



OPEN ACCESS

EDITED BY
Tao Wang,
Northwestern Polytechnical University,
China

REVIEWED BY
Roberta Simeoli,
University of Naples Federico II, Italy
Mohammad Hamdan,
Yarmouk University, Jordan

*CORRESPONDENCE
Yelda Firat
✉ yelda.firat@mudanya.edu.tr

RECEIVED 24 November 2025
REVISED 01 February 2026
ACCEPTED 11 March 2026
PUBLISHED 23 March 2026

CITATION
Firat Y (2026) SHAP-based
explainable AI framework for
autism severity classification
using 3D motor biomarkers.
Front. Psychiatry 17:1751654.
doi: 10.3389/fpsy.2026.1751654

COPYRIGHT
© 2026 Firat. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance
with accepted academic practice. No
use, distribution or reproduction is
permitted which does not comply with
these terms.

SHAP-based explainable AI framework for autism severity classification using 3D motor biomarkers

Yelda Firat*

Department of Computer Engineering, Mudanya University, Bursa, Türkiye

Introduction: Early Autism spectrum disorder (ASD) diagnosis is critical for intervention, yet current methods rely on subjective clinical observations. This study develops objective tools to classify ASD severity using 3D motor movement analysis, addressing motor abnormalities as core diagnostic features.

Methods: A Random Forest (RF) model classified three severity levels using 463 motor features from 25 Kinect V2 joint points. Data from 109 children (50 typical, 50 moderate ASD, 9 severe ASD) were validated via 5-fold cross-validation and two held-out test sets (20% each). Shapley Additive Explanations (SHAP) analysis identified critical motor biomarkers.

Results: The model achieved $84.6 \pm 10.9\%$ accuracy (5-fold cross-validation) and 86.4% accuracy (internal and held-out test sets). For severe ASD, the model achieved 100% classification accuracy on synthetic test data (4/4 cases; 95% CI: 39.8%-100.0%). However, this result represents a methodological proof-of-concept rather than clinical validation, as severe ASD features were synthetically generated from moderate ASD data and the model has not been validated on real Kinect-derived severe ASD motor data. SHAP analysis identified wrist movements, knee trajectories, and elbow-to-foot distances as key motor biomarkers for severity classification.

Discussion: This Kinect-based approach with RF and SHAP offers effective, interpretable ASD severity assessment for typical and moderate ASD classes, with promising methodological foundations for severe ASD pending validation on real data.

KEYWORDS

autism spectrum disorder, motor biomarkers, random forest, SHAP analysis, violence level classification

1 Introduction

Autism spectrum disorder (ASD), affecting 1 in 36 U.S. children (1), is characterized by social communication difficulties, repetitive behaviors, and restricted interests. Current diagnostic methods rely on subjective clinical observation, often delaying diagnosis beyond age four and missing early intervention opportunities. Motor abnormalities—balance, gait, postural control, and stereotypical movements—are core ASD features. Motor impairments may serve as potential early diagnostic biomarkers (2, 3).

Machine learning and 3D motion analysis enable objective assessment of ASD. Microsoft Kinect captures motor characteristics in three dimensions at low cost. Machine

learning achieves high accuracy in classifying ASD patterns. Random Forest (RF) offers performance and flexibility, while Shapley Additive Explanations (SHAP) facilitates clinical interpretation and increases professional confidence (4).

Balance, gait, and motor coordination impairments in children with ASD (3) serve as potential biomarkers. Gait analysis reveals timing variability (5) and differences in spatiotemporal parameters (6) compared to typical development. Machine learning combined with gait deviations enables effective early detection (7), demonstrating motor characteristics as objective diagnostic markers.

3D skeleton-based motion analysis is widely used in ASD detection. Kinect-based approaches include deep learning (8), walking analysis (9), stereotypical movement tracking (hand flapping, rocking, spinning) (10), and real-time action recognition via convolutional neural networks (11). The Kinect V2 dataset (12) provides resources for analyzing motor profiles in ASD and typical development.

Automatic stereotypical motor movement (SMM) detection is important in ASD diagnosis. Studies employed deep learning for video analysis (13), lower-body movement tracking (14), and wearable sensors (15), demonstrating that movement analysis provides objective diagnostic data.

Machine learning in ASD detection has increased rapidly. Studies show promise for gait analysis (16), high accuracy via dual-modal features (17), social interactions (18), and minimal medical data (19), and voice/behavioral data analysis (20), demonstrating effectiveness.

RF is widely used in ASD classification for behavioral aggression (21), cognitive profiling (22), and biomarker discovery across ages and intelligence levels (23), demonstrating high performance and interpretability.

Explainable AI (XAI) is increasingly important in ASD detection. Studies examined XAI applications (4), developed high-accuracy models (24), revealed subtype differences (25), and analyzed functional connectivity (26). SHAP effectively explains predictions and identifies important features.

Motor skills are important in ASD diagnosis. Motor and sensory characteristics improve classification with clinical scales

(27), while atypical kinematics reveal control problems (28). Motor characteristics are fundamental for ASD diagnosis.

Motor analysis and machine learning assess ASD but focus on binary classification, leaving a research gap in severity-level classification and interpretability.

Building on these developments, several studies have demonstrated the potential for objective ASD classification through motor analysis using diverse sensor technologies. Early pioneering work by Crippa et al. (29) achieved 96.7% accuracy using Kinect-based upper limb kinematics with Support Machine Vectors (SVM), while Anzulewicz et al. (30) demonstrated 93% accuracy through tablet-based gesture analysis using ensemble methods. Li et al. (31) explored imitation-based classification using motion capture, achieving 73% accuracy with SVM and RF. More recently, Freud et al. (32) achieved over 84% accuracy in distinguishing autistic from non-autistic young adults using grasping kinematics captured with only two passive markers, demonstrating that minimal sensor configurations can provide robust classification performance. Similarly, Kojovic et al. (33) employed 2D video-based pose estimation combined with convolutional neural networks to achieve 80.4% accuracy in classifying young children with ASD. More advanced approaches include Su et al. (34), who achieved 78.1% accuracy using deep learning (Multilayer Perceptron -MLP) to classify reach-and-place movements captured with Inertial Measurement Unit (IMU) sensors, and Altozano et al. (35), who introduced 3DCNN ResNets for full-body kinematic assessment in virtual reality environments, achieving $85 \pm 3\%$ accuracy. These studies highlight the growing evidence base for motor-based biomarkers in ASD assessment across different age groups, sensor modalities, and motor tasks. A comprehensive summary of machine learning studies for ASD classification using motor features is presented in Table 1.

