Applying Artificial Intelligence for Diagnostic Classification of Korean Autism Spectrum Disorder

Eun Soo Choi^{1*}, Hee Jeong Yoo^{2*}, Min Soo Kang^{3*}, and Soon Ae Kim⁴

¹Department of Medical IT Marketing, Eulji University, Seongnam, Republic of Korea

²Department of Psychiatry, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Republic of Korea

³Department of Medical IT, Eulji University, Seongnam, Republic of Korea

⁴Department of Pharmacology, School of Medicine, Eulji University, Daejeon, Republic of Korea

Objective The primary objective of this study was to predict subgroups of autism spectrum disorder (ASD) based on the Diagnostic Statistical Manual for Mental Disorders-IV Text Revision (DSM-IV-TR) by machine learning (ML). The secondary objective was to set up a ranking of Autism Diagnostic Interview-Revised (ADI-R) diagnostic algorithm items based on ML, and to confirm whether ML can sufficiently predict the diagnosis with these minimum items.

Methods In the first experiment, a multiclass decision forest algorithm was applied, and the diagnostic algorithm score value of 1,269 Korean ADI-R test data was used for prediction. In the second experiment, we used 539 Korean ADI-R case data (over 48 months with verbal language) to apply mutual information to rank items used in the ADI diagnostic algorithm.

Results In the first experiment, the results of predicting in the case of pervasive developmental disorder not otherwise specified as “ASD” were almost three times higher than predicting it as “No diagnosis.” In the second experiment, the top 10 ranking items of ADI-R were mainly related to the quality abnormality of communication.

Conclusion In conclusion, we verified the applicability of ML in diagnosis and found that the application of artificial intelligence for rapid diagnosis or screening of ASD patients may be useful.

Psychiatry Investig 2020;17(11):1090-1095

Key Words Autism spectrum disorder, Autism diagnostic interview-revised, Developmental disorder, Machine learning, Multiclass decision forest.

INTRODUCTION

The application of artificial intelligence (AI) enables fast and diverse access to a large amount of information. In addition, there are aspects that allow us to look at problems from various perspectives based on interpretation methods. To this end, machine learning (ML), as a subfield of AI, is commonly used.¹ Detection and diagnosis of diseases at an early stage is one of the key interests in ML for medical science and it may support

a final diagnosis by a specialist that might be quickly made on the basis of appropriate grounds. For example, Watson for Oncology and Watson Genomics by IBM Watson Health has the goal of advancement in precision medicine by integrating cognitive computing and genome tumor sequencing and clinical data in a maximum of 20 min (<https://www.ibm.com/watson/kr-ko/health/oncology-and-genomics/genomics/>).² Attempts using ML have been performed to achieve outcomes for diagnosis and treatment in multidisciplinary medical fields.

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by limited social communication and social relationships, repetitive behavior and restricted interest.³ As it is a psychiatric condition with complex behavioral phenotypes, the diagnostic process often requires comprehensive history taking and direct observation because there is no definite biomarker, and the highest level of professional expertise is needed for proper diagnosis. Representative diagnostic instruments for ASD include the autism diagnostic observation

Received: June 3, 2020 **Revised:** July 23, 2020

Accepted: September 2, 2020

Correspondence: Soon Ae Kim, MD, PhD

Department of Pharmacology, School of Medicine, Eulji University, 77 Gyeongsang-ro 771 beon-gil, Jung-gu, Daejeon 34824, Republic of Korea

Tel: +82-42-259-1672, **Fax:** +82-42-259-1679, **E-mail:** sakim@eulji.ac.kr

*These authors contributed equally to this work.

