Similarly, association rules can also be used in diagnostic models. For instance, Ucuz et al. investigated the effects of temperament and character traits on ADHD diagnosis. A diagnostic model of ADHD was established based on the classification-based association rules method. Data were collected from 36 children with ADHD and 39 healthy children. The results showed that the diagnostic model based on association rules had good discrimination performance, and temperament personality characteristics can be used for the clinical diagnosis of ADHD (52).

Clustering

Clustering involves dividing a dataset into different classes or clusters based on a set of criteria, to maximize the similarity of data objects within a cluster, while minimizing the difference between data objects that are not in the same cluster. K-means is the most conventional clustering method; it classifies points in n-dimensional space based on the degree of Euclidean distance. Vargason et al. explored ASD complications and the ASD subtypes in the United States between 2000 and 2015 using a database with 3,278 insured children with ASD and 279,693 children with ASD. K-means algorithm was used to identify three subgroups of children with ASD. Meanwhile, there was a strong association between developmental delay and ASD in comorbidities, followed by gastrointestinal problems and immune imbalances. Suggestive clustering results potentially help in screening children with ASD for comorbidities and understanding ASD subgroups (53). In practice, the k-means algorithm has several limitations such as, s specifying the initial number of class clusters and easy overfitting, without obtaining the cluster tree. Therefore, researchers often utilize hierarchical clustering and Gaussian mixed models. For instance, Stevens et al. used hierarchical clustering and Gaussian mixed models to cluster the behavioral phenotypes of ASD and therapeutic outcomes of different phenotypes. This approach provided a scientific reference for personalized interventions (54).

Dimensionality reduction

Clinical data are complex comprising redundant data, which improves the accuracy of model recognition by minimizing dimensionality. At the same time, it also highlights the important structure of data. Of note, principal component analysis (PCA) is the most commonly used linear dimensionality reduction method. The features of origin data points are preserved while data dimensions are reduced (55). For example, N Mashal et al. performed principal component analyses on 37 ASD, 20 LD, and 21 normal children to address the interrelationships between various tests in each group. The results revealed no dichotomy between visual

and verbal metaphors in healthy children. Instead, metaphors were categorized as per their familiarity. In the LD group, visual metaphors were independently categorized as linguistic metaphors. The verbal metaphorical understanding of the ASD group was similar to that of the LD group (56). Additionally, when processing and analyzing a complex image and audio data, Ousts et al. applied PCA technology to minimize data dimensionality, thereby stabilizing the subsequent modeling (57). This suggests that dimensionality reduction methods including PCA should be appropriately used to increase the model stability in processing complex data.

Discussion

In summary, supervised algorithms can be used to develop models for NDDs diagnosis and prediction. Unsupervised algorithms can be applied in exploratory research or optimization of data structures to identify associations between NDDs or key risk factors of a single disorder. Supervised algorithms have varied applicability to different NDDs data structures due to their different algorithm structures. Artificial intelligence has been shown to have good performance on imaging data. For large data samples, ensemble learning often shows fast computing power and performance. In few-shot training, SVM performs well (Table 1). At present, most of the NDDs diagnosis and prediction models built based on ML do not follow the standard *The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis* (TRIPOD) clinical prediction model reporting specifications (68), such as the lack of processing of missing values and outliers in the reporting process, and the failure to report the threshold of the model. This makes the model difficult to reproduce. For model evaluation, multi-dimensional evaluation (e.g., discrimination, calibration, clinical usefulness, etc.) is rarely used, and it is difficult to effectively screen out a model that is truly suitable for samples only from a single discrimination dimension. In terms of model verification, most studies only evaluate the performance of the model on the current sample from the perspective of internal verification, and there is a certain risk of overfitting. Most studies lack the consideration of model generalization ability on external validation based on external data.

Nowadays, several studies have attempted to develop ML clinical diagnostic evaluation tools for NDDs. For example, the ASD diagnosis and assessment tool based on questionnaire data was recently developed by *De novo*. This tool was approved by the Federal Drug Administration for pre-marketing review, which is the first successful application of ML in the early diagnosis and early screening of NDDs (69). More companies, such as ALSOLIFE, are attempting to develop ASD auxiliary diagnostic tools based on ML from imaging data. However, in the field of NDDs research, ML models have numerous

TABLE 1 Advantages and disadvantages of supervised learning and unsupervised learning methods.

