Game-Based Diagnostics for Autism Spectrum Disorder: Integrating Machine Learning with Behavioral and Cognitive Metrics

Ronav Kataria[1, *], Abhiram Gangdhar Gunda[2, *], Victor Yeung[3, *]

[1]*American High School, Fremont, CA, USA*
[2]*American High School, Fremont, CA, USA*
[2]*American High School, Fremont, CA, USA*
**Authors contributed equally*

---***---

Abstract - This study introduces a novel approach to simplifying the diagnosis process of autism by integrating game-based assessments and machine learning technology. We designed an interactive test suite in an easy-to-use, virtual environment for people with autism spectrum disorder (ASD). The system quantifies multiple behavioural metrics: reaction time, task-switching capabilities, and sensory response patterns, while integrating standardized clinical diagnostic data. Our methodology involved 6 male participants aged 15-17: 3 properly diagnosed with autism by official medical professionals (classified as ASD level 1 under testing standards) and 3 neurotypical participants exhibiting normal behavior.Data from gameplay sessions were systematically logged, analyzed, and converted into quantifiable features associated with ASD markers. We evaluated 25 different machine learning classifiers, with AdaBoostClassifier achieving perfect accuracy (100%), while BaggingClassifier and DecisionTreeClassifier provided efficient alternatives at 80% accuracy. Results confirmed established literature on executive function deficits and atypical sensory processing in ASD, though our findings on reaction time contradicted previous research by showing faster responses in ASD participants. This game-based automated diagnostic system demonstrates potential to augment conventional clinical evaluations by offering a low-cost, accessible, and engaging avenue for early ASD detection. While designed as a complementary tool rather than a replacement for formal clinical assessment, our approach addresses significant barriers in current diagnostic practices including resource intensity, time constraints, and specialist dependency. Future improvements will focus on increasing sample diversity, expanding test modules, and refining the user interface to enhance accessibility and reliability across broader populations.

***Key Words*:** Autism Spectrum Disorder, Machine Learning, Game based diagnostics, Cognitive Metrics

1. Introduction

Autism spectrum disorder (ASD) is a heterogeneous neurodevelopmental disorder, characterized by impairments in social communication and interaction, along with limited and repetitive behavioral patterns (American Psychiatric Association, 2013). Current diagnostic procedures are primarily reliant on standardized clinical evaluations—i.e., the Autism Diagnostic Observation Schedule (ADOS-2) and the Childhood Autism Rating Scale (CARS-2)—administered by trained examiners (Lord et al., 2000; Schopler et al., 2010). Though these tests are well-established, they are time-consuming, resource-intensive, and specialist-dependent, which may hinder early diagnosis and intervention.

Emerging digital health technologies provide promising solutions for conventional clinical routines, especially with regard to greater accessibility and lower costs. To this purpose, digital platforms with accessible means of interaction, such as through gaming analogues, have the potential for rapid adoption to complement the crucial, but less accessible, methods of clinical diagnosis. Through their interactivity and high engagement, as well as their ability to collect real-time data in a flexible and standardized way, these platforms demonstrate strong use cases for many potential autism diagnosis and treatment pathways.

By incorporating tasks that assess cognitive and sensory abilities, these sites can prompt clinically meaningful behaviors with minimal stress on participants—particularly important for populations like children with autism, where standard clinical settings can exacerbate anxiety or decrease compliance. With the developing consensus regarding the involvement of atypical reaction times, executive functioning deficits, and sensory sensitivities in Autism Spectrum Disorder (ASD) (Ozonoff et al., 1991; Hill, 2004; Leekam et al., 2007), the incorporation of these measures in a game-based paradigm presents an unprecedented opportunity for early screening and the attainment of ancillary diagnostic data. Coupled with supervised machine learning algorithms that have been trained on behavioral and clinically diagnosed datasets, these platforms have the potential to recognize subtle patterns of ASD and provide real-time feedback.

This paper presents the design, development, and evaluation of a novel game-based diagnostic system for ASD.

2. Material and Methods

A. Participants and Clinical Data Collection

Participants are recruited from collaborating organizations specializing in autism and other clinical settings. Prior to testing, informed consent and assent is obtained from the participants and their parents/guardians, and necessary demographic and clinical data is collected. Clinical diagnoses are made using standardized

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 07 | July - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

evaluation measures, such as the Autism Diagnostic Observation Schedule (ADOS-2) (Lord et al., 2000), the Childhood Autism Rating Scale (CARS-2) (Schopler et al., 2010), and the Autism Spectrum Quotient (AQ) (Baron-Cohen et al., 2001). These standard tools serve as the gold standard for subsequent machine learning studies.