As shown in Table 1, existing machine learning studies for ASD classification using motor features predominantly focus on binary classification (ASD vs. typical development), with limited attention to severity-level differentiation. The specific objectives of this study are: (1) to develop a RF-based classification model for distinguishing three

TABLE 1 Summary of machine learning studies for ASD classification using motor features.

Study	Year	Sensor	Sample	ML method	Accuracy	Key features
Crippa et al.	2015	Kinect	30	SVM	96.7%	Upper limb
Anzulewicz et al.	2016	Tablet	82	Ensemble	93%	Gestures
Li et al.	2017	Motion capture	28	SVM, RF	73%	Imitation
Kojovic et al.	2021	2D video	136	CNN	80.4%	Full-body pose
Simeoli et al.	2024	Review	Various	Various	70-97%	Motor patterns
Su et al.	2024	IMU	41	MLP	78.1%	Reach-place
Altozano et al.	2025	VR	Various	3DCNN	$85 \pm 3\%$	Full-body
Shin et al.	2025	Kinect v2	100	Dual-stream	95.4%	Skeleton
Freud et al.	2025	Motion tracking	60	Multiple	>84%	Grasping
Ganai et al.	2025	Gait	Various	SVM, RF	Various	Gait
Present study	2026	Kinect V2	109	RF	86.4%	3D skeleton, SHAP

ASD severity levels (typical development, moderate ASD, and severe ASD) using 3D motor features; (2) to identify and interpret key motor biomarkers contributing to severity classification through SHAP analysis; and (3) to evaluate the model's performance and generalizability using multiple validation strategies including cross-validation and held-out test sets.

2 Methods

2.1 Dataset

This study utilized the “Three-dimensional dataset combining gait and full-body movement of children with autism spectrum disorders” collected using Kinect V2 and published on Zenodo (12). The dataset contains 3D joint coordinates and motor data from 109 children in a controlled laboratory environment. Participants: typically developing (n=50), moderate ASD (n=50), and severe ASD (n=9). Kinect V2 recorded 25 joint points (head, neck, shoulders, elbows, wrists, hands, spine, hips, knees, ankles, feet) in real time (X, Y, Z coordinates).

Dataset creators pre-calculated 463 motor features in Excel format: inter-joint Euclidean distances, range of motion (ROM) metrics, gait parameters (speed, step length, cycle time), and temporal characteristics. Augmented data (7 geometric transformations: rotation, mirroring, scaling) were available for the Typical and Moderate ASD groups; only original recordings were used.

For Severe ASD, the dataset contains 9 video recordings (Samsung Note 9) without Kinect V2 data. Creators could not set up Kinect V2 in controlled environments for severe cases, so they collected videos in clinical settings. Extracting 463 features from videos was technically impossible (missing 3D coordinates, depth information, and time series data). Therefore, synthetic features were created by averaging Moderate ASD vectors and adding Gaussian noise ($\mu=0$, $\sigma=0.15$), preserving the 463-feature structure for consistent comparisons. Critically, this means the model was trained and evaluated on synthetically generated severe ASD data, not real Kinect-derived motor measurements. Consequently, the model has not been validated on real severe ASD Kinect data, and severe ASD classification performance should be interpreted as methodological rather than clinical.

Data split: training (n=65; 30 Typical, 30 Moderate, 5 Severe), internal test (n=22; 10 Typical, 10 Moderate, 2 Severe), held-out test (n=22; 10 Typical, 10 Moderate, 2 Severe). The held-out test set was reserved solely for final validation. Note that both test sets were drawn from the same original dataset using the same sensor, protocol, and feature extraction pipeline, representing reserved subsets from the same data distribution rather than independent external validation. Ethical approval and parental consent were obtained.

2.2 Data processing

The data preprocessing process has been designed as a systematic pipeline comprising stages that transform raw datasets into data ready for model training, as shown in Figure 1.

As shown in Figure 1, data preprocessing involved six stages. First, raw data from 109 children were extracted. Second, 463 video-level statistical features were extracted per child, omitting timestamps to obtain a single 463-dimensional vector. Third, data were combined into a dataframe (109 × 466) with class labels (Typical=0, Moderate ASD = 1, Severe ASD = 2). Fourth, stratified sampling divided data into training, internal test, and held-out test sets as described in Section 2.1. Fifth, Z-score normalization ($\mu=0$, $\sigma=1$) was applied using training set parameters. Finally, data were saved in NumPy and CSV formats with feature names and metadata.

The class distribution of the data set and its division into subsets are shown in Figure 2.

The class distribution is shown in Figure 2. Figure 2A shows 45.9% Typical (n=50), 45.9% Moderate ASD (n=50), and 8.3% Severe ASD (n=9). Figure 2B displays the stratified sampling distribution across subsets. Figure 2C highlights class imbalance, particularly for Severe ASD. The held-out test set was reserved solely for final validation to evaluate real-world performance and minimize overfitting. No missing data were detected. All operations used Python 3.11.

2.3 Feature extraction

This study utilized 463 pre-calculated motor features from 25 Kinect V2 joint points (head, neck, shoulders, elbows, wrists, hands, fingertips, thumbs, spine levels, hips, knees, ankles, feet). Features are categorized into three types: (1) Raw joint coordinates (X, Y, Z axes for each joint); (2) Inter-joint Euclidean distances between all joint pairs, with systematic naming (e.g., EILTFoL_X: left elbow to left foot X-axis distance; FoRTKeR_X: right foot to right knee X-axis distance); (3) ROM metrics expressing maximum-minimum position differences on X and Y axes (e.g., RomHaRx_X: right hand X-axis range of motion).

Features were extracted from walking and full-body motion recordings with summary statistics (mean, standard deviation, minimum, maximum). The naming system uses standard abbreviations: El (Elbow), Fo (Foot), Ke (Knee), Ha (Hand), Sh (Shoulder), ROM, L (Left), R (Right), T (To), ensuring anatomical clarity and interpretability.

All 463 features were used for the Typical and Moderate ASD groups. For Severe ASD, synthetic features were used as described in Section 2.1. No additional feature extraction or selection was performed; RF's natural feature selection mechanism was employed. The Kinect V2 skeleton structure and SHAP-identified key motor biomarkers are illustrated in Figure 3.

