Journal of Autism and Developmental Disorders
https://doi.org/10.1007/s10803-026-07227-1

ORIGINAL ARTICLE

Machine Learning–Assisted Autism Risk Stratification in Toddlers Using the Vietnamese M-CHAT-R and Perinatal Predictors: A Cross-Sectional Study in Vietnam

Thi Van Vo[1] · Phuong Minh Nguyen[1] · Duy Nhat Nguyen[1] · Trinh Van Nguyen[1] · Dien Minh Thai[2] · Hung Huynh Vinh Ly[1] · Trinh Nguyen Ngoc Luu[1]

Received: 19 August 2025 / Accepted: 8 January 2026

Abstract

Introduction The Modified Checklist for Autism in Toddlers, Revised (M-CHAT-R) is widely used for autism spectrum disorder (ASD) screening; however, evidence on the Vietnamese version and on scalable risk stratification approaches suitable for routine preschool settings remains limited.

Purpose This study aimed to evaluate the internal consistency of the Vietnamese M-CHAT-R and to develop machine learning (ML) models integrating M-CHAT-R–derived features with key obstetric–perinatal predictors to support ASD risk stratification in preschool-based screening.

Methods This cross-sectional study conducted in Ca Mau province, Vietnam, 3,639 children aged 18–36 months were screened using the Vietnamese Ministry of Health–issued M-CHAT-R administered. A mobile clinical team performed onsite DSM-5 evaluations, and the M-CHAT-R/F follow-up was implemented for moderate-risk (scores 3–7) cases.

Results Six ML algorithms, including Random Forest, were trained using a 75/25 train–test split with SMOTE applied to the training set. Model performance was assessed using accuracy, recall, precision, F1-score, and area under the receiver operating characteristic curve (AUC). 75 children met DSM-5 criteria for ASD (2.0%); 53.3% were male and 96.0% were aged 24–36 months. The Vietnamese M-CHAT-R showed good internal consistency (Cronbach's alpha = 0.863). Random Forest with SMOTE achieved the best performance (AUC = 0.983; recall = 0.95; precision = 0.351; F1-score = 0.513) using a refined feature set including M-CHAT-R risk level and a composite biological risk index. High recall with moderate precision reflects a screening-oriented emphasis on sensitivity.

Conclusions ML- assisted risk stratification may support prioritization for follow-up and specialist assessment in preschool-based pathways in low- and middle-income countries.

Keywords Autism spectrum disorder · M-CHAT-R · M-CHAT-R/F · Machine learning · Risk stratification · Vietnam

✉ Trinh Nguyen Ngoc Luu
lnntrinhh@gmail.com

Thi Van Vo
vvthi@ctump.edu.vn

Phuong Minh Nguyen
nmphuong@ctump.edu.vn

Duy Nhat Nguyen
nhatduyshlord136@gmail.com

Trinh Van Nguyen
nguyenvantrinh00021@gmail.com

Dien Minh Thai
tdien365@gmail.com

Hung Huynh Vinh Ly
lhvhung.ctump@gmail.com

[1] Department of Pediatrics, Faculty of Medicine, Can Tho University of Medicine and Pharmacy, No.179, Nguyen Van Cu Street, An Khanh Ward, Ninh Kieu District, Can Tho, Vietnam

[2] Faculty of Medicine, Nam Can Tho University, 168 Nguyen Van Cu Street, An Binh Ward, Can Tho City, Vietnam

Published online: 23 January 2026

Introduction

The consequences of ASD are profound, leading to significant psychological, social, and economic impairments and a substantial burden for both families and society at large. According to the Centers for Disease Control and Prevention (CDC), the current prevalence rate of ASD is estimated at 3.2% in the United States (Shaw et al., 2025) and a discernible upward trend observed in recent years (Dietz et al., 2020), (Autism, Developmental Disabilities Monitoring Network Surveillance Year Principal, Centers for Disease, & Prevention, 2007), (Autism, Developmental Disabilities Monitoring Network Surveillance Year Principal, Centers for Disease, & Prevention, 2009). Beyond the United States, global evidence also indicates substantial heterogeneity in ASD prevalence across regions and study methodologies. A recent systematic review and meta-analysis estimated a pooled global prevalence of 0.77% in children, highlighting wide variability across settings. (Issac et al., 2025)

In Vietnam, nationally representative prevalence estimates are not yet available; existing data primarily come from regional population-based surveys, with reported prevalence varying by study setting and methods. (Hoang et al., 2019; Thi Vui et al., 2021) Data from a systematic review and meta-analysis conducted in several Asian countries, including China, Hong Kong, and Taiwan, also reported substantial variability in ASD prevalence estimates across settings and study methodologies (Sun et al., 2013). An increasing trend in ASD identification has also been reported in Asia (Qiu et al., 2019). However, this prevalence is accompanied by significant health, psychological, educational, and economic burdens, not only for the affected children but also for their families and the broader community. Although no official statistics are currently available for Vietnam, families of children with ASD are widely reported to experience substantial economic and social burdens (Ha et al., 2014).

In Vietnam, diagnostic and intervention services for children with ASD are provided by healthcare units, educational institutions, rehabilitation centers, and psychological counseling services ADDIN EN.CITE (Hoang et al., 2019), (Nguyen et al., 2021). The commonly used tools for ASD screening in Vietnam include the M-CHAT as an initial screener, followed by diagnostic confirmation using standardized criteria (Thi Vui et al., 2021). Evidence on the psychometric performance and scalable implementation of the Vietnamese Ministry of Health–issued M-CHAT-R in routine preschool workflows remains limited, especially outside major cities. In addition, diagnostic and intervention services are concentrated in urban centers, while rural and remote areas often lack specialist capacity for ASD assessment and care. Strengthening early identification and referral pathways for ASD is critical for timely intervention and for informing service planning in under-resourced settings. Scalable approaches to early ASD risk identification may improve children's long-term developmental outcomes. In practice, the diagnostic process often requires comprehensive evaluation by specialists in pediatric neuropsychology or child psychiatry, which poses significant challenges for implementation at the community level, especially in remote and underserved areas. In Vietnam, in particular, awareness and access to child mental health and neurodevelopmental services remain uneven across regions. To address these limitations, early screening tools such as the Modified Checklist for Autism in Toddlers, Revised with Follow-Up (M-CHAT-R/F)-a widely used instrument for children aged 16–30 months—have been developed and validated in multiple countries. However, in Vietnam, there remains a lack of empirical studies evaluating the reliability of the Vietnamese version (Thi Van Vo et al., 2025).

The continuous advancement of artificial intelligence (AI) has significantly contributed to improving healthcare quality. Among its most impactful applications are ML models that support risk stratification and decision support in screening workflows. In the context of ASD, ML approaches may help prioritize follow-up and specialist assessment—especially in settings where sustained delivery of follow-up interviews is constrained by limited healthcare personnel. We integrated a small set of obstetric–perinatal predictors because these variables are often routinely documented around birth and have been consistently associated with ASD diagnosis in epidemiologic studies, although they are not sufficient to infer causality. (Wang et al., 2017; Vui et al., 2023). In real-world preschool programs, the recommended M-CHAT-R/F follow-up interview can be difficult to deliver consistently because it requires trained personnel and additional time. Therefore, we designed the ML component to provide risk ranking/triage using screening-stage information, helping prioritize which children should receive follow-up interviews and specialist assessment first, rather than replacing clinical diagnosis.

When combined with M-CHAT-R–derived features, these predictors may improve discrimination in resource-constrained screening pathways. (Wang et al., 2017; Vui et al., 2023). Although the M-CHAT-R has been validated and widely used as a screening tool for toddlers at risk of ASD in various countries, its performance may be affected by subjective interpretation from caregivers or other assessors. This subjectivity may introduce inconsistencies in screening outcomes, thereby limiting its reliability in large-scale or community-based applications (Khowaja et al., 2015), (Wieckowski et al., 2023). Numerous international studies have shown that ML models, when trained on questionnaire data or medical records, can predict autism risk with

considerable accuracy, approaching the performance of clinical assessments (Bone et al., 2016; Pan et al., 2025).

Despite its critical role, there is a lack of studies in Vietnam evaluating the predictive performance of ML models based on M-CHAT-R–derived data and readily available obstetric–perinatal information to support scalable risk stratification. This study aims to evaluate the internal consistency of the Vietnamese M-CHAT-R and to develop ML-based risk stratification models using M-CHAT-R–derived information together with obstetric–perinatal predictors to support scalable preschool-based screening.

Methods

Study Population and Design

This cross-sectional study evaluated the internal consistency of the Vietnamese M-CHAT-R and developed ML models for ASD risk stratification within a preschool-based screening pathway among toddlers aged 18–36 months in Ca Mau province, Vietnam, from January to September 2022 (n = 3,639).

