Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/67/20240606

A DNN-based diagnosis on autism spectrum disorder in children

Shuai Tan

Software and Engineering, Xi'an Jiaotong University, Xi'an, 710049, China

2206213255@stu.xjtu.edu.cn

Abstract. The prevalence of autism spectrum disorder (ASD) witnesses a sharp increasing in recent years, and early diagnosis and intervention of ASD are critically needed. This study explored the efficacy of Deep Neural Networks (DNN) in diagnosing ASD among children aged 0 to 10. Utilizing the latest dataset derived from the ASDTests mobile application, which encompasses behavioral characteristics of over 2,000 children, we implemented a DNN model to capture complex non-linear patterns indicative of ASD. The results of comparative analysis with traditional machine learning models revealed DNN's superior accuracy in predicting ASD, indicating that the DNN achieved a significant improvement in identifying minority classes post-imbalance learning treatment. The promising results, including the 99.55% accuracy rate, paved the way for future investigations into integrating DNN with multimodal data analysis and other advanced algorithms to enhance early diagnostic processes and intervention strategies for ASD.

Keywords: Deep Neural Networks, Deep Learning, Autism Spectrum Disorder, Machine Learning, Diagnosis.

1. Introduction

Autism Spectrum Disorder (ASD) represents a prevalent neurodevelopmental condition. According to data from the Centers for Disease Control and Prevention (CDC) in 2020, one in every 36 children is diagnosed with ASD [1]. The diagnosis of ASD relies on comprehensive clinical assessments, such as the Autism Diagnostic Observation Schedule (ADOS) [2] and the Autism Diagnostic Interview-Revised (ADI-R) [3]. Although these methods are authoritative, they are time-consuming and depend on the subjective judgement of professionals. Deep Neural Network (DNN), as an emerging diagnostic tool, has the capability to learn features from complex data, thereby enhancing the efficiency and accuracy of diagnoses.

In the domain of ASD diagnosis leveraging machine learning and deep learning techniques, Rahman and Subashini [4] employed DNN models for screening based on the quantitative checklist for autism in toddlers (QCHAT) and QCHAT-10 datasets, achieving accuracies of 78.88% and 99.6%, respectively. Vishal et al. [5] identified that the Naïve Bayes (NB) achieved a high accuracy of 99.6% in predicting ASD. Rawat et al. [6] explored the performance of various models, including SVM, NB, and Convolutional Neural Network (CNN) in ASD diagnosis, and their results indicated the superior performance by CNN which achieved an accuracy of 99.6%. Aslam et al. [7] utilized Large Scale Feature Extraction (LSFE) analysis in conjunction with a Linear Support Vector Machine (LSVM) classifier for ASD identification on ODU and KAU datasets, achieving accuracies of 100% and 95.5%,

DOI: 10.54254/2755-2721/67/20240606

respectively. Alkahtani et al. [8] proposed a system based on deep learning for ASD identification by analyzing facial landmarks in children, where their MobileNetV2 model had an accuracy of 92%, offering an effective image-based method for early ASD screening. These studies collectively highlight the promising application prospects of deep learning and machine learning technologies in the field of ASD diagnosis, providing valuable references for future clinical practice and research. However, these studies failed to focus on children in the critical age range of 0 to 10 years. The current research aims to develop a diagnostic model specifically for children within this age range to enhance the accuracy of early identification and offer more effective guidance for clinical interventions.

2. Methods

2.1. Dataset description

In this research, the dataset employed originated from the mobile app ASD Tests, updated in 2023, and was created by Dr. Fadi Thabtah [9-11]. Constructed from data collected via the mobile application, the dataset comprises autism-related behavioral characteristics of 2,225 children aged 0 to 10 years. These characteristics include the AQ-10-test quotient [12] (A1-A10 score) and six statistical characteristics (age, gender, ethnicity, history of jaundice, family history of autism, and questionnaire tester). The data serve as a valuable resource for training and validating deep learning models aimed at identifying and predicting early signs of ASD.

2.2. Data preprocessing

During the data preprocessing stage, given the absence of missing values, outliers, or duplicate records in the utilized dataset, the focus shifted towards data encoding strategies. For categorical variables, one-hot encoding was implemented to transform unordered categories, ensuring each category is represented by a set of binary features, thus providing clear inputs for the deep learning model. Additionally, for categorical variables with a natural order, Label Encoding was employed to map categories to consecutive integer values in order to preserve the ordinal information between categories. The application of these encoding methods optimized the model training process and enhanced the model's ability to recognize data features.

