Applied Computer Science and Software Engineering (ACSSE)
Vol. 2 No. 2, September 2026 | E-ISSN: 3064-4658
Journal URL: https://journal.zahirastudies.org/index.php/acsse
DOI : https://doi.org/10.63254/acsse.v2i2.49

Knowledge-Based Label Construction and Machine Learning Classification for ASD Technology Recommendation

Jeyarani Periasamy[1], Lukminal Feri Julian[*2]
[1]Faculty of Data Science and Information Technology, INTI Interational University, Malaysia
[2] Department of Information Technology, Faculty of Computer Science, Universitas Amikom Purwokerto, Indonesia
e-mail: jeyarai.periasamy@newinti.edu.my[1], lukminalferi@gmail.com [2]

Article History

Received : 26 June 2026
Revised : 27 August 2026
Accepted : 10 September 2026
Published : 18 September 2026

Keywords: *Autism Spectrum Disorder*, AQ-10, *Knowledge-Based Labeling*, *Machine Learning*, SHAP.

Abstract

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder with heterogeneous behavioral profiles, including impairments in communication, social interaction, repetitive behaviors, and sensory sensitivities. Public ASD datasets only provide binary diagnostic labels without technology intervention recommendations, leaving behavioral profiles underutilized for recommendation purposes. This study proposes a knowledge-based labeling approach using soft scoring to construct five technology recommendation labels: AAC, Serious Game, Social Robot, Sensory App, and Video Modeling, from AQ-10 behavioral profiles derived from verified clinical literature. The dataset combines UCI ML Repository (ID=419) and Kaggle All Ages Combined, totaling 5,505 instances with 14 input features. Seven machine learning algorithms were compared using test-set evaluation, 10-Fold Stratified Cross-Validation, hyperparameter tuning, and SHapley Additive exPlanations (SHAP), with class imbalance handled via Synthetic Minority Oversampling Technique (SMOTE). Tuned Gradient Boosting achieved the best performance with accuracy 96.73%, F1-macro 96.54%, and CV accuracy 94.68% ± 0.79%. SHAP analysis confirmed prediction patterns consistent with the knowledge-based labeling design, while demographic features contributed minimally. The system is designed as a decision-support tool for practitioners, not a substitute for comprehensive clinical assessment.

Corresponding author : email: lukminalferi@gmail.com

1. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by deficits in social communication, repetitive behaviors, and rigid behavioral patterns [1]. The prevalence of ASD continues to rise globally; the WHO reported that 1 in 127 people were diagnosed with ASD in 2021 [2], while the CDC reported that 1 in 36 children in the United States were diagnosed in 2023 [3]. This makes ASD one of the global health priorities requiring targeted treatment and intervention.

ASD is not a uniform condition; each individual has a unique combination of challenges in the domains of communication, social interaction, sensory sensitivity, or repetitive behaviors [1]. Various learning technologies have proven effective as interventions, including *Augmentative and Alternative Communication* (AAC), *serious games*, *social robots*, *sensory apps*, and *video modeling*. However, selecting the right technology depends heavily on the individual's specific behavioral profile; a technology that is effective for one profile may not necessarily be optimal for another [4].

A number of studies have applied *machine learning* to public ASD datasets to support the diagnostic process, but all of them have focused on binary classification (ASD/Non-ASD). Dheyaa and Dheyaa [5] compared seven algorithms with accuracies of up to 97%; Aldrees et al. [6] integrated the Synthetic Minority Oversampling Technique (SMOTE) and SHapley Additive exPlanations (SHAP) to address class imbalance; and Abdelwahab et al. [7] evaluated six models in a pediatric population. The

datasets used in these studies contain ten information-rich AQ-10 behavioral items, but none of them utilize them to recommend technology-based interventions tailored to individual behavioral profiles.

This study proposes *a soft-scoring-based knowledge-based labeling* approach to construct five technology recommendation labels: AAC, *Serious Game*, *Social Robot*, *Sensory App*, and *Video Modeling* from the AQ-10 behavioral profiles in a combined dataset from the UCI ML Repository (ID=419) [8] and Kaggle *All Ages Combined* [9], comprising a total of 5,505 instances. Labeling rules were derived from verified clinical literature by mapping each AQ-10 behavioral domain to the most clinically relevant technology. Seven *machine learning* algorithms were then trained and compared on this labeled dataset, using SHAP analysis [10] to validate that the patterns learned by the models are consistent with the clinical evidence underlying the labeling process.

This study aims to develop a learning technology recommendation system for individuals with ASD based on the AQ-10 behavioral profile, with three main contributions: (1) a *knowledge-based labeling* methodology as a bridge between behavioral data and technology recommendations; (2) a systematic comparison of seven *machine learning* algorithms in the context of multi-class classification for ASD technology recommendations; and (3) SHAP analysis to validate the consistency of model patterns with clinical evidence. The resulting system is designed as *a decision-support tool* for practitioners and therapists, not as a substitute for comprehensive clinical assessment.

2. METHOD

2.1 Research Workflow

This study consists of several main stages conducted systematically, as shown in Figure 1. The stages begin with the collection of the ASD public dataset, followed by preprocessing and *knowledge-based labeling* to construct technology recommendation labels, then the training and evaluation of seven *machine learning* algorithms, and finally an *explainability* analysis using SHAP.

