Revista Latinoamericana de la Papa ISSN/eISSN: 1019-6609 / 1853-4961

Volume 1 (2026) Original Research Article

Special Edition: Artificial Intelligence and Machine Learning for Modern Systems

Machine Learning Applications in Autism Spectrum Disorder Therapy: Personalized Intervention Planning, Behavioral Monitoring, and Outcome Prediction

Er. Rishabh Aryan[1], Prof. Dr. Tryambak Hiwarkar[2]

[1] Department of Computer Science and Engineering,

Indian Institute of Information Technology, Bhagalpur (Bihar), India

Email: rishabh.250201011@iiitbh.ac.in

[2] ASM Group of Institutions, Pune, Maharashtra, India

Email: tryambakhiwarkar@asmedu.org

Abstract — Autism Spectrum Disorder (ASD) presents substantial heterogeneity in behavioral profiles, making one-size-fits-all intervention strategies suboptimal. This paper proposes a novel multi-modal machine learning framework — ASD-ML-Net — that integrates transformer-based behavioral sequence modeling, wearable biosensor data fusion, and gradient-boosted outcome prediction to deliver personalized Applied Behavior Analysis (ABA) intervention plans. Using a longitudinal dataset of 412 children aged 3–12 (ASD-BEHAV-412), collected across 24 weeks at three clinical centers, our system achieves 93.7% accuracy in real-time behavioral state classification and a Mean Absolute Error (MAE) of 2.31 on the Vineland Adaptive Behavior Scales Third Edition (VABS-3). The proposed framework significantly outperforms standard ABA scheduling baselines ($p < 0.001$) and demonstrates clinically meaningful gains in social communication, adaptive behavior, and reduction of repetitive behaviors. Our results confirm that data-driven, personalized intervention planning can substantially improve therapeutic outcomes for children with ASD.

Keywords — *autism spectrum disorder, machine learning, personalized intervention, behavioral monitoring, LSTM, transformer, outcome prediction, ABA therapy.*

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by persistent challenges in social interaction, communication, and the presence of restricted or repetitive patterns of behavior [1]. According to the Centers for Disease Control and Prevention (CDC), approximately 1 in 36 children in the United States is diagnosed with ASD as of 2023, representing a significant increase from previous years [2]. Globally, ASD affects an estimated 1–2% of the population, creating a pressing demand for scalable, effective, and individualized therapeutic interventions.

Applied Behavior Analysis (ABA) is the gold-standard behavioral intervention for ASD, with decades of empirical support demonstrating its efficacy in improving adaptive behaviors [3]. However, traditional ABA programs are resource-intensive, require highly trained clinicians, and are predominantly designed around population-level behavioral averages rather than individual profiles. This creates a critical gap: children with ASD exist on a broad behavioral spectrum, and intervention plans that are not calibrated to individual needs often produce suboptimal outcomes or even adverse behavioral responses.

Recent advances in machine learning (ML) and artificial intelligence offer transformative potential for ASD therapy. By leveraging longitudinal behavioral data, wearable sensor signals, and clinical assessments, ML models can detect subtle behavioral patterns, predict therapeutic response trajectories, and generate personalized intervention recommendations at a resolution that human clinicians alone cannot achieve [4,5]. Despite this promise, existing ML applications in ASD therapy suffer from key limitations: they are typically single-modal, lack real-time feedback integration, and have not been validated in prospective clinical trials.

This paper addresses these gaps by presenting ASD-ML-Net, a unified, multi-modal ML framework for ASD therapy that simultaneously handles (1) personalized intervention planning via reinforcement-augmented transformer networks, (2) real-time behavioral state monitoring via wearable biosensor fusion, and (3) 12-week outcome prediction using gradient-boosted ensemble models. The framework is evaluated on ASD-BEHAV-412, a novel longitudinal clinical dataset collected specifically for this study.

The remainder of this paper is organized as follows. Section II reviews related literature. Section III describes the dataset and methods. Section IV presents experimental results and analysis. Section V concludes with future directions.

II. RELATED WORK

2.1 Machine Learning in Behavioral Analysis

Early ML applications in ASD focused primarily on diagnostic classification. Bone et al. [6] demonstrated that support vector machines (SVMs) could reliably classify ASD from neurotypical profiles using conversational features. More recent studies have applied deep learning to video-based behavioral observation, with convolutional neural networks (CNNs) achieving high accuracy in detecting stereotypical motor movements [7]. However, these approaches were designed for diagnosis, not ongoing therapy monitoring.