2.3. Feature Selection

In this study, a Pearson correlation coefficient matrix, augmented with a heatmap visualization, was employed for feature selection. The linear correlation between two variables is measured by the Pearson correlation coefficient, which has a range of -1 to 1. A value of 1 shows perfect positive correlation, a value of -1 indicates perfect negative correlation, and a value of 0 indicates no linear correlation. By computing the correlation coefficients between features and the target variable, ASD_traits, and visualizing these correlations using the sns.heatmap function, the heatmap succinctly illustrated the degree of linear association between variables. As a visualization tool, heatmaps can display the strength of correlation between features, thereby aiding in the identification of highly correlated or redundant features. By analyzing heatmaps, one can intuitively observe which features exhibit strong linear relationships, and thus decide whether to remove certain features from the model to reduce data redundancy or the risk of overfitting. By examining Figure 1, features with a p-value less than 0.10 were filtered out, as these features (tester, Jaundice and Ethnicity) were considered redundant due to their lower correlation with ASD.

DOI: 10.54254/2755-2721/67/20240606

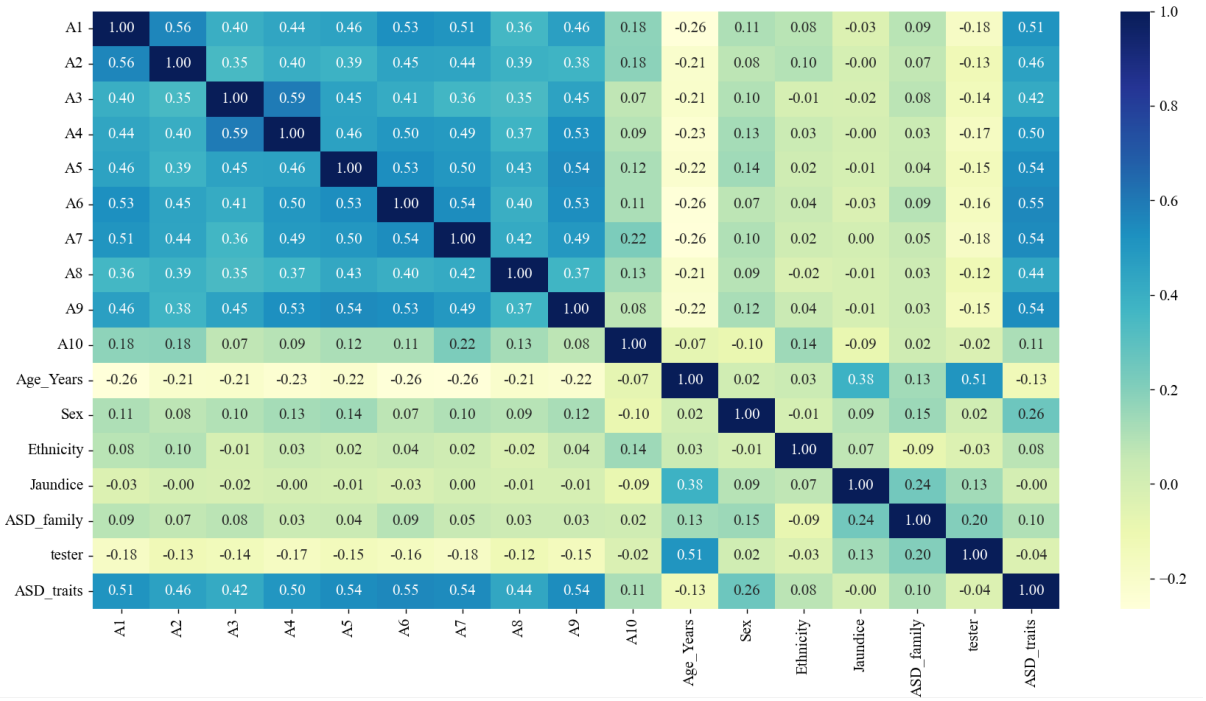

Figure 1. Impact of features on ASD.