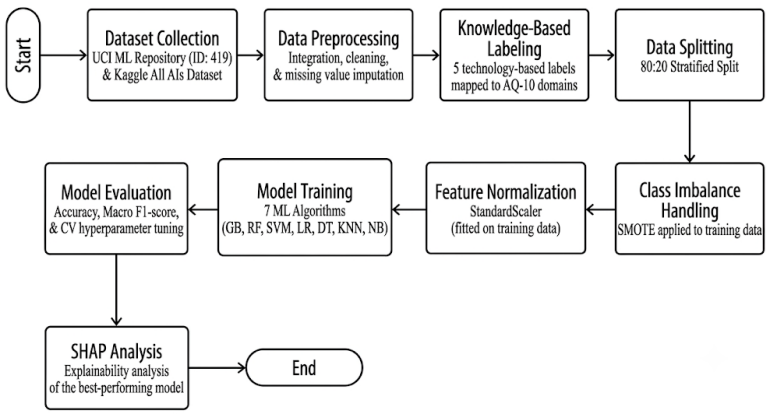

Figure 1. Research Workflow

2.2 Dataset

This study uses two public datasets that were combined into a single integrated dataset: the UCI Machine Learning Repository (ID=419), which contains 292 ASD screening instances for children [8], licensed under CC BY 4.0; and the Kaggle *All Ages* Combined dataset [9], which covers age groups ranging from *toddlers* to adults and contains 6,075 instances. Differences in column names across sources were resolved through standardization prior to merging.

The *data cleaning* process includes investigating *missing values*, removing duplicates, imputing numeric features (A1–A10, age) with the median and categorical features (gender, jaundice, family_asd) with the mode, and validating the final results. The final dataset consists of 5,505 instances with 14 input features: ten binary AQ-10 behavioral items (A1–A10), one continuous numeric feature (age, range 1.5–

80 years), and three binary categorical features (gender, jaundice, family_asd). A summary of the dataset is presented in Table 1.

Table 1. Dataset Summary

Description	Details
Source	UCI ML Repository (ID=419) + Kaggle All Ages Combined
Total Instances (after cleaning)	5,505
Input Features	14 (A1–A10, age, gender, jaundice, family_asd)
Age Range	1.5–80 years
Original Dataset Label	Binary diagnosis (ASD / Non-ASD)
Constructed Labels	5 technology recommendation classes

2.3 Knowledge-Based Labeling

Since public ASD datasets only provide binary diagnostic labels, this study developed a *soft-scoring*, *knowledge-based labeling* approach to construct labels for machine learning models. AQ-10 consists of ten items grouped into four domains: communication (A1, A2), social (A3, A4, A5), repetitive (A6, A7), and sensory (A8, A9, A10) [11]. The mapping of domains to recommended technologies is presented in Table 2.

Table 2. Mapping of AQ-10 Domains to Learning Technologies

Technology	Primary Domain	AQ-10 Items	Clinical References
AAC	Communication	A1, A2	[12] [13]
Serious Game	Repetitive	A6, A7	[14] [15]
Social Robot	Social	A3, A4, A5	[16] [17]
Sensory App / VR	Sensory	A8, A9, A10	[18] [19]
Video Modeling	Mixed	A1–A10	[20] [21]

The labeling mechanism calculates a score for each domain, normalizes it to the range [0,1], and then calculates *a composite score* for each technology based on the domain weights from the literature; the final label is the technology with the highest *composite score*. The *soft scoring* approach was chosen due to the high heterogeneity of ASD profiles; a single individual may have difficulties in more than one domain, so *a* binary *hard threshold* would result in overly rigid labeling [1] [22].

The labeling results yielded a five-class distribution as shown in Figure 2: Video Modeling 1,691 instances (30.7%), Sensory App 1,212 (22.0%), Serious Game 1,033 (18.8%), Social Robot 793 (14.4%), and AAC 776 (14.1%). The dominance of Video Modeling is clinically reasonable because it encompasses mixed behavioral profiles from all domains. Validation via boxplots of domain score distributions (Figure 3) confirms that each class consistently exhibits distinct domain profiles, consistent with the underlying clinical literature.