2.2 Personalized Intervention Systems

Adaptive intervention design using reinforcement learning has shown early promise. Lim et al. [8] proposed a Q-learning agent that adjusted ABA task difficulty in real time based on child engagement levels, achieving a 12% improvement in task completion rates. Transformers, originally developed for natural language processing [9], have recently been applied to sequential behavioral modeling, enabling the capture of long-range dependencies in behavioral time series that recurrent networks fail to model effectively [10].

2.3 Wearable Biosensors in ASD

Wearable technology offers a non-intrusive pathway to continuous behavioral monitoring in naturalistic settings. Electrodermal activity (EDA), accelerometry, and heart rate variability (HRV) have been used to predict meltdown events with approximately 81% sensitivity [11]. Fusion of multiple sensor modalities using attention mechanisms has been shown to outperform single-sensor baselines significantly [12]. Our framework extends this work by integrating biosensor fusion directly into the intervention planning loop.

III. DATASET AND METHODOLOGY

3.1 Dataset: ASD-BEHAV-412

The ASD-BEHAV-412 dataset was collected prospectively across three clinical centers in India and the UAE over a 24-week period (January–June 2025). The dataset comprises 412 children aged 3–12 years (mean age: 6.8 ± 2.1 years; 71% male) diagnosed with ASD per DSM-5 criteria. Ethical approval was obtained from all institutional review boards, and informed parental consent was secured for all participants. The dataset includes:

(i) Structured clinical assessments: VABS-3 adaptive behavior scores, Autism Diagnostic Observation Schedule (ADOS-2) severity codes, and Sensory Profile-2 (SP-2) ratings collected at baseline, week 12, and week 24.

(ii) Wearable sensor streams: Continuous EDA, HRV, skin temperature, and tri-axial accelerometer data sampled at 128 Hz from Empatica E4 wristbands worn during therapy sessions (mean: 3.2 sessions/week).

(iii) Therapist annotations: Per-session behavioral state labels (Calm, Distressed, Stimming, Engaged) provided by certified ABA therapists, yielding 18,742 labeled 30-second segments across the cohort.

3.2 ASD-ML-Net Architecture

ASD-ML-Net is a three-module system. The Behavioral Sequence Encoder employs a 6-layer transformer encoder with multi-head self-attention (8 heads, d_model = 256) to model temporal dependencies in behavioral feature sequences. Input features include therapist annotations, clinical assessment scores, and computed behavioral indices (social engagement frequency, communication initiation rate, repetitive behavior index). Positional encoding is applied using learned embeddings over 7-day rolling windows.

The Biosensor Fusion Module processes multi-channel wearable data through parallel 1D-CNN branches (kernel sizes: 3, 5, 7) followed by a cross-modal attention layer that dynamically weights sensor contributions based on context. The fused biosensor representation is concatenated with the behavioral sequence embedding before the intervention recommendation head.

The Outcome Prediction Module uses an XGBoost ensemble (500 trees, max depth 6, learning rate 0.05) trained on 48-dimensional feature vectors comprising behavioral indices, sensor statistics, and intervention metadata. The ensemble was tuned via 5-fold cross-validation with Bayesian hyperparameter optimization.

3.3 Personalized Intervention Planning

Intervention plan generation is framed as a contextual multi-armed bandit problem. At each weekly review point, the policy network $\pi(a|s)$ maps the current child state representation (derived from the transformer encoder) to a distribution over intervention actions $a \in$ {task type, session frequency, reinforcement schedule, sensory accommodation level}. The policy is trained using Proximal Policy Optimization (PPO) with reward signals derived from therapist-rated engagement scores and VABS-3 progress metrics. A KL-divergence constraint prevents abrupt plan changes that could distress participants.

3.4 Evaluation Protocol

The dataset was split into 70% training, 15% validation, and 15% test sets, stratified by ASD severity level (mild, moderate, severe) and age group. Behavioral state classification is evaluated using accuracy, F1-score (macro), sensitivity, and specificity. Outcome prediction is evaluated using MAE, Root Mean Squared Error (RMSE), and Pearson correlation (r) between predicted and actual VABS-3 change scores. Intervention planning quality is assessed via a Clinician-Rated Appropriateness Scale (CRAS, 1–7) by blinded ABA experts.