	Advantages	Disadvantages
Supervised learning		
Regression analysis (25, 26, 58–61)	1. Simple modeling and strong interpretability 2. Independent of parameter adjustment, the same model and data can usually calculate the unique result	1. Sensitive to missing and abnormal values 2. Logistic regression is difficult to deal with nonlinear problems 3. It is difficult to deal with multicollinearity problems 4. There is risk of over fitting 5. Sensitive to unbalanced data
Decision tree (28–30, 62–66)	1. The model is highly interpretable 2. The model can be described graphically 3. Insensitive to continuous and discrete data 4. It can process multi classification data 5. Insensitive to missing and abnormal values 6. It does not depend on background knowledge and can be modeled directly	1. The decision tree without pruning has the risk of over fitting 2. Sensitive to unbalanced data 3. The performance of the model is generally weaker than that of ensemble learning and regression analysis
SVM (31)	1. It has complete theoretical support, especially suitable for small sample research 2. The computational complexity depends on the support vector, which avoids the disaster of dimensionality to a certain extent 3. A few support vectors determine the final result, reducing the impact of miscellaneous samples on the model 4. Insensitive to outliers	1. It is difficult to train on big data samples 2. It is difficult to solve the multi classification problem 3. The model depends on parameter selection
ANN (38, 39, 67)	1. Strong nonlinear mapping ability 2. Have the ability to associate input information, self-learning and adaptive ability 3. Have a strong ability to distinguish training samples 4. Convolution algorithm can recognize imaging data well	1. The model has the risk of over fitting 2. Large amount of model calculation 3. Complex imaging data analysis
Ensemble learning (40–42, 44, 48)	1. The performance of the model is improved to a certain extent compared with the weak classifier 2. Insensitive to outliers 3. High performance on large samples 4. It can deal with nonlinear problems 5. Random forest is not sensitive to unbalanced data 6. Little possibility of over fitting	1. The model is difficult to explain, and there is a black box problem 2. Normalization is required 3. Some models are sensitive to missing values
Unsupervised learning		
Association rule (49, 50)	1. The algorithm principle is simple and easy to implement 2. It is not restricted by dependent variables, and the association between data can be found in big data	1. There are many output rules and a lot of useless information
Clustering (53, 54, 61)	1. The principle is relatively simple, the implementation is also very easy, and the convergence speed is fast 2. Be able to handle big data problems 3. Strong interpretability	1. The model is sensitive to outliers 2. The model is sensitive to unbalanced data 3. Local optimal solutions are often obtained
Dimensionality reduction (55, 56)	The model is fast, simple and effective	Poor interpretability of the model

limitations. For example, the heterogeneity of ASD in phenotype and pathological mechanism leads to inconsistent performance and result interpretation of ML models on different training samples (14), and it is impossible to obtain a ML model suitable for the entire ASD population. In addition, the training of supervised ML models relies on existing samples, and for NDDs, there is no database of existing samples. Currently, numerous diagnostic models based on clinical imaging data (32, 33, 38), have been reported such as Rs-fMRI and EEG. However, the cost of obtaining these data is high and this imposes a huge economic burden on the patient's family. Even if the ML model has excellent performance on these data, its application in the diagnosis of NDDs is challenging.

There are several limitations of this review article. First of all, this paper focuses on the applicability of ML in the diagnosis and treatment of NDDs, so the subject content of the cited literature is reviewed. Some literature did not present the data in full, so it was impossible to strictly checked the data quality of the cited literature. Second, NDDs are a class of diseases, and the pathogenesis, clinical manifestations, treatment options and prognosis of each disease in NDDs are different. At the same time, the obtained data also have various degrees of difference, and the analysis of different diseases still needs to be combined with the characteristics of the disease data. Currently, there is no single ML method or model that works for all data types. At present, the application of ML in a certain NDDs has been reviewed, and this kind of research is also very meaningful. Finally, since NDDs are a current research hotspot, some of the views in this paper may become incomplete as ML applications in the field further increase.

In conclusion, the benefits of ML in the diagnosis and intervention of NDDs are taking shape with its excellent performance and interpretability. Integration of medical big data and ML may be an effective strategy to guide the diagnosis, intervention, and prognosis of NDDs. Collecting clinical big data of NDDs and constructing models scientifically are the work that can be set out now.