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

schedule (ADOS) and autism diagnostic interview-revised (ADI-R)⁴ questionnaire. ADI-R is a semi-structured interview developed by Lord et al.⁵ in 1994 for mutually complementary use with the ADOS. This instrument is used in the diagnosis of ASD through the process of interviewing parents or caregivers of children. ADI-R is designed to comprehensively describe and evaluate the aspects of development and behavior of children from infancy to present time using the following questions: a total of 93 items in 3 domains, including communication, social interaction, and restrictions of behavior and interest, as included in the standards for diagnosis of autistic disorder. The extent of severity of ASD-related behavior for each item is coded on a scale of 0–3, which is then applied to the diagnostic algorithm to determine whether it corresponds to the cutoff for diagnosis of autism.⁵ ADI-R is composed to assign scores to each item on the basis of contents of the interview and questionnaire, and to determine the diagnostic category in accordance with the diagnosis algorithm. Although this is an essential instrument for diagnosis, it requires more than 2 hours for completion by each individual. Accordingly, some of the diversely collected phenotype data remains unused.⁶ Required specialist training and the prolonged time of the ADI-R diagnostic process act as bottlenecks in clinical settings.

To overcome administrator bias during assessments, many have attempted to use AI technology to create a more concise algorithm for ASD diagnosis.⁷ Several studies have attempted to classify items from assessment of ASD diagnostic instruments that are most predictive of the diagnosis to make the process less time-consuming. Wall et al.⁸ applied 16 ML algorithms with data in multiple ASD data repositories, including the Autism Genetic Resource Exchange (AGRE) and Autism Consortium (AC), and they suggested that ML enables the classification of autism and non-autism with 100% accuracy using tree algorithms such as ADTree. Furthermore, the same group also suggested a few important ADI-R diagnostic items by applying the ADTree algorithm with datasets including multiple ASD population data.^{8,9} However, considering ethnic or sociocultural differences, it is not known whether these results could be replicable after applying to other populations, including Koreans.

In this study, we apply a machine learning algorithm using the Korean ADI-R diagnostic algorithm dataset. The purpose of this research was to examine the following questions: 1) whether the diagnostic algorithm domain scores of in ADI-R can predict the results of diagnosis of autism spectrum disorder. We used the data classified based on the diagnostic criteria of Diagnostic and Statistical Manual (DSM)-IV, to examine whether ML distinguishes even the pervasive developmental disorder not otherwise specified (PDD-NOS) category based on the ASD group and 2) whether machine learning can sug-

gest the minimum and reliable items from the ADI-R diagnostic algorithm for Korean ASD diagnosis by the rankings related algorithm.

METHODS

Subjects and data

In this study, we use the coding data and scores for ADI-R diagnostic algorithm item of 1,269 anonymous patients with ASD and their unaffected siblings, collected at the Bundang Seoul National University Hospital for the first experiment. Each subject was given clinical best estimate diagnosis by two board certified child psychiatrists, based on ADI-R, Autism Diagnostic Observation Schedule (ADOS) and clinical interviews. For the first experiment set, the total scores of each domain in diagnostic algorithm in ADI-R (namely, “A: Qualitative abnormalities in reciprocal social interaction,” “B. Qualitative abnormalities in communication,” “C. Restricted, repetitive, and stereotyped patterns of behavior” and “D. Abnormality of Development Evident at or Before 36 Months”) which suggested the cut-off value for diagnosis were used. Cases for which the scores in any one of these domains did not exceed the ASD cutoff are categorized as PDD-NOS. If the subject did not meet any criteria, she/he was assigned to “No Diagnosis” group.

In the second experiment set, we selected 539 cases, older than 48 months with verbal language out of first dataset, to apply mutual information to rank 78 items in the total ADI-R diagnostic algorithm dataset. It is used not only the score for “current” but also the score for “ever” or “most abnormal between 4.0–5.0 years of age” of each item in the domains. The item “86. Age When Abnormality First Evident” and “87. Interviewer’s Judgement on Age When Abnormalities First Manifest” in D domain were not used in this study because those two are summary items having different scoring system with other items. In this second set, coding value 991–999 (not specified) were treated as missing values, and 7, 8 and 9 (not typical/not available/not known) were also used. Supplementary Table 1 (in the online-only Data Supplement) shows the basic characteristics of the first dataset used for learning. The use of data for this study was approved by the Institutional Bioethics Committee of Bundang Seoul National University Hospital (IRB No. B-1803-459-103).