This provincial program was implemented through a multi-sector collaboration involving Can Tho University of Medicine and Pharmacy, the Ca Mau Department of Health, the Ca Mau Department of Education and Training, and Ca Mau Obstetrics and Pediatrics Hospital to support coordinated screening, onsite DSM-5 evaluation for the full screened cohort, and referral pathways.

The study was conducted across 117 preschools and kindergartens in Ca Mau province. These participating institutions represented all registered preschools/kindergartens enrolling toddlers aged 18–36 months in the province during the study period, as listed by the provincial Department of Education and Training. Eligible participants were children aged 18–36 months enrolled in participating institutions whose parent or legal guardian provided written informed consent. Children with incomplete M-CHAT-R questionnaires or missing key covariate information were excluded. Key covariates included age, sex, maternal age, and the prespecified obstetric–perinatal variables used in the primary analyses; records missing these fields were excluded from the analytic dataset. We attempted to include all eligible toddlers enrolled in participating preschools/kindergartens during the study period. Because the sampling frame did not include non-enrolled toddlers, findings (including the proportion classified as ASD in the screened cohort) should be interpreted as estimates for a preschool-enrolled screened cohort rather than population-based community prevalence.

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki. The study protocol was approved by the Ethics Committee of Can Tho University of Medicine and Pharmacy (Approval No. 261/PCT-HĐĐĐ; 28 August 2020). Written informed consent was obtained from each child's parent or legal guardian prior to participation. Parents/guardians received detailed information about the study objectives and procedures. Participation was voluntary, and parents/guardians could decline or withdraw at any time without any effect on the child's access to routine care.

Participant Selection

All toddlers aged 18–36 months enrolled in the 117 participating preschools/kindergartens during January–September 2022 were invited. Of 3,842 initially identified children, 3,639 were included in the final analytic sample after excluding records with incomplete M-CHAT-R questionnaires and/or missing key covariate information (Fig. 1).

Screening personnel, training, and Data Collection

Preschool teachers served as the primary screening personnel and were trained in a standardized protocol prior to data collection. The research team (pediatricians and child mental health/psychiatry clinicians from Can Tho University of Medicine and Pharmacy), in collaboration with the provincial Department of Education and Training, delivered nine district-level training workshops (one per district) for participating teachers. Training included core knowledge on autism spectrum disorder, standardized administration of the Vietnamese Ministry of Health–issued M-CHAT-R, item-by-item guidance with worked examples, and practical case scenarios to improve consistency across sites. Teachers selected for training were those directly responsible for classrooms with toddlers aged 18–36 months to ensure regular contact with children and the opportunity to observe behaviors in daily classroom routines. The provincial Department of Education and Training coordinated the list of participating schools and teachers. Each school designated at least one principal and one staff member to coordinate screening activities, disseminate study materials, and oversee adherence to standardized procedures.

The M-CHAT-R was completed by teachers in collaboration with parents/guardians. Because caregiver literacy varies in the study setting, teachers acted as trained facilitators who explained items using standardized instructions and recorded responses after combining classroom observation with caregiver input obtained through direct, face-to-face discussions. When information was unclear at the time of

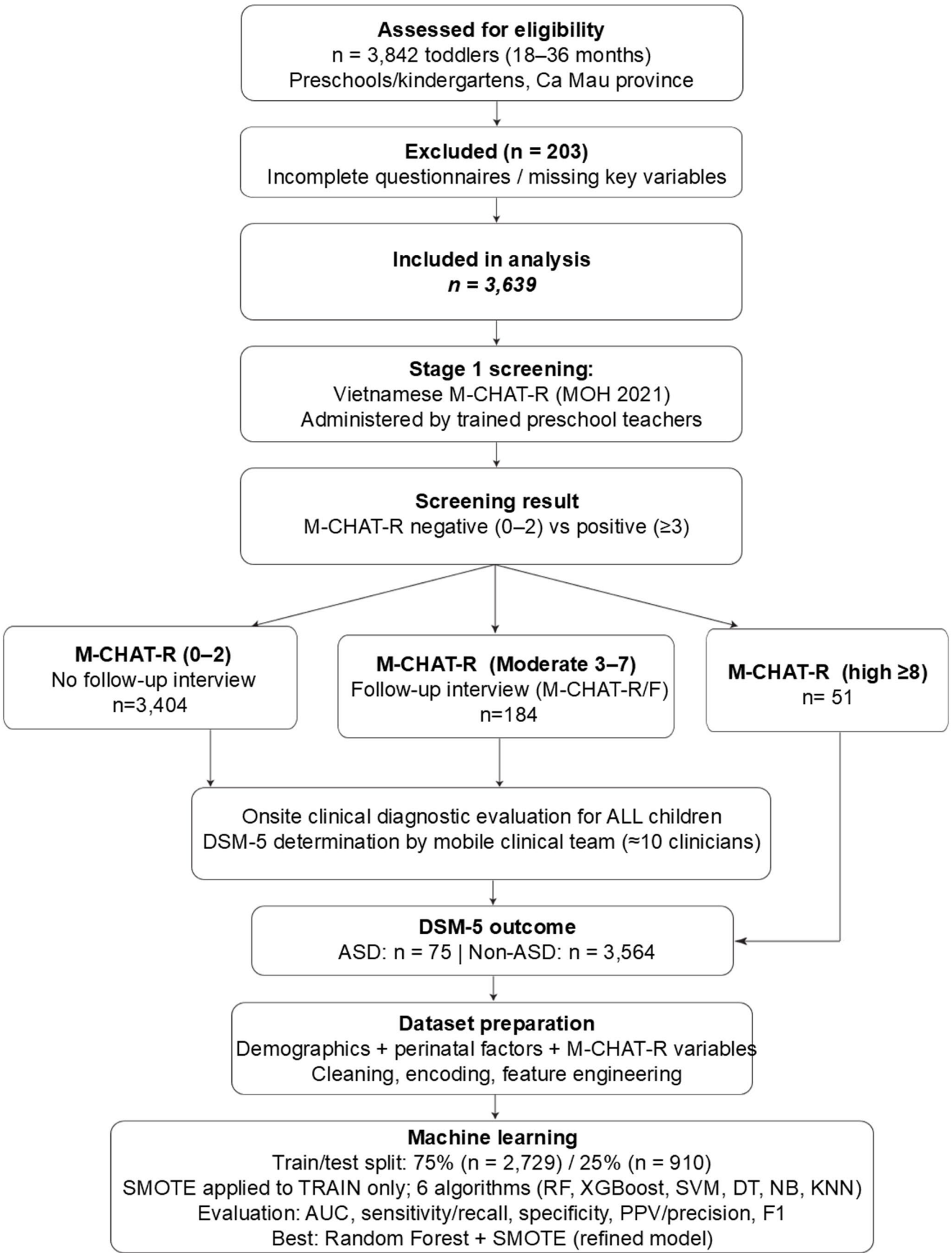

Fig. 1 Research flowchart

screening, it was documented and revisited during onsite clinical visits.

Information on obstetric and perinatal variables was collected through structured, face-to-face interviews with parents/guardians, with verification where feasible using available child health booklets/records (e.g., birth or neonatal documentation). If an item could not be reliably confirmed, it was recorded as "unknown/missing" rather than inferred. During onsite visits, the mobile clinical team reviewed unclear histories with caregivers and performed additional verification where possible to minimize misclassification. Children with incomplete M-CHAT-R questionnaires or missing key covariate information required for main analyses were excluded according to predefined eligibility criteria.

Instrument

The Vietnamese version of the Modified Checklist for Autism in Toddlers, Revised (M-CHAT-R) used in this study was standardized and officially issued by the Ministry of Health of Vietnam in 2021. The M-CHAT-R is a 20-item screening checklist designed to identify toddlers at elevated likelihood for autism spectrum disorder (ASD). The M-CHAT-R/F (Modified Checklist for Autism in Toddlers, Revised with Follow-Up) is the recommended two-stage screening procedure in which an initial M-CHAT-R screen is followed by a structured follow-up interview for selected cases to reduce false positives. Although the M-CHAT-R/F was originally developed for children aged 16–30 months, in this provincial preschool screening program it was applied to toddlers aged 18–36 months, consistent with the local implementation context.

M-CHAT-R scoring and risk categories.

The M-CHAT-R comprises 20 items. Responses are scored as follows: a response of "YES" to items 2, 5, and 12, or "NO" to all other items, is scored as 1 (failed item). The total score ranges from 0 to 20. During data processing and analysis, M-CHAT-R total scores were categorized into three risk levels:

(1) Low risk: 0–2 points.
(2) Moderate risk: 3–7 points.
(3) High risk: 8–20 points.

