A Novel Feature Optimization Approach for Accurate Autism Spectrum Disorder Prediction

K. Arpitha
Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology, Bengaluru, India | Visvesvaraya Technological University, Belagavi, Karnataka, India
arpitha.anilkumar.14@gmail.com (corresponding author)

N. Nandini
Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology, Bengaluru, India | Visvesvaraya Technological University, Belagavi, Karnataka, India
Nandu_8449@rediffmail.com

Received: 22 May 2025 | Revised: 13 June 2025 and 18 July 2025 | Accepted: 23 July 2025

ABSTRACT

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by challenges in communication, social interaction, and behavior. Timely detection is critical for early intervention, yet traditional diagnostic practices are often subjective, time-consuming, and prone to inaccuracies. This study addresses the limitations of existing Machine Learning (ML) and Deep Learning (DL) models in the prediction of ASD, particularly suboptimal performance caused by irrelevant or redundant features. The primary objective was to develop a robust and accurate ASD prediction framework using a novel feature optimization approach called CNN-ET-XGB. The proposed model integrates Convolutional Neural Networks (CNNs) to extract high-level abstract features from behavioral questionnaire data, Extra Trees (ET) to select the most relevant and discriminative features, and Extreme Gradient Boosting (XGB) for final classification. The model was evaluated on the UCI ASD children dataset, with a 50:50 train-test split, achieving 99.992% accuracy, outperforming existing models such as Random Forest, AlexNet CNN, and other approaches. The CNN-ET-XGB framework demonstrates significant potential for real-world applicability in clinical pre-screening tools and early ASD detection systems. Its layered feature optimization strategy enhances both accuracy and efficiency, providing a solution for assisting healthcare professionals in early ASD prediction.

Keywords-Autism Spectrum Disorder (ASD); machine learning; deep learning; feature selection; Convolutional Neural Networks (CNN); Extra Trees (ET); Extreme Gradient Boosting (XGB); ASD prediction

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by persistent deficits in social interaction, communication, and the presence of restricted or repetitive behaviors. Manifesting in early childhood, ASD can significantly affect a child's ability to function across various aspects of life [1]. One of the major challenges faced by children with ASD pertains to sensory processing. Many children show atypical responses to sensory stimuli, ranging from hypersensitivity to hyposensitivity to sounds, lights, textures, and smells, which trigger anxiety, discomfort, or behavioral outbursts in overwhelming environments [2]. In addition to sensory issues, behavioral dysregulation is a common concern. Repetitive behaviors, resistance to changes, and emotional dysregulation hinder participation in social, academic, and daily life settings [3]. These impairments have a profound effect on both the mental and physical well-being of affected children. Research has shown elevated instances of anxiety, depression, and social withdrawal in children with ASD, driven by their difficulties in interpreting and responding to social cues [4]. The World Health Organization (WHO) estimates that ASD affects approximately 1 in 100 children worldwide, with this prevalence continuing to rise due to improved diagnostic tools and awareness [5]. However, despite increasing global attention, many cases remain undiagnosed or diagnosed too late, preventing timely intervention [5].

Various screening and diagnostic tools have been developed to support the early diagnosis of ASD [6-9]. These tools mainly involve behavioral observations, structured assessments, and parent-reported questionnaires to identify ASD symptoms. Although these assessments have proven useful in clinical settings, they are largely manual and subjective and vulnerable to human error and interpretation bias. This can result in a delayed or inaccurate diagnosis, limiting the child's access to early intervention. To address these limitations, researchers have turned to Machine Learning (ML) and Deep Learning (DL) techniques, which offer the potential to improve diagnostic accuracy and efficiency [10]. However, existing ML/DL models typically use full feature sets without considering the relevance or redundancy of each feature, leading to inefficiencies and reduced accuracy [11-18]. Although recent approaches have attempted feature selection to improve performance, they still fall short of achieving optimal accuracy.

