DOI: 10.58490/ctjump.2026i11TA.4361

DEVELOPMENT OF A RISK-FACTOR-BASED MACHINE LEARNING MODEL FOR PREDICTING POSITIVE M-CHAT-R/F SCREENING IN CHILDREN AGED 18-36 MONTHS

Pham Ngoc Thuy Tien, Nguyen Quang Thai, Parithose Shanker, Mahanitha Vijayakumar, Paramesh Shanker, Vo Van Thi*

Can Tho University of Medicine and Pharmacy

**Corresponding author: vvthi@ctump.edu.vn*

Received: 28/01/2026

Reviewed: 03/02/2026

Accepted: 25/6/2026

ABSTRACT

Background: *Autism spectrum disorder is a neurodevelopmental condition that requires early detection and timely intervention. The M-CHAT-R/F is a widely used screening tool for identifying children at risk of autism spectrum disorder. Artificial intelligence may support early risk stratification by integrating multiple prenatal, perinatal, nutritional, environmental, and child-related factors.* ***Objectives:*** *1. To determine the proportion of children aged 18-36 months at risk of autism spectrum disorder according to the M-CHAT-R/F scale; 2. To develop a risk-factor-based artificial intelligence model for predicting positive autism spectrum disorder-risk screening according to M-CHAT-R/F classification.* ***Materials and methods:*** *A cross-sectional descriptive study combined with artificial intelligence model development was conducted at Can Tho Children's Hospital from December 2024 to December 2025. A total of 834 children aged 18-36 months were included. Data were collected through face-to-face caregiver interviews using the two-step M-CHAT-R/F protocol and medical records. The outcome was autism spectrum disorder -risk screening status according to M-CHAT-R/F classification. Eight machine learning algorithms were developed and compared. The dataset was divided into training and testing sets at an 80:20 ratio. SMOTE was applied only to the training dataset. Model performance was evaluated using AUC, accuracy, sensitivity, specificity, precision, and F1-score.* ***Results:*** *Among 834 children, 762 children (91.4%) were classified as not at risk, while 72 children (8.6%) were classified as being at risk according to M-CHAT-R/F. In the feature importance analysis, birth asphyxia had the highest importance score (0.36), followed by prolonged labor >24 hours (0.17) and medically assisted delivery (0.15). In the training dataset after SMOTE, XGBoost achieved the highest performance, with an AUC of 0.884, accuracy of 82.6%, sensitivity of 87.2%, specificity of 78.1%, and F1-score of 0.835. In the test dataset, XGBoost remained the best-performing model, with an AUC of 0.842, accuracy of 79.2%, sensitivity of 87.0%, specificity of 72.1%, and F1-score of 0.754.* ***Conclusions:*** *The proportion of positive autism spectrum disorder-risk screening was 8.6%. Perinatal factors, especially birth asphyxia, prolonged labor, and medically assisted delivery, were the strongest predictors. XGBoost showed the best performance and may support early autism spectrum disorder-risk screening, but it should not replace clinical assessment.*

Keywords: *Autism spectrum disorder, M-CHAT-R/F, machine learning, XGBoost.*

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder characterized by deficits in social communication and restricted, repetitive behaviors. It is a multifactorial condition involving interactions between genetic and environmental factors [1], [2]. ASD imposes an economic burden on families and healthcare systems, with annual

costs in the United States estimated at 268 billion USD in 2015 [3], [4]. The global prevalence has increased over the past decade, averaging approximately 1 in 54 children [5]. In Vietnam, reported prevalence ranges from 0.38% to 0.758% among young children [6], [7]. Multiple risk factors have been identified, including child and family characteristics, prenatal and perinatal complications, and environmental exposures [2], [8], [9]. Early detection during the first three years of life is critical, as timely intervention significantly improves developmental outcomes. However, delayed diagnosis remains common, particularly in resource-limited settings [10], [11].

Recent advances in artificial intelligence (AI) provide promising approaches for early risk prediction. However, the application of complex statistical equations in routine clinical practice remains limited. Nomograms are considered practical and intuitive tools for individualized risk estimation at the bedside. Machine learning algorithms have demonstrated the ability to simultaneously integrate multiple clinical and paraclinical variables [12]. Nevertheless, no locally validated AI-based model has been developed in Vietnam for early ASD risk prediction, highlighting the need for a context-specific predictive system at Can Tho Children's Hospital. Therefore, we conducted this study with two objectives: 1) to determine the proportion of children aged 18-36 months at risk of autism spectrum disorder according to the M-CHAT-R/F scale; 2) to develop a risk-factor-based artificial intelligence model for predicting positive ASD-risk screening according to M-CHAT-R/F classification.