Author contributions

CS conceived the study and critically revised the article. Z-QJ, DL, and L-LW performed literature search and drafted the manuscript. All authors contributed to the study and approved the final version to be submitted.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Parenti I, Rabaneda LG, Schoen H, Novarino G. Neurodevelopmental disorders: from genetics to functional pathways. *Trends Neurosci.* (2020) 43:608–21. doi: 10.1016/j.tins.2020.05.004

2. Battle DE. Diagnostic and statistical manual of mental disorders (DSM). *Codas.* (2013) 25:191–2. doi: 10.1590/s2317-17822013000200017

3. Niemi MEK, Martin HC, Rice DL, Gallone G, Gordon S, Kelemen M, et al. Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature.* (2018) 562:268–71. doi: 10.1038/s41586-018-0566-4

4. Gilissen C JY, Thung DT. Genome sequencing identifies major causes of severe intellectual disability. *Nature.* (2014) 511:344–7. doi: 10.1038/nature13394

5. Maenner M, Shaw K, Baio J, Washington A, Dietz P. Prevalence of autism spectrum disorder among children aged 8 years — autism and developmental

disabilities monitoring network, 11 sites, United States, 2016. *Ment Retard Dev Disabil Res Rev.* (2020) 69:1–12. doi: 10.15585/mmwr.ss6904a1

6. Maenner MJ, Shaw KA, Bakian AV, Bilder DA, Durkin MS, Esler A, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2018. *Ment Retard Dev Disabil Res Rev.* (2021) 70:1–23. doi: 10.15585/mmwr.ss7011a1

7. Thomas R, Sanders S, Doust J, Beller E, Glasziou P. Prevalence of attention-deficit/hyperactivity disorder: a systematic review and meta-analysis. *Pediatrics.* (2015) 135:e994–1001. doi: 10.1542/peds.2014-3482

8. Leonard H,Wen X. The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment Retard Dev Disabil Res Rev.* (2002) 8:117–34. doi: 10.1002/mrdd.10031

9. Law J, Boyle J, Harris F, Harkness A, Nye C. Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature. *Int J Lang Comm Dis.* (2000) 35:165–88. doi: 10.1080/136828200247133

10. Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. *J Speech Lang Hear Res.* (1997) 40:1245–60. doi: 10.1044/jslhr.4006.1245

11. Peñuelas-Calvo I, Palomar-Ciria N, Porras-Segovia A, Miguélez-Fernández C, Baltasar-Tello I, Perez-Colmenero S, et al. Impact of ADHD symptoms on family functioning, family burden and parents' quality of life in a hospital area in Spain. *Eur J Psychiatry.* (2021) 35:166–72. doi: 10.1016/j.ejpsy.2020.10.003

12. Lopez K, Reed J, Magaña S. Associations among family burden, optimism, services received and unmet need within families of children with ASD. *Child Youth Serv Rev.* (2019) 98:105–12. doi: 10.1016/j.childyouth.2018.12.027

13. Bölte S, Girdler S, Marschik P. The contribution of environmental exposure to the etiology of autism spectrum disorder. *Cell Mol Life Sci.* (2019) 76:1275–97. doi: 10.1007/s00018-018-2988-4

14. Bhat S, Acharya UR, Adeli H, Bairy GM, Adeli A. Autism: cause factors, early diagnosis and therapies. *Rev Neurosci.* (2014) 25:841–50. doi: 10.1515/revneuro-2014-0056

15. Falkmer T, Anderson K, Falkmer M, Horlin C. Diagnostic procedures in autism spectrum disorders: a systematic literature review. *Eur Child Adolesc Psychiatry.* (2013) 22:329–40. doi: 10.1007/s00787-013-0375-0

16. Dosreis S, Weiner CL, Johnson L, Newschaffer CJ. Autism spectrum disorder screening and management practices among general pediatric providers. *J Dev Behav Pediatr.* (2006) 27:S88–94. doi: 10.1097/00004703-200604002-00006

17. Antezana L, Scarpa A, Valdespino A, Albright J, Richey JA. Rural trends in diagnosis and services for autism spectrum disorder. *Front Psychol.* (2017) 8:590. doi: 10.3389/fpsyg.2017.00590