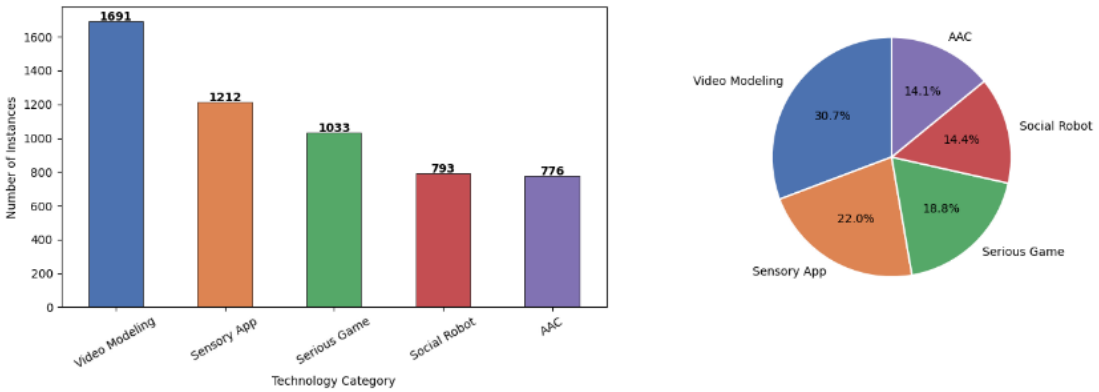

Figure 2. Distribution of Technology Recommendation Labels

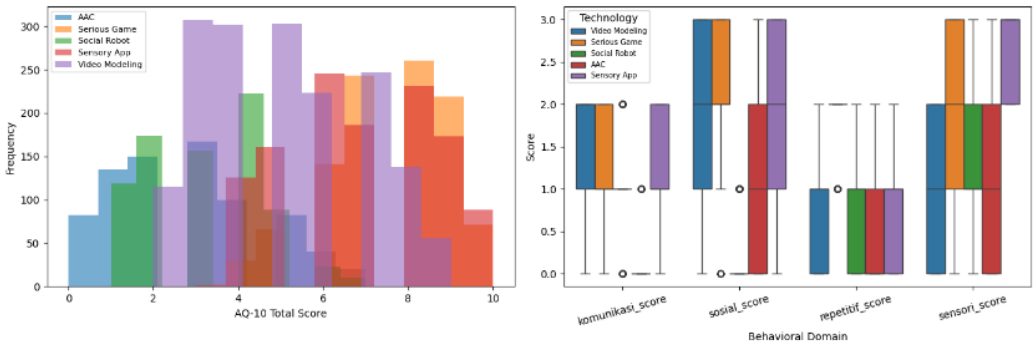

Figure 3. Distribution of Domain Scores by Technology Category

2.4 Data Split and Class Imbalance Handling

The dataset was split using an 80:20 stratified split, resulting in 4,404 training instances and 1,101 test instances, with class proportions preserved in both subsets. The split was performed before *oversampling* to prevent *data leakage,* a condition in which information from the test data indirectly influences the training process, which can lead to invalid performance estimates [23] .

The training data exhibited class imbalance with a ratio of 0.459, so it was addressed using SMOTE, which generates synthetic samples through linear interpolation between minority instances and their neighbors in the feature space [24]. Mathematically, synthetic samples x_{new} are generated using the following equation:

$$x_{new} = x_i + \lambda \times (x_{zi} - x_i) \quad (1)$$

Where x_i is an instance of the minority class, x_{zi} is one of the k-nearest neighbors x_i , and λ is a random value in the range 0 to 1 [24]. After SMOTE, the training data increased to 6,765 instances with a balanced distribution (1,353 instances per class), while the test data remained unchanged so that it continued to reflect the original distribution. After SMOTE, all input features were normalized using *StandardScaler* to ensure each feature has a mean of 0 and a standard deviation of 1 [25]. Normalization was applied by *fitting the model* to the training data and transforming both the training and test data, rather than *fitting the model* to the entire dataset to prevent statistical information from leaking from the test data into the training process.

2.5 Machine Learning Algorithms

This study compares seven *machine learning* algorithms representing three groups of approaches: decision tree-based (*ensemble* and single), function-based, and distance-based. All algorithms were trained on SMOTE-processed training data and evaluated on unmodified test data using the Scikit-learn implementation with *default* parameters as the initial configuration [25]. A summary of each algorithm's characteristics is presented in Table 3.

Table 3. Summary of Machine Learning Algorithms

Algorithm	Group	Key Characteristics	Reference
Gradient Boosting (GB)	Ensemble Boosting	Builds trees sequentially, correcting errors from previous trees through gradient optimization; robust against nonlinear interactions	[26]
Random Forest (RF)	Ensemble bagging	Building trees in parallel on random subsets; resists overfitting and produces *feature importance* estimates	[27]
Support Vector Machine (SVM)	Function-based	Finds the optimal *hyperplane* with maximum margin; effective on high-dimensional data using the *one-vs-rest* strategy	[28]

Logistic Regression (LR)	Function-based	A linear model with a sigmoid function; serves as a robust and easily interpretable *baseline*	[28]
Decision Tree (DT)	Single tree	Splits data recursively based on *information gain*; transparent but prone to *overfitting* without pruning	[28]
K-Nearest Neighbors (KNN)	Distance-based	Classification based on voting among the k nearest neighbors; a *lazy learner* that is sensitive to feature scale	[28]
Naive Bayes (NB)	Probability-based	Applies Bayes' theorem under the assumption of feature independence; computationally efficient	[28]

2.6 Model Evaluation

Model performance evaluation was conducted using four complementary approaches: evaluation on test data, *cross-validation*, *hyperparameter tuning*, and SHAP-based *explainability* analysis.

2.6.1 Evaluation Metrics

The primary metrics used are *accuracy* and *macro-averaged F1-score*. *Accuracy* measures the proportion of correct predictions relative to all test data instances, formulated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

For multi-class classification, *the macro-averaged F1-score* is used as the primary metric because it calculates the F1-score for each class independently and then averages them, thereby giving equal weight to each class without being affected by distribution imbalances [29]. The F1-score is the harmonic mean of *precision* and *recall*:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{5}$$

$$F1_{macro} = \frac{1}{K} \sum_{k=1}^{K} F\,1_k \tag{6}$$

Where K is the number of classes ($K = 5$ in this study) and $F1_k$ is the F1-score for the class k [29].

