A Cross-Dataset Evaluation of Machine Learning Approaches for Autism Spectrum Disorder Across Age Groups

Lola Anančevska and Hristijan Gjoreski
Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius University in Skopje,
Rugjer Boshkovik 18, 1000 Skopje, N.Macedonia
anancevskal@gmail.com

Keywords: Autism Spectrum Disorder (ASD), Early Detection, Clinical Data, Machine Learning.

Abstract: Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition that requires timely and accurate detection for effective intervention. This study evaluates the performance of six machine learning models – Logistic Regression, Support Vector Machine (SVM), Decision Tree, Random Forest, XGBoost, and k-Nearest Neighbors (kNN) – in detecting ASD across three age groups, using three separate datasets: children, adolescents, and adults. Using a comprehensive dataset segmented by age, each model was assessed through stratified 10-fold cross-validation based on five key performance metrics: accuracy, precision, recall, F1 score, and area under the curve (AUC). The findings indicate that ensemble and linear models, particularly XGBoost and Logistic Regression, consistently deliver the most reliable results across all age groups and the combined dataset, with high classification accuracy and balanced precision-recall tradeoffs. In contrast, kNN and Decision Tree models displayed inconsistent performance, often struggling with both false positives and false negatives. This analysis supports the application of advanced machine learning methods, especially ensemble techniques, for developing robust and generalizable ASD detection systems across diverse age demographics.

1 INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that typically manifests within the first three years of life **Помилка! Джерело посилання не знайдено.**. It is primarily characterized by challenges in social interaction, communication difficulties, restricted interests and repetitive behaviors **Помилка! Джерело посилання не знайдено.**. Individuals with ASD often struggle to understand others' emotions and thoughts, which can make interpersonal communication difficult. According to the World Health Organization (WHO), approximately 1 in 270 people globally are affected by ASD **Помилка! Джерело посилання не знайдено.**. Each individual spectrum presents a unique set of characteristics – while some face significant challenges, others may exhibit exceptional abilities in areas such as visual processing, academics or music. Early diagnosis and intervention are crucial in managing ASD effectively. Detecting ASD early enables more timely and targeted treatments, which can significantly improve outcomes and reduce the long-term impact of the disorder. Traditionally, symptoms are identified through clinical observation, which can be time-consuming and labor-intensive.

In recent years, machine learning techniques have emerged as powerful tools for analyzing symptoms associated with complex conditions such as heart disease, diabetes, and cancer. These methods are now increasingly being applied to ASD detection, with researchers **Помилка! Джерело посилання не знайдено.**, **Помилка! Джерело посилання не знайдено.** developing a range of models to improve the accuracy and efficiency of diagnosis and help mitigate the effects of ASD more effectively.

The prevalence of ASD has been increasing rapidly across all age groups in recent years. Early detection of this neurodevelopmental condition is crucial, as it can significantly improve the individual's mental and physical well-being. With the growing application of machine learning in healthcare, early diagnosis of various neurological and physiological conditions is becoming increasingly feasible. This progress has fueled growing interest in utilizing these technologies to enhance the detection and analysis of complex disorders, thereby enabling more effective treatment strategies. However, diagnosing ASD remains challenging due to symptom overlap with other mental health conditions. These similarities often complicate the diagnostic process, making accurate and timely identification of ASD a difficult task. This underscores the need for advanced computational tools, such as machine learning models, to support clinicians in distinguishing ASD from related disorders **Помилка! Джерело посилання не знайдено.**.