18. Randall M, Egberts KJ, Samtani A, Scholten RJ, Hooft L, Livingstone N, et al. Diagnostic tests for autism spectrum disorder (ASD) in preschool children. *Cochrane Database Syst Rev.* (2018) 7:CD009044. doi: 10.1002/14651858.CD009044.pub2

19. Ertel W. Machine learning and data mining. *SpringerPlus.* (2011) 42:175–243. doi: 10.1007/978-3-319-58487-4_8

20. Duda M, Ma R, Haber N, Wall DP. Use of machine learning for behavioral distinction of autism and ADHD. *Transl Psychiat.* (2016) 6:732–7. doi: 10.1038/tp.2015.221

21. Tenev A, Markovska-Simoska S, Kocarev L, Pop-Jordanov J, Müller A, Candrian G. Machine learning approach for classification of ADHD adults. *Int J Psychophysiol.* (2014) 93:162–6. doi: 10.1016/j.ijpsycho.2013.01.008

22. Pahwa A, Aggarwal G, Sharma A. A machine learning approach for identification & diagnosing features of Neurodevelopmental disorders using speech and spoken sentences. In: *Int Conf Comput.* Greater Noida (2016). doi: 10.1109/CCAA.2016.7813749

23. Wang J, Zhou X, Wei X, Sun CH, Wu LJ, Wang JL. Autism awareness and attitudes towards treatment in caregivers of children aged 3–6years in Harbin, China. *Soc Psych Psych Epid.* (2012) 47:1301–8. doi: 10.1007/s00127-011-0438-9

24. Burd L, Li Q, Kerbeshian J, Klug M, Freeman R. Tourette syndrome and comorbid pervasive developmental disorders. *J Child Neurol.* (2009) 24:170–5. doi: 10.1177/0883073808322666

25. Bertoncelli C, Altamura P, Vieira ER, Bertoncelli D, Thummler S, Solla F. Identifying factors associated with severe intellectual disabilities in teenagers with cerebral palsy using a predictive learning model. *J Child Neurol.* (2019) 34:221–9. doi: 10.1177/0883073818822358

26. Openneer TJC, Huyser C, Martino D, Schrag A; EMTICS Collaborative Group, Hoekstra PJ, et al. Clinical precursors of tics: an EMTICS study. *J Child Psychol Psychiatry.* (2021) 63:305–14. doi: 10.1111/jcpp.13472

27. Quinlan J. Induction of decision trees. *Mach Learn.* (1986) 1:81–106. doi: 10.1007/BF00116251

28. Rostami M, Farashi S, Khosrowabadi R, Pouretemad H. Discrimination of ADHD subtypes using decision tree on behavioral, neuropsychological and neural markers. *Basic Clin Neurosci.* (2019) 11:359–67. doi: 10.32598/bcn.9.10.115

29. Jiao Y, Chen R, Ke X, Cheng L, Chu K, Lu Z, et al. Predictive models for subtypes of autism spectrum disorder based on single-nucleotide polymorphisms and magnetic resonance imaging. *Adv Med.* (2011) 56:334–42. doi: 10.2478/v10039-011-0042-y

30. Hanc T, Szwed A, SOpien A, Wolanczyk T, Dmitrzak-Weglarz M, Ratajczak J. Perinatal risk factors and ADHD in children and adolescents: a hierarchical structure of disorder predictors. *J Atten Disord.* (2016) 22:855–63. doi: 10.1177/1087054716643389

31. Cortes C, Vapnik V, Llorens C, Vapnik V, Cortes C, Côrtes M. Support-vector networks. *Mach Learn.* (1995) 20:273–97. doi: 10.1007/BF00994018

32. Conti E, Retico A, Palumbo L, Spera G, Bosco P, Biagi L, et al. Autism spectrum disorder and childhood apraxia of speech: early language-related hallmarks across structural MRI study. *J Pers Med.* (2020) 10:359–67. doi: 10.3390/jpm10040275

33. Agastinose Ronicko JF, Thomas J, Thangavel P, Koneru V, Langs G, Dauwels J. Diagnostic classification of autism using resting-state fMRI data improves with full correlation functional brain connectivity compared to partial correlation. *J Neurosci Methods.* (2020) 345:108884. doi: 10.1016/j.jneumeth.2020.108884