Multiclass decision forest

Multiclass decision forest is an ensemble model that rapidly builds a multitude of decision trees. The biggest feature of decision forest is that the trees have different characteristics due to randomness. As the prediction model learns through multiple trees, it is possible to expect generalization of the prediction results. In addition, robust prediction results can be

expected for data containing noise.¹⁰ The representative parameter of Decision Tree is Number of tree, which is a parameter that defines the number of times the prediction model is trained. In the case of Decision Forest, it can be observed increased accuracy of the prediction model as the number of tree increasing, but it does not increase more than a certain value.¹¹ The accuracy does not increase more than a certain value, it is important to find the minimum number of tree parameter values with maximum accuracy. Ensemble learning can solve machine learning problems by generating and combining multiple prediction models.¹² Research on decision forests has started in early 1990s, but it only started gaining attention in mid-1990s.¹³ Decision forests are used in various fields, including medicine, engineering, and information retrieval.¹⁴ Various types of decision forest methods exist, including bootstrap aggregating (bagging) and adaptive boosting (AdaBoost) as representative methods.¹⁵ Bagging is a simple and effective method for generating an ensemble model in decision forests. Each decision tree in the forest is trained on a data subset taken as a replacement of the original training dataset. All trees are trained using the same learning algorithm. The final prediction for untrained data is determined by a majority vote on the predicted results of the individual trees. The advantage of bagging is that it can be easily implemented in a parallel model by training several ensemble classifiers.¹⁶ Boosting is a way to improve the performance of weak learners. AdaBoost is the most popular boosting algorithm.¹⁷ The main idea of AdaBoost is to assign higher weights to samples that were misclassified during the previous iteration of the training. Specifically, weights of correctly classified samples are decreased, and weights of misclassified samples are increased. As a result, the classifier focuses on “difficult” samples. This iterative learning procedure provides a series of decision trees that complement one another.¹⁸

Mutual information

Feature Selection is a method for efficiently learning various machine learning and data mining. It is a method to select the feature that best represents the dataset among high-dimensional datasets. Information Gain is one of the feature selection methods, Information Theoretical based method.¹⁹ Information gain is also called Mutual Information Maximization (MIM). A formula for calculating the information gain is following equation.

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

Information Gain measures how much information a feature gives us about the class. $\text{Value}(A)$ means feature. The larger the information gain value, the more likely the feature is describing the label.²⁰ Entropy value is between 0 to 1. Also

$\text{Pr}(X=1)$ value is between 0 to 1. Entropy is impurity and purity in an arbitrary collection of examples.

$$\text{Entropy}(S) = - \sum_{i=1}^c p_i \log_2 p_i$$

The above equation is a formula for calculating the entropy. P_i is the ratio of the set S of binary data sets belonging to class i .^{2,4}

The experiments were conducted using Microsoft Azure Machine Learning Studio (<https://studio.azureml.net>). To apply machine learning, this study used a classification algorithm—specifically, a multiclass decision forest at the first data setting. Bagging, which is one of the techniques of multiclass decision forest, was applied as a resampling method.²¹ In addition, to select important diagnostic items from the second data set, we used a Mutual Information. To evaluate, Multiclass decision forest was also applied to the selected 5 diagnostic items. In all experiment sets, and the ratio of learning and test data was 90:10 and Number of trees for training was set at 1, 2, 4, and 8.²²

RESULTS

Supplementary Figure 1 (in the online-only Data Supplement) shows the results of the tree that illustrates the results of learning on the data of the ADI-R diagnostic algorithm first dataset. Figure 1 shows the accuracy of classification for the set of 127 patients used for verification of the prediction model. Although the rate of positive prediction of ASD reduced from 98.5% to 97.0% as the number of trees increased from 1 to 4, the rate of negative prediction increased from 94.9% to 97.4%. In the case of PDD-NOS, the lowest prediction rate of 71.4% was observed when the number of trees was 2, and the prediction rate was the same at 81.0% when the number of trees was 4 or 8. Supplementary Table 2 (in the online-only Data Supplement) shows the accuracy, precision, and recall of predictive models that have learned four trees in the 1st set. The average accuracy was 96.3%, macro-averaged precision was 92.7%, and macro-averaged recall was 91.8%.