2 RELATED WORK

Recent studies have explored the use of wearable sensor-based devices for the detection of Autism Spectrum Disorder (ASD)**Помилка! Джерело посилання не знайдено.Помилка! Джерело посилання не знайдено.**. The integration of intelligent systems has introduced more cost-effective and scalable solutions for identifying ASD in children, adolescents and adults **Помилка! Джерело посилання не знайдено.**. A wide range of methods has been applied in the literature, including structural magnetic resonance imaging (MRI) **Помилка! Джерело посилання не знайдено.**, neural networks, machine learning (ML) techniques, deep learning, transfer learning, and Internet of Things (IoT) – based frameworks **Помилка! Джерело посилання не знайдено.**. While these approaches have shown promising results in ASD detection with reasonable accuracy, they are often hindered by data acquisition challenges. Many healthcare institutions are reluctant to share patient data due to organizational policies and regional data protection regulations, making data privacy, security, and availability major obstacles in the development of effective intelligent diagnostic systems. Furthermore, even when data access is granted, transmitting large volumes of data over networks introduces additional concerns such as network congestion, latency and potential data breaches.

Wong et al. **Помилка! Джерело посилання не знайдено.** utilized data mining tools to investigate the impact of various autism treatments on behavioral outcomes. Their approach enabled the prediction and better understanding of autistic children's behavior, providing a framework for differentiating between socially acceptable and unacceptable conduct. Similarly, Kaur et al. **Помилка! Джерело посилання не знайдено.** conducted a comprehensive review of 45 research studies that employed supervised machine learning and classification algorithms in the context of ASD diagnosis and analysis. Among the most frequently used models were Support Vector Machine (SVM), Neural Networks, Random Forests, Decision Tree, Least Absolute Shrinkage, and Selection Operator (LASSO), various regression methods, Conditional Forests (CF), Naïve Bayes (NB), Elastic Net regression (Enet), Random Tree, and Flex Tree classifiers. Furthermore Koumpouros et al. **Помилка! Джерело посилання не знайдено.** carried out a systematic review of 83 studies published after the year 2000, focusing on interventions for ASD that leveraged wearable technologies and computation methods to support treatment and behavioral monitoring.

3 DATA

The following section describes the dataset preprocessing, model selection, hyperparameter tuning, and performance evaluation procedures used to assess the effectiveness of various machine learning models for predicting ASD.

Three datasets are used for the purpose of this paper, one for children, one for adolescents, and one for adults. Every dataset consists of 20 features, 10 of which (A1-A10) are behavioral features that have a binary value (value that is 0 or 1). The other 10 features are shown in Table 1 **Помилка! Джерело посилання не знайдено.-Помилка! Джерело посилання не знайдено.Помилка! Джерело посилання не знайдено.**.

Table 1: Dataset features.

Feature	Description
age	Age in years
gender	Male or female
ethnicity	List of common ethnicities in text format
jaundice	Whether the person was born with jaundice
autism	Whether any immediate family member has ASD
country_of_res	List of countries in text format
used_app_before	Whether the user has used a screening app
result	The final score obtained based on the scoring algorithm of the screening method used. This was computed in an automated manner
age_desc	The age span of the patient
relation	Who diagnosed the patient

For this paper, in addition to the analysis done on these datasets, containing the same features, we constructed a dataset that is a combination of all three.

The analysis of the combined dataset gives another view of the results.

The target variable, Class/ASD, was categorical, representing the presence or absence of ASD. To enable machine learning algorithms to process this variable, Label Encoding **Помилка! Джерело посилання не знайдено.** is applied, where the classes were transformed into binary numeric values: 0 for negative and 1 for positive cases.

4 METHODS

For this study, six machine learning models were selected for comparison:

- Logistic Regression: A linear classifier suitable for binary classification tasks **Помилка! Джерело посилання не знайдено.**;
- Support Vector Machine (SVM): A non-linear classifier that maximizes the margin between classes using kernels **Помилка! Джерело посилання не знайдено.**;
- Decision Tree: A recursive partitioning method that constructs decision boundaries based on features. **Помилка! Джерело посилання не знайдено.**;
- Random Forests: An ensemble model that aggregates the predictions of multiple Decision Trees **Помилка! Джерело посилання не знайдено.**;
- XGBoost: A gradient boosting method known for its robustness and performance **Помилка! Джерело посилання не знайдено.**;
- k-Nearest Neighbors (kNN): A non-parametric method that assigns labels based on the majority vote of neighboring points **Помилка! Джерело посилання не знайдено.**.