34. Bi X, Yang W, Shu Q, Sun Q, Qian X. Classification of autism spectrum disorder using random support vector machine cluster. *Front Genet.* (2018) 9:18. doi: 10.3389/fgene.2018.00018

35. Fulceri F, Grossi E, Contaldo A, Narzisi A, Apicella F, Parrini I, et al. Motor skills as moderators of core symptoms in autism spectrum disorders: preliminary data from an exploratory analysis with artificial neural networks. *Front Psychol.* (2018) 9:2683. doi: 10.3389/fpsyg.2018.02683

36. Todd PM, Loy G. Machine tongues XII: neural networks. *MIT Press.* (1989) 13:28–40. doi: 10.2307/3680009

37. Hossain M, Kabir M, Anwar A, Islam M. Detecting autism spectrum disorder using machine learning techniques. *Health Inf Sci Syst.* (2021) 9:17. doi: 10.1007/s13755-021-00145-9

38. Thomas RM, Gallo S, Cerliani L, Zhutovsky P, Wingen GV. Classifying autism spectrum disorder using the temporal statistics of resting-state functional MRI data with 3D convolutional neural networks. *Front Psychiatry.* (2020) 11:440. doi: 10.3389/fpsyt.2020.00440

39. Khullar V, Singh HP, Bala M. Deep neural network-based handheld diagnosis system for autism spectrum disorder. *Neurol India.* (2021) 69:66–74. doi: 10.4103/0028-3886.310069

40. Breiman L. Bagging predictors. *Mach Learn.* (1996) 24:123–40. doi: 10.1007/BF00058655

41. Freund Y, Schapire RE. *Experiments With a New Boosting Algorithm.* Citeseer (1996). p. 148–56.

42. Breiman L. Stacked regressions. *Mach Learn.* (1996) 24:49–64. doi: 10.1007/BF00117832

43. Cao Y, Miao Q, Liu J, Gao L. Advance and prospects of AdaBoost Algorithm. *ACTA.* (2013) 39:745–58. doi: 10.1016/S1874-1029(13)60052-X

44. Putra PU, Shima K, Alvarez SA, Shimatani K. Identifying autism spectrum disorder symptoms using response and gaze behavior during the Go/NoGo game CatChicken. *Sci Rep.* (2021) 11:22012. doi: 10.1038/s41598-021-01050-7

45. Liaw A, Wiener M. Classification and regression by randomForest. *R News.* (2002) 2:18–22. doi: 10.1057/9780230509993

46. Feczko E, Balba N, Miranda-Dominguez O, Cordova M, Karalunas S, Irwin L, et al. Subtyping cognitive profiles in Autism Spectrum Disorder using a random forest algorithm. *Neuroimage.* (2017) 172:674–88. doi: 10.1016/j.neuroimage.2017.12.044

47. Gao X, Xi W, Zhao H, Luo X, Yang Y. Depicting the composition of gut microbiota in children with tic disorders: an exploratory study. *J Child Psychol Psychiatry.* (2021) 62:1246–54. doi: 10.1111/jcpp.13409

48. Luo Y, Alvarez TL, Halperin JM, Li X. Multimodal neuroimaging-based prediction of adult outcomes in childhood-onset ADHD using ensemble learning techniques. *Neuroimage Clin.* (2019) 26:102238. doi: 10.1101/785766

49. Du X, Akifumi. A fast algorithm for mining of association rules. *Comput Eng Appl.* (2002) 15: 619–24. doi: 10.1007/BF02948845

50. Kim L, Myoung S. Comorbidity study of Attention-deficit Hyperactivity Disorder (ADHD) in children: applying Association Rule Mining (ARM) to Korean National Health Insurance Data. *Iran J Public Health.* (2018) 47:481–8.