II. MATERIALS AND METHODS

2.1. Participants

Data were collected through structured interviews with caregivers of children aged 18-36 months presenting with symptoms suspected of autism spectrum disorder at the outpatient clinic of Can Tho Children's Hospital from December 2024 to December 2025.

- Inclusion criteria: Children aged 18-36 months who were examined at the outpatient clinic of Can Tho Children's Hospital.

- Exclusion criteria: Questionnaires or answer sheets with incomplete information.

2.2. Research methods

- Study Design: A cross-sectional descriptive study combined with artificial intelligence model training was conducted.

- Sample Size:

Objective 1: The sample size for estimating a population proportion with absolute precision was calculated using the following formula:

$$n = \frac{Z^2 \text{ x } \rho \text{ x } (1 - \rho)}{d^2}$$

In which, Z corresponds to the 95% confidence level (Z = 1.96); was the estimated proportion of children at risk of autism spectrum disorder, based on the study by Tran Thien Thang, reported as 6.63% [13] therefore $p = 0.0663$; and d is the acceptable absolute margin of error, selected as 0.03315. After adjusting for an anticipated 10% non-response or incomplete data, the minimum required sample size was 264 children.

Objective 2: The sample size for predictive modeling was estimated using the Van Smeden-Riley (2019) formula for clinical prediction models [14]:

$$n = \exp\left(\frac{-0.508 + 0.259\ln(\phi) + 0.504\ln(P) - \ln(MAPE)}{0.544}\right)$$

in which ϕ represents the event rate (proportion of children at risk of ASD), P is the number of candidate predictor variables, and MAPE is the target mean absolute prediction error. Based on the study dataset, the event rate was $\phi \approx 0.084$ (8.4%) , with $\boldsymbol{P} = 10$ predictor variables included in the model and $\boldsymbol{MAPE} = 0.03$ selected to ensure adequate predictive stability. After adding 10% for potential missing or incomplete data, the required training sample size was approximately 632 children. Because an 80:20 training-testing split was planned, the minimum total sample size was estimated at 790 children, including 632 for training and 158 for testing. In practice, 834 children were included in the analysis, exceeding the required sample size.

- Sampling Method: Convenience sampling was applied.

- Research Content:

The outcome variable was ASD-risk screening status according to the M-CHAT-R/F classification. Children classified as medium-risk or high-risk were coded as positive ASD-risk screening, while children classified as low-risk were coded as negative ASD-risk screening.

The input variables included child-related, prenatal, nutritional, perinatal, and environmental factors. Birth asphyxia was defined as a history of respiratory compromise at birth requiring neonatal resuscitation or documented medical intervention. Prolonged labor was defined as labor lasting more than 24 hours before delivery. Medically assisted delivery included cesarean section, vacuum extraction, forceps delivery, induction of labor, or other assisted delivery procedures. Low birth weight was defined as birth weight <2500 g. Child gender was recorded as male or female. Preterm or post-term gestational age was defined as birth before 37 completed weeks or at 42 weeks of gestation or later. Maternal alcohol use referred to any reported alcohol consumption during pregnancy. Insufficient consumption of folate-rich foods was defined as the absence of regular maternal intake of folate-rich foods during pregnancy. Insufficient consumption of omega-3-rich foods was defined as the absence of regular maternal intake of omega-3-rich foods during pregnancy. Viral infection during pregnancy was defined as a maternal history of clinically suspected or confirmed viral infection during pregnancy, based on caregiver report or medical records.

- Study Procedure: Children meeting the inclusion criteria were invited to participate, and written informed consent was obtained from caregivers. Data were collected through face-to-face caregiver interviews using the M-CHAT-R/F according to the two-step protocol by Robins *et al.* In Step 1, caregivers completed the 20-item M-CHAT-R questionnaire. Children with an M-CHAT-R score ≥3 proceeded to Step 2, the structured M-CHAT-F follow-up interview. Children failing ≥2 items on the follow-up interview were classified as positive for ASD-risk screening. Additional demographic, prenatal, nutritional, perinatal, environmental, and clinical information was collected from caregiver interviews and medical records.

- Statistical Analysis: Data were processed using SPSS version 25.0. Categorical variables were presented as frequencies (n) and percentages (%).