2.4. Data Visualization

In the preliminary phase of the experiment, an in-depth visual analysis of the dataset was conducted to reveal its intrinsic distribution and the relationships among features. By plotting histograms, box plots, and scatter plots, the distribution of data could be observed intuitively, including potential outliers and issues of data imbalance. These visual results provided a crucial basis for subsequent data preprocessing and feature engineering efforts. The distribution of categorical variables within the dataset was depicted. (Figure 2 and Figure 3)

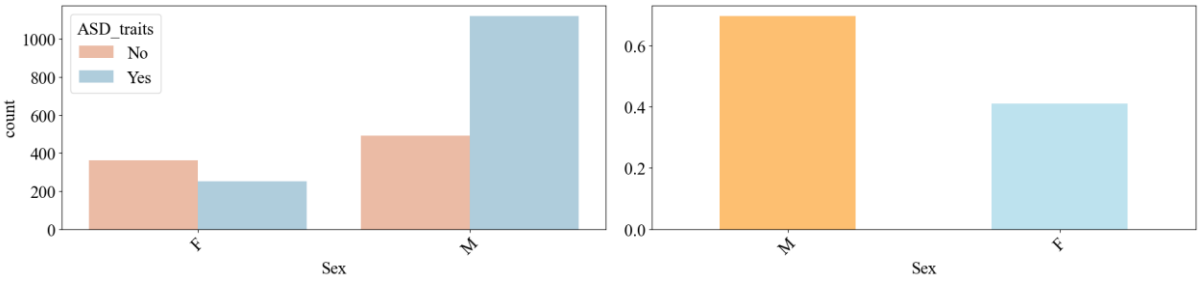

Figure 2. The distribution of sex on ASD traits (**left**) and the distribution of sex on the dataset (**right**).

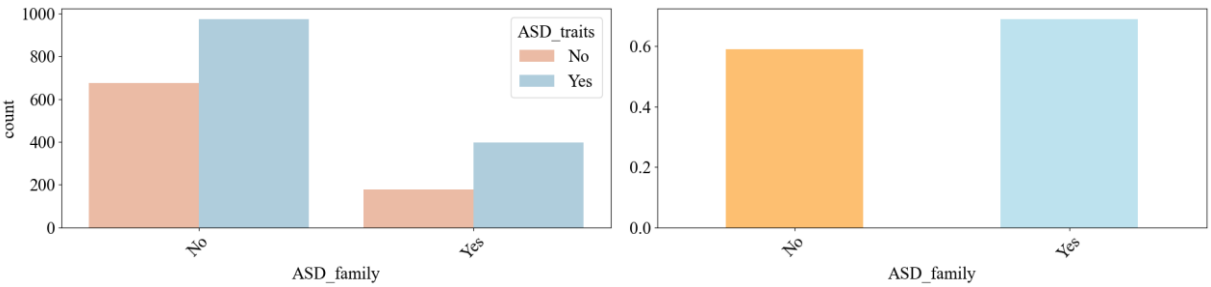

Figure 3. The distribution of family history on ASD traits (**left**) and the distribution of family history on the dataset (**right**).

DOI: 10.54254/2755-2721/67/20240606

2.5. Deep Neural Network (DNN)

DNN is a category of artificial neural networks composed of multiple processing layers, capable of learning complex patterns and features from data. The layers within a DNN can be categorized into input layer, hidden layers, and output layer, with each layer being fully connected. This structure enables the model to process and analyze data at various levels of abstraction, significantly enhancing its predictive and analytical capabilities. DNN is trained through the forward propagation and backpropagation algorithms, and gradient descent is used to optimize weights.

In the forward propagation process, each neuron receives inputs from the previous layer, which are then transformed through a weighted sum, followed by a nonlinear activation function to produce an output. This process can be represented by the equation:

$$z^{[l]} = \sigma\left(W^{[l]}a^{[l-1]} + b^{[l]}\right) \tag{1}$$

For any given layer l, its output, denoted as $z^{[l]}$, is computed by applying the activation function σ to the sum of the weighted input and the bias term for that layer. Here, the weighted input is obtained by multiplying the output from the previous layer, $a^{[l-1]}$, with the weight matrix $W^{[l]}$ the current layer. The bias term, $b^{[l]}$, serves to introduce an additional adjustment, facilitating the optimization of model performance.