Children with 0–2 points were considered screen negative. Children with 3–7 points were considered screen positive (moderate risk) and proceeded to the M-CHAT-R/F follow-up interview within the project workflow. Children with 8–20 points were considered screen positive (high risk); consistent with the standard M-CHAT-R/F algorithm, these children were referred directly for diagnostic evaluation without requiring the follow-up interview.

M-CHAT-R/F follow-up interview

The M-CHAT-R/F is a structured follow-up interview administered for moderate-risk screen-positive cases to clarify at-risk responses. A follow-up result was classified as:

M-CHAT-R/F positive: the child fails two or more follow-up items (i.e., total follow-up score ≥ 2).

M-CHAT-R/F negative: the child fails fewer than two follow-up items (i.e., total follow-up score < 2).

For clarity, throughout this manuscript, "M-CHAT-R" refers to the initial 20-item checklist and its total score/risk level, whereas "M-CHAT-R/F" refers specifically to the follow-up interview outcome (positive vs. negative) when applied.

In this program, the M-CHAT-R/F was used as a workflow tool to reduce false positives and support triage for follow-up and referral; DSM-5–based onsite clinical evaluation served as the reference standard outcome for analyses.

Socio-Demographic Questionnaire

A structured study questionnaire was used to collect sociodemographic information on the child and family. Collected variables included age, sex, parental age at the time of birth, occupation, family medical history, socioeconomic characteristics, and other relevant information. Information was obtained through face-to-face caregiver interviews.

Procedure

This study was conducted in six main steps: planning and approvals, pilot testing, training of personnel, preschool-based screening, diagnostic evaluation, and ML model development (Fig. 1). The project was implemented in two phases: a pilot phase and a provincial-wide implementation phase.

Step 1: Research planning and approvals.

Can Tho University of Medicine and Pharmacy developed the study plan in collaboration with the Ca Mau Department of Education and Training, Ca Mau Department of Health, and Ca Mau Department of Science and Technology. Ca Mau Obstetrics and Pediatrics Hospital coordinated clinical logistics and referral pathways. The Department of Education and Training issued official communications to education offices in all nine districts/cities to coordinate implementation at participating preschools/kindergartens. This multi-sector structure enabled standardized training,

data collection, onsite clinical assessment, and linkage to local services.

Step 2: Implementation of the pilot phase.

The pilot phase was conducted in an initial sample (n = 497) to evaluate feasibility of the screening workflow, refine training materials, and finalize data collection procedures prior to province-wide rollout.

Step 3: Training of screening personnel.

Nine district-level training classes were delivered by the research clinical team (pediatricians and mental health clinicians). Teachers nominated by each preschool/kindergarten specifically those directly responsible for age-eligible classrooms, received standardized instruction on ASD, administration and scoring of the Vietnamese Ministry of Health–issued M-CHAT-R, observation of child behaviors, and procedures for collecting caregiver input. Standardized manuals and examples were provided to improve consistency across sites.

Step 4: Provincial-wide screening with the M-CHAT-R and project workflow for follow-up.

All toddlers aged 18–36 months enrolled in participating preschools/kindergartens were screened using the M-CHAT-R. Teachers completed the checklist based on daily observation and structured discussion with parents/guardians. Because caregiver literacy varies in the study setting, teachers acted as trained facilitators to explain items using standardized instructions and record responses. To reduce rater fatigue and improve data quality, each teacher screened no more than 10 children per day; teachers received project support for time spent on screening activities. Children with moderate-risk M-CHAT-R scores (3–7) received the M-CHAT-R/F follow-up interview within the project workflow, whereas children with high-risk scores (8–20) were referred directly for diagnostic evaluation without follow-up, consistent with the standard M-CHAT-R/F algorithm. When information was unclear at the time of screening, it was documented and revisited during onsite clinical visits. Regardless of screening results, all children in the screened cohort subsequently underwent onsite DSM-5 evaluation by the mobile clinical team.

Step 5: Diagnostic evaluation according to DSM-5.

A mobile clinical team (n = 10 clinicians; pediatricians and mental health/psychiatry clinicians) conducted onsite clinical evaluations at participating preschools/kindergartens. All screened children in the analytic cohort (n = 3,639) received an onsite DSM-5–based diagnostic evaluation, enabling classification of ASD vs. non-ASD across the full screened cohort. Diagnostic determination was made according to DSM-5 criteria based on direct clinical assessment and caregiver interview. During onsite visits, clinicians reviewed screening forms, clarified unclear responses with caregivers, and reviewed obstetric/perinatal histories with verification where feasible using available child health booklets/records. Items that could not be reliably confirmed were recorded as missing/unknown rather than inferred to minimize misclassification.

Step 6: Development and evaluation of the machine-learning models.

The ML target outcome was DSM-5 diagnostic status (ASD vs. non-ASD). Candidate predictors included M-CHAT-R–derived features (e.g., total score and/or risk category) together with selected obstetric–perinatal variables available at the time of screening. Data were randomly split into training (75%) and test (25%) sets. Given the marked class imbalance (~2% ASD), SMOTE was applied to the training set only. Six algorithms (Random Forest, XGBoost, support vector machine, decision tree, Naïve Bayes, and k-nearest neighbors) were trained and evaluated on the held-out test set using accuracy, recall (sensitivity), precision (positive predictive value), F1-score, and the area under the receiver operating characteristic curve (AUC). The models were developed to support scalable risk stratification in preschool settings—particularly where sustained delivery of M-CHAT-R/F follow-up interviews by healthcare personnel may be difficult—by prioritizing children for intensified follow-up and specialist assessment. In addition to the hold-out evaluation, we generated cohort-level, out-of-fold predictions using stratified k-fold cross-validation (with SMOTE applied within each training fold only) to construct an unbiased confusion matrix for the full screened cohort. The held-out test set results were used for primary model comparison (Tables 3 and 5), while out-of-fold predictions from stratified k-fold cross-validation were used to summarize cohort-level classification counts for the full sample (Table 4).

Operational Definitions

Target Variable

The target variable for the ML models was ASD diagnostic status, determined by clinician assessment according to DSM-5 criteria (1 = ASD; 0 = non-ASD). The "M-CHAT-R-only" baseline model used the M-CHAT-R total score (0–20) as the sole predictor; it was trained and evaluated under the same data split (75/25) and performance metrics as all other models to ensure fair comparison.

Candidate Predictors

Predictors were defined a priori based on clinical relevance and availability at screening. Child sex was coded as male/female. Child age was categorized as 18≤24 vs. 24–36 months. Maternal age at delivery was categorized as ≥ 35

vs. <35 years. Obstetric–perinatal variables were coded as binary (yes/no): medically assisted delivery, abnormal labor, preterm birth, perinatal asphyxia, neonatal jaundice, and neonatal seizures. M-CHAT-R information was represented as (i) total score (0–20) and/or (ii) categorical risk level (low 0–2, moderate 3–7, high 8–20). When the follow-up interview was used in the program workflow, its outcome (M-CHAT-R/F positive vs. negative) was treated as a separate screening-stage result and was not assumed to be equivalent to the M-CHAT-R total score.

Operational Definitions

Operational definitions. Obstetric–perinatal variables were defined a priori using caregiver report and, where available, documentation in child health booklets/records. Preterm birth was defined as delivery before 37 completed weeks of gestation. Medically assisted delivery referred to delivery requiring medical intervention, including operative vaginal delivery (vacuum/forceps) or cesarean section. Abnormal labor referred to labor complicated by clinically documented events such as prolonged/obstructed labor, fetal distress, or need for labor augmentation/assisted delivery, as reported by caregivers and/or recorded in birth documentation. Perinatal asphyxia was defined as a history of birth asphyxia requiring resuscitation and/or recorded low Apgar score at 5 min (when available in records); when documentation was unavailable, caregiver report of "birth asphyxia/need for resuscitation" was used and otherwise recorded as unknown/missing.