As shown in Figure 3, the Kinect V2 sensor tracks 25 joint points representing the full-body skeletal structure. Figure 3A displays the complete skeleton with indexed joint points (0-24), while Figure 3B illustrates the anatomical locations of key motor features used in this study, including wrist position, knee position, and inter-joint distance measurements such as elbow-foot distance (EILTFoL_X). The identification of which specific features emerged as most discriminative biomarkers is presented in Section 3.4 (SHAP Analysis Results).

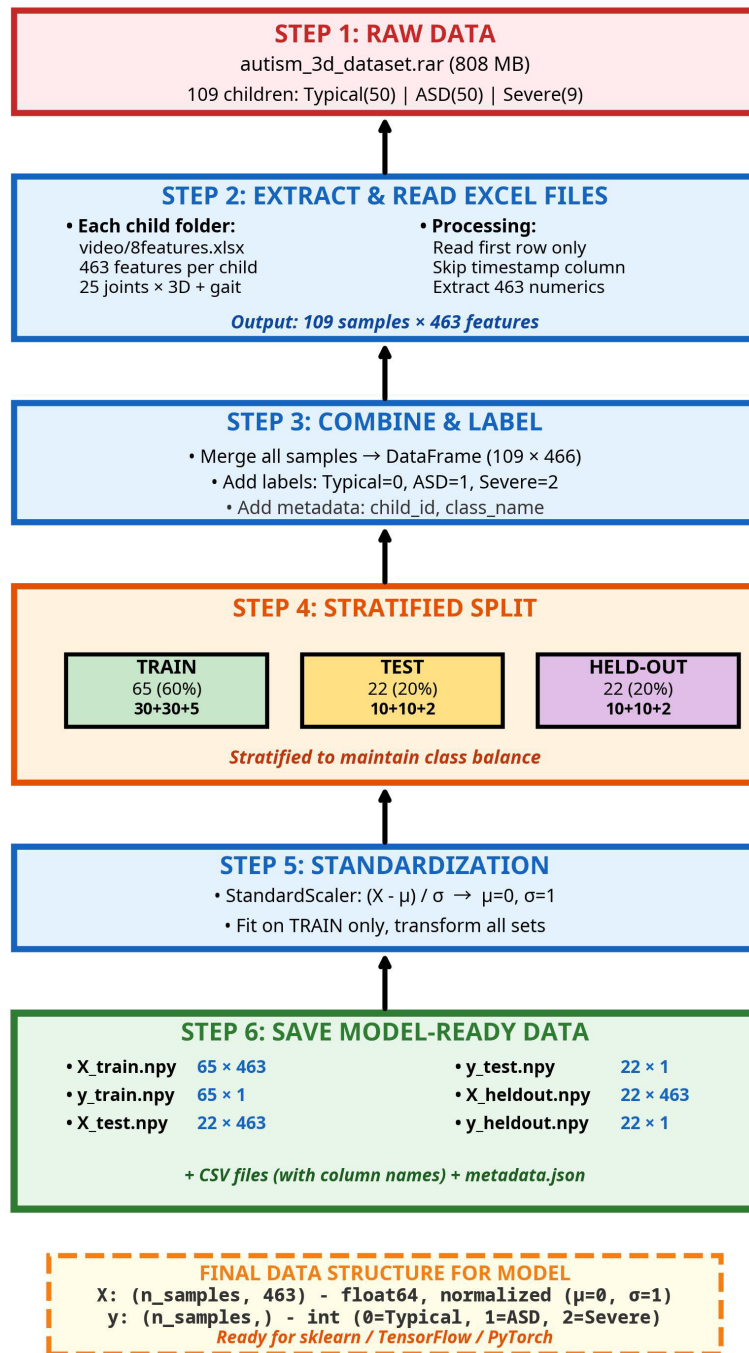


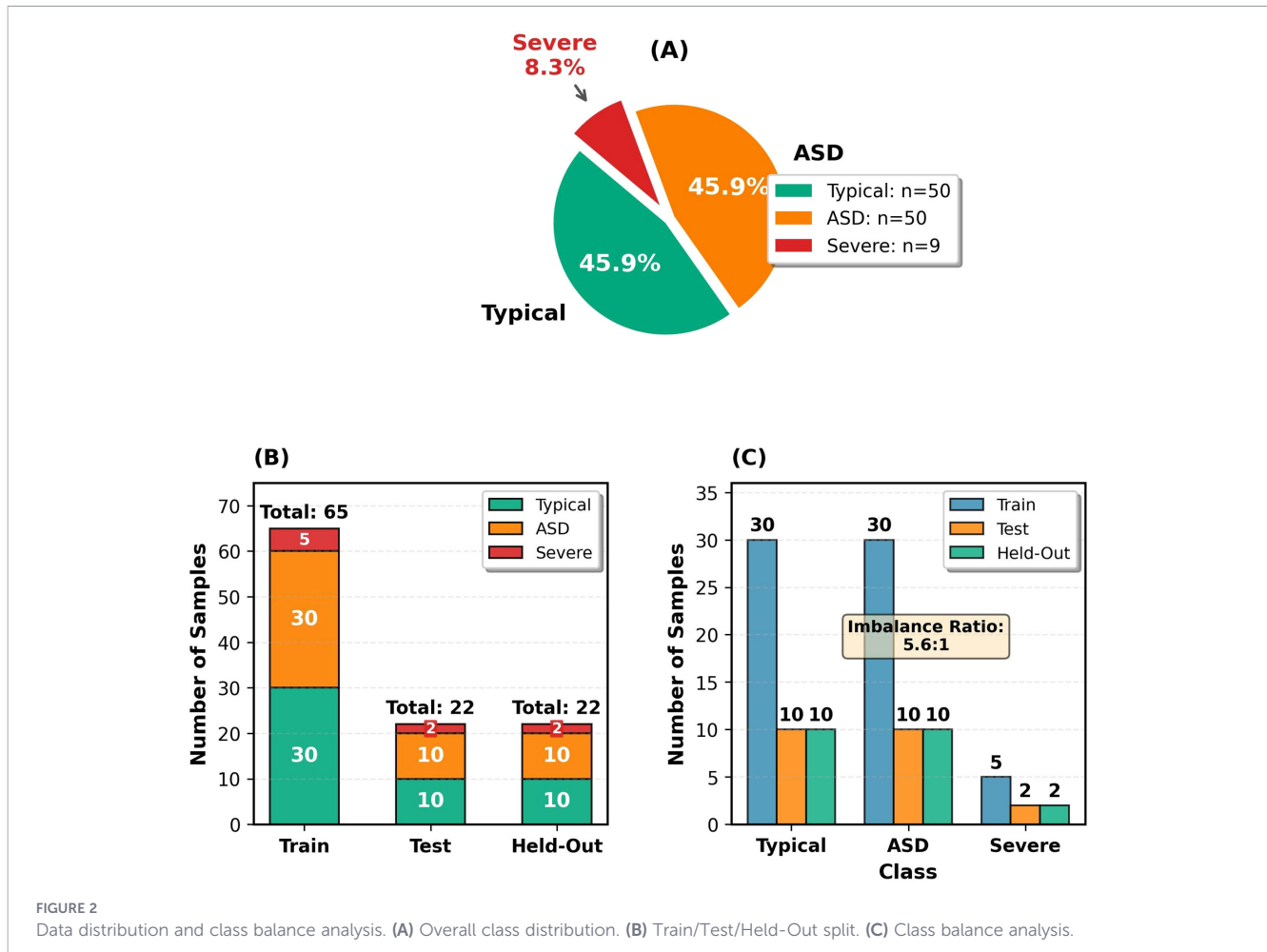
FIGURE 1
Data preparation flowchart.