The top-ranking ADI-R interview items in the diagnostic algorithm are selected by AI in the second data setting (Table 1). In the top 10 items, 7 items were included in the communication domain and 3 items were included in the Reciprocal Social Interaction domain. As a result of applying properties from 1st to 5th using 10% test beds, the accuracy of “No diagnosis” was 100% (specificity) and “ASD” was 97.6% (sensitivity). Misclassification from ASD to no diagnosis was 2.4%. As a result of applying Top 5, the four confusion matrices were the same from the number of trees 1 to 8 (Figure 2). Supplementary Table 2 (in the online-only Data Supplement) also shows the accuracy, precision, recall and Youden index of predictive

models that have learned four trees in the 2nd set. The average accuracy was 98.14%, macro-averaged precision was 96.42%, and macro-averaged recall was 98.78%.

DISCUSSION

Several studies have reported the application of machine

Figure 1. Confusion Matrix of the predictive model (number of tree 1, 2, 4, and 8) for 127 new patients (10% the test bed) in the first experimental set. ASD: autism spectrum disorder, PDD-NOS: pervasive developmental disorder not otherwise specified.

Table 1. Top 10 ranked items of ADI-R diagnostic algorithm

Rank	ADI-R item number and title	Subdomain in diagnostic algorithm
1	35. Reciprocal conversation (recent 3 months)	B2: relative failure to initiate or sustain conversational interchange
2	35. Reciprocal conversation	B2: relative failure to initiate or sustain conversational interchange
3	49. Imaginative play with peers	A2: failure to develop peer relationships
4	64. Group play with peers	A2: failure to develop peer relationships
5	61. Imitative social play	B4: lack of varied spontaneous make-believe or social imitative play
6	34. Social verbalization/chat	B2: relative failure to initiate or sustain conversational interchange
7	48. Imaginative play	B4: lack of varied spontaneous make-believe or social imitative play
8	45. Conventional/instrumental gesture	B1: lack of, or delay in, spoken language and failure to compensate through gesture
9	42. Pointing to express interest	B1: lack of, or delay in, spoken language and failure to compensate through gesture
10	57. Range of facial expressions used to communicate	A1: failure to use nonverbal behaviors to regulate social interaction

ADI-R: autism diagnostic interview-revised

Figure 2. Confusion Matrix of the predictive model (number of tree 1, 2, 4, and 8) for 54 new patients (10% the test bed) with top 5 ranked attributes in the second experimental set. ASD: autism spectrum disorder.

learning to ASD diagnosis. Wall et al.⁸ summarized 29 items of Module 1 of the ADOS Data Set, provided by the AGRE and AC, as 8 items by applying the ADTree algorithm. A total of 1,000 training steps were executed, with the reported distinction between “Autism” and “Non-Autism.” Duda et al.²³ tested and evaluated the performance of this algorithm by simultaneously applying the observation-based classifier (OBC) algorithm simultaneously. Although previous studies only distinguished between “Autism” and “Non-Autism,” the first part of this study took a step further to predict the subgroup of PDD-NOS in the DSM-IV-TR diagnostic system. The results of this analysis show that machine learning can predict the boundary between ASD and PDD-NOS relatively accurately. As a result of learning data in this study, only 2.6% of cases of normal children were predicted to have PDD-NOS. In addition, in the case of PDD-NOS, the results of predicting it as ASD were almost 3 times higher than those being predicted as “No diagnosis.” Moreover, while the previous studies used datasets with diversified cultural and racial backgrounds, this study only used the data collected from Korean population for the learning process and made a verification for the same population sample. Therefore, it was determined that it was possible to achieve an accuracy, precision, and recall of greater than 90%.