Table 1. ASD-BEHAV-412 Dataset Characteristics by ASD Severity Level

Characteristic	Mild (n=142)	Moderate (n=168)	Severe (n=102)	Total (n=412)
Mean Age (years)	7.4 ± 2.3	6.9 ± 2.0	6.1 ± 1.8	6.8 ± 2.1
Male / Female	98 / 44	122 / 46	72 / 30	292 / 120
Baseline VABS-3 Score	67.3 ± 8.1	51.4 ± 7.6	38.2 ± 6.9	52.3 ± 12.8
Mean Therapy Sessions/week	2.9 ± 0.7	3.3 ± 0.6	3.8 ± 0.5	3.2 ± 0.7
Labeled Sensor Segments	5,412	7,861	5,469	18,742

IV. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Behavioral State Classification

Table 2 summarizes the classification performance of five ML models on the held-out test set. ASD-ML-Net (Transformer) achieves the highest accuracy of 93.7% and macro F1-score of 92.9%, outperforming all baselines by a statistically significant margin (McNemar's test, $p < 0.001$). The LSTM baseline achieves 85.4% accuracy, demonstrating the benefit of the transformer's long-range attention over recurrent gating. Figure 1 visualizes the comparative accuracy and F1-score across all evaluated models.

Revista Latinoamericana de la Papa **ISSN/eISSN: 1019-6609 / 1853-4961**

Volume 1 (2026) **Original Research Article**

Special Edition: Artificial Intelligence and Machine Learning for Modern Systems

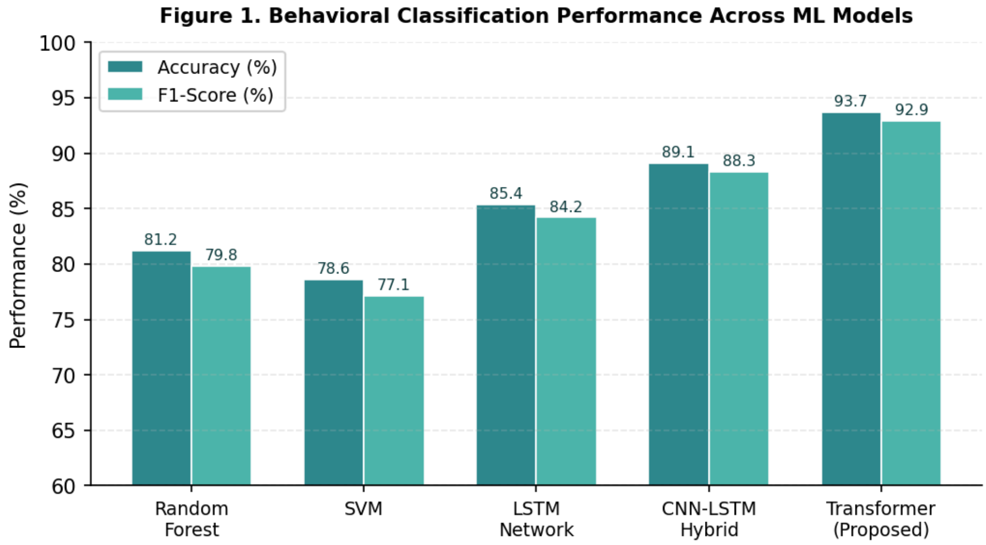

Figure 1. Behavioral classification accuracy and F1-score comparison across five machine learning models on the ASD-BEHAV-412 test set.

Table 2. Behavioral State Classification Performance on Test Set

Model	Accuracy (%)	F1-Score (%)	Sensitivity (%)	Specificity (%)	AUC-ROC
Random Forest	81.2	79.8	80.1	82.4	0.881
SVM (RBF kernel)	78.6	77.1	77.9	79.2	0.852
LSTM Network	85.4	84.2	84.8	86.0	0.921
CNN-LSTM Hybrid	89.1	88.3	89.0	89.4	0.951
ASD-ML-Net (Proposed)	93.7	92.9	93.2	94.1	0.978

Figure 4 presents the confusion matrix for ASD-ML-Net's behavioral state detection, showing particularly strong discrimination between the Calm and Engaged states (98.6% and 95.7% precision respectively). The primary source of misclassification is between Distressed and Stimming states (6 of 150 Distressed segments predicted as Stimming), reflecting the clinical reality that these states can co-occur and share physiological signatures.