A comprehensive review of the current literature reveals various efforts toward predicting ASD. In [11], multiple ML classifiers were utilized with diverse feature scaling and selection methods, but struggled to maintain consistent high accuracy across datasets. In [12], accuracy was improved by merging datasets across age groups, but did not evaluate individual subsets. In [13], the Cuckoo Search Algorithm (CSA) was applied for feature reduction, but the accuracy remained suboptimal. In [14], traditional ML classifiers achieved high accuracy but ignored feature selection, which limits efficiency. Similarly, in [15], hyper-tuned classifiers were employed but omitted feature reduction strategies. In [16], Recursive Feature Elimination (RFE) was successfully applied to specific geographic datasets, although it lacked cross-dataset validation. In [17], an optimization-based approach was introduced for feature selection and classification, achieving impressive results for adults and adolescents but underperforming on the children dataset. In [18], a sophisticated multi-modal framework combined text, numerical, and facial features using DL models, but its reliance on facial datasets limits general applicability, particularly in low-resource environments. From this review a clear research gap emerges, as most existing ASD prediction models either rely on full feature sets without selection or employ feature selection methods that are not robust enough to capture the most informative data, especially for children. This highlights a critical need for an approach that not only selects relevant features efficiently but also integrates them into a powerful classification model for improved prediction accuracy.

In addition, current ASD diagnostic methods are often manual, subjective, and limited in scalability. Although ML/DL-based solutions have improved predictive accuracy, many still rely on unoptimized feature sets, leading to inefficiencies and reduced model performance, particularly for children's ASD datasets. Hence, there is a need for an accurate and automated approach that performs effective feature extraction and selection for ASD prediction in children. To bridge the identified research gap, this study presents a novel hybrid approach, combining Convolutional Neural Networks (CNNs) for feature extraction, Extra-Trees (ETs) for feature selection, and Extreme Gradient Boosting (XGB) for classification. The primary contributions of this study are as follows:

- Combines 1D-CNNs and ETs to enhance feature representation, leading to a more informative and reduced feature space.
- The CNN-ET captures sensory and behavioral feature correlations effectively from structured ASD datasets.
- Identifies the most relevant features, improving efficiency and accuracy.
- Trained on optimized features, XGB provided high accuracy while minimizing overfitting.

II. METHODOLOGY

A. Architecture

ASD detection is usually a binary classification issue, where a classifier aims to predict if a child suffers from ASD or not based on the input features considered. Hence, to detect ASD, this work presents an ML classifier with a feature optimization approach to identify whether a child suffers from ASD and help with taking preventive measures to provide better care. Figure 1 presents the architecture of the proposed model. The architecture first considers the input ASD dataset consisting of various sensory and behavioral features. The dataset is first preprocessed. Then, a feature optimization approach uses a CNN to extract features from the preprocessed training ASD data, which are then passed on to the ET, where the best features are selected. These features are used to train the XGB for binary classification.

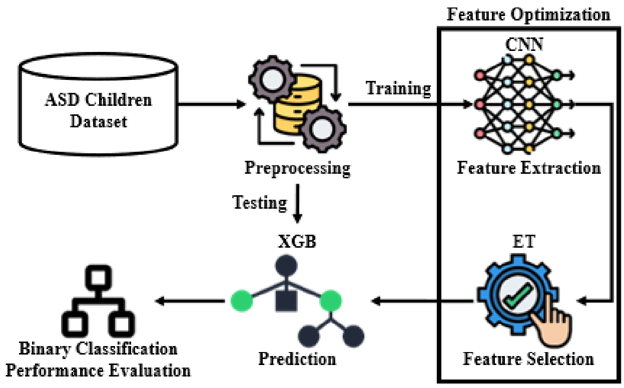

Fig. 1. CNN-ET-XGB ASD prediction model.

B. Dataset

This study used the UCI ASD children dataset [19], which consists of 292 instances, comprising 141 children diagnosed with ASD and 151 non-ASD (control group). Each record includes a variety of demographic, behavioral, and sensory features, structured to facilitate early detection of ASD through screening questionnaires. A key component of the dataset is the set of 10 Q-Chat items (A1-A10), behavioral and sensory-related questions derived from Q-Chat, a clinically validated ASD screening tool. These items are scored based on parental responses and are used to compute the Total Q-Chat Score, which provides a composite indication of ASD risk. In addition to these core features, the dataset includes demographic and biological attributes such as age, gender, ethnicity, presence of neonatal jaundice, previous usage of a mobile screening app, relation of the respondent to the child, and country of residence. The dataset includes a combination of integer, binary, and categorical data types.