Table 1 Characteristics of children (n=3639)

Variables	Total (n=3639)	Autism (n=75)	Non-autistic (n=3564)	p-value
Sample size (%)	3639 (100)	75 (2.0)	3564 (98.0)	
Sex				0.505
Male	1801 (49.5)	40 (53.3)	1761 (49.4)	
Female	1838 (50.5)	35 (46.7)	1803 (50.6)	
Age				0.530
18≤24 months	206 (5.7)	3 (4.0)	203 (5.7)	
24–36 months	3433 (94.3)	72 (96.0)	3361 (94.3)	
Maternal age				0.161
<35 years	3074 (84.5)	59 (78.7)	3015 (84.6)	
≥ 35 years	565 (15.5)	16 (21.3)	549 (15.4)	
Medically assisted delivery				<0.001
Yes	928 (25.5)	39 (52.0)	889 (24.9)	
No	2711 (74.5)	36 (48.0)	2675 (75.1)	
Abnormal labor				0.005
Yes	144 (4.0)	12 (16.0)	132 (3.7)	
No	3495 (96.0)	63 (84.0)	3432 (96.3)	
Preterm birth				0.021
Yes	188 (5.2)	11 (14.7)	177 (5.0)	
No	3451 (94.8)	64 (85.3)	3387 (95.0)	
Perinatal asphyxia				0.042
Yes	28 (0.8)	5 (6.7)	23 (0.6)	
No	3611 (99.2)	70 (93.3)	3541 (99.4)	
Neonatal jaundice				0.556
Yes	145 (4.0)	2 (2.7)	143 (4.0)	
No	3494 (96.0)	73 (97.3)	3421 (96.0)	
Seizure				0.739
Yes	35 (1.0)	1 (1.3)	34 (1.0)	
No	3604 (99.0)	74 (98.7)	3530 (99.0)	
M-CHAT-R				<0.001
Positive	235 (6.4)	74 (98.7)	161 (4.5)	

Statistical Analysis

Categorical variables were presented as frequencies and percentages. The chi-squared test was used to compare categorical variables. Statistical analyses and data visualization were conducted using R software, version 4.4.1 (R Foundation for Statistical Computing, Vienna, Austria). The main R packages used for model fitting and performance evaluation included: caret, randomForest, rpart, e1071, recipes, dplyr, themis, xgboost, klaR, and pROC. A p-value less than 0.05 was considered statistically significant.

Results

Study Characteristics

From January to September 2022, the study included a total of 3,639 children aged 18 to 36 months attending preschools and kindergartens in Ca Mau province. The demographic characteristics of the study participants are summarized in Table 1. Using DSM-5 classification as the reference standard for the full screened cohort, M-CHAT-R screening positivity (score ≥ 3) showed a sensitivity of 98.67% (74/75) and a specificity of 95.48% (3403/3564). The positive predictive value was 31.49% (74/235) and the negative predictive value was 99.97% (3403/3404), consistent with a screening strategy prioritizing sensitivity. The study identified 75 children diagnosed with ASD, accounting for 2.0% of the screened preschool-enrolled cohort. Most children were in the 24 - 36-month age group (94.3%), while 5.7% were aged 18≤24 months. The gender distribution in the overall sample was nearly equal, with 49.5% male and 50.5% female. Several factors showed statistically significant differences between the ASD and non-ASD groups ($p < 0.05$), including medically assisted delivery (52.0% in the ASD group vs. 24.9% in the non-ASD group), abnormal labor (16.0% vs. 3.7%), preterm birth (14.7% vs. 5.0%), and perinatal asphyxia (6.7% vs. 0.6%). Consistent with these indices, 98.7% of children with ASD screened positive on the M-CHAT-R compared with 4.5% of children without ASD ($p < 0.001$). Other variables, including child's

sex, maternal age, neonatal jaundice, and history of seizure, did not show statistically significant differences between the two groups (Table 1).

Within the group of children diagnosed with ASD (n=75), the majority were aged 24–36 months (96.0%), and 53.3% were male. A notable 21.3% of the children had mothers aged 35 years or older. Regarding obstetric factors, 52.0% had medically assisted delivery, 16.0% experienced abnormal labor, 14.7% were born prematurely, and 6.7% had a history of perinatal asphyxia. Neonatal jaundice and a history of seizures were recorded in 2.7% and 1.3% of cases, respectively. In terms of screening results, 98.7% of these children had a positive M-CHAT-R screening result, whereas only 1.3% had a negative result.

In the non-ASD group (n=3,564), 94.3% of the children were aged 24–36 months, and males constituted 49.4% of this group. The proportion of mothers aged 35 or older was 15.4%. Regarding obstetric factors, 24.9% had medically assisted delivery, 3.7% had abnormal labor, 5.0% were born prematurely, and 0.6% experienced perinatal asphyxia. Neonatal jaundice and a history of seizures were reported in 4.0% and 1.0% of the cases, respectively. The positive screening rate in this group was only 4.5%, with the majority (95.5%) having a negative result. Screening positivity was substantially higher among children with ASD than among those without ASD.

Table 2 Internal consistency of the 20 items of the M-CHAT-R

Item-total statistics (n=497)		
	Corrected item-total correlation	Cronbach's alpha if item deleted
Item 1	0.537	0.856
Item 2	0.409	0.859
Item 3	0.521	0.855
Item 4	0.125	0.870
Item 5	0.164	0.876
Item 6	0.448	0.858
Item 7	0.572	0.854
Item 8	0.714	0.848
Item 9	0.642	0.849
Item 10	0.613	0.852
Item 11	0.630	0.854
Item 12	0.308	0.874
Item 13	0.456	0.859
Item 14	0.646	0.851
Item 15	0.674	0.852
Item 16	0.481	0.856
Item 17	0.728	0.845
Item 18	0.594	0.853
Item 19	0.461	0.857
Item 20	0.167	0.865

Assessment of the Cronbach's Alpha of the M-CHAT-R Scale in Southern Vietnam

As part of the project preparation, a pilot phase was conducted in 2021 in Ca Mau City (n = 497) to assess the internal consistency of the Vietnamese M-CHAT-R prior to province-wide implementation. Cronbach's alpha for the 20-item M-CHAT-R was 0.863, indicating good internal consistency. Most items showed acceptable corrected item-to-total correlations. Item deletion did not materially change the overall internal consistency, and all 20 items were retained for the province-wide implementation to preserve content coverage.

Item 12 had the lowest corrected item-to-total correlation (r = 0.308); when removed, Cronbach's alpha increased to 0.874, suggesting that Item 12 may contribute less to internal consistency in this sample. Items 4, 5, and 20 also showed relatively low corrected item-to-total correlations (0.125, 0.164, and 0.167, respectively), warranting further review in future validation work (Table 2).

Machine Learning Model

We developed machine learning (ML) models to support early ASD risk stratification in young children. Because ASD cases were rare in the sample (~2%), the dataset was highly imbalanced; therefore, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the training set only prior to model fitting, while the held-out test set remained unchanged. The full dataset was randomly split into a training set (n = 2,729; 75%) and a testing set (n = 910; 25%). Overall, the training set contained 53/75 (70.7%) children with ASD and the testing set contained 22/75 (29.3%) children with ASD. The ML models were trained using predictors available at screening, including demographic variables (e.g., child age, maternal age), obstetric–perinatal variables, and caregiver-reported clinical history (e.g., family history of neurodevelopmental disorders, multiple birth, gestational complications, neonatal conditions). The M-CHAT-R total score was treated as a discrete quantitative variable. Model performance metrics-including accuracy, recall, precision, F1-score, and area under the receiver operating characteristic curve (AUROC)-were computed and compared across six ML algorithms (Table 3).

Table 3 compares the predictive performance of six ML algorithms trained with SMOTE on the training set and evaluated on the held-out test set. Overall, the top-performing models achieved high discrimination (AUC 0.97–0.982 for the best-performing algorithms). We selected the Random Forest + SMOTE model for subsequent analyses because it showed strong overall discrimination and offered a stable

Table 3 Predictive performance of six ML algorithms (SMOTE applied to training set only)

Model	Accuracy	Recall	Precision	F1-score	AUC
Decision Tree + SMOTE	0.957	1.0	0.319	0.484	0.978
Random Forest + SMOTE	0.964	0.773	0.333	0.466	0.981
XGBoost + SMOTE	0.963	0.864	0.339	0.487	0.982
SVM + SMOTE	0.962	0.636	0.292	0.407	0.97
Naive Bayes + SMOTE	0.978	0.045	0.25	0.077	0.981
KNN + SMOTE	0.94	0.727	0.211	0.327	0.899

Table 4 Confusion matrix based on out-of-fold predictions from stratified k-fold cross-validation (n = 3639)

	Predicted: ASD	Predicted: No ASD
Actual: ASD	TP = 58	FN = 17
Actual: No ASD	FP = 117	TN = 3447

Table 5 Incremental value of feature reduction and logic-informed feature engineering (Random forest framework)

Metric	M-CHAT-*R*-only model	Full model	Reduced model	Refined model
Precision	0.305	0.333	0.333	0.351
Recall	0.947	0.773	0.864	0.95
F1-score	0.461	0.466	0.481	0.513
AUC	0.955	0.981	0.983	0.983

framework for feature reduction and rule-informed refinement (Table 3).