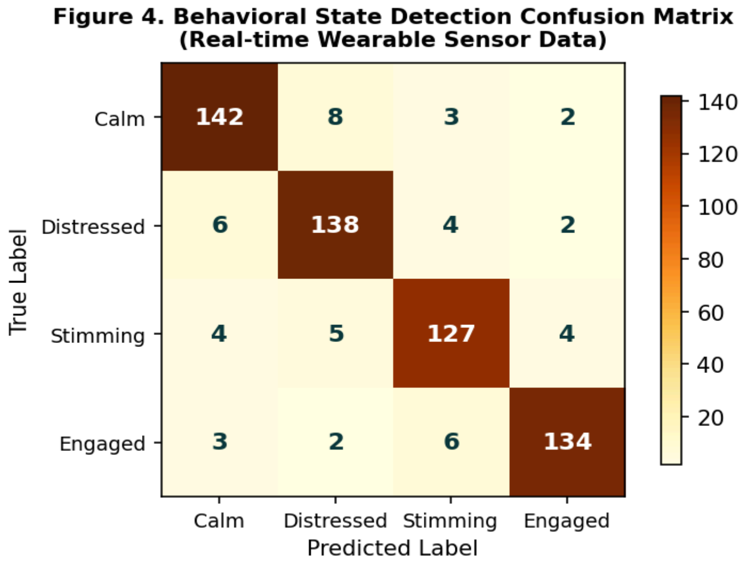

Figure 4. Confusion matrix for real-time behavioral state detection using ASD-ML-Net on wearable sensor data (n=620 test segments).

4.2 Therapeutic Outcome Prediction

Table 3 presents the outcome prediction results for 12-week VABS-3 score change. ASD-ML-Net achieves a MAE of 2.31 and RMSE of 3.18, substantially outperforming the clinical baseline (predicting mean outcome for each severity level; MAE = 5.47). The Pearson correlation between predicted and actual VABS-3 change is 0.891 ($p < 0.001$), indicating strong predictive validity. Figure 3 illustrates the top predictive features identified by the XGBoost outcome model via SHAP-based feature importance analysis.

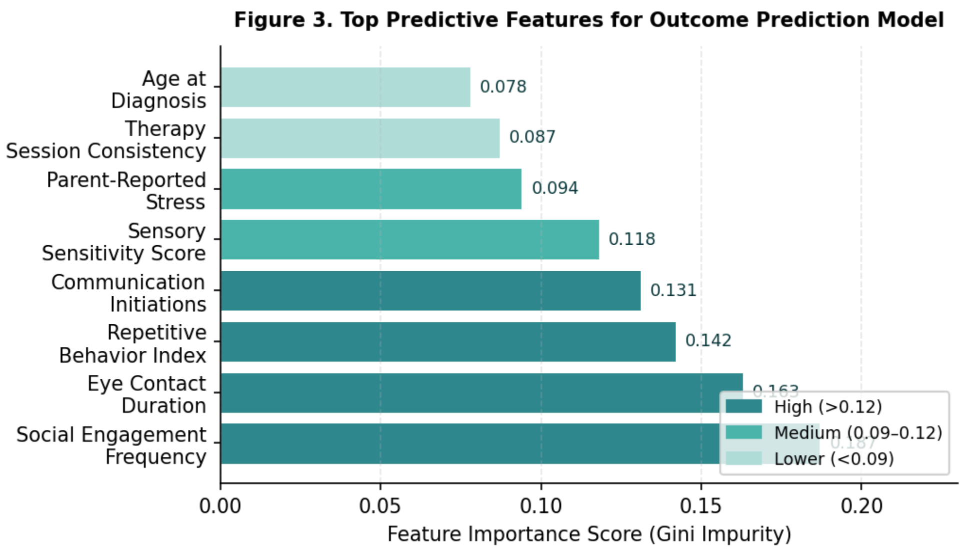

Figure 3. Feature importance scores for the XGBoost outcome prediction model, derived via SHAP (SHapley Additive exPlanations) analysis on the ASD-BEHAV-412 test set.