5 EXPERIMENTAL SETUP

Initially, a traditional train-validation-test split was implemented to evaluate model performance **Помилка! Джерело посилання не знайдено.**. However, this method resulted in inconsistent and unreliable evaluation metrics, likely due to the limited size and potential variability within the dataset. To address these limitations, the study adopted a 10-fold cross-validation strategy, which better maintains class proportions across folds and provides more stable and generalizable evaluation results **Помилка! Джерело посилання не знайдено.**.

In 10-fold Stratified Cross-Validation, the dataset was divided into ten equal-sized folds, ensuring that each fold preserved the same class distribution as the overall dataset. The model was trained on nine folds and tested on the remaining fold, and this process was repeated 10 times. The final reported metrics were the average values across all folds.

5.1 Feature Selection

Feature selection is performed using a Random Forest Classifier, to identify the most important predictors of ASD. Random Forest, an ensemble method, computes the importance of each feature by measuring how much they contribute to the reduction of impurity in the Decision Trees **Помилка! Джерело посилання не знайдено.**. The topmost important feature was identified and subsequently dropped to assess the effect of excluding this feature on model performance. This step was essential for ensuring that the model was not overly reliant on a single feature, which could lead to overfitting.

5.2 Hyperparameter Tuning

To optimize the performance of the models, hyperparameter tuning was performed using GridSearchCV **Помилка! Джерело посилання не знайдено.**. This method exhaustively searches through a specified parameter grid to identify the best combination of hyperparameters for each model. The hyperparameters tuned for each model included regularization strength (c) for Logistic Regression and SVM, maximum tree depth and minimum samples for Decision Trees and Random Forests, number of neighbors for kNN and learning rate, tree depth, and number of estimators for XGBoost. The hyperparameters were selected based on accuracy as the scoring metric.

5.3 Evaluation Metrics

The metrics that were computed to assess the performance of each model are:

- Accuracy – Measures the proportion of correctly classified instances among all instances **Помилка! Джерело посилання не знайдено.**;
- Precision – Measures the proportion of correctly predicted positive instances among all predicted positives **Помилка! Джерело посилання не знайдено.**;
- Recall – Measures the proportion of correctly predicted positive instances among all actual

positives **Помилка! Джерело посилання не знайдено.**;

- F1 Score – Represents the harmonic mean of precision and recall, balancing the two **Помилка! Джерело посилання не знайдено.**;
- Area Under the ROC Curve (AUC) – Evaluates model's ability to distinguish between classes across all threshold settings **Помилка! Джерело посилання не знайдено.**.

These metrics were selected to provide a comprehensive evaluation of predictive ability, especially considering the potential imbalance in class distributions **Помилка! Джерело посилання не знайдено.**.

6 RESULTS

This section presents the results of the descriptive analysis and model evaluation. It explores the distribution of demographic and target features across children, adolescent and adult datasets, followed by performance of six classification algorithms. The analysis is then replicated on a merged dataset to assess model behavior across a combined population.

6.1 Exploratory Data Analysis

To understand the class balance of the dataset across different age groups, the distribution of the target variable (Class/ASD) was visualized separately for children, adolescents and adults.

The class distribution varies across datasets. The children dataset is relatively balanced, with nearly equal numbers of ASD-positive and negative cases, which minimizes bias during model training. In contrast, the adolescent dataset shows a mild skew toward ASD-positive cases, which may influence sensitivity and false positive rates. The adult dataset is strongly imbalanced, with a much higher proportion on ASD-negative cases. This imbalance can hinder the detection of positive cases and may require techniques such as stratified cross-validation or class weighting.

The approximate values of the target variable distribution are illustrated in Table 2 below.

Table 2: Target variable distribution.

Dataset	NO	YES
Children	49%	51%
Adolescents	38%	62%
Adults	72%	28%

To better understand the demographics of each dataset and assess potential biases, the distribution of age was visualized for the children, adolescent, and adult datasets (Figure 1).