Table 4 presents the confusion matrix based on out-of-fold predictions from stratified k-fold cross-validation in the full analytic cohort. Of 75 children meeting DSM-5 criteria for ASD, the model correctly identified 58 (true positives) and missed 17 (false negatives), yielding a recall (sensitivity) of 0.773. Among 175 children predicted as ASD, 58 were true positives and 117 were false positives, corresponding to a precision (positive predictive value) of 0.333. Despite this moderate precision-reflecting a higher false-positive rate typical of screening-oriented workflows—the model maintained high overall accuracy (0.964) and excellent discrimination (AUC = 0.981), supporting its utility for prioritizing follow-up and specialist assessment rather than replacing diagnostic evaluation.

Table 5 summarizes the impact of model simplification and feature engineering within the Random Forest framework. The full model showed precision = 0.333, recall = 0.773, F1-score = 0.466, and AUC = 0.981. Reducing predictors to an 8-variable set improved recall to 0.864 while maintaining precision at 0.333, with AUC increasing to 0.983 and F1-score to 0.481, indicating that a parsimonious model preserved (and in this dataset improved) discrimination. Finally, a refined model incorporating two engineered predictors-(i) M-CHAT-R risk level (derived from total score; low/moderate/high) and (ii) a composite biological risk index derived from preterm birth, perinatal asphyxia, and medically assisted delivery-further increased recall to 0.95 with a modest gain in precision (0.351) and F1-score (0.513), while maintaining AUC at 0.983. Together, these findings suggest that a simpler, logic-informed feature representation can enhance sensitivity and support scalable triage in preschool-based screening pathways, acknowledging that confirmatory clinical assessment remains essential given the moderate precision.

Discussion

Principal Findings

This study evaluates the internal consistency of the Vietnamese M-CHAT-R and the performance of ML models for ASD risk stratification within a preschool-based screening pathway in Ca Mau province, Vietnam. In this screened preschool-enrolled cohort, 2.0% of children met DSM-5 criteria for ASD; this proportion should not be interpreted as population prevalence. The study revealed a nearly balanced gender distribution among children with ASD, demonstrating a male-to-female ratio of 1.1:1 (53.3% male vs. 46.7% female). The majority of identified cases (96%) fell within the 24 - 36 month age range. Several obstetric–perinatal variables differed between children with and without ASD (medically assisted delivery, abnormal labor, preterm birth, and perinatal asphyxia); these should be interpreted as non-causal correlates rather than causal obstetric–perinatal predictors.

The Vietnamese M-CHAT-R showed good internal consistency (Cronbach's alpha = 0.863). Corrected item–total correlations were generally acceptable; however, Items 4, 5, and 20 were low and Item 12 was borderline, indicating that these items may contribute less to internal consistency in this sample. Deleting Item 12 would increase alpha to 0.874, but all 20 items were retained to preserve content coverage. For machine learning, the Random Forest + SMOTE model achieved strong discrimination (AUC = 0.981) with moderate sensitivity (recall = 0.773) but relatively low precision (0.333), consistent with a screening-oriented approach prioritizing case detection. Importantly, the ML models are intended to support triage/risk ranking in preschool workflows (where sustained M-CHAT-R/F follow-up may be difficult), rather than to replace diagnostic assessment.

Compare With Previous Studies

The proportion of children meeting DSM-5 criteria for ASD in our screened cohort (2.0%) is comparable to the

most recent estimates in high-income countries, such as the United States (2.27%) reported by Maenner (Maenner et al., 2023). This proportion is also higher than earlier Vietnamese studies, such as that by Le Thi Vui (Thi Vui et al., 2021; 2020), which reported a prevalence of 0.76% using DSM-IV criteria; this difference may reflect variation in diagnostic criteria (DSM-IV vs. DSM-5), screening/ascertainment procedures, and study sampling frames. Differences likely reflect diagnostic criteria, sampling frame, and ascertainment within a structured screening pathway.

Moreover, a meta-analysis showed that 0.77% of children globally are diagnosed with ASD (Issac et al., 2025). This highlights substantial heterogeneity in ASD estimates across settings and methodologies, and suggests that observed proportions in screened samples may depend on the screening pathway, diagnostic ascertainment, and service availability in each context. The male-to-female ratio in our ASD cohort was approximately 1.1:1 (53.3% male vs. 46.7% female). This pattern may vary across studies due to differences in sampling frames, case ascertainment, and age distribution. The majority of children in our study who were diagnosed with ASD were aged 24–36 months, accounting for 96% of cases. This finding is consistent with a multi-center study in China involving over 1,500 children with ASD, which reported a median age at diagnosis of 29 months, with 76.3% of cases falling within this age range (Long et al., 2022).

Obstetric–perinatal variables that differed between children with and without ASD in our cohort included medically assisted delivery, abnormal labor, preterm birth, and perinatal asphyxia; these findings should be interpreted as correlates/predictors rather than causal obstetric–perinatal predictors, given potential confounding and inter-correlation among perinatal events. Our results are broadly consistent with evidence from Vietnam (Vui et al., 2023) and pooled analyses reporting associations between adverse obstetric/perinatal conditions and ASD (Wang et al., 2017). Prior studies have also reported higher odds of ASD-related outcomes among children born preterm compared with full-term peers (Joseph et al., 2017), (Leviton et al., 2018), (Agrawal et al., 2018). In large population-based data, Gregory et al. (2013) found that induction or augmentation of labor was associated with a small but statistically significant increase in ASD risk (OR 1.13–1.23), particularly in male children; proposed mechanisms include altered oxytocin signaling and transient fetal hypoxia during excessive uterine contractions, although these remain hypotheses and do not establish causality (Weisman et al., 2012).

In our cohort, abnormal labor and preterm birth were associated with ASD classification; however, these findings should be interpreted as correlates rather than evidence of causality, because perinatal events may co-occur and reflect shared underlying pathways or confounding. Biologically plausible mechanisms have been proposed linking adverse perinatal conditions to later neurodevelopmental vulnerability, including hypoxia–ischemia–related neuroinflammation and white-matter injury, which may affect myelination and cortical connectivity (Back, 2015). Preterm birth may also increase susceptibility to hypoxia–ischemia, inflammation, and oxidative stress during a period of immature brain development (Back, 2015). In addition, preterm birth is frequently linked to maternal intrauterine infection and inflammatory pathways that have been associated with adverse neurodevelopmental outcomes (Goldenberg et al., 2000, 2008; Nadeau et al., 2016). From a clinical perspective, these observations support the importance of regular neurodevelopmental surveillance in preterm infants during early childhood to facilitate timely identification and referral when developmental concerns arise.

Overall, our findings align with prior epidemiologic evidence, supporting inclusion of a small set of routinely documented perinatal variables for risk stratification in resource-constrained screening pathways. Accordingly, documenting these readily available obstetric–perinatal variables may help strengthen risk stratification in screening pathways and guide closer developmental surveillance and timely referral for children who screen positive or present with developmental concerns.

In our study, neonatal jaundice and history of seizure among children with ASD were uncommon (2.7% and 1.3%, respectively) and did not differ significantly from the non-ASD group ($p > 0.05$). This contrasts with the study by Vui et al. (2023), which identified neonatal jaundice and seizures as statistically significant predictors; differences in sample characteristics, outcome ascertainment, and operational definitions of neonatal conditions may partly explain this inconsistency (Vui et al., 2023).

In this study, the Vietnamese version of the M-CHAT-R demonstrated a Cronbach's alpha coefficient of 0.863, and most items achieved an acceptable threshold for corrected item–total correlations. These results are consistent with international trends: the M-CHAT-R/F has been standardized in large, low-risk populations, demonstrating stable psychometric properties, and its two-step procedure (questionnaire plus follow-up interview) helps maintain high sensitivity while improving specificity compared with the original version ADDIN EN.CITE (Robins et al., 2014). The M-CHAT-R proved effective in distinguishing children with a potential risk of ASD from those with typical development. This result is consistent with findings from previously published international studies. A study by Ariffin (2024) in Malaysia reported a Cronbach's alpha of 0.906 ADDIN EN.CITE (Ariffin et al., 2024). Similarly, a study by Coelho-Medeiros et al. (2019) in Chile reported a Cronbach's alpha

of 0.889 ADDIN EN.CITE (Coelho-Medeiros et al., 2019). A study by Tsai (2019) in Taiwan reported both test-retest reliability and internal consistency values greater than 0.8 ADDIN EN.CITE (Tsai et al., 2019).