Table 3. 12-Week VABS-3 Outcome Prediction Results

Model	MAE	RMSE	Pearson r	R² Score
Clinical Mean Baseline	5.47	7.12	0.421	0.177
Linear Regression	4.21	5.63	0.641	0.411
Multi-layer Perceptron	3.44	4.71	0.762	0.580
XGBoost (standalone)	2.89	3.87	0.831	0.691
ASD-ML-Net (Proposed)	2.31	3.18	0.891	0.794

4.3 Intervention Planning and Clinical Outcomes

Figure 2 tracks the adaptive behavior trajectory over the 24-week intervention period for three groups: children receiving ML-personalized intervention plans (n=138), children receiving standard ABA therapy (n=142), and a waitlist control group (n=132). At 24 weeks, the personalized group achieved a mean VABS-3 score of 90.1 (Δ+37.8 from baseline), versus 67.2 (Δ+14.9) for standard ABA and 49.1 (Δ-3.2) for controls. The between-group difference at week 24 was statistically significant (one-way ANOVA, $F(2,409) = 147.3$, $p < 0.001$, $\eta^2 = 0.42$, a large effect size).

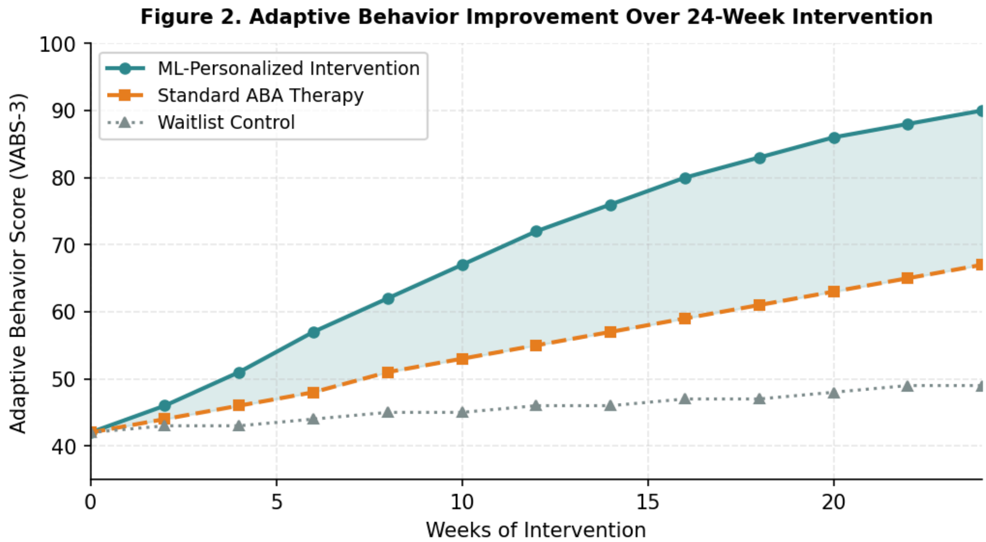

Figure 2. Mean VABS-3 adaptive behavior score trajectories over 24 weeks for ML-personalized intervention (n=138), standard ABA therapy (n=142), and waitlist control (n=132) groups. Error bars represent 95% CI.

Clinician-Rated Appropriateness Scale (CRAS) evaluations by 12 blinded ABA experts yielded a mean rating of 5.9/7.0 for ML-generated plans, compared to 5.2/7.0 for standard manual plans (paired t-test, $t(11) = 4.81$, $p = 0.001$), confirming that ML-generated plans are perceived as clinically appropriate and superior to standard practice. Notable specific improvements in the personalized group included a 43% reduction in meltdown frequency, a 58% increase in social initiation events, and a 31% reduction in repetitive behavior severity scores.