2.6.2 Cross-Validation

To measure the stability and generalization of the model more robustly, *10-fold stratified cross-validation* was performed on all seven algorithms. The data was stratified into 10 *folds*; the model was trained on 9 *folds* and tested on 1 *fold* in turn, and then the mean and standard deviation of the accuracy were reported. A low standard deviation indicates a stable model that is not sensitive to variations in data partitioning [30].

2.6.3 Hyperparameter Tuning

Hyperparameter tuning was performed on the two best models, Random Forest and Gradient Boosting, using *GridSearchCV* with 5-fold cross-validation. These two models were selected based on their status as the top-performing models in the test data evaluation (Table 3); therefore, tuning was focused on the models most likely to yield significant improvements.

For Random Forest, the grid parameters include n_estimators ∈ {100, 200, 300}, max_depth ∈ {5, 8, 10, 15}, max_features ∈ {*sqrt*, *log2*}, min_samples_split ∈ {5, 10, 15}, and min_samples_leaf ∈ {3, 5, 8}, resulting in 216 combinations. For Gradient Boosting, the parameter grid includes n_estimators ∈ {100, 200, 300}, max_depth ∈ {3, 5, 7}, learning_rate ∈ {0.05, 0.1, 0.2}, subsample ∈ {0.8, 1.0}, and max_features ∈ {*sqrt*}, resulting in 108 combinations. *GridSearchCV* was chosen because *exhaustive* search in a well-defined parameter space is more reliable than *random search* for moderate grid sizes [25].

2.6.4 SHAP Explainability

To understand the contribution of each feature to the model's prediction, we used SHapley Additive exPlanations (SHAP), developed by Lundberg and Lee [10]. SHAP measures the marginal contribution of each feature to the prediction by calculating the Shapley value from cooperative game theory. The SHAP value ϕ_i for the feature i is formulated as:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad (7)$$

Where F is the set of all features, S is a subset of features excluding the feature i , and $f(S)$ is the model's prediction using the feature subset S. SHAP was applied to a *tuned* Random Forest model using TreeExplainer, which is optimized for tree-based models and thus capable of calculating SHAP values exactly with polynomial complexity [31]. The SHAP analysis was conducted at two levels: global feature importance to examine the average contribution of each feature overall, and per-class SHAP values to understand which features most influence the predictions for each technology class.

3. RESULTS AND DISCUSSION

3.1 Model Performance Comparison

Seven machine learning algorithms were evaluated on a test dataset of 1,101 unmodified instances. The complete evaluation results are presented in Table 4.

Table 4. Performance Comparison of 7 Machine Learning Models

Model	Accuracy	F1-Macro	Precision-Macro	Recall-Macro	Train Accuracy	Overfitting Gap
Gradient Boosting	95.46%	95.53%	95.91%	95.21%	95.93%	0.0048
Random Forest	91.19%	90.89%	91.90%	90.28%	95.28%	0.0409
SVM	89.74%	89.54%	89.57%	89.55%	93.29%	0.0355
Logistic Regression	88.83%	88.18%	88.23%	88.20%	89.37%	0.0054
Decision Tree	88.65%	88.39%	88.36%	88.47%	93.27%	0.0463
KNN	79.84%	79.56%	79.36%	80.55%	91.19%	0.1135
Naive Bayes	74.75%	74.71%	74.31%	75.82%	78.82%	0.0407

Gradient Boosting ranked highest with *an accuracy* of 95.46% and *an F1-macro* of 95.53%, followed by Random Forest (91.19%) and SVM (89.74%). Five of the seven models exceeded the 85%

threshold set as the minimum performance threshold for the multiclass classification task [28], while KNN (79.84%) and Naive Bayes (74.75%) fell below that *threshold*. In terms of generalization, Gradient Boosting also showed the smallest *overfit gap* (0.0048), indicating a stable model that does not merely memorize the training data [26]. Meanwhile, KNN had the highest *overfit gap* (0.1135) due to its sensitivity as *a lazy learner* to the data distribution [28]. Logistic Regression is worth noting: although its accuracy is lower than that of SVM and Random Forest, its *overfit* gap is only 0.0054, nearly on par with Gradient Boosting, demonstrating high consistency despite its simple model [25].

3.2 Overfitting Analysis

Overfitting was detected by comparing *the training accuracy* and *test accuracy* of each model, as shown in Figure 4. Gradient Boosting and Logistic Regression exhibited the smallest *overfit gaps* (0.0048 and 0.0054), indicating good generalization, while KNN recorded the largest gap (0.1135) due to its sensitivity as *a lazy learner* [28]. Overall, none of the models experienced severe *overfitting*, and the exclusive application of SMOTE to the training data proved effective in preventing information leakage [23].