In the item-level analysis, Item 12 ("sensitivity to noise"), Item 4 ("likes climbing"), and Item 5 ("unusual finger movements near eyes") of the Vietnamese version of the M-CHAT-R showed low item–total correlation coefficients. The underlying reasons may stem from the content characteristics, statistical factors, and cultural influences. In terms of content, Item 12 reflects auditory hypersensitivity- a manifestation of sensory processing disorder commonly observed in ASD, but it can be affected by living environment and exposure to noise. Children in rural areas or with limited exposure to loud noises may not exhibit pronounced reactions, leading to variability in responses when items are interpreted by caregivers and/or teachers in different home and classroom contexts. Items 4 and 5 involve motor or sensory- motor behaviors, which can also appear in typically developing children at certain developmental stages, thereby reducing their ability to discriminate between at-risk and non–at-risk groups. From a statistical perspective, both Item 12 and Item 5 are reverse-scored in the M-CHAT-R, which often results in statistical properties different from other items. In addition, cultural factors play an important role, as Vietnamese parents may observe and interpret sensory–motor behaviors differently from those in Western contexts where the original tool was developed- for example, the behavior "likes climbing" may sometimes be regarded as normal activeness rather than a sign of atypical development. The combination of these factors likely contributed to the lower item–total correlation coefficients of these three items in our study sample.

The consistency of results across studies from different countries suggests that this tool has broad applicability in various regions. This is contingent upon accurate translation, culturally and linguistically appropriate adaptations, and implementation by trained personnel. This finding underscores the potential for a standardized, yet adaptable, screening method for global use.

The role of the screener is also a critical factor that warrants careful consideration in future studies. Further research should aim to evaluate the influence of the individual screener on the screening outcomes and to ensure consistency across different application contexts. This will help to strengthen the reliability and generalizability of the screening protocol.

Because Cronbach's alpha reflects internal consistency of the 20-item M-CHAT-R questionnaire (not the follow-up interview), our alpha estimate (0.863) was higher than that reported in some prior studies, although other work has observed increases in internal consistency predictorss after follow-up procedures that clarify ambiguous responses. Conversely, a number of studies have reported a low Cronbach's alpha for the M-CHAT-R scale, which subsequently increased after the foll/estimateow-up procedure was implemented. This suggests that the reliability of the M-CHAT-R can be highly variable and may be influenced by methodological factors and the specific population being studied.

A large-scale study in Turkey reported a Cronbach's alpha of 0.67 during the initial screening phase, increasing to 0.82 after the follow-up interview stage (Oner & Munir, 2020). However, the value increased to 0.82 after the follow-up interview stage (M-CHAT-F) ADDIN EN.CITE (Oner & Munir, 2020). A study by Robins (2014) conducted on over 16,000 children in the United States reported an initial Cronbach's alpha coefficient of 0.63 for the M-CHAT-R. However, this value increased to 0.79 after a follow-up assessment, highlighting the crucial role of information verification in enhancing the measurement's reliability ADDIN EN.CITE (Robins et al., 2014).

This demonstrates the crucial role of the follow-up procedure in clarifying inaccurate or ambiguous answers. This process, in turn, enhances the internal consistency and overall reliability of the measurement scale. This difference may be attributed to factors such as language variations, study implementation methods, respondent characteristics, and the quality of evaluator training. These factors may have contributed to our study achieving a higher Cronbach's alpha score, even without the inclusion of a follow-up step.

Our reliability value falls within the range reported by regional adaptation studies ($\alpha \approx 0.86$–0.91) and is consistent with pooled evidence indicating optimal performance of the M-CHAT-R when the two-step procedure with follow-up interview is applied. Therefore, we recommend retaining the original 20-item structure while conducting longitudinal studies and refining language or cultural examples for certain sensory–motor items to enhance discriminative capacity while preserving content validity.

Our findings indicate that Random Forest combined with SMOTE provided the best overall performance among the evaluated algorithms in this dataset, with high discrimination and screening-oriented sensitivity. Random Forest + SMOTE achieved excellent discrimination (AUC 0.981) with moderate sensitivity (recall 0.773) but modest precision (0.333), reflecting the expected sensitivity–PPV tradeoff in low-prevalence screening. Our findings are consistent with the study by Wingfield et al. (2020), as both research efforts identified the Random Forest model, combined with the SMOTE technique, as the optimal approach for community-based screening of ASD. In the study by Wingfield et al. (2020), this model achieved an AUC of 0.98 and a recall of up to 96% (Wingfield et al., 2020). Our findings suggest that the Random Forest+SMOTE model is an effective solution

for binary classification problems with imbalanced datasets. This is particularly relevant in the field of ASD screening research, where the prevalence of children with ASD constitutes a very small proportion of the overall population. Despite using different screening tools and input data, the consistency and high performance across these studies confirm that the Random Forest + SMOTE model is a reliable choice for community-based ASD screening. The selection of the Random Forest model also allows for an evaluation of the contribution of each feature. This helps in identifying the strongest predictive factors and lays the groundwork for streamlining the questionnaire if necessary. In contrast, Wall (2012) utilized ML to build an alternating decision tree model based on the ADI-R - The input data for this model was sourced from the Autism Genetic Research Exchange archive, which contains data from a repository in Boston..

The alternating decision tree model, which contained only 7 items, was able to achieve highly accurate diagnostic results. This represents a 93.3% reduction in the number of items compared to the full ADI-R. The model was shown to be effective in diagnosing autism in children aged 13 to 48 months, with an impressive accuracy of 99.97% (Wall et al., 2012). A study by Aldrees et al. (2024) reported that XGBoost 2.0 was the most effective predictive model. That model achieved an accuracy of 99.29%, recall of 99%, F1-score of 99%, and an almost perfect AUC ADDIN EN.CITE (Aldrees et al., 2024). In addition to XGBoost 2.0, other models such as Support Vector Machine (SVM) and Logistic Regression also demonstrated high levels of accuracy and sensitivity, both exceeding 98%. These differences may be attributed to variations in study context, sample size, and feature selection techniques, which can influence model performance by identifying the most relevant predictors, reducing noise, and minimizing computational complexity. This comparison highlights that there is no single best ML model that can be universally applied to all situations.

The choice of an appropriate model should be based on the application goals, the nature of the input data, and the practical implementation requirements of a specific region or country. In the context of our research, which aimed to develop a feasible, effective, and easily deployable screening tool for community use, the Random Forest + SMOTE model demonstrated a strong balance between accuracy, sensitivity, and practical applicability.

In our comparison, XGBoost + SMOTE showed slightly higher recall/F1 (recall 0.864; F1 0.487; AUC 0.982), whereas Random Forest + SMOTE offered comparable discrimination (AUC 0.981) with simpler deployment and greater interpretability. We selected Random Forest because it provides greater interpretability, requires fewer computational resources, and is easier to implement in primary healthcare. These advantages make it particularly suitable for community-based ASD screening in Vietnam, ensuring a balance between accuracy, transparency, and practicality. Although the precision of our Random Forest + SMOTE model was relatively low, it still outperformed similar work such as Pan (2025) ADDIN EN.CITE (Pan et al., 2025). Moreover, the reduction in the number of variables while maintaining stable F1-score and AUC values demonstrates the principle of "simple yet effective," supporting feasibility in real-world deployment at the primary care level. By contrast, the SVM + undersampling model in Pan's study yielded a PPV of only 10.5%, highlighting how variations in population characteristics, model choice, and imbalance-handling strategies can substantially influence results. In our study, the use of obstetric–perinatal predictors and M-CHAT-R scores - while useful - also contributed to some false positives, as these factors are associated with ASD but not fully specific. This limitation is expected in screening contexts, where sensitivity is prioritized over precision. Consequently, despite its modest precision, the Random Forest + SMOTE model remains acceptable for our objective of capturing as many at-risk children as possible in the community.

The Random Forest model combined with SMOTE was selected for its high classification performance, achieving an F1-score of 0.466 and an AUC of 0.983. After feature reduction, all key metrics remained stable, demonstrating that the model preserved both stability and generalizability despite fewer input variables, consistent with the principle of simplifying models while maintaining accuracy. The refined logic-based model, developed from clinically significant variables supported by pathophysiological evidence, further improved Recall (0.95–1.00) while maintaining moderate Precision (0.328–0.351), an F1-score of 0.488, and an AUC of approximately 0.98, thereby enhancing the ability to detect potential ASD cases while balancing true detection and minimizing false alarms. In community-based screening, prioritizing high Recall is particularly appropriate for developmental disorders such as ASD, where missing high-risk children may result in lost opportunities for early intervention. Although Precision remained moderate, this limitation is acceptable in initial screening prior to specialized assessments. Moreover, building the refined model on clinically meaningful predictors enhances transparency and interpretability, aligning with current recommendations that emphasize explainable models in clinical decision support systems.

The current model is suitable for use as an initial community-based screening tool, followed by clinical assessment and the M-CHAT-R/F interview to reduce false positives. Multicenter studies with large sample sizes and real-world implementation evaluations are needed to optimize the classification threshold, adapt input variables to cultural

contexts, and enhance positive predictive value before widespread application.