B. Game Design and User Interface

The diagnostic tool comprises several interactive games designed to capture specific behavioral and cognitive markers associated with autism:

- Reaction Time Test: Participants are prompted to respond to a visual cue ("GO!") after a randomized delay. This test is used to measure motor response times, an aspect which has been associated with atypical cognitive processing in autism (Ozonoff et al., 1991; Happé & Frith, 2006).

- Task Switching Test: This test evaluates cognitive flexibility by prompting participants to provide the opposite response to a displayed word (YES/NO). The ability to switch responses efficiently is linked to executive functioning deficits observed in autism (Hill, 2004).

- Sensory Test**:** A simulated sensory stimulus (a "loud noise") is used to gauge sensory sensitivity by recording the response latency when the participant indicates discomfort. Sensory processing differences are well-documented in autism (Leekam et al., 2007; Green et al., 2012).

The user interface is designed around themes that emphasize clarity and ease of navigation, with large buttons, clear instructions, and a visually streamlined color scheme to ensure a high degree of accessibility for autistic individuals.

C. Data Collection and Processing

Gameplay data—including reaction times, task-switching performance, and sensory response times—are automatically recorded in real time and linked to a unique, anonymized participant ID. The data collection pipeline involves the following steps:

- Data Logging: Each interaction during gameplay is time-stamped and stored in a secure database.
- Data Cleaning: Incomplete or inconsistent data entries are flagged and removed to maintain dataset integrity.
- Feature Extraction: Key performance metrics (e.g., mean reaction time, variability, error rates) are computed from the raw data.
- Normalization: Techniques such as Z-score standardization are applied to ensure comparability across participants.

This preprocessed dataset, combined with clinical diagnosis labels, forms the input for the machine learning model. Figure 1 shows the details of the dataset.

	clinical_diagnosis	reaction_time	task_switching_time	task_switching_errors	sensory_reaction_time
0	ASD	0.57	6.11	6	0.60
1	ASD	0.50	3.17	6	0.45
2	ASD	0.53	3.26	3	0.66
3	ASD	1.15	3.07	6	0.53
4	ASD	0.50	3.35	5	0.45
5	TD	0.47	4.07	4	0.36
6	TD	0.40	2.85	1	0.39
7	TD	0.35	3.45	0	0.39
8	TD	0.37	2.66	3	0.36
9	TD	0.39	2.30	2	0.37
10	ASD	1.42	4.70	0	0.52
11	ASD	0.48	3.45	0	0.50
12	ASD	0.47	3.04	1	0.42
13	ASD	0.43	5.92	0	0.44
14	ASD	0.49	3.31	0	1.42
15	ASD	0.43	5.12	0	0.34
16	ASD	0.42	2.91	0	0.35
17	ASD	0.35	2.77	2	0.30
18	ASD	0.38	2.61	0	0.32
19	ASD	0.38	3.34	0	0.31
20	TD	0.36	1.93	0	0.53
21	TD	0.42	2.06	0	0.42
22	TD	1.56	1.96	0	0.43
23	TD	0.58	3.30	0	0.43
24	TD	1.37	1.97	0	0.03

Fig -1: Dataset Info

D. Machine Learning and Diagnostic Modeling

The preprocessed behavioral data are used to train a supervised machine learning model aimed at classifying individuals along the autism spectrum. The model development process includes:

- Model Selection: Various algorithms, including Random Forests and Gradient Boosting Machines are evaluated to determine the best performance.
- Training and Validation: The model is trained on the input dataset with validation and scoring assigned to each compute model. Separate runs of the model are used to verify final accuracy statistics and ensure consistent results across different subject cases.
- Integration and Deployment: The final model is deployed within the diagnostic system to provide real-time feedback based on new gameplay data, thereby assisting clinicians with the diagnostic process.

E. Ethical Considerations and Data Security

All procedures align with ethical guidelines in research with vulnerable groups. Data is anonymized and securely stored according to GDPR and HIPAA guidelines. The diagnostic tool is intended to complement—not replace—usual clinical assessment to ensure professional clinical judgment retains its critical position in the diagnostic process.

F. Descriptive Statistics

After finalizing data collection and preprocessing, a total of 5 participants' game interactions (reaction times, task-switching metrics, and sensory response patterns) were analyzed. The sample included all male subjects in the Bay Area, ranging from ages 15 to 17. All the subjects who are part of the autism dataset were classified as ASD level 1 ("requiring support") under the ADOS-2 testing standard.