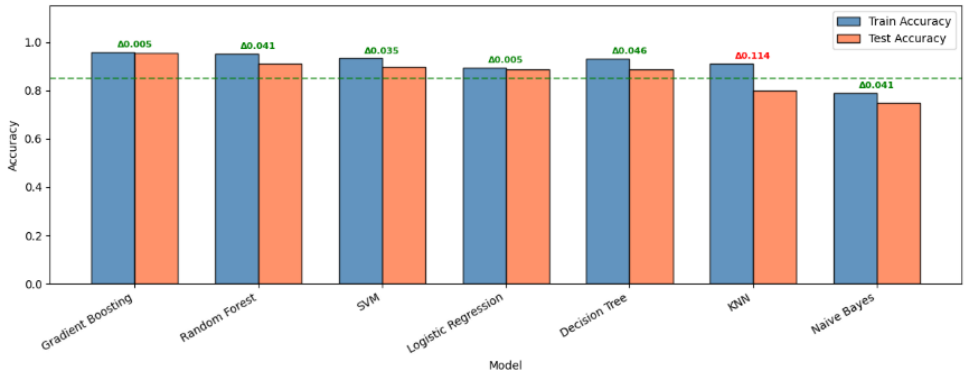

Figure 4. Comparison of Train vs. Test Accuracy for Overfitting Detection

3.3 Cross-Validation

The results of the 10-fold stratified cross-validation are presented in Table 5. Gradient Boosting once again ranked first with a CV accuracy of 94.68% ± 0.79%, followed by Random Forest at 92.42% ± 0.73%. These results are consistent with the evaluation on the test set, confirming that Gradient Boosting's performance is not a coincidence resulting from a specific data partition but rather reflects the model's true capability [30].

Table 5. Results of 10-Fold Stratified Cross-Validation

Model	CV Accuracy	Std Dev
Gradient Boosting	94.68%	±0.79%
Random Forest	92.42%	±0.73%
SVM	90.85%	±0.85%
Decision Tree	89.80%	±1.10%
Logistic Regression	89.02%	±0.59%
KNN	85.82%	±0.71%
Naive Bayes	78.79%	±1.19%

The standard deviation of all models falls within a narrow range (0.59% to 1.19%), indicating good stability across all *folds*. Logistic Regression has the lowest standard deviation (±0.59%), consistent with its small *overfit* gap. It should be noted that the CV result for KNN (85.82%) is higher than its *test* accuracy (79.84%) because CV was run on SMOTE-balanced training data, while the test data still reflects the original unbalanced distribution; KNN's high *overfit gap* (0.1135) indicates that this model is sensitive to the composition of the data used [28].

3.4 Confusion Matrix Analysis

The confusion matrices for the two best models are shown in Figure 5. Gradient Boosting exhibits high diagonal values across all classes, with the highest error occurring in the Sensory App class, which was predicted as Video Modeling, a pattern that can be clinically explained because these two classes have the greatest domain overlap. Random Forest shows more errors, particularly in the Sensory App and Social Robot classes, which are often predicted as Serious Game and Video Modeling, confirming that Gradient Boosting is superior at distinguishing classes with clinical overlap.

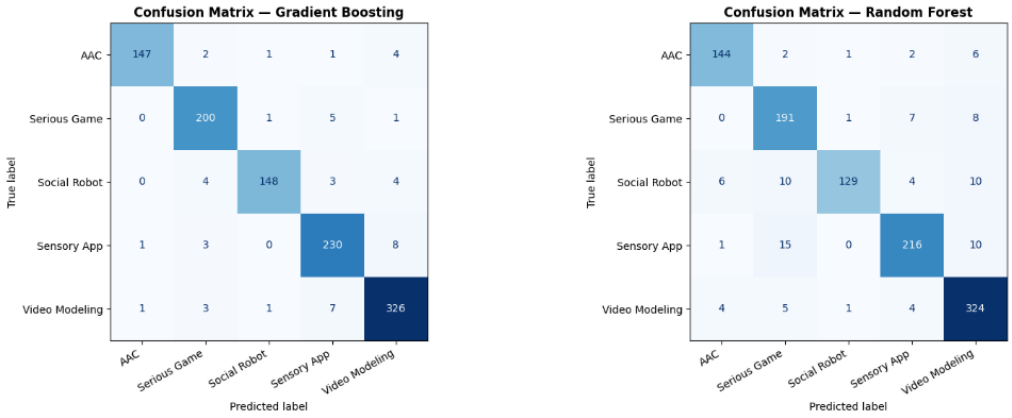

Figure 5. Confusion Matrix of the Top 2 Best Models

3.5 Hyperparameter Tuning for Random Forest and Gradient Boosting

Hyperparameter tuning was performed on the two best models using GridSearchCV with 216 parameter combinations for Random Forest and 108 combinations for Gradient Boosting, each with 5-fold cross-validation. The tuning results are presented in Table 6.

Table 6. Hyperparameter Tuning Results for Random Forest and Gradient Boosting

Configuration	Accuracy	F1-Macro	Overfit Gap
Random Forest (default)	91.19%	90.89%	0.0409
Random Forest (tuned)	93.19%	92.90%	0.0415
Gradient Boosting (default)	95.46%	95.53%	0.0048
Gradient Boosting (tuned)	96.73%	96.54%	0.0247

The tuned Random Forest achieved an accuracy of 93.19% and an F1-macro of 92.90%, an increase of 2.00 percentage points from the baseline of 91.19%. The optimal parameters found were max_depth=15, max_features=sqrt, min_samples_leaf=3, min_samples_split=5, and n_estimators=300. The overfit gap increased slightly from 0.0409 to 0.0415, but remained in *the good fit* category[27] . Tuned Gradient Boosting achieved an accuracy of 96.73% and an F1-macro of 96.54%, an increase of 1.27 percentage points from the baseline of 95.46%. The optimal parameters found were learning_rate=0.1, max_depth=5, max_features=sqrt, n_estimators=200, and subsample=0.8. Although the overfit gap increased from 0.0048 to 0.0247, this value is still well below the threshold for significant *overfitting* and remains in *the good fit* category [25]. This increase in the overfit gap is a natural consequence of parameter optimization, which encourages the model to further explore patterns in the training data; however, the model's generalization capability remains intact, as reflected in the high test accuracy.