The age distribution varies notably across the three datasets. In the children group (ages 4-11), the highest concentration of participants is at age 4 (n≈78), after which the numbers decline steadily with age. This pattern reflects the priority given to early ASD screening, often beginning in preschool years **Помилка! Джерело посилання не знайдено.**.

Among adolescents (ages 12-16), the distribution is relatively even, with each year represented (n≈14-29), and a slight peak at age 16. This indicates that diagnoses continue into adolescence, often due to increased social and academic pressure **Помилка! Джерело посилання не знайдено.**.

In adults (ages 18-80), the distribution peaks in early adulthood (ages 18-25, n≈90) before gradually tapering off with age. This suggests that while most adult diagnoses occur in early adulthood there is still a presence of individuals diagnosed later in life, possibly due to retrospective recognition of lifelong traits or increasing public and clinical awareness of ASD [35].

Analysis of ethnicity across age groups revealed distinct demographic patterns. In the children dataset, White-European participants make up nearly half the sample, followed by Asian and Middle Eastern groups, with minority representation remaining limited underscoring persistent disparities in early ASD diagnosis across ethic lines **Помилка! Джерело посилання не знайдено.**.

Among adolescents, White-European representation decreases slightly, while "Others," Black, and Hispanic groups increase, suggesting broader diagnostic access and improved outreach **Помилка! Джерело посилання не знайдено.**.

The adult dataset shows the most diversity, with White-Europeans below 40% and strong representation from Asian, Middle Eastern, South Asian, Black, and Latino groups, reflecting greater awareness and inclusivity in adult diagnoses.

The gender distribution across the three age groups – children, adolescents and adults – was evaluated to understand the demographic composition of the dataset (Table 3). The gender distribution for children, exhibits a noticeable imbalance, with the number of males being significantly higher than females. This disproportionate representation aligns with the existing research which indicates a higher prevalence of ASD diagnoses among male children compared to female **Помилка! Джерело посилання не знайдено.**.

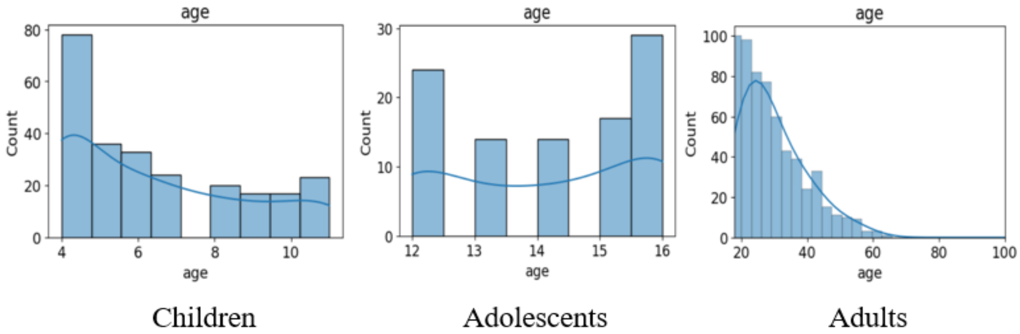

Figure 1: Age distribution across age groups.

In the adolescent group, the gender distribution achieves near parity, with 50 males, and 50 females represented. This balanced composition ensures that the models trained on this dataset are less likely to exhibit gender-related biases, enabling a fair evaluation of ASD detection across both genders in this age group.

The adult dataset demonstrates a relatively balanced gender distribution, albeit with a slight skew. While not perfectly balanced, this distribution provides sufficient representation of both genders to minimize bias in the analysis of ASD detection among adults.

Table 3: Gender distribution across age groups.