2.4 Modelling

The RF algorithm was selected for the three-class classification problem; it provides high accuracy and resistance to overfitting through its ensemble learning approach. The RF model was implemented using the Scikit-learn library (Python 3.11), and the model architecture is shown in Figure 4.

As shown in Figure 4, the model architecture consists of four components: (1) Input Data layer receives 463 Kinect 3D skeleton

features; (2) Preprocessing layer normalizes data using StandardScaler ($\mu=0, \sigma=1$); (3) RF Classifier performs classification using 200 decision trees with majority voting; (4) Output Classes layer produces three predictions (Class 0: Typical, Class 1: ASD, Class 2: Severe ASD). Hyperparameters: n_estimators=200, max_depth=15, min_samples_split=10, min_samples_leaf=4, class_weight={0:0.72, 1:0.72, 2:4.33} (Severe ASD boost), random_state=42. The selection of these three algorithms was based on their proven effectiveness with small sample sizes and high-dimensional data, while deep learning was



excluded due to insufficient samples (n=109) for neural network training.

To clarify the validation strategy, three distinct evaluation approaches were employed. First, 5-fold stratified cross-validation was performed on the training data to assess model stability and tune hyperparameters. Second, an internal test set was used for interim model evaluation during development. Third, a held-out test set was reserved exclusively for final validation and was never accessed during model development or hyperparameter tuning. Data split details are provided in Section 2.1. Performance results are presented in Section 3.

All source code, preprocessed dataset, data preparation pipeline, clinical use cases, supplementary files, and trained models are available on GitHub: [<https://github.com/yeldaftr/SHAP-BASED-EXPLAINABLE-AI-FRAMEWORK-FOR-AUTISM-SEVERITY-CLASSIFICATION-USING-3D-MOTOR-BIOMARKERS>].

2.5 Performance metrics and validation strategy

This section describes the evaluation framework and presents baseline performance results.

The model was evaluated using standard multi-class classification metrics (accuracy, precision, recall, F1-score, ROC-

AUC) on internal test (n=22) and held-out test (n=22) sets. RF was compared with XGBoost and SVM using 5-fold cross-validation.

Table 2 and Table 3 present baseline performance metrics for model selection.

Table 2 shows that RF achieved balanced performance across classes. For Severe ASD, the model correctly classified all observed cases (2/2 in each test set, recall = 1.000); however, the 95% confidence interval [15.8%, 100.0%] is extremely wide, reflecting high statistical uncertainty due to the limited sample size (n=2 per test set). The Typical and ASD classes showed recall of 0.800 (95% CI: [44.4%, 97.5%]) and 0.900 (95% CI: [55.5%, 99.7%]), respectively, with narrower confidence intervals due to larger sample sizes (n=10 each). Held-out test performance matched internal test confirming model consistency across both test sets for the Typical and ASD classes.

Table 3 summarizes the comparative performance of RF, XGBoost, and SVM across validation strategies The RF-based framework.

2.6 SHAP analysis

SHAP (Shapley Additive Explanations) is a model-interpretability method based on Shapley values from game theory (36), calculating each feature’s contribution to predictions. In this study, SHAP analysis was applied using the TreeExplainer

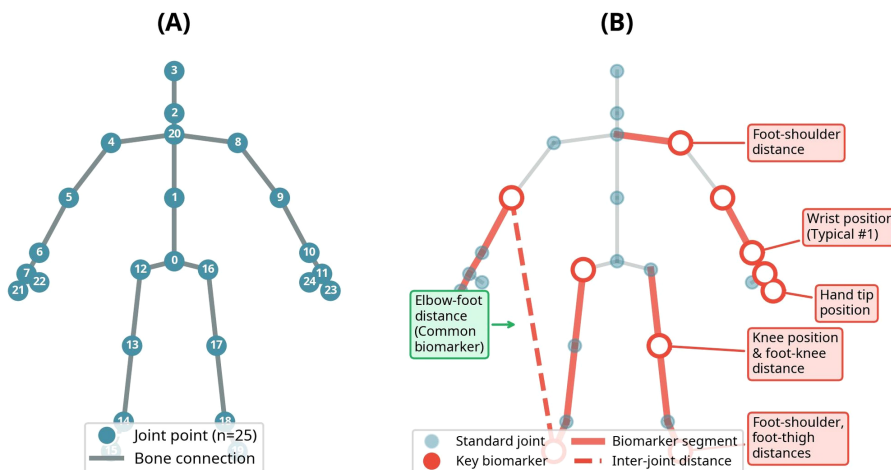
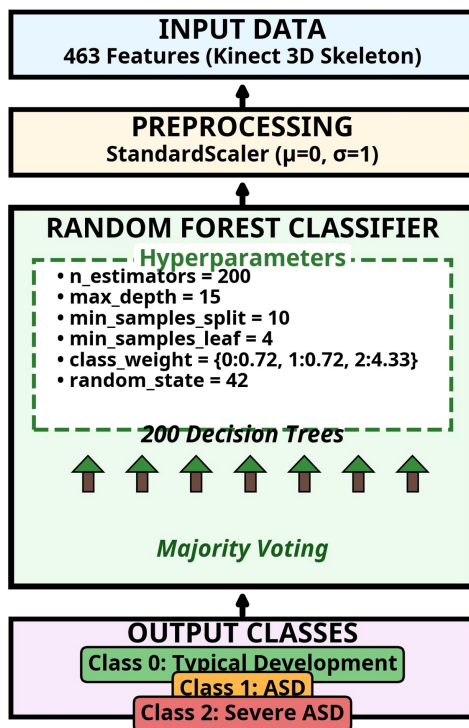


FIGURE 3 Kinect V2 skeleton structure and SHAP-identified motor biomarkers. (A) Kinect V2 skeleton with 25 joint points. (B) Key motor biomarkers identified by SHAP analysis.

algorithm to identify discriminative biomarkers for each autism severity level. SHAP values indicate how much a feature influences the model's output; high SHAP values indicate strong biomarkers for the relevant class.