Overall, Gradient Boosting emerged as the best-performing model in this study, with an accuracy of 96.73% and an F1-macro of 96.54%, confirming that the boosting mechanism, which gradually corrects the errors of previous models, provides a consistent advantage even after the parameter optimization process for both models [26].

3.6 SHAP Analysis

3.6.1 Global Feature Importance

SHAP analysis was applied to the tuned Random Forest model (93.19%) as a simpler *ensemble* representation, with a performance difference of about two percentage points from Gradient Boosting (95.46%), ensuring that the identified dominant feature patterns remain representative.

The five features with the highest SHAP contributions, in order, are A1 (0.0703), A2 (0.0607), A7 (0.0563), A6 (0.0556), and A9 (0.0488). The dominance of A1–A2 is consistent with the role of the communication domain as the primary classifier for AAC, while A6–A7 (repetitive) and A9 (sensory) follow next. The demographic features gender, family_asd, and jaundice show contributions below 0.005, confirming that recommendations are more strongly determined by the AQ-10 behavioral profile than by demographic characteristics [11]. A comparison with *impurity-based importance* showed consistency for A1, A2, A7, and A6, with the only notable difference being the "age" feature, which was overestimated in RF importance compared to SHAP [29].

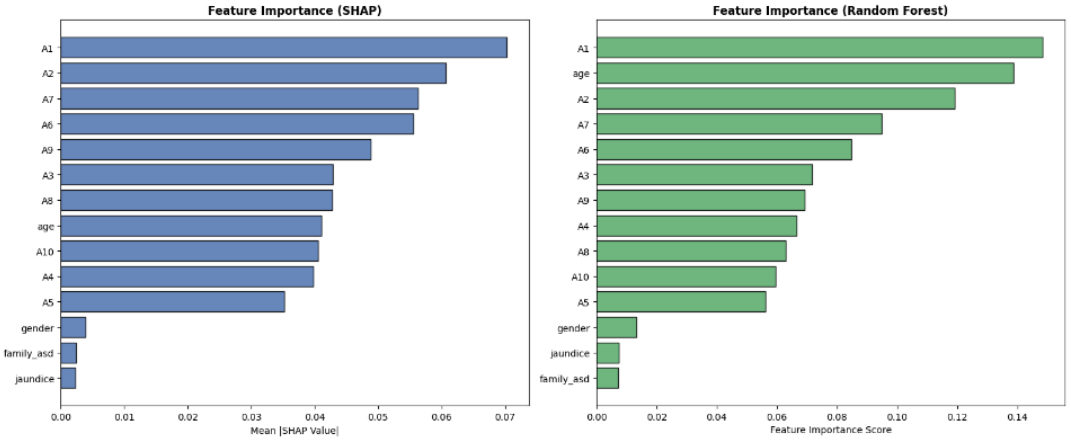

Figure 6. Feature Importance Comparison: SHAP vs. Random Forest

3.6.2 SHAP by Technology Class

The SHAP analysis by class, as shown in Figure 7, reveals distinct patterns of dominant features for each class, consistent with the *knowledge-based labeling* design. The AAC class is dominated by A1 and A2 (communication), the Serious Game class by A6 and A7 (repetitive), and the Social Robot class by A3, A4, and A5 (social); all three indicate that high values on key domain features (red dots) drive predictions toward the corresponding class. The Sensory App class exhibits a different pattern: A9, A8, and A10 (sensory) actually drive predictions toward this class when their values *are low* (red dots), rather than high, indicating that the sensory profile in this dataset is characterized by low scores on these items. The Video Modeling class has a combination of features from various domains (A2, A1, A9, A10, A6), consistent with its characteristics as a mixed behavioral profile class. Across all classes, the demographic features gender, family_asd, and jaundice consistently ranked at the bottom, confirming the global SHAP findings in subsection 3.6.1 that model decisions are dominated by the AQ-10 behavioral profile rather than demographic characteristics.