Dataset	Male	Female
Children	70%	30%
Adolescents	50%	50%
Adults	53.3%	46.7%

Jaundice, particularly neonatal, has been studied for its potential association with neurodevelopmental disorders, including ASD. Neonatal jaundice results from elevate bilirubin levels, which, if untreated, may lead to kernicterus – a condition that can cause brain damage. **Помилка! Джерело посилання не знайдено.** In the dataset, most children do not have a history of jaundice. For adolescents, jaundice is even rarer, while in adults, the majority also report no history (Table 4). Overall jaundice is an uncommon feature across all age groups, however, its presence in early life has been linked to potential neurodevelopmental impacts, which could have implications for ASD detection **Помилка! Джерело посилання не знайдено.**.

Table 4: Jaundice distribution across age groups.

Dataset	YES	NO
Children	22.2%	77.8%
Adolescents	18.4%	81.6%
Adults	8.1%	91.9%

6.2 Model Evaluation by Age Group

In this section, the evaluation results of the machine learning models used to predict ASD, across different age groups are presented. The performance of each model is assessed using several key metrics, including accuracy, precision, recall, F1 Score and AUC. These metrics provide a comprehensive understanding of how well each model performs in terms of classification accuracy, the ability to detect true ASD cases and the balance between false positives and false negatives.

Table 5: Model comparison for the children database.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC (%)
Log Reg	95.58	95.64	96.09	95.77	95.54
SVM	95.2	94.3	96.86	95.43	95.16
kNN	84.35	79.05	96.79	86.59	84.23
Decision Tree	87.52	85.07	92.05	88.24	87.53
Random Forests	94.77	94.89	95.38	94.95	94.84
XGBoost	95.57	95.49	96.09	95.69	95.57

As shown in Table 5, for the children's dataset shows that Logistic Regression and XGBoost perform very similarly across all metrics, making them strong candidates for the children's dataset. SVM has high recall but lower precision, indicating a tendency toward false positives. kNN performs the worst overall, with low accuracy, precision, and F1 score despite high recall. The Decision Tree is balanced but lags behind the top models in precision and F1 score, indicating room for improvement. Random Forest is reliable and comparable to XGBoost but slightly lower in accuracy, precision, and F1.

Table 6: Model comparison for the adolescent dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC (%)
Log Reg	90	92.74	91.9	91.99	89.29
SVM	89.89	88.87	96.67	92.25	87.5
kNN	73.56	71.4	98.57	82.68	63.87
Decision Tree	74.78	80.77	80.95	79.6	72.14
Random Forests	88.78	87.78	96.67	91.69	85.42
XGBoost	86.78	88.12	91.9	89.62	84.29

According to the evaluation results shown in Table 6, Logistic Regression is the top-performing model for the adolescent dataset, with high precision (92.74%) and recall (91.9%), making it a balanced and reliable choice. SVM has strong recall but lower precision, leading to more false positives. kNN has high recall (98.57%) but low precision and F1, reducing overall effectiveness. Decision Tree is balanced but suboptimal, while Random Forest and XGBoost perform well but slightly behind Logistic Regression in precision and F1.

Table 7: Model comparison for the adult dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC (%)
Log Reg	97.53	96.59	95	95.75	96.8
SVM	96.22	94.42	92.78	93.5	95.22
kNN	84.35	84.03	86.11	84.94	89.56
Decision Tree	88.99	82.44	80.56	81.09	86.54
Random Forests	96.06	95.58	90.56	92.83	94.46
XGBoost	97.53	96.66	95	95.76	96.8

Table 7, shows that Logistic Regression and XGBoost are the top models for the adult dataset, with the highest accuracy (~97.53%) and excellent precision, recall, F1, and AUC, indicating strong and consistent classification of ASD cases. SVM performs well but slightly below the top models. Random Forest is solid but less consistent, kNN shows moderate performance with lower precision and F1, and Decision Tree performs the weakest overall.

6.3 Analysis on the Combined Dataset

The models were also evaluated on a combined dataset that merges data from children, adolescents, and adults to assess overall performance across age groups. As with age-specific analyses, 10-fold stratified cross-validation was used, and performance was measured using accuracy, precision, recall, F1 score, and AUC. The evaluation results are shown in Table 8.