The backpropagation algorithm's basic idea is to minimize the loss function by computing the gradient of the loss function with respect to the weights of the network and updating the weights accordingly. The gradient computation for the weights $W^{[l]}$ and biases $b^{[l]}$ of the l layer is as follows:

The gradient at the output layer:

$$\frac{\partial L}{\partial z^{(L)}} = \frac{\partial L}{\partial a^{(L)}} \cdot \frac{\partial a^{(L)}}{\partial z^{(L)}} \tag{2}$$

$$\frac{\partial L}{\partial W^{(l)}} = \frac{\partial L}{\partial z^{(l)}} \cdot a^{(l-1)} \tag{3}$$

$$\frac{\partial L}{\partial b^{(L)}} = \frac{\partial L}{\partial z^{(L)}} \tag{4}$$

The gradient at the hidden layers:

$$\frac{\partial L}{\partial z^{(l)}} = \left(W^{(l+1)T} \cdot \frac{\partial L}{\partial z^{(l+1)}}\right) \cdot \frac{\partial z^{(l)}}{\partial a^{(l)}} \tag{5}$$

$$\frac{\partial L}{\partial W^{(l)}} = \frac{\partial L}{\partial z^{(l)}} \cdot a^{(l-1)T} \tag{6}$$

$$\frac{\partial L}{\partial b^{(l)}} = \frac{\partial L}{\partial z^{(l)}} \tag{7}$$

The gradient of L with respect to $a^{[L]}$ is the derivative of the activation function with respect to its input, and $W^{(l+1)T}$ is the transpose of the weight matrix of the $l + 1$ layer. Once the gradient is obtained, gradient descent or other optimization algorithms can be used to update the network's weights. The weight update formula can be represented as follows:

$$W_{new} = W_{old} - \alpha \frac{\partial L}{\partial W} \tag{8}$$

where, α is the learning rate. The DNN used in this study had an input layer, four hidden layers, and an output layer as its overall network structure. The four hidden layers used the ReLU activation function and had, respectively, 256, 128, 64, and 32 neurons. The output layer used the Sigmoid activation function.

DOI: 10.54254/2755-2721/67/20240606

2.6. Comparable Model Selection

This study tested the effectiveness of DNN with a range of machine learning models, including Convolutional Neural Networks (CNN), Gradient Boosting Decision Trees (GBDT), Random Forests (RF), Logistic Regression (LR), Decision Trees (DT), AdaBoost, and XGBoost.LR is a linear model suited for binary classification problems, where parameter estimation is performed through maximum likelihood estimation. DT is non-parametric supervised learning models that build a tree structure by recursively selecting the best feature. RF is an ensemble method of DTs, increasing the precision of the model by building several decision trees and combining their forecasts via voting or averaging, and GBDT is another ensemble method, which minimizes the loss function by sequentially adding decision trees. As a boosting method, AdaBoostconstructs a strong learner by combining multiple weak learners (such as decision trees), while XGBoost is an optimized version of GBDT, which improves performance by optimizing tree structure and splitting strategies. CNN is a category of deep learning models specifically designed to process data with a grid-like topology, commonly including convolutional layers, pooling layers, and fully connected layers.

2.7. Model Training

To improve the stability and efficiency of model training, the feature data in this study were normalized by deleting the minimum value and dividing by the variation between the highest and lowest numbers. The dataset was then split into two parts: a training set (80%) and a test set (20%). To balance the dataset, the Synthetic Minority Over-sampling Technique (SMOTE) was utilized to oversample the minority class and the RandomUnderSampler technique was used to undersample, lowering the quantity of samples in the class with a majority. The dataset was randomly divided into ten subsets, nine of which were utilized for training the model and the remaining subset for testing, before the chosen models were subjected to 10-fold cross-validation. Ten times through this procedure, a different subset was used as the test set and the remaining subset as the training set. This helped to decrease the degree of unpredictability in the evaluation findings and produced more consistent and trustworthy performance measures. Each model was trained on the training set and then made predictions on the test set. The DNN and CNN models were trained for 150 epochs, with a batch size set to 100. Using Adam optimizer to merge the benefits of momentum and RMSprop algorithms, effectively adjusting weights during training to expedite convergence.

2.8. Model evaluation

To evaluate the models' performance, metrics such as accuracy, area under the Receiver Operating Characteristic (ROC) curve (AUC), recall, precision, and F1 score were calculated. These metrics provided insights into the models' performance from different aspects, aiding in a comprehensive understanding of the models' predictive capabilities.