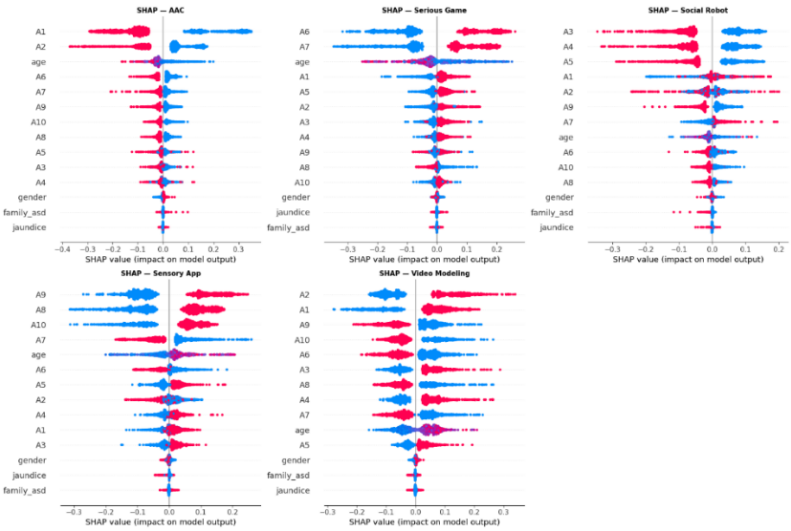

Figure 7. SHAP Values by Technology Class

3.7 Discussion

The performance of the tuned Gradient Boosting model (accuracy 96.73%, overfit gap 0.0247) is explained by the boosting mechanism, which iteratively corrects residual errors, making it effective on datasets with predominantly binary features and overlapping class boundaries [26]. Random Forest is more susceptible to ambiguities in similar domain profiles, as seen in the classification errors for Social Robot → Serious Game and Sensory App → Video Modeling in the confusion matrix. This advantage also confirms that the *soft scoring* mechanism, which preserves the ambiguity of mixed profiles in the Video Modeling label, is successfully represented statistically by the model.

SHAP findings indicate that the AQ-10 behavioral signal is strong enough to distinguish technology profiles without domain-specific information being explicitly included as an additional feature [10], meaning that labels constructed through *knowledge-based labeling* are statistically significant and can be learned by the model. In practical terms, this system has the potential to serve as *an* initial *decision-support tool* for therapists prior to in-depth clinical assessments. A direct comparison with Dheyaa and Dheyaa [5] or Abdelwahab et al. [7] is not possible due to fundamental differences in task objectives: previous studies addressed binary diagnostic classification, whereas this study addresses multi-class technology recommendation classification using self-constructed labels.

The study's limitations include labels that have not been validated by ASD practitioners, a dataset from an international population that limits local representation, and SHAP analysis that does not yet include Gradient Boosting. Further research is recommended to involve clinical validation, the development of local datasets, and the expansion of *explainability* analysis on the best model.

4. CONCLUSION

This study successfully constructed five technology recommendation labels; AAC, *Serious Games*, *Social Robots*, *Sensory App*, and *Video Modeling* from AQ-10 behavioral profiles using *a soft-scoring-based knowledge-based labeling* approach on a public ASD dataset that had previously been used only for binary diagnostic classification. Of the seven algorithms compared across 5,505 instances, tuned Gradient Boosting achieved the best performance with *an accuracy* of 96.73%, *an F1-macro* of 96.54%, and *a* CV *accuracy* of 94.68% ± 0.79% with *an overfit gap* of 0.0247, indicating good generalization ability. SHAP analysis confirmed that the model's prediction patterns were consistent with the *knowledge-based labeling* design, indicating that the AQ-10 behavioral signals were strong enough to statistically distinguish technology profiles. Limitations of the study include labels that have not been validated by ASD practitioners, a dataset from an international population, and SHAP analysis that does not yet include Gradient Boosting; future research is recommended to involve clinical validation, the development of local datasets, and the expansion *of explainability* for the best-performing model.

ACKNOWLEDGMENTS

The author would like to express his deepest gratitude to Amikom University Purwokerto, particularly the Information Technology Study Program, for providing facilities, an academic environment, and guidance throughout this research process. The author also thanks all the faculty members who provided scientific guidance and motivation in the preparation of this article. Thanks are also extended to the UCI Machine Learning Repository and Kaggle for providing the public datasets used in this research.

BIBLIOGRAPHY

[1] H. Hodges, C. Fealko, and N. Soares, "Autism spectrum disorder : definition , epidemiology , causes , and clinical evaluation," vol. 9, no. 8, 2020, doi: 10.21037/tp.2019.09.09.

[2] World Health Organization, "Autism Spectrum Disorders." Accessed: Jun. 04, 2026. [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/autism-spectrum-disorders

[3] D. Control, "Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network , 11 Sites , United States , 2020," vol. 72, no. 2, 2023, [Online]. Available: https://doi.org/10.15585/mmwr.ss7202a1

[4] I. Scarcella *et al.*, "Information and communication technologies-based interventions for children with autism spectrum conditions : a systematic review of randomized control trials from a positive technology perspective", doi: 10.3389/fpsyt.2023.1212522.

[5] D. Dheyaa and S. Dheyaa, "Measurement : Sensors The classification of autism spectrum disorder by machine learning methods on multiple datasets for four age groups," *Meas. Sensors*, vol. 27, no. March, p. 100774, 2023, doi: 10.1016/j.measen.2023.100774.

[6] A. Aldrees *et al.*, "Data-centric automated approach to predict autism spectrum disorder based on selective features and explainable artificial intelligence", doi: 10.3389/fncom.2024.1489463.

[7] M. M. Abdelwahab, K. A. Al-karawi, E. M. Hasanin, and H. E. Semary, "Autism Spectrum Disorder Prediction in Children Using Machine Learning," vol. 3, pp. 1–9, 2024, doi: 10.57197/JDR-2023-0064.