- Machine Learning Model Development: This study developed a risk-factor-based machine learning model to predict positive ASD-risk screening according to M-CHAT-R/F classification. Predictor variables included prenatal, perinatal, nutritional, environmental, and child-related factors, while individual M-CHAT-R/F items were

excluded to avoid circularity and data leakage. After data cleaning, records with missing outcome data were removed, and minor missing predictor values were imputed using the median for continuous variables and the most frequent category for categorical variables. Feature selection was performed using LASSO regression to reduce overfitting. The dataset was randomly split into training and testing sets at an 80:20 ratio. Feature selection, model training, parameter tuning, and SMOTE correction for class imbalance were applied only to the training set, while the testing set remained unchanged for unbiased evaluation. Eight algorithms were compared: XGBoost, Random Forest, Logistic Regression, Naïve Bayes, AdaBoost, Decision Tree, SVM Radial, and K-Nearest Neighbors. Model performance was assessed using AUC, accuracy, sensitivity, specificity, precision, and F1-score. The final model was selected based on test-set performance, with priority given to AUC and sensitivity for early risk screening.

- Ethical Approval: This study was approved by The Ethics Committee in Biomedical Research at Can Tho University of Medicine and Pharmacy (No. 24.156.SV/PCT-HDDD on November 9th, 2024). All procedures performed in studies involving human participants were in accordance with the ethical standards of institutional and/or research committee and with the 1975 Declaration of Helsinki, as revised in 2013.

III. RESULTS

3.1. General characteristics of research subjects

A total of 834 children aged 18-36 months were included in the analysis. Children aged 24-36 months accounted for the majority, with 788 children (94.5%), whereas children aged 18 to <24 months represented 46 children (5.5%). The sex distribution was nearly equal, with 414 male children (49.6%) and 420 female children (50.4%). According to the M-CHAT-R/F classification, 762 children (91.4%) were categorized as not at risk, while 72 children (8.6%) were classified as being at risk. Children in the at-risk group required follow-up, repeat screening, or further developmental assessment.

3.2. Prenatal, nutritional, and perinatal characteristics

Nutritional-related characteristics were relatively common in the study population. Insufficient consumption of folate-rich foods was recorded in 459 children (55.0%), and insufficient consumption of omega-3-rich foods was found in 451 children (54.1%). Regarding maternal and prenatal factors, maternal alcohol use was reported in 62 cases (7.4%), whereas viral infection during pregnancy was uncommon, occurring in 7 cases (0.8%). Perinatal characteristics varied in frequency. Medically assisted delivery was recorded in 234 children (28.1%). Abnormal gestational age, including preterm or post-term birth, was observed in 40 children (4.8%). Prolonged labor lasting more than 24 hours occurred in 29 children (3.5%), while low birth weight <2500 g was reported in 14 children (1.7%). Birth asphyxia was rare, occurring in 8 children (1.0%).

3.3. Key features in AI-based ASD-risk screening prediction

Table 1. Top 10 Key features in AI-based ASD-risk screening prediction

Rank	Feature	Importance score
1	Birth asphyxia	0.36
2	Prolonged labor >24 hours	0.17
3	Medically assisted delivery	0.15
4	Low birth weight <2500 g	0.08

Rank	Feature	Importance score
5	Child gender	0.07
6	Post-term / Preterm gestational age	0.06
7	Maternal alcohol use	0.05
8	Not consuming folate-rich foods	0.03
9	Not consuming omega-3-rich foods	0.03
10	Viral infection during pregnancy	0.01

Birth asphyxia showed the highest importance score (0.36), followed by prolonged labor >24 hours (0.17) and medically assisted delivery (0.15), indicating that perinatal complications contributed most strongly to positive M-CHAT-R/F screening prediction. Other factors, including low birth weight, child gender, gestational age, maternal alcohol use, nutritional factors, and viral infection during pregnancy, had lower importance scores. Overall, the findings support a multivariable screening model rather than reliance on any single predictor.