In the second set of this study, all 10 highly ranked ADI-R diagnostic algorithm items were classified only in the Reciprocal Social Interaction and Communication domain. In addition, repetitive behavior and restricted interest domain and stereotyped, repetitive, or idiosyncratic speech subdomains, that are related to repetitive and stereotypical behavior and speech, were not included. Because we used data that are se-

lected from patients over four years old with oral communication, it can be possible that “Reciprocal conversation” item was mainly selected which is used for only persons with oral communication. In a previous statistical analysis study with a Korean ASD population, Jin et al.²⁴ reported the factor analysis results of 380 Korean ADI-R diagnostic algorithm data and suggested that 16 items may be useful for classification of ASD using a decision tree model. It was reported that a factor (factor 3) including top ranked item (“Reciprocal conversation”) had over 90% sensitivity and 5 items from the top 10 items in this study were included in 16 items that were previously suggested by Jin et al.²⁴ This could be considered since we did not use the score data (0, 1, 2) of the diagnostic algorithm but the coding value data (0, 1, 2, 3, 7, 8, 9) of each item for the second analysis set, and it gave us more variables for machine learning in this study. In addition, although they used 42 items including domain D (“D. Abnormality of Development Evident at or Before 36 Months”), it did not use items 86 and 87 in domain D in this study.

In another study using AI, Dennis P. Wall summarized 93 items of the ADI-R Data Set, provided by AGRE, Simons Simplex Collection, and AC, as 7 items by applying the ADTree algorithm. ADTree is an algorithm for which the boosting technique of ensemble learning is applied, and it is possible to train an optimized prediction model through the prescribed quantity of learning.²⁵ Similar to other research, a total of 1,000 training steps were executed, and a distinction between “Autism” and “Non-Autism” was made.⁹ In the selected 7 items, “Reciprocal conversation,” “Imaginative play,” “Imaginative play with peers” and “Group play with peers” items were used also in this study as top 5 ranked items. “Comprehension of simple language; answer most abnormal between 4 and 5” is not included in the diagnostic algorithm items and “Age when abnormality first evident” in the domain D was not used in this study. “Direct gaze” item, which was selected in the previous study, was not ranked and “Imitative Social Play” was a new ranked item in this study. Most items were shared with each other between studies, and the differences might be induced by genetic and socio-cultural environmental differences in the study population.

Although it may be induced by the relatively small sample size ($n=54$) in the 10% test bed, the result with 100% specificity and over 97% sensitivity suggested the possibility of improving the huge ADI-R assay, which requires long time and large efforts by physicians, patients, and caregivers to be simple and effective. In addition, owing to the fact that the confusion matrix after applying the Ranked data were formed identically even after additional learning occurs (Figure 2), it can be determined that the ASD can be sufficiently diagnosed with the ranked attributes without much learning. It can be estimated

that the attribute ranked by mutual information is somewhat important for Korean ASD diagnosis. Although, it is not always necessary to reduce diagnostic items because a large amount of information is needed for accurate disease diagnosis, it can be useful for the development of screening tools for populations. Furthermore, the application of AI with more clinical and biological information, such as other diagnostic tools, facial expression, and genetic data for rapid and accurate diagnosis or screening of ASD patients seems to be useful.

Though, the use of previous guidelines (DSM-IV) and family data could be a limitation in the application of decision support for diagnosis, this study showed the possibility of reducing time for diagnosis and suggested the top-ranked items for diagnosis of ASD in a Korean population. As such, AI in medical science may provide for a quick diagnosis, and the final diagnosis can then be arrived at on the basis of appropriate grounds and the specialist's decision.

Supplementary Materials

The online-only Data Supplement is available with this article at <https://doi.org/10.30773/pi.2020.0211>.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant (NRF-2017R1D1A3B03033533 and NRF-2020R1A2C1009499) and Institute of Information & Communications Technology Planning & Evaluation (IITP) grant (2019-0-00330) funded by the Korea government (MSIT).