3. Result and discussion

3.1. Training results

During the training process of the DNN model, changes in loss and accuracy were meticulously recorded and visualized as Figure 4. By plotting the training and validation loss curves, as well as the corresponding accuracy curves, the learning dynamics of the model were monitored. This approach also ensured the model's generalization ability and prevented the occurrence of overfitting. These visualization results not only revealed the evolution of the model's performance during training but also provided a scientific basis for adjusting the model's training parameters.

DOI: 10.54254/2755-2721/67/20240606

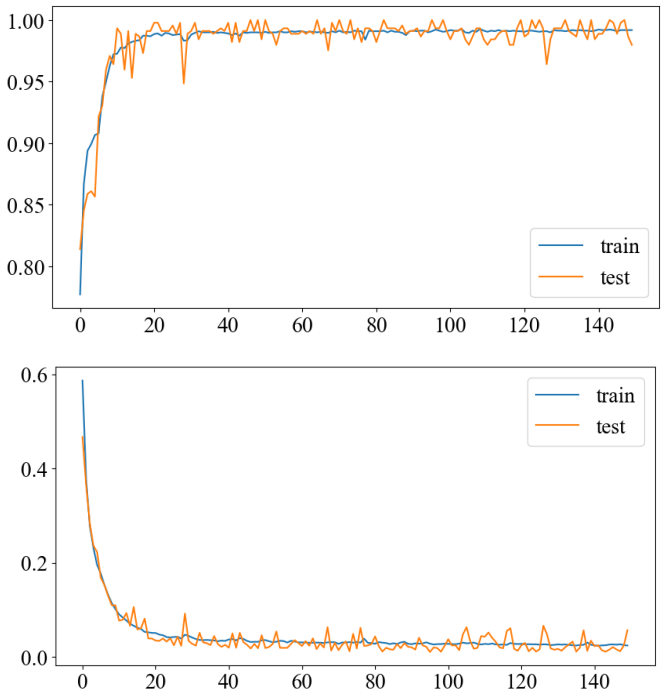

Figure 4. Accuracy curve (**left**) and loss curve (**right**) of DNN.

3.2. Testing results

During the testing results phase, a comprehensive performance evaluation was carried out on the trained models. This involved a range of evaluation metrics, including accuracy, precision, recall, and F1 score, to quantitatively compare the performance of each model. The performance metrics for all models are displayed in Table 1. Among them, the DNN model exhibited superior performance, achieving high metrics of 99.55% of Accuracy, 100% of AUC, 100% of Precision, 99.64% of Recall, and 99.82% of F1 score. Conversely, the LR model demonstrated the least favorable outcomes, with an Accuracy of only 84.68% and a Recall of merely 78.25%.

Table 1. Performance Metrics of Different Models.

Models	Accuracy (%)	AUC (%)	Precision (%)	Recall (%)	F1-score (%)
LR	84.68	92.71	94.32	78.54	85.71
DT	95.19	95.44	96.98	93.45	95.19
RF	97.35	99.72	98.16	97.09	97.62
GDBT	95.19	98.45	96.47	89.45	92.83
AdaBoost	86.34	94.14	92.71	83.27	87.74
XGBoost	98.47	99.83	98.89	97.45	98.17
DNN	**99.55**	**100**	**100**	**99.64**	**99.82**
CNN	99.42	100	100	98.55	99.27

3.3. Discussion

Early diagnosis is crucial for children with ASD, as it facilitates intervention at the initial stages of the disorder, thereby enhancing the intellectual development of children with ASD [13]. Timely intervention is known to significantly contribute to improved outcomes in cognitive growth for these children. In the present study, compared to traditional machine learning models, the DNN model used in this study shows significant advantages in various metrics, especially in predicting minority classes after imbalance learning treatment. In the context of binary classification for ASD diagnosis, DNN has

DOI: 10.54254/2755-2721/67/20240606

demonstrated exceptional performance, achieving a perfect score of 100% on both the AUC and precision metrics. Furthermore, the model has attained high accuracy rates of 99.55% for accuracy, 99.64% for recall, and 99.82% for the F1 Score. These results indicated that the DNN maintained a significant balance in predicting both positive and negative classes, thereby providing a highly accurate performance for the diagnosis of ASD. The deep structure and extensive parameters of DNN enable them to learn complex nonlinear relationships and patterns from data.