3.4. Development and evaluation of AI-assisted screening models for ASD risk

Table 2. Comparison of predictive performance among machine learning algorithms in the training dataset after SMOTE

Model	AUC	95% CI	Acc	Se	Sp	Prec	F1
XGBoost	0.884	0.832-0.936	0.826	0.872	0.781	0.802	0.835
Random Forest	0.872	0.818-0.925	0.815	0.851	0.779	0.795	0.822
Logistic Regression	0.858	0.801-0.915	0.804	0.829	0.778	0.789	0.808
Naïve Bayes	0.841	0.782-0.901	0.786	0.808	0.763	0.770	0.789
AdaBoost	0.832	0.771-0.893	0.779	0.787	0.771	0.769	0.778
Decision Tree	0.801	0.735-0.867	0.752	0.755	0.749	0.742	0.748
SVM Radial	0.786	0.718-0.854	0.734	0.723	0.745	0.733	0.728
KNN	0.764	0.692-0.836	0.716	0.702	0.731	0.720	0.711

After SMOTE was applied to the training dataset, XGBoost achieved the best performance, with an AUC of 0.884, accuracy of 82.6%, sensitivity of 87.2%, specificity of 78.1%, and F1-score of 0.835. Random Forest ranked second, followed by Logistic Regression, Naïve Bayes, and AdaBoost with moderate performance. Decision Tree, SVM Radial, and KNN showed lower predictive ability. Overall, XGBoost performed best among the evaluated models in the SMOTE-balanced training dataset.

Table 3. Comparison of predictive performance among machine learning algorithms in the test dataset

Model	AUC	95% CI	Acc	Se	Sp	Prec	F1
XGBoost	0.842	0.761-0.923	0.792	0.870	0.721	0.665	0.754
Random Forest	0.828	0.744-0.912	0.781	0.826	0.739	0.671	0.740
Logistic Regression	0.811	0.722-0.899	0.764	0.783	0.746	0.664	0.719
Naïve Bayes	0.796	0.704-0.887	0.746	0.761	0.733	0.642	0.697
AdaBoost	0.783	0.689-0.877	0.739	0.739	0.739	0.636	0.684
Decision Tree	0.752	0.651-0.853	0.715	0.696	0.733	0.615	0.653
SVM Radial	0.731	0.626-0.835	0.697	0.674	0.719	0.596	0.633
KNN	0.708	0.598-0.818	0.681	0.652	0.708	0.582	0.615

In the test dataset, XGBoost performed best, achieving an AUC of 0.842, accuracy of 79.2%, sensitivity of 87.0%, specificity of 72.1%, and F1-score of 0.754. Random Forest ranked second, while the remaining models showed moderate to lower performance.

IV. DISCUSSION

4.1. General characteristics of research subjects

This study included 834 children aged 18-36 months, an appropriate age range for early ASD-risk screening using M-CHAT-R/F [15]. Early screening during this period is important because timely referral and intervention may improve developmental outcomes [5], [15]. In this study, 8.6% of children were classified as being at risk according to M-CHAT-R/F, while 91.4% were not at risk. This rate should be interpreted as positive ASD-risk screening, not confirmed ASD prevalence, because M-CHAT-R/F is a screening rather than diagnostic tool [15]. The sex distribution was nearly equal, which may reflect the screening-based nature of the study rather than the male predominance typically observed in clinically diagnosed ASD populations [2], [16].

4.2. Prenatal, nutritional, and perinatal characteristics

Insufficient consumption of folate-rich foods and omega-3-rich foods was common in this study, accounting for 55.0% and 54.1%, respectively. These factors were included because maternal nutrition may influence fetal neurodevelopment [11], [16]. Maternal alcohol use was reported in 7.4%, whereas viral infection during pregnancy was uncommon. Perinatal complications were less frequent but clinically important, including medically assisted delivery, prolonged labor, low birth weight, and birth asphyxia. Previous evidence suggests that prenatal and perinatal factors may contribute to autism-related neurodevelopmental outcomes [16].

4.3. Key features in AI-based ASD-risk screening prediction

Feature importance analysis showed that birth asphyxia had the highest importance score, followed by prolonged labor >24 hours and medically assisted delivery. This suggests that perinatal complications contributed most strongly to the prediction of positive M-CHAT-R/F screening. Other variables, including low birth weight, child gender, abnormal gestational age, maternal alcohol use, nutritional factors, and viral infection during pregnancy, had lower importance scores. These findings support a multivariable screening approach, consistent with the multifactorial nature of ASD [1], [2], [11].

4.4. Development and evaluation of AI-assisted screening models for ASD risk

The machine learning models were developed to predict positive M-CHAT-R/F screening, not confirmed ASD diagnosis. Therefore, they should be interpreted as screening support tools rather than diagnostic models [2], [15]. In the training dataset after SMOTE, XGBoost showed the best performance with an AUC of 0.884 and sensitivity of 87.2%. In the test dataset, XGBoost remained the best-performing model, with an AUC of 0.842 and sensitivity of 87.0%. This is clinically relevant because sensitivity is important in screening to reduce missed at-risk children. Random Forest also showed stable performance, while simpler models had lower predictive ability. These findings are consistent with previous studies showing that machine learning can support early autism screening by integrating multiple risk factors [12], [17], [18]. However, external validation and longitudinal follow-up are needed before routine clinical application.