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Soon Ae Kim. Data curation: Min Soo Kang. Formal analysis: Eun Soo Choi. Funding acquisition: Soon Ae Kim. Investigation: Eun Soo Choi. Methodology: Eun Soo Choi. Project administration: Soon Ae Kim. Resources: Min Soo Kang. Software: Min Soo Kang. Supervision: Soon Ae Kim. Validation: Hee Jeong You. Visualization: Eun Soo Choi. Writing—original draft: Eun Soo Choi, Soon Ae Kim. Writing—review & editing: Hee Jeong Yoo, Min Soo Kang.

ORCID iDs

Eun Soo Choi	https://orcid.org/0000-0002-7493-6763
Hee Jeong Yoo	https://orcid.org/0000-0003-0521-2718
Min Soo Kang	https://orcid.org/0000-0002-3931-7430
Soon Ae Kim	https://orcid.org/0000-0002-9831-0511

REFERENCES

- Deo RC. Machine learning in medicine. *Circulation* 2015;132:1920-1930.
- Doyle-Lindrud S. Watson will see you now: a supercomputer to help clinicians make informed treatment decisions. *Clin J Oncol Nurs* 2015; 19:31-32.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edition. Arlington: American Psychiatric Association publishing; 2013. <https://doi.org/10.1176/appi.books.9780890425596>.
- Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* 2000;30:205-223.
- Lord C, Petkova E, Hus V, Gan W, Lu F, Martin DM, et al. A multisite study of the clinical diagnosis of different autism spectrum disorders. *Arch Gen Psychiatry* 2012;69:306-313.
- Lord C, Rutter M, Le Couteur A. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 1994;24:659-685.
- Song DY, Kim SY, Bong G, Kim JM, Yoo HJ. The use of artificial intelligence in screening and diagnosis of autism spectrum disorder: a literature review. *J Korean Acad Child Adolesc Psychiatry* 2019;30:145-152.
- Wall DP, Kosmicki J, Deluca TF, Harstad E, Fusaro VA. Use of machine learning to shorten observation-based screening and diagnosis of autism. *Transl Psychiatry* 2012;2:e100.
- Wall DP, Dally R, Luyster R, Jung JY, Deluca TF. Use of artificial intelligence to shorten the behavioral diagnosis of autism. *PLoS One* 2012; 7:e43855.
- Opitz D, Maclin R. Popular ensemble methods: an empirical study. *J Artif Intell Res* 1999;11:169-198.
- Oshiro TM, Perez PS, Baranauskas JA. How Many Trees in a Random Forest? In: Perner P, Editor. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Berlin, Heidelberg: Springer; 2012. p.154-168.
- Rokach L. Decision forest: twenty years of research. *Inf Fusion* 2016; 27:111-125.
- Hastie TJ. *Statistical Models in S*. New York: Routledge; 2017.
- Oza NC, Tumer K. Classifier ensembles: select real-world applications. *Inf Fusion* 2008;9:4-20.
- Marsland S. *Machine Learning?: An Algorithmic Perspective*. Boca Raton: CRC Press; 2014.
- Breiman L. Bagging predictors. *Mach Learn* 1996;24:123-140.
- Freund Y, Freund Y, Schapire RE. Experiments with a New Boosting Algorithm (1996). Available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.133.1040>. Accessed March 13, 2020.
- Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 1997;55: 119-139.
- Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, et al. Feature selection: a data perspective. *ACM Computing Surv* 2017;50:94.
- Quinlan JR. Induction of decision trees. *Mach Learn* 1986;1:81-106.
- Dietterich TG. Experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Mach Learn* 2000;40:139-157.
- Criminisi A, Shotton J. *Decision Forests for Computer Vision and Medical Image Analysis*. London: Springer-Verlag; 2013.
- Duda M, Kosmicki JA, Wall DP. Testing the accuracy of an observation-based classifier for rapid detection of autism risk. *Transl Psychiatry* 2011;4:e424.
- Jin S, Yoo HJ, Park M. Discriminant model for diagnosing Korean autism spectrum disorders. *J Korean Data Anal Soc* 2013;15:669-676.
- Kelarev AV, Abawajy J, Stranieri A, Jelinek HF. Empirical investigation of decision tree ensembles for monitoring cardiac complications of diabetes. *Int J Data Warehouse Min* 2013;9:1-18.