C. Preprocessing

The dataset was first analyzed for any null values, finding that it did not have any. Then, a manual analysis found that in the ethnicity column, some values were of "?", which were manually replaced with "Unknown". Unnecessary columns, such as the residing country and age description, were removed, as they did not provide any specific information for ASD prediction. Finally, the data was converted into binary for consistent training and feature extraction, and the column names were renamed.

D. Feature Optimization

For feature optimization, the preprocessed dataset was considered, which can be represented mathematically as:

$$D = \{(x_{ij}, y_i)\}_{i=1}^{N} \quad (1)$$

where x_{ij} denotes a feature vector for the i^{th} ASD sample having j behavioral features, y_i denotes the target label, i.e., non-ASD or ASD, $y_i \in \{0,1\}$, and N denotes the total samples in the dataset. The dataset was divided to training-testing set, considering a stratified random sampling technique to ensure equal class proportions for training and testing. This study considered 50% data for training and 50% data for testing. Further, to standardize both training and testing data, this work applied Z-score normalization to transform the data, making it easier to compare different features. Z-score normalization was applied using:

$$\tilde{x}_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (2)$$

where $\tilde{x}_{ij}$ is Z-score normalization for the x_{ij}, μ_j denotes the mean of feature j, and σ_j denotes the standard deviation for feature j. This transformation ensures that every feature has a zero mean and a specific unit variance, which is important for feature extraction using CNN.

The normalized data was passed on to the feature optimization layer, where the CNN extracted high-level feature representations from the normalized input features. The CNN was shaped to 1D for input feature extraction, with each convolutional layer applying convolution using:

$$F_{i,j} = \sum_{k=1}^{h} W_k X_{i+k-1,j} + b \quad (3)$$

where $F_{i,j}$ denotes the convolution output for the i^{th} ASD sample and j^{th} behavioral features, W_k denotes the convolutional kernel weights, h denotes the filter size, and b denotes the bias. The initial CNN layer had 64 filters with three kernels.

To avoid overfitting during the feature extraction process, L2 regularization was used, represented as $W \sim \mathcal{N}(0, \sigma^2)$, where σ^2 denotes a small variance for controlling weight initialization. The Leaky Rectified Linear Unit (L-ReLU) was used as an activation function, as it addresses the issue of dying ReLU, preventing CNN neurons from becoming permanently inactive. The L-ReLU is applied as:

$$f(x_{ij}) = \max(\alpha x_{ij}, x_{ij}) \quad (4)$$

where α is a constant variable that allows small gradients for negative values during the feature extraction process.

During the max-pooling operation in the CNN, the pooling layer reduced the feature map size using:

$$P_{i,j} = \max_{k \in \{1,2,\ldots,h\}} F_{i+k-1,j} \quad (5)$$

Using the max-pooling operation (5), the training dataset representation was down-sampled while retaining essential features.

In CNNs, After the pooling layer, the feature map is usually flattened. In this work, instead of flattening the feature map, Global-Average Pooling (GAP) was used, calculating the average of every feature channel using:

$$Z_j = \frac{1}{M} \sum_{i=1}^{M} F_{ij} \quad (6)$$

where M denotes output channels. Using (6), a compact feature vector Z is attained. Finally, in the Fully-Connected Layer (FCL), a dense layer was used with ReLU activation to extract deep features using:

$$H = ReLU(W_f Z_j + b_f) \quad (7)$$

where W_f and b_f denote trainable weights, and the final extracted feature vector H is passed on to ET for feature selection. ET is used for identifying the best features for training the prediction classifier. In ET, the features are selected using an ensemble of Decision Trees (DTs) trained recursively. During the split of every tree in DT, this work uses the Gini impurity index for splitting:

$$G(p) = 1 - \sum_{y_i} p_{y_i}^2 \quad (8)$$

where p_{y_i} denotes the proportion of the prediction for the class y_i. In ET, every feature importance score is evaluated using:

$$I_j = \sum_{t \in i} p_t (1 - p_t) \quad (9)$$

where p_t denotes the probability of every i^{th} ASD sample. From the ET, the top ten most important features were selected. Further, the final features extracted from the CNN and the selected features from ET were concatenated, creating a final feature representation that is passed on to the XGB classifier.