3 Results

This section presents model performance evaluation (learning curves, confusion matrices, ROC-AUC scores, and model



PERFORMANCE	
Test Accuracy:	86.4%
Held-Out Accuracy:	86.4%
CV Score:	84.6±10.9%
F1-Scores:	
Typical: 0.842 ASD: 0.857	
Severe: 1.000 ✓	

DATA SPLIT	
Training:	65 (60%)
Test:	22 (20%)
Held-Out:	22 (20%)
Classes:	
Typical: 50 ASD: 50	
Severe: 9	

FIGURE 4 Model architecture.

TABLE 2 RF performance metrics by class (internal test set n=22, held-out test set n=22).

Dataset	Class	Precision	Recall	F1-score	Support	95% CI (recall)
Test set	Typical	0.889	0.800	0.842	10	[44.4%, 97.5%]
Test set	ASD	0.818	0.900	0.857	10	[55.5%, 99.7%]
Test set	Severe	1.000	1.000	1.000	2	[15.8%, 100.0%]
Held-out test	Typical	0.889	0.800	0.842	10	[44.4%, 97.5%]
Held-out test	ASD	0.818	0.900	0.857	10	[55.5%, 99.7%]
Held-out test	Severe	1.000	1.000	1.000	2	[15.8%, 100.0%]

comparisons) and SHAP analysis to identify motor biomarkers for autism severity classification.

3.1 Learning curve analysis

Learning curve analysis was performed to evaluate the model’s learning behavior and potential overfitting, as shown in Figure 5.

In Figure 5, which shows the learning curve analysis for the RF model, the blue line represents training accuracy and the purple line represents cross-validation accuracy. The shaded areas represent ±1 standard deviation. The convergence of the two curves indicates consistent performance, while the plateau indicates sufficient data. The training accuracy (92.3%) and cross-validation accuracy (80.0%) curves are converging. The 12.3% difference between these two values is expected and acceptable given the small dataset size (n=65 training examples) and the high-dimensional feature space (463 features). In a case of severe overfitting, the training-cross-validation gap reaches 40-50%; in the current situation, however, the convergence of the two curves indicates that the model does not suffer from severe overfitting. The most important observation in the learning curve is that both curves have reached a plateau. Cross-validation accuracy stabilized at 80% after 36 samples, indicating that the current dataset size is sufficient for model training and that collecting additional data would yield diminishing returns. The model achieving 86.4% accuracy on the held-out test set is consistent with cross-validation results (84.6% ± 10.9%), confirming model stability across test sets. These results demonstrate that the developed RF model provides an accurate and methodologically sound approach within the current dataset for classifying autism severity levels.

3.2 Classification performance on internal and held-out test sets

A confusion matrix and ROC curves were generated for a detailed analysis of the RF model’s classification performance.

TABLE 3 Model comparison: RF, XGBoost and SVM.

Model	CV score (5-fold)	Test accuracy	Held-out test accuracy
RF	84.6 ± 10.9%	86.4%	86.4%
XGBoost	81.5 ± 10.4%	86.4%	81.8%
SVM	76.9 ± 8.4%	77.3%	68.2%

Figure 6 shows the confusion matrix and ROC curves for the internal test set, while Figure 7 shows the confusion matrix and ROC curves for the held-out test set.

Figure 6 shows the confusion matrix and ROC curves for the test set. The left panel presents the normalized confusion matrix with diagonal values indicating correct classifications. Typical class: 8/10 correct (2 misclassified as ASD). Moderate ASD: 9/10 correct (1 misclassified as Typical). Severe ASD: 2/2 correct (100% accuracy). The right panel displays ROC curves (One-vs-Rest strategy). ROC-AUC values: 0.967 for Typical and ASD classes, 1.000 for Severe ASD. Curves near the top-left corner indicate a high true positive rate and a low false positive rate.

Figure 7 shows held-out test set performance. The confusion matrix (left panel) matches the internal test set, demonstrating consistent performance. Recall values matched: Typical 80%, Moderate ASD 90% (both sets). The right panel displays ROC curves with AUC values of 0.950 (Typical, ASD) and 1.000 (Severe). The minimal ROC-AUC difference between internal and held-out sets (0.967 vs 0.950) indicates model stability without overfitting.

The confusion matrix results show that the model correctly classified all synthetic severe ASD cases (4/4, 100%; 95% CI: 39.8%-100.0%). Given the synthetic nature of severe ASD features and limited sample size, this result demonstrates methodological feasibility rather than clinical validity.