Table 8: Model comparison for the combined dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC (%)
Log Reg	97.38	97.02	96.19	96.58	97.16
SVM	96.12	94.64	95.38	94.99	95.98
kNN	82.2	74.17	83.13	78.28	82.37
Decision Tree	87.85	83.28	85.87	84.46	87.48
Random Forests	95.19	95.05	93.4	93.64	94.67
XGBoost	98.22	97.85	97.54	97.68	98.09

The findings from the combined dataset underscore key factors in developing a unified model effective across all age groups. XGBoost proved to be the most effective, achieving high sensitivity and specificity, which is crucial in clinical screening where minimizing both false positives and false negatives is essential. Its strong performance across varied age-related features demonstrates robustness and adaptability, Likewise, Logistic Regression showed consistent results, confirming its reliability for classification tasks involving structured clinical data.

7 DISCUSSION

The comparative analysis of the six machine learning models revealed distinct performance patters across age groups, prompting further interpretation of their mechanisms and suitability for ASD detection.

The superior performance of Logistic Regression and XGBoost can be attributed to their model design advantages. Logistic Regression performs well because the dataset's features show strong, mostly linear relationship with ASD diagnosis, allowing for a simple and generalizable decision boundary.

In contrast XGBoost, effectively captured complex, non-linear feature interaction through its ensemble of decision trees, providing strong adaptability and robustness.

The comparatively weaker and less stable performance of kNN and Decision Tree aligns with expectations. Decision Tree models are prone to overfitting, while kNN's distance-based approach struggles with high-dimensional data, leading to inconsistent performance.

Error analysis showed that the models struggled with individuals whose features were borderline for ASD or who showed mild symptoms, leading to more false negatives, especially among adults, where learned social behaviors can mask ASD traits. Future work should include more detailed feature analysis to improve accuracy in these borderline cases.

8 CONCLUSIONS

This study evaluated six machine learning models for detecting ASD across children, adolescents, and adults, using Accuracy, Precision, Recall, F1 Score, and AUC as performance metrics.

The study concludes that machine learning models, specifically Logistic Regression and XGBoost, provide a robust and generalizable foundation for early non-invasive ASD screening across all age groups. However, the limitations of this work must be acknowledged. Firstly, the models are trained solely on structured clinical questionnaire data, and do not incorporate time-series behavioral, genetic, or brain-imaging data. Secondly, the model performance, particularly on the adult dataset strongly influenced by the demographic and cultural background of the data, so clinical validation on diverse populations is essential before wider use.

In practice clinicians could input patient assessment data into the model, which would provide a risk estimate to support further evaluation or referral, supporting informed clinical decisions without replacing professional judgment.

.

REFERENCES

[1] R. J. Landa, A. L. Gross, E. A. Stuart, and A. Faherty, "Developmental trajectories in children with and without autism spectrum disorders: The first 3 years," Child Development, vol. 84, no. 2, pp. 429–442, Mar.–Apr. 2013, doi: 10.1111/j.1467-8624.2012.01870.x.

[2] H. Hodges, C. Fealko, and N. Soares, "Autism spectrum disorder: Definition, epidemiology, causes, and clinical evaluation," Translational Pediatrics, vol. 9, suppl. 1, pp. S55–S65, Feb. 2020, doi: 10.21037/tp.2019.09.09.

[3] GBD 2019 Diseases and Injuries Collaborators, "Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019," The Lancet, vol. 396, no. 10258, pp. 1204–1222, Oct. 2020, doi: 10.1016/S0140-6736(20)30925-9.

[4] M. B. Usta et al., "Use of machine learning methods in prediction of short-term outcome in autism spectrum disorders," Psychiatry and Clinical Psychopharmacology, vol. 29, no. 3, pp. 320–325, 2019, doi: 10.1080/24750573.2018.1545334.

[5] K. Hyde, M. Novack, N. LaHaye, C. Parlett-Pelleriti, R. Anden, D. Dixon, and E. Linstead, "Applications of supervised machine learning in autism spectrum disorder research: A review," Review Journal of Autism and Developmental Disorders, vol. 6, 2019, doi: 10.1007/s40489-019-00158-x.