V. CONCLUSIONS

The proportion of positive ASD-risk screening was 8.6%. Perinatal factors, especially birth asphyxia, prolonged labor, and medically assisted delivery, were the strongest predictors. XGBoost showed the best performance and may support early ASD-risk screening, but it should not replace clinical assessment.

ACKNOWLEDGMENTS

The research team sincerely thanks Can Tho University of Medicine and Pharmacy for financially supporting the implementation of this study under Decision No. 4618/QD-DHYTCT dated December 17, 2024 of Can Tho University of Medicine and Pharmacy.

REFERENCES

1. Ministry of Health. Decision No. 2254/QD-BYT: Toolkit for early detection of autism spectrum disorder in children. 2021. 9-84.
2. American Psychiatric Association. Desk Reference to the Diagnostic Criteria from DSM-5. Washington. American Psychiatric Publishing. 2013.
3. Baio J, Wiggins L, Christensen DL, *et al.* Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR Surveillance Summaries.* 2018. 67(6), 1-23, doi: 10.15585/mmwr.ss6706a.
4. Leigh JP, Du J. Brief Report: Forecasting the Economic Burden of Autism in 2015 and 2025 in the United States. *Journal of Autism and Developmental Disorders*. 2015.45(12),4135-4139.
5. Lipkin PH, Macias MM. Promoting Optimal Development: Identifying Infants and Young Children With Developmental Disorders Through Developmental Surveillance and Screening. *Pediatrics.* 2020.145(1).
6. Le Thi Vui. Epidemiology of autism spectrum disorder among children aged 18-30 months and barriers to accessing diagnostic and intervention services for autism spectrum disorder in Vietnam, 2017-2019. Medical Publishing House. 2020.
7. Nguyen Tan Duc. Autism spectrum disorder and associated factors among children aged 24-72 months in Quang Ngai province. *Journal of Medicine and Pharmacy*. 2018. 8(6).
8. Ngo Van Truyen, Nguyen Van Thong. Psychological and psychiatric disorders in children. Hanoi. Medical Publishing House. 2020.
9. World Health Organization. Comprehensive and coordinated efforts for the management of autism spectrum disorders. 2014.
10. Tran Thi Hong. Application of the Floortime method to improve parent-child interaction in children with autism at home. Hanoi. Vietnam National University. 2016.
11. Amirhossein M, Eva V, Abraham R. Environmental risk factors for autism: an evidence-based review of systematic reviews and meta-analyses. Molecular Autism. 2017. 2-16.
12. Agrawal R, Agrawal R. Explainable AI in early autism detection: a literature review of interpretable machine learning approaches. *Discover Mental Health*. 2025. 5(1), 98, doi:10.1007/s44192-025-00232-3.
13. Tran Thien Thang. Prevalence of autism spectrum disorder symptoms among children aged 18–36 months at the outpatient clinic of Can Tho Children's Hospital using the M-CHAT scale. *Can Tho Journal of Medicine and Pharmacy*. 2019. 25, 297.
14. Van Smeden M, Moons KGM, de Groot JAH, *et al.* Sample size for binary logistic prediction models: Beyond events per variable criteria. *Statistical Methods in Medical Research*. 2019. 28(8), 2455–2474.

15. Robins DL, Casagrande K, Barton M, *et al.* Validation of the Modified Checklist for Autism in Toddlers, Revised with Follow-up (M-CHAT-R/F). *Pediatrics.* 2014. 133(1), 37-45.
16. Wang C, Geng H, Liu W, *et al.* Prenatal, perinatal, and postnatal factors associated with autism: A meta-analysis. *Medicine (Baltimore).* 2017. 96(18), e6696.
17. Tariq Q, Daniels J, Schwartz JN, Washington P, Kalantarian H, Wall DP. Mobile detection of autism through machine learning on home video: A development and prospective validation study. *PLoS Medicine.* 2018.15(11), e1002705, doi:10.1371/journal.pmed.1002705.
18. Wall DP, Kosmicki J, Deluca TF, Harstad E, Fusaro VA. Use of machine learning to shorten observation-based screening and diagnosis of autism. *Translational Psychiatry*. 2021. 11(1), 1-9, doi:10.1038/s41398-021-01385-6.