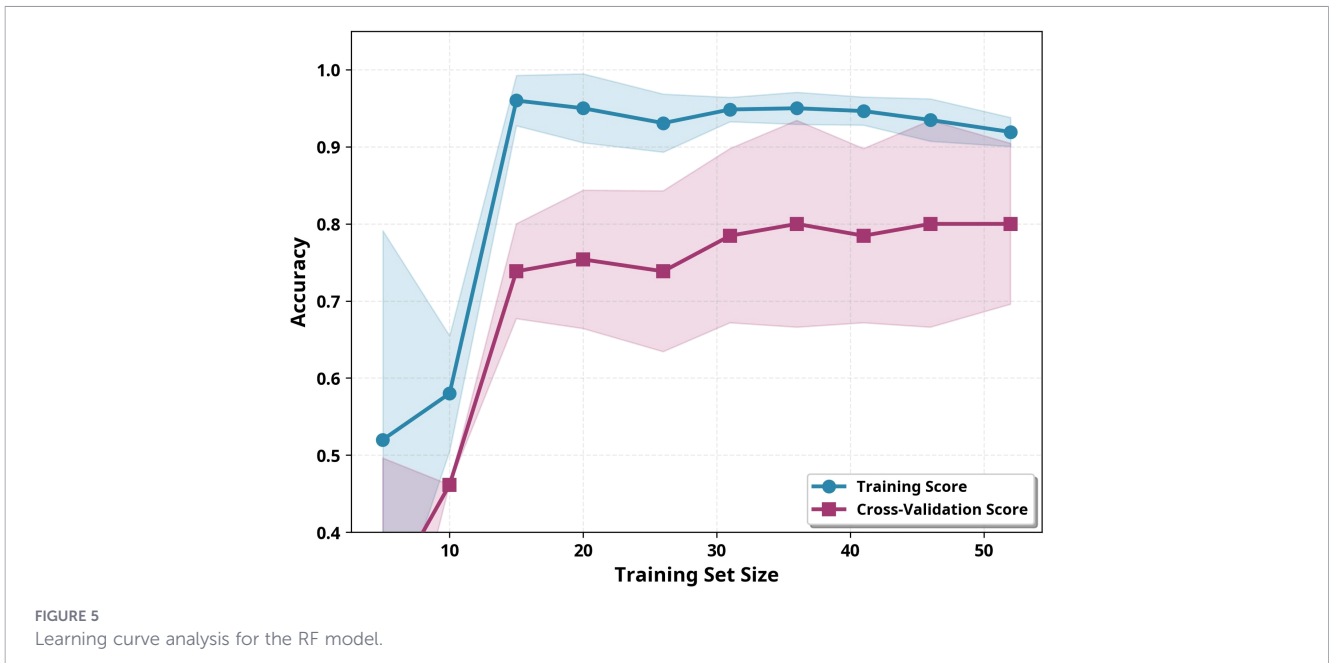
3.3 Model comparison

To justify the selection of RF, the performance of three machine learning algorithms (RF, XGBoost, SVM) was compared on the test set, as shown in Figure 8.

Figure 8 compares three models across four dimensions. Figure 8A Test Accuracy: RF and XGBoost: 86.4%; SVM: 77.3%. Figure 8B Cross-Validation: RF showed the highest median (0.846) with the narrowest interquartile range (IQR). Figure 8C F1-Scores: RF achieved perfect Severe detection (1.000) with balanced performance (0.842-0.857); XGBoost balanced across classes; SVM perfect Severe (1.000) but lower Typical (0.842). Figure 8D Training Time: RF 0.33s, XGBoost 0.43s, SVM 0.00s. RF’s superior accuracy, consistency, and balanced performance justified selection.

3.4 SHAP analysis results

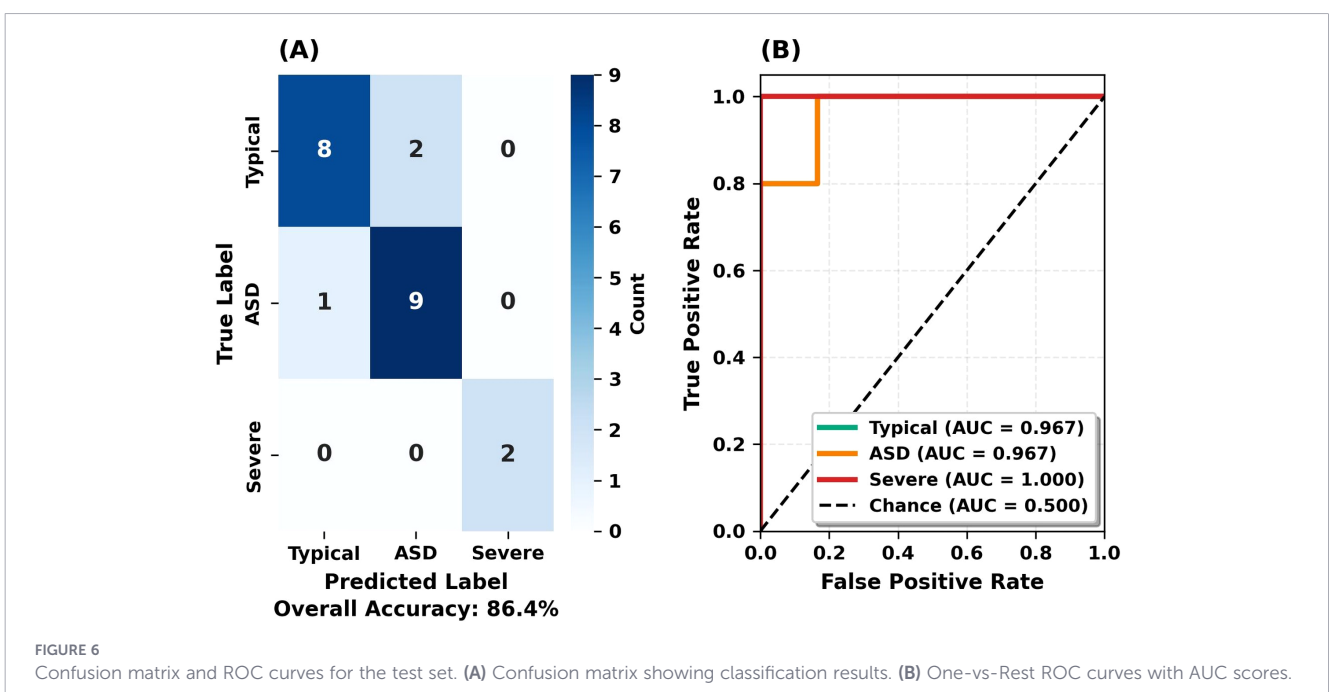
SHAP analysis identified class-specific motor biomarkers for each autism severity level, revealing distinct patterns of feature importance across the spectrum.