E. XGB Classifier

XGB is an ensemble ML classifier that builds ensemble gradient-boosting DTs. In XGB, the main objective function for any binary prediction is [20]:

$$\mathcal{L}(\theta) = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{k=1}^{T} \Omega(f_k) \quad (10)$$

where y_i denotes the target label, $\hat{y}_i$ denotes the predicted target label using XGB, $\Omega(f_k)$ is regularization function that regularizes tree complexity, and $l(y_i, \hat{y}_i)$ is log loss, calculated using:

$$l(y_i, \hat{y}_i) = -[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (11)$$

The XGB was trained using the feature-optimized training data and tested using the other half of the data. Accuracy, precision, recall, and F1-score were used for performance evaluation, as presented in (12-15), where FN denotes false-negatives, FP denotes false-positives, TN denotes true-negatives, and TP denotes true-positives.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (13)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (14)$$

$$\text{F1} - \text{score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (15)$$

III. RESULTS AND DISCUSSION

The experimental setup involved the CNN-ET-XGB pipeline implemented on a workstation equipped with an AMD Ryzen 5 CPU, NVIDIA GTX 1650 GPU, and 16 GB of RAM, ensuring efficient computation and reproducibility. The CNN-ET-XGB was implemented in Python and executed using the Anaconda environment. Feature extraction was performed using a 1D-CNN comprising two convolutional layers (64 and 128 filters, kernel size=3, L2 regularization=0.001), followed by batch normalization, LeakyReLU activations, max pooling, global average pooling, a 128-unit dense layer with 0.5 dropout, and a 64-unit feature output layer. ET used 100 estimators and default hyperparameters for feature importance scoring to identify the top-10 features. The final classifier, XGB, was optimized via grid-tuning of hyperparameters within the following ranges: n_estimators: 50-200, learning-rate: 0.001-0.1, max-depth: 6–12, subsample and colsample-bytree: 0.7-1.0, and reg-lambda and reg-alpha of 0.5-2.0. The best model used n-estimators = 50, learning-rate = 0.003, max-depth = 12, subsample=0.95, colsample-bytree = 0.95, and both regularizers set to 1.5. Class imbalance was addressed using scale-pos-weight.

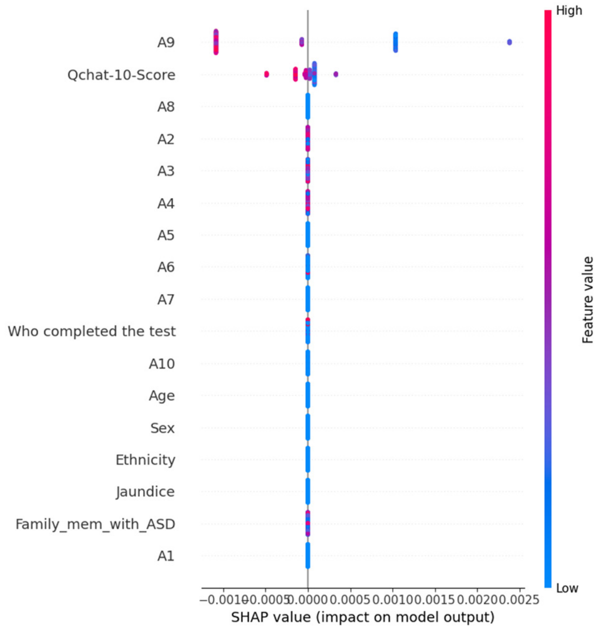

Fig. 2. Features extracted using CNN.

Figures 2 and 3 present the feature optimization process, i.e., feature extraction and feature selection for the prediction of ASD. Figure 2 shows the features extracted using the CNN and their Shapley Additive Explanation (SHAP) values. The SHAP summary plot shows that most of the features have very little impact on the prediction of ASD, with only a few significantly contributing, such as the scores of A1-A10 and the Qchat-10 score. This little impact is because the dataset is small. Figure 3 presents the top 10 features selected using ET, where it can be seen that the Qchat 10-score is the most important feature, followed by A4, A9, A10, and A5. From this analysis, it can be seen that during the prediction, the questionnaire response plays an important role in the prediction of ASD.