[8] F. Thabtah, "Autism Screening for Children." [Online]. Available: https://archive.ics.uci.edu/dataset/419

[9] F. Thabtah, "ASD Screening Data for Toddler, Child, Adolescent and Adult." [Online]. Available: https://www.kaggle.com/datasets/fabdelja/asd-screening-data-toddler-child-adoles-adult

[10] S. M. Lundberg and S. Lee, "Predictions," no. Section 2, pp. 1–10, 2017, [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[11] C. Allison, B. Auyeung, and S. Baron-cohen, "Toward Brief ' Red Flags ' for Autism Screening : The Short Autism Spectrum Quotient and the Short Quantitative," *JAAC*, vol. 51, no. 2, pp. 202-212.e7, 2012, doi: 10.1016/j.jaac.2011.11.003.

[12] D. Mavritsakis, "Augmentative and alternative communication in autism spectrum disorder : transitioning from letter board to iPad – a case study," no. May, pp. 1–10, 2024, doi: 10.3389/fpsyt.2024.1345447.

[13] T. Iacono, D. Trembath, and S. Erickson, "The role of augmentative and alternative communication for children with autism : current status and future trends," pp. 2349–2361, 2016, doi: 10.2147/NDT.S95967.

[14] S. Review and M. G. Filipe, "Serious Games for Developing Social Skills in Children and Adolescents with Autism Spectrum Disorder : A Systematic Review," 2024, doi: 10.3390/healthcare12050508.

[15] K. Song, "Rehabilitation therapy for children with autism based on interactive VR-motion serious game intervention : a," no. September, pp. 1–10, 2025, doi: 10.3389/fpubh.2025.1628741.

[16] E. Y. Chung, K. K. Sin, and D. H. Chow, "Qualitative outcomes and impact of a robotic intervention on children with autism spectrum disorder : A multiple embedded case study," 2024,

doi: 10.1177/03080226241252272.
[17] R. Ramnauth, A. S. Arora, A. Arora, and K. Sivakumar, “Managing social-educational robotics for students with autism spectrum disorder through business model canvas and customer discovery,” no. April, pp. 1–17, 2024, doi: 10.3389/frobt.2024.1328467.
[18] C. Failla *et al.*, “Virtual reality for autism : unlocking learning and growth”, doi: 10.3389/fpsyg.2024.1417717.
[19] B. Lyu, Y. Ba, D. Ma, N. Liu, and L. Fu, “Effectiveness of sensory integration-based intervention in autistic children , focusing on Chinese children : a systematic review and meta-analysis,” no. November, 2025, doi: 10.3389/fpsyt.2025.1623149.
[20] C. Bloh, L. Bacon, B. Begel, K. Madara, and B. Koller, “Comparing Human Video Modeling to Animated Video Modeling for Learners with Autism,” *Anal. Verbal Behav.*, vol. 41, no. 2, pp. 262–279, 2025, doi: 10.1007/s40616-025-00224-y.
[21] B. Bengtson, B. Bengtson, and J. Christensen, “The Repository at St . Cloud State A Literature Review : Video Modeling and Autism Spectrum Disorder by,” 2024, [Online]. Available: https://repository.stcloudstate.edu/sped_etds/219
[22] J. Herrero-mart, “Development of personalized pro fi les of students with autism spectrum disorder for interactive interventions with robots to enhance language and social skills,” no. November, pp. 1–10, 2024, doi: 10.3389/fpsyt.2024.1455627.
[23] S. Kapoor and A. Narayanan, “Article Leakage and the reproducibility crisis in machine-learning-based science Leakage and the reproducibility crisis in machine-learning-based science,” *Patterns*, vol. 4, no. 9, p. 100804, 2023, doi: 10.1016/j.patter.2023.100804.
[24] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE : Synthetic Minority Over-sampling Technique,” vol. 16, pp. 321–357, 2002, [Online]. Available: https://www.jair.org/index.php/jair/article/view/10302
[25] F. Pedregosa, R. Weiss, and M. Brucher, “Scikit-learn : Machine Learning in Python,” vol. 12, pp. 2825–2830, 2011, [Online]. Available: https://jmlr.org/papers/v12/pedregosa11a.html
[26] J. H. Friedman, “Greedy Function Approximation: A Gradient Boosting Machine,” *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451.
[27] A. Cutler, D. R. Cutler, and J. R. Stevens, “Random Forests,” in *Ensemble Machine Learning: Methods and Applications*, Y. Zhang, Cha; Ma, Ed., Springer, 2012, pp. 157–175. doi: 10.1007/978-1-4419-9326-7_5.
[28] V. Sheth, U. Tripathi, and A. Sharma, “A Comparative Analysis of Machine Learning Algorithms for Classification Purpose,” *Procedia Comput. Sci.*, vol. 215, pp. 422–431, 2022, doi: 10.1016/j.procs.2022.12.044.
[29] E. Bagli and G. Visani, “Metrics for Multi-Class Classification: an Overview,” pp. 1–17, 2020, doi: 10.48550/arXiv.2008.05756.
[30] R. Kohavi and S. Edu, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” pp. 1137–1143, 1993.
[31] S. M. L. et Al., “From local explanations to global understanding with explainable AI for trees,” *Nat. Mach. Intell.*, vol. 2, no. 1, pp. 56–67, 2020, doi: 10.1038/s42256-019-0138-9.