51. Tai Y-M, Chiu H-W. Comorbidity study of ADHD: applying association rule mining (ARM) to National Health Insurance Database of Taiwan. *Int J Med Inform.* (2009) 78:e75–83. doi: 10.1016/j.ijmedinf.2009.09.005

52. Ucuz I, Cicek AU, Cansel N, Kilic B, Colak C, Yazici IP, et al. Can temperament and character traits be used in the diagnostic differentiation of children with ADHD? *J Nerv Ment Dis.* (2021) 209:905–10. doi: 10.1097/NMD.0000000000001395

53. Vargason T, Frye RE, Mcguinness DL, Hahn J. Clustering of co-occurring conditions in autism spectrum disorder during early childhood: a retrospective analysis of medical claims data. *Autism Res.* (2019) 12:1272–85. doi: 10.1002/aur.2128

54. Stevens E, Dixon DR, Novack MN, Granpeesheh D, Linstead E. Identification and analysis of behavioral phenotypes in autism spectrum

disorder via unsupervised machine learning. *Int J Med Inform.* (2019) 129:29–36. doi: 10.1016/j.ijmedinf.2019.05.006

55. Jolliffe IT, Cadima J. Principal component analysis: A review and recent developments. *Philos Trans A Math Phys Eng Sci.* (2016) 374:20150202. doi: 10.1098/rsta.2015.0202

56. Mashal N, Kasirer A. Principal component analysis study of visual and verbal metaphoric comprehension in children with autism and learning disabilities. *Res Dev Disabil.* (2012) 33:274–82. doi: 10.1016/j.ridd.2011.09.010

57. Ouss L, Palestra G, Saint-Georges C, Gille ML, Afshar M, Pellerin H, et al. Behavior and interaction imaging at 9 months of age predict autism/intellectual disability in high-risk infants with West syndrome. *Transl Psychiatry.* (2020) 10:608–21. doi: 10.1038/s41398-020-0743-8

58. Subasi A, Erçelebi E. Classification of EEG signals using neural network and logistic regression. *Comput Methods Programs Biomed.* (2005) 78:87–99. doi: 10.1016/j.cmpb.2004.10.009

59. Slinker BK, Glantz SA. Multiple regression for physiological data analysis: the problem of multicollinearity. *Am J Physiol.* (1985) 249:R1–12. doi: 10.1152/ajpregu.1985.249.1.R1

60. Dingemans AJM, Hinne M, Jansen S, van Reeuwijk J, de Leeuw N, Pfundt R, et al. Phenotype based prediction of exome sequencing outcome using machine learning for neurodevelopmental disorders. *Genet Med.* (2022) 24:645–53. doi: 10.1016/j.gim.2021.10.019

61. Rahman HAA, Wah YB, He H, Bulgiba A. Comparisons of ADABOOST, KNN, SVM Logistic Regression in Classification of Imbalanced Dataset. In: MW Berry, A Mohamed, BW Yap, editors. *Soft Computing in Data Science.* Springer (2015). p. 54–64.

62. Coadou Y. Boosted decision trees and applications. In: *EPJ Web of Conferences.* Autrans (2013). p. 55 doi: 10.1051/epjconf/20135502004

63. Brodley CE, Utgoff PE. Multivariate decision trees. *Mach Learn.* (1995) 19:45–77. doi: 10.1007/BF00994660

64. Khoshgoftaar TM, Allen EB. Controlling overfitting in classification-tree models of software quality. *Empir Softw Eng.* (2001) 6:59–79. doi: 10.1023/A:1009803004576

65. Yap BW, Rani KA, Rahman HAA, Fong S, Khairudin Z, Abdullah NN. An application of oversampling, et al. In: T Herawan, MM Deris, J Abawajy, editors. *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013).* Springer Singapore: Singapore (2014). p. 13–22. doi: 10.1007/978-981-4585-18-7_2

66. Pandey P, Prabhakar R. "An analysis of machine learning techniques (J48 & AdaBoost)-for classification," in *2016 1st India International Conference on Information Processing (IICIP).* (2016). p. 1–6. doi: 10.1109/IICIP.2016.7975394

67. Wang S-C. Artificial Neural Network. In: Wang SC, editor. *Interdisciplinary Computing in Java Programming,* Boston, MA: Springer US (2003). p. 81–100. doi: 10.1007/978-1-4615-0377-4_5

68. Collins GS, Reitsma JB, Altman DG, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg.* (2015) 102:148–58. doi: 10.1002/bjs.9736

69. Dwyer D, Koutsouleris N. Annual research review: translational machine learning for child and adolescent psychiatry. *J Child Psychol Psychiatry.* (2022) 63:421–43. doi: 10.1111/jcpp.13545