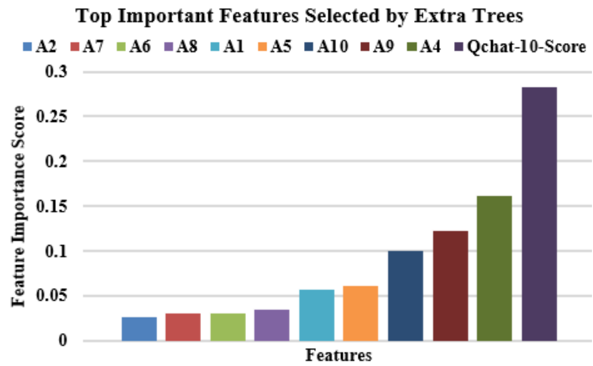

Fig. 3. Top 10 features selected by ET.

Figure 4 shows the performance metrics of CNN-ET-XGB, which achieved 99.992% accuracy, 99.985% recall, and 99.986% precision and F1-score. These results were achieved due to the feature selection and feature extraction approach using the CNN and ET, respectively.

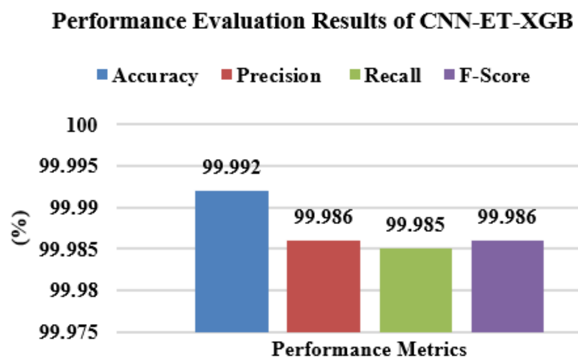

Fig. 4. Performance evaluation of CNN-ET-XGB on the ASD children dataset.

TABLE I. COMPARATIVE STUDY

Ref	Model	Accuracy	Precision	Recall	F1-score
[11]	AB	97.95	94.02	96.84	97.02
[12]	RF	99.75	-	99.33	-
[13]	L-R	96.23	97.9	94.5	94.7
[14]	LR	94.3	93.1	94.4	93.8
[15]	RF	95.9	95.8	95.8	95.8
[16]	XGB	89	-	91	-
[17]	CBBOAFS	98.97	-	98.58	98.93
[18]	AlexNet CNN	99.2	98.75	99	-
Proposed	**CNN-ET-XGB**	**99.992**	**99.986**	**99.985**	**99.986**

Table I presents a comparative study, where the CNN-ET-XGB model is compared with existing ASD prediction approaches. It can be seen that the proposed CNN-ET-XGB model achieved better performance compared to other approaches for all the performance metrics.

Table II presents the effectiveness of feature selection for the improvement of ASD prediction. The RF approach in [12] used AWFS for selecting features and attained 99.75% accuracy. Similarly, CBBOAFS [17] used CBOA, achieving 98.9% accuracy, and AlexNet CNN [18] used SBiLSTM, achieving 99.2% accuracy. It can be observed that the models that did not select any features before prediction, such as RF [14] and LR [15], showed less accuracy, i.e., 95.9% and 94.3%, respectively. Moreover, the XGB [16], even after selecting features using RFE, failed to achieve higher accuracy, showing that the selection of important features is very important. The CNN-ET-XGB model achieved better accuracy considering the ET for feature selection, thereby providing a better ASD prediction model.