Supplementary Table 1. Basic characteristics of the K-ADI-R diagnostic algorithm data

1) 1st set	Autism	PDD	No diagnosis	Total
Number of participants (%)	708 (55.8)	177 (13.9)	384 (30.3)	1,269 (100)
Sex (male:female)	605:103	135:42	184:200	924:345
Age (month; mean $\pm$ SD)	95.61 $\pm$ 62.80	64.18 $\pm$ 59.22	87.84 $\pm$ 60.27	88.87 $\pm$ 62.40
Total score for “A. Qualitative Abnormalities in Reciprocal Social Interaction” domain (mean $\pm$ SD)	21.09 $\pm$ 5.61	10.59 $\pm$ 6.08	2.14 $\pm$ 2.34	13.87 $\pm$ 9.86
Total score for “B. Qualitative Abnormalities in Communication” domain (for verbal patients; mean $\pm$ SD)	15.62 $\pm$ 4.51	7.79 $\pm$ 4.76	1.45 $\pm$ 1.80	10.11 $\pm$ 7.58
Total score for B domain (for non-verbal patients; mean $\pm$ SD)	11.23 $\pm$ 2.62	6.05 $\pm$ 3.96	1.62 $\pm$ 1.70	8.57 $\pm$ 4.64
Sum of subdomain score for non-verbal only in B domain (mean $\pm$ SD)	10.77 $\pm$ 2.99	5.54 $\pm$ 3.80	1.14 $\pm$ 1.50	7.1 $\pm$ 5.15
Total score for “C. Restricted, Repetitive, and Stereotyped Patterns of Behavior” domain (mean $\pm$ SD)	5.78 $\pm$ 2.33	2.16 $\pm$ 1.85	0.50 $\pm$ 0.86	3.68 $\pm$ 3.10
Total score for “D. Abnormality of Development Evident at or Before 36 Months” domain (mean $\pm$ SD)	3.74 $\pm$ 1.12	2.57 $\pm$ 1.51	0.15 $\pm$ 0.60	2.49 $\pm$ 1.91
2) 2nd set	ASD	No diagnosis	Total	
Number of participants (%)	338 (62.7)	201 (37.3)	539 (100)	
Male (%)	301 (79.2)	79 (20.8)	380 (100)	
Female (%)	37 (23.3)	122 (76.7)	159 (100)	
Age (month; mean $\pm$ SD)	132.25 $\pm$ 67.04	116.30 $\pm$ 55.97	126.30 $\pm$ 63.55	

ADI-R: autism diagnostic interview-revised, PDD: pervasive developmental disorder, SD: standard deviation, ASD: Autism spectrum disorder

Supplementary Figure 1. Decision tree production in the first experimental set.

Supplementary Table 2. Accuracy, precision and recall value of prediction model

Evaluation measures	Ratio (%)
1) 1st set, tree=4	
Overall accuracy	0.9448 (94.48)
Average accuracy	0.9632 (96.32)
Micro-averaged precision	0.9448 (94.48)
Macro-averaged precision	0.9267 (92.67)
Micro-averaged recall	0.9448 (94.48)
Macro-averaged recall	0.9180 (91.8)
2) 2nd set, tree=4, 5 attribute	
Overall accuracy	0.9814 (98.14)
Average accuracy	0.9814 (98.14)
Micro-averaged precision	0.9814 (98.14)
Macro-averaged precision	0.9642 (96.42)
Micro-averaged recall	0.9814 (98.14)
Macro-averaged recall	0.9878 (98.78)
Overall youden index	0.9285