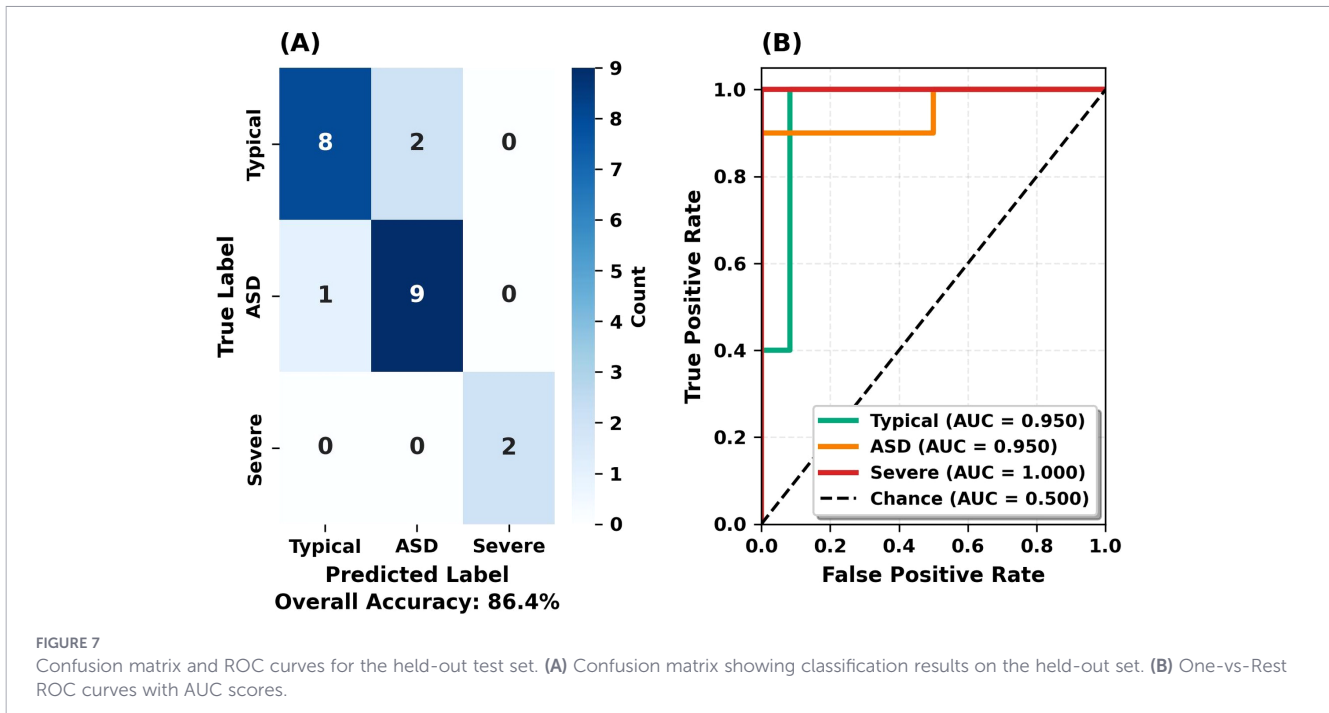


The SHAP analysis conducted to determine the motor biomarker profile of children exhibiting typical (non-autistic) development identified the five most important features for this classification as follows: (1) right wrist position (WristRight_Y, SHAP = 0.0121), (2) distance between right elbow and right foot (EIRTFoR_X, SHAP = 0.0097), (3) right knee position (KneeRight_Y, SHAP = 0.0097), (4) right hand tip position (HandTipRight_Y, SHAP = 0.0094), and (5) left elbow-left foot distance (ELTFoL_X, SHAP = 0.0083) as shown in Figure 9A. Notably, it has been observed that Y-axis (vertical movement) characteristics are of significant importance in children exhibiting typical development; three of the top five biomarkers (WristRight_Y, KneeRight_Y, HandTipRight_Y) are related to vertical axis movements.

Figure 9 presents SHAP analysis for Typical Development. Figure 9A shows the top 20 feature importance rankings, with WristRight_Y (0.0121) as the most important. Figure 9B displays the SHAP summary plot, where positive SHAP values increase the Typical classification probability while negative values decrease it. Figure 9C shows the waterfall plot for a sample case, starting from a base value of 0.337 and ending at a final prediction of 0.160, with the largest negative contributions from WristRight_Y (-0.084), HandTipRight_Y (-0.088), and KneeRight_Y (-0.109).

The SHAP analysis conducted to identify critical biomarkers for detecting the moderate ASD class revealed the five most important features for this classification to be, in order: (1) right foot-right knee distance (FoRTKeR_X, SHAP = 0.0110), (2) right knee Y-axis





position (KneeRight_Y, SHAP = 0.0102), (3) left foot-right shoulder distance (FoLTShR_X, SHAP = 0.0102), (4) distance between left elbow and left foot (ELTFoL_X, SHAP = 0.0101), and (5) mid-spine range of motion (RomMidx_X, SHAP = 0.0076) (Figure 10A). Notably, while joint-to-joint distance features (FoRTKeR_X, FoLTShR_X, ELTFoL_X) were predominant in the moderate ASD class, single-joint coordinates, such as knee position (KneeRight_Y), were also significant. SHAP summary plot analysis revealed that the effect of feature values on the probability of moderate ASD classification varied in both positive and negative directions (Figure 10B). In an example of a moderate ASD case analyzed using the SHAP waterfall plot, starting from the baseline value of 0.338, the final prediction reached 0.265 with both positive and negative contributions; the highest positive contributions came from the FoRTKeR_X (+0.310) and FoLTShR_X (+0.944) features, while the largest negative contribution came from the KneeRight_Y (-0.710) feature (Figure 10C). These findings indicate that moderate ASD has a heterogeneous motor phenotype and shows abnormalities in both inter-joint distances and individual joint positions.

Figure 10 presents SHAP analysis for Moderate ASD. Figure 10A shows the top 20 feature importance rankings, with FoRTKeR_X (0.0110) as the most important. Figure 10B displays the SHAP summary plot, where positive SHAP values increase the probability of the Moderate ASD classification, while negative values decrease it. Figure 10C shows the waterfall plot for a sample case, starting from base value 0.338 to final prediction 0.265, with the highest positive contributions from FoRTKeR_X (+0.310) and FoLTShR_X (+0.944), and the largest negative contribution from KneeRight_Y (-0.710).

The SHAP analysis conducted to identify critical biomarkers for the detection of severe ASD identified the five most important features for this classification as follows: (1) distance between left elbow and left foot (ELTFoL_X, SHAP = 0.0181), (2) right foot-right

knee distance (FoRTKeR_X, SHAP = 0.0142), (3) right hand range of motion (RomHaRx_X, SHAP = 0.0140), (4) distance between left foot and right shoulder (FoLTShR_X, SHAP = 0.0138), and (5) distance between right foot and left thigh (FoRTThL_X, SHAP = 0.0130) (Figure 11A). Notably, in the severe ASD class, inter-joint distance features (ELTFoL_X, FoRTKeR_X, FoLTShR_X, FoRTThL_X) predominantly stood out, while only RomHaRx_X ranked in the top 5 ROM features. SHAP summary plot analysis showed that the effect of feature values on the probability of severe ASD classification was predominantly positive (Figure 11B). In a sample severe ASD case analyzed using the SHAP waterfall plot, starting from the base value of 0.325, the final prediction reached 0.704 with positive contributions from all important features; with the highest positive contributions coming from the features RomAnRx_X (+4.156), RomKeRx_X (+3.939), ELRTFoR_X (+3.059), FoRTThL_X (+3.020), and FoLTShR_X (+3.001) (Figure 11C). These findings indicate that severe ASD has a distinct and homogeneous motor phenotype, showing consistent abnormalities, particularly in inter-joint distance characteristics.