TABLE II. FEATURE SELECTION AND ACCURACY COMPARISON

Ref	Model	Feature selection	Feature selector	Accuracy
[11]	AB	Yes	CAE, RFAE, GRAE, IGAE	97.95
[12]	RF	Yes	AWFS	99.75
[13]	L-R	Yes	CSA	96.23
[14]	LR	No	NIL	94.3
[15]	RF	No	NIL	95.9
[16]	XGB	Yes	RFE	89
[17]	CBBOAFS	Yes	CBOA	98.97
[18]	AlexNet CNN	Yes	SBiLSTM	99.2
This	**CNN-ET-XGB**	**Yes**	**ET**	**99.992**

The CNN-ET-XGB model shows better performance in ASD prediction due to its hybrid approach of combining deep feature extraction with robust feature selection. Using CNN, latent patterns and complex relationships across questionnaire and demographic inputs are effectively captured, while ET selects the most informative features, eliminating noise and redundancy. This dual-stage optimization leads to highly discriminative feature representations, which are further exploited by the XGB classifier to achieve an accuracy of 99.992%. The model's ability to integrate both domain-relevant features (e.g., Qchat-10 score, A1-A10) and deep abstracted features ensures that critical signals for ASD detection are retained, while less impactful inputs (e.g., age, sex) are down-weighted, as observed in SHAP analysis. The synergy between CNN-based feature extraction, ET-driven selection, and XGB classification is key to the strong predictive power of the proposed model. However, the relatively small dataset remains a limiting factor, which can be resolved by incorporating larger, more diverse datasets, fine-tuning the CNN architecture, or using ensemble learning with multiple feature selection techniques.

IV. CONCLUSION

The proposed CNN-ET-XGB model effectively achieves the core objectives of this study, i.e., enhancing the accuracy of ASD prediction through a novel feature optimization strategy. By combining CNN-based feature extraction with ET-based feature selection and XGB classification, the model addresses key limitations in existing ML approaches, such as overfitting and reduced accuracy due to irrelevant features. The high performance achieved, i.e., 99.992% accuracy, demonstrates the efficiency of the proposed method in extracting and utilizing meaningful features for ASD prediction. This shows that integrating feature extraction and selection significantly contributes to reliable diagnostic outcomes. The comparative analysis further reinforces the model's superiority over existing approaches, validating the effectiveness of the CNN-ET-XGB pipeline. For future work, the ET-based feature selection can be explored with various other classifiers to evaluate its impact across different learning algorithms and datasets. Additionally, incorporating physiological signal data, such as EEG, eye-tracking, or heart rate variability, can further enrich the feature space and potentially enhance the model's generalizability and diagnostic precision.

REFERENCES

[1] R. V. Bidwe, S. Mishra, S. Kamini Bajaj, and K. Kotecha, "Attention-Focused Eye Gaze Analysis to Predict Autistic Traits Using Transfer Learning," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, May 2024, Art. no. 120, https://doi.org/10.1007/s44196-024-00491-y.

[2] O. Patil and M. Kaple, "Sensory Processing Differences in Individuals With Autism Spectrum Disorder: A Narrative Review of Underlying Mechanisms and Sensory-Based Interventions," *Cureus*, Oct. 2023, https://doi.org/10.7759/cureus.48020.

[3] A. S. Mills, P. Tablon-Modica, C. A. Mazefksy, and J. A. Weiss, "Emotion dysregulation in children with autism: A multimethod investigation of the role of child and parent factors," *Research in Autism Spectrum Disorders*, vol. 91, Mar. 2022, Art. no. 101911, https://doi.org/10.1016/j.rasd.2021.101911.

[4] M. C. Hunsche, M. Cervin, E. A. Storch, P. C. Kendall, J. J. Wood, and C. M. Kerns, "Social functioning and the presentation of anxiety in children on the autism spectrum: A multimethod, multiinformant analysis.," *Journal of Psychopathology and Clinical Science*, vol. 131, no. 2, pp. 198–208, Feb. 2022, https://doi.org/10.1037/abn0000724.

[5] N. Salari *et al.*, "The global prevalence of autism spectrum disorder: a comprehensive systematic review and meta-analysis," *Italian Journal of Pediatrics*, vol. 48, no. 1, Dec. 2022, Art. no. 112, https://doi.org/10.1186/s13052-022-01310-w.

[6] D. Stevanović, "Quantitative Checklist for Autism in Toddlers (Q-CHAT): A psychometric study with Serbian Toddlers," *Research in Autism Spectrum Disorders*, vol. 83, May 2021, Art. no. 101760, https://doi.org/10.1016/j.rasd.2021.101760.