Implication

Findings from this preschool-based screening cohort underscore the need to strengthen early ASD screening and diagnostic capacity in Vietnam, particularly in under-resourced settings. Our results highlight the importance of allocating resources for scalable early screening pathways and timely diagnostic evaluation, especially in rural and underserved areas where specialist services are limited. The observed associations between selected obstetric–perinatal variables and ASD diagnosis should be interpreted as non-causal predictors; however, they may be useful for risk stratification to prioritize follow-up in resource-constrained screening programs. These findings support integrating early ASD screening into routine child health services and preschool settings, with clear referral pathways for confirmatory assessment when screening is positive. Complementary efforts such as caregiver education on early developmental warning signs and standardized follow-up procedures may improve the effectiveness of screening programs and facilitate earlier access to intervention. In this context, an ML-assisted triage model that leverages M-CHAT-R results plus routinely available perinatal information may help prioritize referrals and optimize use of limited specialist resources.

Further Work

Although this preschool-based cohort supports the internal consistency of the Vietnamese M-CHAT-R and suggests that ML-assisted risk stratification may help prioritize follow-up and referral in resource-constrained settings, additional evidence is needed before broader implementation. Future multicenter, prospective studies across urban, rural, and remote regions of Vietnam should externally validate the ML models and the screening pathway, reporting calibration, decision thresholds, and real-world predictive value. In routine programs where universal diagnostic assessment is not feasible, these studies should document follow-up completion and diagnostic attendance and, where feasible, evaluate a subset of screen-negative children to strengthen key psychometric estimates (e.g., sensitivity, specificity, PPV/NPV) for both the two-step M-CHAT-R/F pathway and the ML-assisted workflow while minimizing partial verification bias. Finally, implementation research should assess acceptability, training requirements for teacher-administered screening, integration with health services, cost-effectiveness, and the impact of ML-enabled triage on time-to-diagnosis and access to early intervention.

Strengths and Limitations

This study leveraged a large preschool-based screening program (n = 3639) to evaluate the internal consistency of the Vietnamese M-CHAT-R and to test whether ML-assisted risk stratification can support scalable triage in a resource-limited setting. Using DSM-5 classification as the reference outcome within the screened cohort, we observed the expected screening trade-off: very high sensitivity with only moderate PPV, underscoring that the tool is intended to prioritize follow-up rather than replace diagnosis. Within the Random Forest framework, performance remained stable after feature reduction, and a logic-informed refinement (M-CHAT-R risk level plus a composite biological risk index) further increased sensitivity while simplifying inputs—features that enhance feasibility for community implementation.

These findings should be interpreted within the study design: the 2.0% figure reflects a screened, preschool-enrolled cohort and should not be generalized as population prevalence, and obstetric–perinatal variables should be framed as non-causal correlates rather than independent causal obstetric–perinatal predictors. External validation across diverse Vietnamese regions and prospective implementation studies are needed to assess calibration, decision thresholds, data quality for perinatal variables, and real-world impact on referral efficiency and time to diagnosis.

Conclusion

This study shows that the Vietnamese M-CHAT-R has good internal consistency (Cronbach's alpha = 0.863), supporting its use for community-based early autism screening in preschool settings. In this screened cohort, ML models combining M-CHAT-R information with selected obstetric–perinatal predictors achieved high discrimination (AUC up to 0.983) and high sensitivity, although precision remained modest—consistent with a screening-oriented approach prioritizing case detection and follow-up. To our knowledge, this is the first study in Vietnam to develop and evaluate an ML-assisted screening risk-stratification model that integrates M-CHAT-R data with obstetric–perinatal variables. However, the cross-sectional design, single-province setting, and lack of external validation limit generalizability. Future multicenter studies with prospective implementation and external validation are needed to assess calibration, optimize decision thresholds, and confirm real-world utility for referral and early intervention pathways.

Acknowledgements To complete this study, we would like to express our sincere gratitude to Can Tho University of Medicine and Pharmacy and Can Tho Pediatric Hospital for their invaluable support. We also

extend our heartfelt thanks to the pediatric patients and their families who participated in this research.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by all authors. The first draft of the manuscript was written by T.V.V., H.H.V.L, and T.N.N.L. and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. Conceptualization: P.M.N. and T.V.V.; Methodology: P.M.N. and T.V.V.; Formal analysis and investigation: D.N.N., D.M.T., and T.V.N.; Writing - original draft preparation: all authors; Writing - review and editing: all authors; Funding acquisition: none; Resources: P.M.N. and T.V.V.; Supervision: P.M.N.

Funding This research received no external funding.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical Approval The research protocol received approval from the Ethics Committee of Can Tho University of Medicine and Pharmacy (Approval No. 261/PCT-HĐĐĐ, dated 28/8/2020).

References

Agrawal, S., Rao, S. C., Bulsara, M. K., & Patole, S. K. (2018). Prevalence of autism spectrum disorder in preterm infants: A meta-analysis. *Pediatrics*. https://doi.org/10.1542/peds.2018-0134

Aldrees, A., Ojo, S., Wanliss, J., Umer, M., Khan, M. A., Alabdullah, B., & Innab, N. (2024). Data-centric automated approach to predict autism spectrum disorder based on selective features and explainable artificial intelligence. *Frontiers in Computational Neuroscience, 18*, Article 1489463. https://doi.org/10.3389/fncom.2024.1489463

Ariffin, R. A., Ismail, J., Abd Rahman, F. N., Wan Ismail, W. S., Ahmad, N., Abdul Ghafar, A., & Kamal Nor, N. (2024). Malay translation and validation of modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Frontiers in Pediatrics, 12*, Article 1384292. https://doi.org/10.3389/fped.2024.1384292

Autism, Developmental Disabilities Monitoring Network Surveillance Year, Principal, I., Centers for Disease, C., & Prevention (2007). Prevalence of autism spectrum disorders--autism and developmental disabilities monitoring network, six sites, united States, 2000. *MMWR Surveill Summ*, *56*(1), 1–11.

Autism, Developmental Disabilities Monitoring Network Surveillance Year, Principal, I., Centers for Disease, C., & Prevention (2009). Prevalence of autism spectrum disorders - Autism and developmental disabilities monitoring Network, united States, 2006. *MMWR Surveill Summ*, *58*(10), 1–20.

Back, S. A. (2015). Brain injury in the preterm infant: New horizons for pathogenesis and prevention. *Pediatric Neurology, 53*(3), 185–192. https://doi.org/10.1016/j.pediatrneurol.2015.04.006

Bone, D., Bishop, S. L., Black, M. P., Goodwin, M. S., Lord, C., & Narayanan, S. S. (2016). Use of machine learning to improve autism screening and diagnostic instruments: Effectiveness, efficiency, and multi-instrument fusion. *Journal of Child Psychology and Psychiatry, 57*(8), 927–937. https://doi.org/10.1111/jcpp.12559

Coelho-Medeiros, M. E., Bronstein, J., Aedo, K., Pereira, J. A., Arrano, V., Perez, C. A., & Bedregal, P. (2019). M-CHAT-R/F validation as a screening tool for early detection in children with autism spectrum disorder. *Revista Chilena de Pediatría, 90*(5), 492–499. https://doi.org/10.32641/rchped.v90i5.703

Dietz, P. M., Rose, C. E., McArthur, D., & Maenner, M. (2020). National and state estimates of adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 50*(12), 4258–4266. https://doi.org/10.1007/s10803-020-04494-4

Goldenberg, R. L., Hauth, J. C., & Andrews, W. W. (2000). Intrauterine infection and preterm delivery. *The New England Journal of Medicine, 342*(20), 1500–1507. https://doi.org/10.1056/NEJM200005183422007

Goldenberg, R. L., Culhane, J. F., Iams, J. D., & Romero, R. (2008). Epidemiology and causes of preterm birth. *Lancet, 371*(9606), 75–84. https://doi.org/10.1016/S0140-6736(08)60074-4

Gregory, S. G., Anthopolos, R., Osgood, C. E., Grotegut, C. A., & Miranda, M. L. (2013). Association of autism with induced or augmented childbirth in North Carolina Birth Record (1990–1998) and Education Research (1997–2007) databases. *JAMA Pediatrics*, *167*(10), 959–966. https://doi.org/10.1001/jamapediatrics.2013.2904

Ha, V. S., Whittaker, A., Whittaker, M., & Rodger, S. (2014). Living with autism spectrum disorder in Hanoi, Vietnam. *Social Science & Medicine, 120*, 278–285. https://doi.org/10.1016/j.socscimed.2014.09.038

Hoang, V. M., Le, T. V., Chu, T. T. Q., Le, B. N., Duong, M. D., Thanh, N. M., Tac Pham, V., Minas, H., & Bui, T. T. H. (2019). Prevalence of autism spectrum disorders and their relation to selected socio-demographic factors among children aged 18-30 months in northern Vietnam, 2017. *International Journal of Mental Health Systems, 13*, Article 29. https://doi.org/10.1186/s13033-019-0285-8