Preliminary inspection of the data revealed:

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 07 | July - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Reaction Time Test: The mean reaction times for the combined, neurotypical, and ASD datasets are 0.60 s ± 0.36 s, 0.66 s ± 0.47 s, and 0.58 s ± 0.31 s, respectively. The ranges for the combined, neurotypical, and ASD datasets are 1.21 s, 1.21 s, and 1.07 s, respectively. These findings contradict the existing literature, showing that the subjects with ASD have lower variability and faster response times, whereas the inverse is typically indicated in generalized studies (Ozonoff et al., 1991).

Task Switching Test: The average error rate for the combined, neurotypical, and ASD datasets are 26%, 17%, and 32% respectively. The error rate ranged from 100% to 0%. The mean task switching times for the combined, neurotypical, and ASD datasets are 3.34 s ± 1.12 s, 2.76 s ± 0.81 s , and 3.74 s ± 1.14 s respectively. The range for the task switching time for the combined, neurotypical, and ASD datasets are 4.18 s, 2.14 s, and 3.5 s respectively. These error rates correlate with existing literature indicating increased difficulty for individuals with ASD with response switching (Hill, 2004). Similar to the reaction time data, individuals with ASD performed better in terms of response speed, which although may contradict existing data, would be offset in this instance due to the substantially higher error rate.

Sensory Test: The mean response latencies for the combined, neurotypical, and ASD datasets are 0.45 s ± 0.23 s, 0.45 s ± 0.23 s, and 0.51 s ± 0.27 s, respectively. The ranges for the combined, neurotypical, and ASD datasets are 1.39 s, 1.39 s, and 1.12 s, respectively. This data tracks with the research consensus that individuals with ASD have greater variability and slower sensory response times (Leekam et al., 2007).

These results aligned with existing literature indicating variability and difficulty for task switching (Hill, 2004), and sensory response in individuals on the spectrum (Leekam et al., 2007). The reaction time data did not show the same variability and delay in response time that the general research consensus agrees upon (Ozonoff et al., 1991).

3. Results and Discussions

Our experiment compared 25 different machine learning classifiers on the same dataset as shown in Figure 2. The models showed varying performance across several metrics including accuracy, balanced accuracy, F1 score, and computation time.

AdaBoostClassifier demonstrated the best overall performance with perfect scores of 1.0 for accuracy, balanced accuracy, and F1 score. It also had a relatively fast computation time of 0.11158 seconds.

The ensemble methods generally performed well, with RandomForestClassifier and BaggingClassifier both achieving 0.80000 accuracy and F1 scores. BaggingClassifier was notably more efficient, completing in 0.04480 seconds compared to RandomForestClassifier's 0.16301 seconds.

DecisionTreeClassifier also achieved 0.80000 accuracy but had a slightly lower F1 score of 0.78095. However, it was the fastest high-performing model with a time of only 0.02506 seconds.

Several models including LinearDiscriminantAnalysis, RidgeClassifierCV, CalibratedClassifierCV, RidgeClassifier, ExtraTreesClassifier, and Perceptron all achieved moderate performance with 0.60000 accuracy and F1 scores.

The lowest performing models were the various SVC implementations, GaussianNB, QuadraticDiscriminantAnalysis, LabelSpreading, and LabelPropagation, which all achieved only 0.20000 accuracy and F1 scores.

Table 1: Comparison of Models

	Accuracy	Balanced Accuracy	F1 Score	Time Taken
Model				
AdaBoostClassifier	1.00	1.00	1.00	0.12
RandomForestClassifier	0.80	0.83	0.80	0.18
BaggingClassifier	0.80	0.83	0.80	0.05
DecisionTreeClassifier	0.80	0.75	0.78	0.02
LinearDiscriminantAnalysis	0.60	0.58	0.60	0.05
RidgeClassifierCV	0.60	0.58	0.60	0.03
CalibratedClassifierCV	0.60	0.58	0.60	0.07
RidgeClassifier	0.60	0.58	0.60	0.04
ExtraTreesClassifier	0.60	0.58	0.60	0.13
Perceptron	0.60	0.58	0.60	0.02
LGBMClassifier	0.60	0.50	0.45	0.09
DummyClassifier	0.60	0.50	0.45	0.02
NearestCentroid	0.40	0.42	0.40	0.02
KNeighborsClassifier	0.40	0.33	0.34	0.03
LogisticRegression	0.40	0.33	0.34	0.03
PassiveAggressiveClassifier	0.40	0.33	0.34	0.03
ExtraTreeClassifier	0.40	0.33	0.34	0.02
BernoulliNB	0.40	0.33	0.34	0.08
SGDClassifier	0.40	0.33	0.34	0.02
LinearSVC	0.40	0.33	0.34	0.03
LabelSpreading	0.20	0.17	0.20	0.02
LabelPropagation	0.20	0.17	0.20	0.03
NuSVC	0.20	0.17	0.20	0.02
GaussianNB	0.20	0.17	0.20	0.02
QuadraticDiscriminantAnalysis	0.20	0.17	0.20	0.02
SVC	0.20	0.17	0.20	0.02