The present study complemented the research conducted by Rahman and Subashini [4], despite differences in the architecture of the constructed DNN models, it substantiated the efficacy of DNN in the diagnosis of ASD in children aged 0 to 3 years and extended the applicability to children up to 10 years of age. Additionally, the investigation encompassed the examination of other machine learning models, including XGBoost and AdaBoost, indicating a persistent dominance of DNN. This research contributed empirical evidence on the use of machine learning in the area of ASD diagnosis in children, which is pivotal for early identification of the condition.

There are some limitations in this study. At first, although the dataset has been expanded to three times the size of previous research, challenges of data imbalance and limited sample sizes persist. Then, only text data was used at present, multimodal data, such as visual information (eye-tracking experiments), facial expression recognition, and the distribution of *E. coli* colonies (exploring the gut-brain axis), should be integrated to strengthen diagnostic capabilities and develop more precise, personalized early diagnosis and intervention strategies. Finally, future research will explore the integration of DNN with attention mechanism or transformers module to improve diagnostic accuracy and efficiency.

4. Conclusion

This study delved into DNN for diagnosing ASD in children aged 0 to 10, demonstrating its superior performance over traditional machine learning models. DNN model achieved promising results, with the performance metrics exceeding 99.5% across various performance metrics. By showcasing superior performance in accuracy, recall, precision, and F1 scores, DNN confirmed its robust capability in processing complex nonlinear relationships and underscore the effectiveness of early ASD diagnosis.

References

[1] Maenner MJ, Warren Z, Williams AR, et al. 2023 Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2020. MMWR Surveill. Summ. 72 1–14.

[2] Bastiaansen J A, Meffert H, Hein S, Huizinga P, Ketelaars C, Pijnenborg M, Bartels A, Minderaa R, Keysers C and De Bildt 2010 Diagnosing Autism Spectrum Disorders in Adults: the Use of Autism Diagnostic Observation Schedule (ADOS) Module 4 J. Autism. Dev. Disord. 41 1256–66

[3] De Bildt A et al. 2015 Autism Diagnostic Interview-Revised (ADI-R) Algorithms for Toddlers and Young Preschoolers: Application in a Non-US Sample of 1,104 Children J. Autism. Dev. Disord. 45 2076–91

[4] Rahman K K M, and Subashini M M 2021 A Deep Neural Network-Based Model for Screening Autism Spectrum Disorder Using the Quantitative Checklist for Autism in Toddlers (QCHAT) J. Autism. Dev. Disord. 52 2732–46

[5] Vishal V 2022 A Comparative Analysis of Prediction of Autism Spectrum Disorder (ASD) using Machine Learning. In Proceedings of the Sixth International Conference on Trends in Electronics and Informatics 1355-1358

[6] P Rawat, M Bajaj, S Vats and V. Sharma 2022 ASD Diagnosis in Children, Adults, and Adolescents using Various Machine Learning Techniques In International Conference on Device Intelligence, Computing and Communication Technologies 625-630

DOI: 10.54254/2755-2721/67/20240606

[7] Aslam A R, Hafeez N, Heidari H and Altaf M A B 2022 Channels and Features Identification: A Review and a Machine-Learning Based Model With Large Scale Feature Extraction for Emotions and ASD Classification Front. Neurosci. 16 844851
[8] Alkahtani H, Aldhyani T H H and Alzahrani M Y 2023 Deep Learning Algorithms to Identify Autism Spectrum Disorder in Children-Based Facial Landmarks Appl. Sci. 13 4855
[9] Thabtah F and Peebles D 2019 Early Autism Screening: A Comprehensive Review Int. J. Environ. Res. Public. Health 16 3502
[10] Thabtah F, Abdelhamid N and Peebles D 2019 A machine learning autism classification based on logistic regression analysis Health Inf. Sci. Syst. 7 12
[11] Thabtah F, Mampusti E, Peebles D, Herradura R and Varghese J 2019 A Mobile-Based Screening System for Data Analyses of Early Dementia Traits Detection J. Med. Syst. 44 24
[12] Booth T, Murray A L, McKenzie K, Kuenssberg R, O'Donnell M and Burnett H 2013 Brief Report: An Evaluation of the AQ-10 as a Brief Screening Instrument for ASD in Adults J. Autism. Dev. Disord. 43 2997–3000
[13] Estes A, Munson J, Rogers S J, Greenson J, Winter J and Dawson G 2015 Long-Term Outcomes of Early Intervention in 6-Year-Old Children With Autism Spectrum Disorder J. Am. Acad. Child. Adolesc. Psychiatry 54 580-7