Issac, A., Halemani, K., Shetty, A., Thimmappa, L., Vijay, V. R., Koni, K., & Kapoor, V. (2025). The global prevalence of autism spectrum disorder in children: A systematic review and meta-analysis. *Osong Public Health and Research Perspectives, 16*(1), 3–27. https://doi.org/10.24171/j.phrp.2024.0286

Joseph, R. M., O'Shea, T. M., Allred, E. N., Heeren, T., Hirtz, D., Paneth, N., Leviton, A., Kuban, K. C. K., & Kuban, K. C. (2017). Prevalence and associated features of autism spectrum disorder in extremely low gestational age newborns at age 10 years. *Autism Research, 10*(2), 224–232. https://doi.org/10.1002/aur.1644

Khowaja, M. K., Hazzard, A. P., & Robins, D. L. (2015). Sociodemographic barriers to early detection of autism: Screening and evaluation using the M-CHAT, M-CHAT-R, and follow-up. *Journal of Autism and Developmental Disorders, 45*(6), 1797–1808. https://doi.org/10.1007/s10803-014-2339-8

Leviton, A., Joseph, R. M., Allred, E. N., Fichorova, R. N., O'Shea, T. M., Kuban, K. K. C., & Dammann, O. (2018). The risk of neurodevelopmental disorders at age 10years associated with blood concentrations of interleukins 4 and 10 during the first postnatal month of children born extremely preterm. *Cytokine*, *110*, 181–188. https://doi.org/10.1016/j.cyto.2018.05.004

Long, D., Yang, T., Chen, J., Dai, Y., Chen, L., Jia, F., & Li, T. (2022). Age of diagnosis and demographic factors associated with autism spectrum disorders in Chinese children: A Multi-Center survey. *Neuropsychiatric Disease and Treatment*, *18*, 3055–3065. https://doi.org/10.2147/NDT.S374840

Maenner, M. J., Warren, Z., Williams, A. R., Amoakohene, E., Bakian, A. V., Bilder, D. A., Durkin, M. S., Fitzgerald, R. T., Furnier, S. M., Hughes, M. M., Ladd-Acosta, C. M., McArthur, D., Pas, E. T., Salinas, A., Vehorn, A., Williams, S., Esler, A., Grzybowski, A., Hall-Lande, J., … Shaw, K. A. (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. Surveillance Summaries, 72*(2), 1–14. https://doi.org/10.15585/mmwr.ss7202a1

Nadeau, H. C., Subramaniam, A., & Andrews, W. W. (2016). Infection and preterm birth. *Seminars in Fetal & Neonatal Medicine, 21*(2), 100–105. https://doi.org/10.1016/j.siny.2015.12.008

Nguyen, P. M., Tran, T., & Tran, T. V. (2021). Clinical characteristics and associated socio-demographic factors of autism spectrum disorder in Vietnamese children. *Current Pediatric Research, 25*(1), 308–312.

Oner, O., & Munir, K. M. (2020). Modified checklist for autism in toddlers revised (MCHAT-R/F) in an urban metropolitan sample of young children in Turkey. *Journal of Autism and Developmental Disorders*, *50*(9), 3312–3319. https://doi.org/10.1007/s10803-019-04160-4

Pan, N., Chen, L., Wu, B., Chen, F., Chen, J., Huang, S., Guo, C., Wu, J., Wang, Y., Chen, X., Yang, S., Jing, J., Weng, X., Lin, L., Liang, J., & Wang, X. (2025). Developing a simplified measure to predict the risk of autism spectrum disorders: Abbreviating the M-CHAT-R using a machine learning approach in China. *Psychiatry Research, 344*, Article 116353. https://doi.org/10.1016/j.psychres.2025.116353

Qiu, S., Lu, Y., Li, Y., Shi, J., Cui, H., Gu, Y., Li, Y., Zhong, W., Zhu, X., Liu, Y., Cheng, Y., Liu, Y., & Qiao, Y. (2019). Prevalence of autism spectrum disorder in Asia: A systematic review and meta-analysis. *Psychiatry Research*, *284*, 112679. https://doi.org/10.1016/j.psychres.2019.112679

Robins, D. L., Casagrande, K., Barton, M., Chen, C. M., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the modified checklist for Autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics, 133*(1), 37–45. https://doi.org/10.1542/peds.2013-1813

Shaw, K. A., Williams, S., Patrick, M. E., Valencia-Prado, M., Durkin, M. S., Howerton, E. M., Ladd-Acosta, C. M., Pas, E. T., Bakian, A. V., Bartholomew, P., Nieves-Muñoz, N., Sidwell, K., Alford, A., Bilder, D. A., DiRienzo, M., Fitzgerald, R. T., Furnier, S. M., Hudson, A. E., Pokoski, O. M., … Maenner, M. J. (2025). Prevalence and early identification of autism spectrum disorder among children aged 4 and 8 years - Autism and Developmental Disabilities Monitoring Network, 16 Sites, United States, 2022. *MMWR. Surveillance Summaries, 74*(2), 1–22. https://doi.org/10.15585/mmwr.ss7402a1

Sun, X., Allison, C., Matthews, F. E., et al. (2013). Prevalence of autism in mainland China, Hong Kong and Taiwan: A systematic review and meta-analysis. *Molecular Autism*, *4*, 7. https://doi.org/10.1186/2040-2392-4-7

Thi Vui, L., Duc, D. M., Thuy Quynh, N., Giang, N. T. H., Mai, V. T. T., Ha, B. T. T., & Van Minh, H. (2021). Early screening and diagnosis of autism spectrum disorders in Vietnam: A population-based cross-sectional survey. *Journal of Public Health Research*. https://doi.org/10.4081/jphr.2021.2460

Tsai, J. M., Lu, L., Jeng, S. F., Cheong, P. L., Gau, S. S., Huang, Y. H., & Wu, Y. T. (2019). Validation of the modified checklist for autism in toddlers, revised with follow-up in Taiwanese toddlers. *Research in Developmental Disabilities, 85*, 205–216. https://doi.org/10.1016/j.ridd.2018.11.011

Van Thi, P. M., Nguyen, D. T., Lu, Q. C., Ngo, M. H., Le, Dien Minh Thai, & Luu, T. N. N. (2025). Prevalence and key perinatal risk factors of autism spectrum disorder among toddlers in Ca Mau province, Vietnam. *Pediatric Medicine*. https://doi.org/10.21037/pm-25-63

Vui, L. T. (2020). *Epidemiology of autism spectrum disorder in children aged 18-30 months and barriers to accessing autism spectrum disorder diagnosis and intervention services in Vietnam, 2017-2019. (PhD)*. Ha Noi University of Public Health.

Vui, L. T., Duc, D. M., Quynh, C. T. T., Tuan, D. K., Huong, N. M., Thanh, N. T. M., Hung, N. M., Minh, H. V., & Ha, B. T. T. (2023). Ante-, peri-, and neonatal factors associated with autism spectrum disorders in Vietnam: A population-based cross-sectional survey. *Iranian Journal of Public Health, 52*(5), 950–959. https://doi.org/10.18502/ijph.v52i5.12711

Wall, D. P., Dally, R., Luyster, R., Jung, J. Y., & Deluca, T. F. (2012). Use of artificial intelligence to shorten the behavioral diagnosis of autism. *PLoS One, 7*(8), Article e43855. https://doi.org/10.1371/journal.pone.0043855

Wang, C., Geng, H., Liu, W., & Zhang, G. (2017). Prenatal, perinatal, and postnatal factors associated with autism: A meta-analysis. *Medicine (Baltimore), 96*(18), Article e6696. https://doi.org/10.1097/MD.0000000000006696

Weisman, O., Zagoory-Sharon, O., & Feldman, R. (2012). Oxytocin administration to parent enhances infant physiological and behavioral readiness for social engagement. *Biological Psychiatry*, *72*(12), 982–989. https://doi.org/10.1016/j.biopsych.2012.06.011

Wieckowski, A. T., Williams, L. N., Rando, J., Lyall, K., & Robins, D. L. (2023). Sensitivity and specificity of the modified checklist for autism in toddlers (Original and Revised): A systematic review and meta-analysis. *JAMA Pediatrics, 177*(4), 373–383. https://doi.org/10.1001/jamapediatrics.2022.5975

Wingfield, B., Miller, S., Yogarajah, P., Kerr, D., Gardiner, B., Seneviratne, S., & Coleman, S. (2020). A predictive model for paediatric autism screening. *Health Informatics Journal, 26*(4), 2538–2553. https://doi.org/10.1177/1460458219887823