The findings hold established, as well as unexpected, trends regarding certain ASD behaviors. For instance, the Task Switching Test (error rates and switching times) and the Sensory Test (latency of response) are in accordance with much of the previous literature demonstrating both executive function and atypical sensory responses (Hill, 2004; Leekam et al., 2007); the Reaction Time Test demonstrated faster and less variable responses from those with ASD participants-an observation that sharply contrasts with previous studies, which defined slower or more variable motor responses within this population (Ozonoff et al., 1991). This contrast may indicate some unique behavioral characteristics in this subgroup.

From the perspective of classification, the highest accuracy (100%) was achieved by AdaBoostClassifier, although it was more computationally intensive than both BaggingClassifier (80% accuracy) and Decision Tree Classifier (80% accuracy). These results show that adaptive boosting methods take in distinguishing features from small datasets, while ensemble and tree-based methods could be performed faster, with still effective results. All in all, this pilot study shows promise in game-based assessment and supervised learning to begin identifying markers for ASD and enhances the argument for behavior-computational model integration for early detection and intervention.

4. Conclusions

Clinical Relevance

The potential for real-time feedback within the gaming environment stands to make early detection more accessible, particularly in settings with limited clinical resources. Although the algorithms showed robust validation metrics, it is critical to emphasize that the system is intended to complement (not replace) formal clinical evaluations.

Limitations

Sample Size and Diversity:
The relatively small sample size used, along with the homogeneity of the subjects' demographics, leaves room for improvement in terms of cohort size and diversity. Such improvements would be necessary to generalize results and provide deeper understanding for refining the prediction model.

Lab vs. Naturalistic Settings: The controlled game environment may obscure some of the real-life complexities of patients with ASD, possibly skewing some of the data points and offsetting natural variance. While a real-life testing environment could provide more holistic and pertinent data, lacking standardization could adversely affect data reliability.

Our preliminary findings demonstrate the viability of using machine learning in conjunction with a game-based approach to help detect ASD characteristics early. Future research directions include:

Expanding participant demographics to test generalizability across different age groups and cultural contexts.

Integrating additional game modules to provide additional testing variables, such as cognition or language-based tasks.

Refining the user interface for improved accessibility and engagement.

In summary, combining interactive, user-friendly platforms with clinically validated metrics may speed up diagnosis, lower costs, and possibly enhance results by enabling earlier interventions.

References

1. American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). *American Psychiatric Publishing*.
2. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, and Parents of Children with Autism. *Journal of Autism and Developmental Disorders, 31*(1), 5–17.
3. Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies. *Developmental Neuropsychology, 27*(3), 403–424.
4. Fournier, K. A., Hass, C. J., Naik, S. K., Lodha, N., & Cauraugh, J. H. (2010). Motor coordination in autism spectrum disorders: A synthesis and meta-analysis. *Journal of Autism and Developmental Disorders, 40*(10), 1227–1240.
5. Green, S. A., Ben-Sasson, A., Soto, T. W., & Carter, A. S. (2012). Anxiety and sensory over-responsivity in toddlers with autism spectrum disorders: Bidirectional effects across time. *Journal of Autism and Developmental Disorders, 42*(6), 1112–1119.
6. Happé, F., & Frith, U. (2006). The Weak Central Coherence Account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 36*(1), 5–25.
7. Hill, E. L. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences, 8*(1), 26–32.
8. Leekam, S. R., Nieto, C., Libby, S. J., Wing, L., & Gould, J. (2007). Describing the sensory abnormalities of children and adults with autism. *Journal of Autism and Developmental Disorders, 37*(5), 894–910.
9. Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function in autism. *Journal of Child Psychology and Psychiatry, 32*(6), 913–931.
10. Pierce, K., Marinero, S., Hazin, R., McKenna, B., Barnes, C. C., & Malige, A. (2016). Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder subtype associated with increased symptom severity. *Biological Psychiatry, 79*(8), 657–666.
11. Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). *Childhood Autism Rating Scale* (2nd ed.). Western Psychological Services.

BIOGRAPHIES

Ronav Kataria is High School Student in American High School.

Abhiram Gangdhar Gunda is High School Student in American High School.

Victor Yeung is High School Student in American High School.