



Kaggle Toddler Dataset: A publicly available dataset containing 1,050 records of toddlers aged 18-36 months. It includes 18 attributes, such as AQ-10 scores and parental observations, and is balanced between ASD-positive and ASD-negative cases.

UCI Child Dataset: Sourced from the UCI Machine Learning Repository, this dataset has 292 records of children between 4 and 11 years old. It features 21 attributes, including AQ-10 responses, gender, and ethnicity. After cleaning, 242 instances were used.

UCI Adolescent Dataset

UCI Adult Dataset

Real-World Child Dataset: A proprietary dataset collected anonymously from mainstream and autism schools in Bangladesh, comprising 250 samples. It includes AQ-10 responses, demographic information, and behavioral observations recorded by professionals.[not publicly available but dataset can be made available on request], labels based on clinical diagnosis

Saudi Arabia Toddler Dataset:

Polish QCHAT-25 Dataset: Clinical diagnosis labels

Create a table in which columns should be dataset name, dataset link, no of instances, age group, brief description about attributes, whether labels are based on screening cutoff or clinical diagnosis for each of the dataset above

Below is a consolidated table for the datasets listed. Notes:

- Links are included where these datasets are commonly hosted or described publicly. For proprietary/on-request datasets, I've indicated that status.
- "Screening cutoff" refers to labels derived from AQ-10/Q-CHAT scoring rules; "Clinical diagnosis" refers to labels obtained from clinician-confirmed diagnosis (e.g., via ADOS/ADI-R or medical records).

Dataset name	Dataset link	No. of instances	Age group	Brief description of attributes	Labels source
Kaggle Toddler Dataset	Kaggle "Autism screening data for toddlers" (public benchmark)	~1,050	18–36 months (toddlers)	~18 attributes: AQ-10 toddler items (parent-report), parental observations, demographics (e.g., age, gender, ethnicity), screening result	Screening cutoff (AQ-10-based)
UCI Child Dataset	UCI ML Repository "Autism Screening Children"	292 (242 after cleaning, per common usage in studies)	4–11 years (children)	~21 attributes: AQ-10 child items, gender, ethnicity, family/medical history flags, screening result	Screening cutoff (AQ-10-based)
UCI Adolescent Dataset	UCI ML Repository "Autism Screening Adolescents"	~1,000+ (varies by mirrored versions; core UCI set is smaller)	Adolescents (typically 12–17 years)	AQ-10 adolescent items, demographics (gender, age, ethnicity), family/medical flags, screening result	Screening cutoff (AQ-10-based)
UCI Adult Dataset	UCI ML Repository "Autism Screening Adult"	~700–1,100 (commonly cited ~704)	Adults (16+ years)	AQ-10 adult items (A1–A10), demographics (age, gender, ethnicity), social/medical history flags, screening outcome	Screening cutoff (AQ-10-based)
Real-World Child Dataset (Bangladesh)	Proprietary; available on request from study team	250	Children (school-age)	AQ-10 responses, demographics, professional behavioral observations collected in mainstream and autism schools	Clinical diagnosis
Saudi Arabia Toddler Dataset	Research dataset referenced in multicountry QCHAT-10 studies (contact authors)	Varies by study (hundreds typical)	18–36 months (toddlers)	Q-CHAT/QCHAT-10 items (parent-report), demographics; sometimes includes healthcare history	Often clinical diagnosis in curated research cohorts; confirm per release

Dataset name	Dataset link	No. of instances	Age group	Brief description of attributes	Labels source
Polish QCHAT-25 Dataset	Public repository (e.g., Mendeley Data: "Q-CHAT scores of Polish toddlers...") and related publications	~250 (commonly cited 252)	Toddlers	Q-CHAT-25 item scores (parent-report), demographics; used in validation and ML studies	Clinical diagnosis

Important clarifications and tips:

- UCI and Kaggle AQ-10 datasets are primarily screening datasets; the "ASD" label typically indicates screening-positive based on cutoffs, not a clinician-confirmed diagnosis. This is appropriate for screening-model work but not for training a diagnostic classifier without caveats.
- The Polish QCHAT-25 and many curated toddler datasets used in multicountry studies report labels tied to clinical diagnosis; always check the dataset's readme/data dictionary for the exact label provenance.
- For the Saudi Arabia toddler dataset, multiple research teams have published on Q-CHAT/QCHAT-10 cohorts; availability and label source can vary by study. If clinical labels are essential, contact the corresponding authors of the specific study that used clinician-confirmed diagnoses.
- Instance counts shown for UCI/Kaggle are the typical figures cited in literature; mirrored or cleaned versions may differ slightly (e.g., after deduplication or missing-data filtering).

If desired, I can tailor direct retrieval steps and exact links for each public dataset, and draft request emails for the on-request datasets to obtain clinically labeled versions.