[6] F. Thabtah, F. Kamalov, and K. Rajab, "A new computational intelligence approach to detect autistic features for autism screening," International Journal of Medical Informatics, vol. 117, pp. 112–124, Sep. 2018, doi: 10.1016/j.ijmedinf.2018.06.009.

[7] S. Alhassan, A. Soudani, and M. Almusallam, "Energy-efficient EEG-based scheme for autism spectrum disorder detection using wearable sensors," Sensors, vol. 23, no. 4, Art. no. 2228, 2023, doi: 10.3390/s23042228.

[8] N. Ali, "Autism spectrum disorder classification on electroencephalogram signal using deep learning algorithm," IAES Int. J. Artificial Intelligence, vol. 9, no. 1, pp. 91–99, 2020, doi: 10.11591/ijai.v9.i1.pp91-99.

[9] D. S. Sujana and D. P. Augustine, "Diagnosis of autism spectrum disorder: A review of three focused interventions," SN Computer Science, vol. 4, Art. no. 139, 2023, doi: 10.1007/s42979-022-01584-1.

[10] Y. ElNakieb et al., "Understanding the role of connectivity dynamics of resting-state functional MRI in the diagnosis of autism spectrum disorder: A comprehensive study," Bioengineering, vol. 10, no. 1, Art. no. 56, 2023, doi: 10.3390/bioengineering10010056.

[11] M. S. Farooq, R. Tehseen, M. Sabir, and Z. Atal, "Detection of autism spectrum disorder (ASD) in children and adults using machine learning," Scientific Reports, vol. 13, Art. no. 9605, 2023, doi: 10.1038/s41598-023-35910-1.

[12] R. Y.-Y. Chan, C. M. V. Wong, and Y. N. Yum, "Predicting behavior change in students with special education needs using multimodal learning analytics," IEEE Access, vol. 11, pp. 63238–63251, 2023, doi: 10.1109/ACCESS.2023.3288695.

[13] N. Kaur, A. Kaur, N. Dhiman, A. Sharma, and R. Rana, "A systematic analysis of detection of autism spectrum disorder: IoT perspective," Int. J. Innovative Science and Modern Engineering, vol. 6, no. 6, pp. 10–13, 2020, doi: 10.35940/ijisme.F1219.046620.

[14] S. Dedgaonkar, R. K. Sachdeva-Bedi, K. Kothari, R. Loya, and S. Godbole, "Role of IoT and ML for autistic people," Int. J. Future Generation Communication and Networking, vol. 13, no. 3s, pp. 773–781, 2020.

[15] F. Thabtah, "Autistic spectrum disorder screening data for children," UCI Machine Learning Repository, 2017.

[16] F. Thabtah, "Autism screening adult," UCI Machine Learning Repository, 2017.

[17] F. Thabtah, "Autistic spectrum disorder screening data for adolescent," UCI Machine Learning Repository, 2017.

[18] J. Brownlee, "Train-test split for evaluating machine learning algorithms," Machine Learning Mastery, 2020.

[19] J. Wei and Z. Hou, "Binary encoding for label," 2022, doi: 10.36227/techrxiv.21353763.

[20] J. Obregon and J.-Y. Jung, "Explanation of ensemble models," in Human-Centered Artificial Intelligence. Academic Press, 2022, pp. 51–72, doi: 10.1016/B978-0-323-85648-5.00011-6.

[21] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, "Hyperparameter optimization for machine learning models based on Bayesian optimization," Journal of Electronic Science and Technology, vol. 17, no. 1, pp. 26–40, 2019, doi: 10.11989/JEST.1674-862X.80904120.

[22] H. Jeong and B. Zou, "Linear classifier models for binary classification," 2023.

[23] M. Awad and R. Khanna, "Support vector machines for classification," 2015, doi: 10.1007/978-1-4302-5990-9_3.