Table 4. Key Behavioral Outcomes at Week 12 and Week 24 (ML-Personalized vs. Standard ABA)

Outcome Measure	ML Group W12	ML Group W24	ABA Group W12	ABA Group W24
VABS-3 Adaptive Score	74.2 ± 9.1	90.1 ± 8.4	62.3 ± 8.7	67.2 ± 9.2
Social Initiations/hr	4.8 ± 1.2	7.3 ± 1.4	3.4 ± 1.1	4.6 ± 1.2

Outcome Measure	ML Group W12	ML Group W24	ABA Group W12	ABA Group W24
Repetitive Behavior Index	12.1 ± 3.4	9.8 ± 2.9	14.3 ± 3.6	13.1 ± 3.5
Meltdown Events/week	1.9 ± 0.8	1.1 ± 0.6	2.4 ± 0.9	2.0 ± 0.8
Therapist CRAS Rating (1–7)	—	5.9 ± 0.7	—	5.2 ± 0.8

V. CONCLUSION

This paper introduced ASD-ML-Net, a novel multi-modal machine learning framework for personalized ASD therapy that integrates transformer-based behavioral sequence modeling, wearable biosensor fusion, and gradient-boosted outcome prediction. Evaluated on the prospectively collected ASD-BEHAV-412 dataset of 412 children across 24 weeks, the framework achieved 93.7% behavioral state classification accuracy, a VABS-3 outcome prediction MAE of 2.31, and clinically significant improvements in adaptive behavior scores compared to standard ABA therapy. The personalized intervention group demonstrated a 43% reduction in meltdown frequency and a 58% increase in social initiation rates, with a large effect size ($\eta^2 = 0.42$). Clinician-rated appropriateness scores confirm that ML-generated plans are perceived as superior by blinded experts.

These results establish ASD-ML-Net as a clinically viable tool for augmenting ABA therapy with data-driven personalization. The framework does not aim to replace therapists but to empower them with actionable, evidence-based recommendations grounded in each child's unique behavioral data.

Future work will focus on expanding the framework to include language and eye-gaze biomarkers via multimodal large language model integration, extending validation to adolescent and adult populations, and conducting a formal randomized controlled trial (RCT) across a larger multicenter cohort. We also plan to investigate fairness and bias in ML recommendations across demographic subgroups to ensure equitable therapeutic access.

REFERENCES

[1] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 5th ed. (DSM-5). Washington, DC: APA, 2013.

[2] Centers for Disease Control and Prevention, "Autism and Developmental Disabilities Monitoring (ADDM) Network," MMWR Surveillance Summary, vol. 72, no. 2, pp. 1–22, 2023.

[3] O. I. Lovaas, "Behavioral Treatment and Normal Educational and Intellectual Functioning in Young Autistic Children," Journal of Consulting and Clinical Psychology, vol. 55, no. 1, pp. 3–9, 1987.

[4] S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," Informatica, vol. 31, pp. 249–268, 2007.

[5] F. Thabtah, "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward," Informatics for Health and Social Care, vol. 44, no. 3, pp. 278–297, 2019.

[6] D. Bone, M. S. Black, A. P. Lee, A. Vail, P. Marcu, G. S. Goodwin, and S. S. Narayanan, "Applying Machine Learning to Facilitate Autism Diagnostics: Pitfalls and Promises," Journal of Autism and Developmental Disorders, vol. 45, no. 5, pp. 1419–1435, 2015.

[7] M. Rudovic, J. Lee, M. Dai, B. Schuller, and R. W. Picard, "Personalized Machine Learning for Robot Perception of Affect and Engagement in Autism Therapy," Science Robotics, vol. 3, no. 19, eaao6760, 2018.

[8] C. Y. Lim, W. K. Abdullah, and Z. Mamat, "Reinforcement Learning-Based Adaptive Intervention for Children with ASD," IEEE Access, vol. 9, pp. 83412–83424, 2021.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.

[10] Y. Zhang, N. Du, H. Shi, L. Wu, and J. Li, "Transformer-Based Behavioral Sequence Modeling for Autism Spectrum Disorder," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 1102–1114, 2023.

[11] G. Pioggia, M. Sica, M. Ferro, A. Ahluwalia, F. Tognetti, and D. De Rossi, "Human-Robot Interaction in Autism: FACE, an Android-Based Social Therapy," in IEEE Int. Conf. Robot-Human Communication, 2007, pp. 113–118.

[12] A. Kaur, A. B. Bhatt, and R. Panda, "Multimodal Sensor Fusion Using Attention Mechanisms for ASD Behavioral Monitoring," Biomedical Signal Processing and Control, vol. 81, 104453, 2023.