[7] A. Shahid, K. Wilkinson, S. Marcu, and C. M. Shapiro, "Modified Checklist for Autism in Toddlers (M-CHAT)," in *STOP, THAT and One Hundred Other Sleep Scales*, A. Shahid, K. Wilkinson, S. Marcu, and C. M. Shapiro, Eds. Springer New York, 2011, pp. 225–227.

[8] D. Nakamura *et al.*, "Predictive model using autism diagnostic observation schedule, second edition for differential diagnosis between schizophrenia and autism spectrum disorder," *Frontiers in Psychiatry*, vol. 15, Dec. 2024, Art. no. 1493158, https://doi.org/10.3389/fpsyt.2024.1493158.

[9] K. Yoshinaga, J. Egawa, Y. Watanabe, H. Kasahara, A. Sugimoto, and T. Someya, "Usefulness of the autism spectrum quotient (AQ) in screening for autism spectrum disorder and social communication disorder," *BMC Psychiatry*, vol. 23, no. 1, Nov. 2023, Art. no. 831, https://doi.org/10.1186/s12888-023-05362-y.

[10] F. J. Rajam and B. R. K. Swakkin, "MAPLE: A Novel Processing Technique for Adult Autism Prediction," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 23901–23906, Jun. 2025, https://doi.org/10.48084/etasr.10864.

[11] S. M. M. Hasan, M. P. Uddin, M. A. Mamun, M. I. Sharif, A. Ulhaq, and G. Krishnamoorthy, "A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders," *IEEE Access*, vol. 11, pp. 15038–15057, 2023, https://doi.org/10.1109/ACCESS.2022.3232490.

[12] M. M. Abdelwahab, K. A. Al-Karawi, E. M. Hasanin, and H. E. Semary, "Autism Spectrum Disorder Prediction in Children Using Machine Learning," *Journal of Disability Research*, vol. 3, no. 1, Jan. 2024, https://doi.org/10.57197/JDR-2023-0064.

[13] R. K. Reghunathan, P. N. P. Venkidusamy, R. G. Kurup, B. George, and N. Thomas, "Machine Learning-Based Classification of Autism Spectrum Disorder across Age Groups," in *CC 2023*, Mar. 2024, Art. no. 12, https://doi.org/10.3390/engproc2024062012.

[14] P. Bawa, V. Kadyan, A. Mantri, and H. Vardhan, "Investigating multiclass autism spectrum disorder classification using machine learning techniques," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 8, Jun. 2024, Art. no. 100602, https://doi.org/10.1016/j.prime.2024.100602.

[15] C. J. Pae and N. H. Pae, "Advanced Predictive Autism Spectrum Disorder Screening Method," *International Journal of High School Research*, vol. 6, no. 2, 2024.

[16] L. J. Sollis, D. P. Wall, and P. Y. Washington, "Evaluating Multicultural Autism Screening for Toddlers Using Machine Learning on the QCHAT-10." *Health Informatics*, Nov. 13, 2024, https://doi.org/10.1101/2024.11.12.24317211.

[17] A. Ramakrishnan, R. Ramalingam, P. Ramalingam, V. Ravi, T. J. Alahmadi, and S. S. Maidin, "A novel chaotic binary butterfly optimization algorithm based feature selection model for classification of autism spectrum disorder," *International Journal of Applied Mathematics and Computer Science*, vol. 34, no. 4, 2024, https://doi.org/10.61822/amcs-2024-0043.

[18] M. Sha, H. Al-Dossary, and M. P. Rahamathulla, "Multimodal Data Fusion Framework for Early Prediction of Autism Spectrum Disorder," *Human Behavior and Emerging Technologies*, vol. 2025, no. 1, Jan. 2025, Art. no. 1496105, https://doi.org/10.1155/hbe2/1496105.

[19] F. Thabtah, "Autistic Spectrum Disorder Screening Data for Children." UCI Machine Learning Repository, 2017, https://doi.org/10.24432/C5659W.

[20] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, Aug. 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.