[24] L. Rokach and O. Maimon, "Decision trees," 2005, doi: 10.1007/0-387-25465-X_9.

[25] Z. Ali et al., "Extreme gradient boosting algorithm with machine learning: A review," Academic Journal of Nawroz University, vol. 12, pp. 320–334, 2023, doi: 10.25007/ajnu.v12n2a1612.

[26] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in Proc. OTM Confederated Int. Conf., LNCS, vol. 2888, 2003, doi: 10.1007/978-3-540-39964-3_62.

[27] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," PLoS One, vol. 10, no. 3, Art. no. e0118432, 2015, doi: 10.1371/journal.pone.0118432.

[28] Ž. Đ. Vujović, "Classification model evaluation metrics," Int. J. Advanced Computer Science and Applications, vol. 12, no. 6, 2021, doi: 10.14569/IJACSA.2021.0120670.

[29] S. Swaminathan and B. R. Tantri, "Confusion matrix-based performance evaluation metrics," African Journal of Biomedical Research, vol. 27, pp. 4023–4031, 2024, doi: 10.53555/AJBR.v27i4S.4345.

[30] R. Diallo, C. Edalo, and O. O. Awe, "Machine learning evaluation of imbalanced health data: A comparative analysis of balanced accuracy, MCC, and F1 score," in Practical Statistical Learning and Data Science Methods. Springer, 2025, doi: 10.1007/978-3-031-72215-8_12.

[31] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms," IEEE Trans. Knowledge and Data Engineering, vol. 17, no. 3, pp. 299–310, Mar. 2005, doi: 10.1109/TKDE.2005.50.

[32] T. M. Dutschmann et al., "Large-scale evaluation of k-fold cross-validation ensembles for uncertainty estimation," Journal of Cheminformatics, vol. 15, Art. no. 49, 2023, doi: 10.1186/s13321-023-00709-9.

[33] S. L. Hyman, S. E. Levy, and S. M. Myers, "Identification, evaluation, and management of children with autism spectrum disorder," Pediatrics, vol. 145, no. 1, Art. no. e20193447, 2020, doi: 10.1542/peds.2019-3447.

[34] M. C. Lai and S. Baron-Cohen, "Identifying the lost generation of adults with autism spectrum conditions," The Lancet Psychiatry, vol. 2, no. 11, pp. 1013–1027, 2015, doi: 10.1016/S2215-0366(15)00277-1.

[35] F. Happé, R. A. Charlton, and B. Maughan, "Autism and age: Findings from the SNAP cohort," Journal of Child Psychology and Psychiatry, vol. 57, no. 9, pp. 1025–1033, 2016, doi: 10.1111/jcpp.12559.

[36] D. S. Mandell et al., "Racial/ethnic disparities in the identification of children with autism spectrum disorders," American Journal of Public Health, vol. 99, no. 3, pp. 493–498, 2009, doi: 10.2105/AJPH.2007.131243.

[37] K. A. Wantchekon, M. Satterthwaite-Freiman, and A. J. Umaña-Taylor, "Ethnic-racial identity among White adolescents: Patterns and links to adjustment," Journal of Applied Developmental Psychology, vol. 98, Art. no. 101772, 2025, doi: 10.1016/j.appdev.2025.101772.

[38] M. C. Lai, M. V. Lombardo, and S. Baron-Cohen, "Autism," The Lancet, vol. 383, no. 9920, pp. 896–910, 2014, doi: 10.1016/S0140-6736(13)61539-1.

[39] C. Cordero et al., "Neonatal jaundice in association with autism spectrum disorder and developmental disorder," Journal of Perinatology, vol. 40, no. 2, pp. 219–225, 2020, doi: 10.1038/s41372-019-0452-4.

[40] R. D. Maimburg, B. H. Bech, M. Væth, B. Møller-Madsen, and J. Olsen, "Neonatal jaundice, autism, and other disorders of psychological development," Pediatrics, vol. 126, no. 5, pp. 872–878, 2010, doi: 10.1542/peds.2010-0052.