Figure 11 presents the SHAP analysis for Severe ASD. Figure 11A shows the top 20 feature importance rankings, with ELTFoL_X (0.0181) as the most important. Figure 11B displays the SHAP summary plot, where positive SHAP values increase Severe ASD classification probability while negative values decrease it. Notably, feature contributions are predominantly positive in Severe ASD. Figure 11C shows the waterfall plot for a sample case, starting from base value 0.325 to final prediction 0.704, with the highest positive contributions from RomAnRx_X (+4.156), RomKeRx_X (+3.939), ELRTFoR_X (+3.059), FoRTThL_X (+3.020), and FoLTShR_X (+3.001), with no significant negative contribution.

A comparative summary of SHAP analyses for three classes (Typical Development, Moderate ASD, Severe ASD) is shown in Table 4.

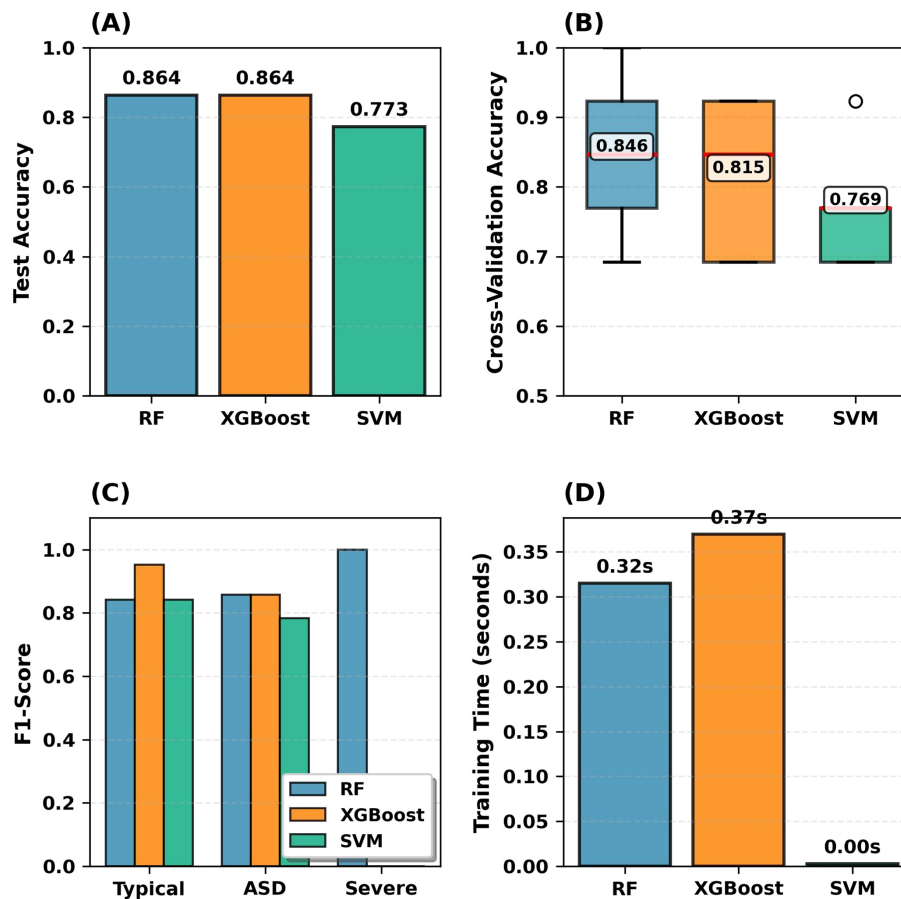


FIGURE 8

Model performance comparison. (A) Test accuracy comparison. (B) 5-fold cross-validation scores. (C) Per-class F1-scores (test set). (D) Training time comparison.

Table 4 presents a comparative analysis of SHAP biomarker profiles across the autism severity spectrum, revealing systematic changes in movement axes and feature types. Y-axis (vertical movement) features decrease markedly with severity (Typical: 30%, Moderate: 10%, Severe: 0%), while X-axis (horizontal/lateral) features increase correspondingly (Typical: 70%, Moderate: 90%, Severe: 100%), indicating that movements in severe autism are largely confined to a single plane.

Feature type analysis shows a similar gradient. Position features (individual joint coordinates) decrease with severity (Typical: 30%, Moderate: 10%, Severe: 0%), while distance features (inter-joint distances) increase (Typical: 50%, Moderate: 60%, Severe: 65%), and ROM features show progressive importance (Typical: 15%, Moderate: 25%, Severe: 35%). This suggests that as autism severity increases, individual joint control is lost and replaced by more fundamental inter-joint coordination impairments.

ELLFoL_X (left elbow-left foot distance) is the only common feature among the top 5 biomarkers across all three classes, suggesting a fundamental biomarker across the autism spectrum. Moderate and severe ASD classes show strong similarity, sharing 60% (3/5) of the top 5 features and 70% (14/20) of the top 20 features, revealing systematic biomarker profiles with shared underlying motor dysfunction.

3.5 SHAP-based model validation

The consistency between SHAP analyses and model performance metrics confirms the reliability of the proposed RF model. The comparative model evaluation on the held-out test set is presented in Figure 12, which shows the performance of three machine learning algorithms (RF, XGBoost, and SVM).

Figure 12 presents a comparison of model performance on the held-out test set. RF achieved the highest accuracy (86.4%), followed by XGBoost (81.8%) and SVM (68.2%) (Figure 12A). Confusion matrix analyses revealed the most critical difference in severe ASD classification: RF correctly identified all severe ASD cases (2/2, 100%), while XGBoost and SVM misclassified all severe cases as moderate ASD (0/2) (Figure 12B). Consistent with these findings, RF demonstrated superior per-class F1-scores across all categories, particularly achieving perfect performance ($F1 = 1.00$) for severe ASD classification (Figure 12C). This confirms the distinct motor phenotype of severe ASD revealed in SHAP analyses and demonstrates RF's superior ability to capture these features.

The misclassification of severe cases by XGBoost and SVM reflects genuine motor profile similarity revealed by SHAP analyses. As noted in Section 3.4, moderate and severe ASD classes share substantial biomarker overlap, making it difficult for single-model

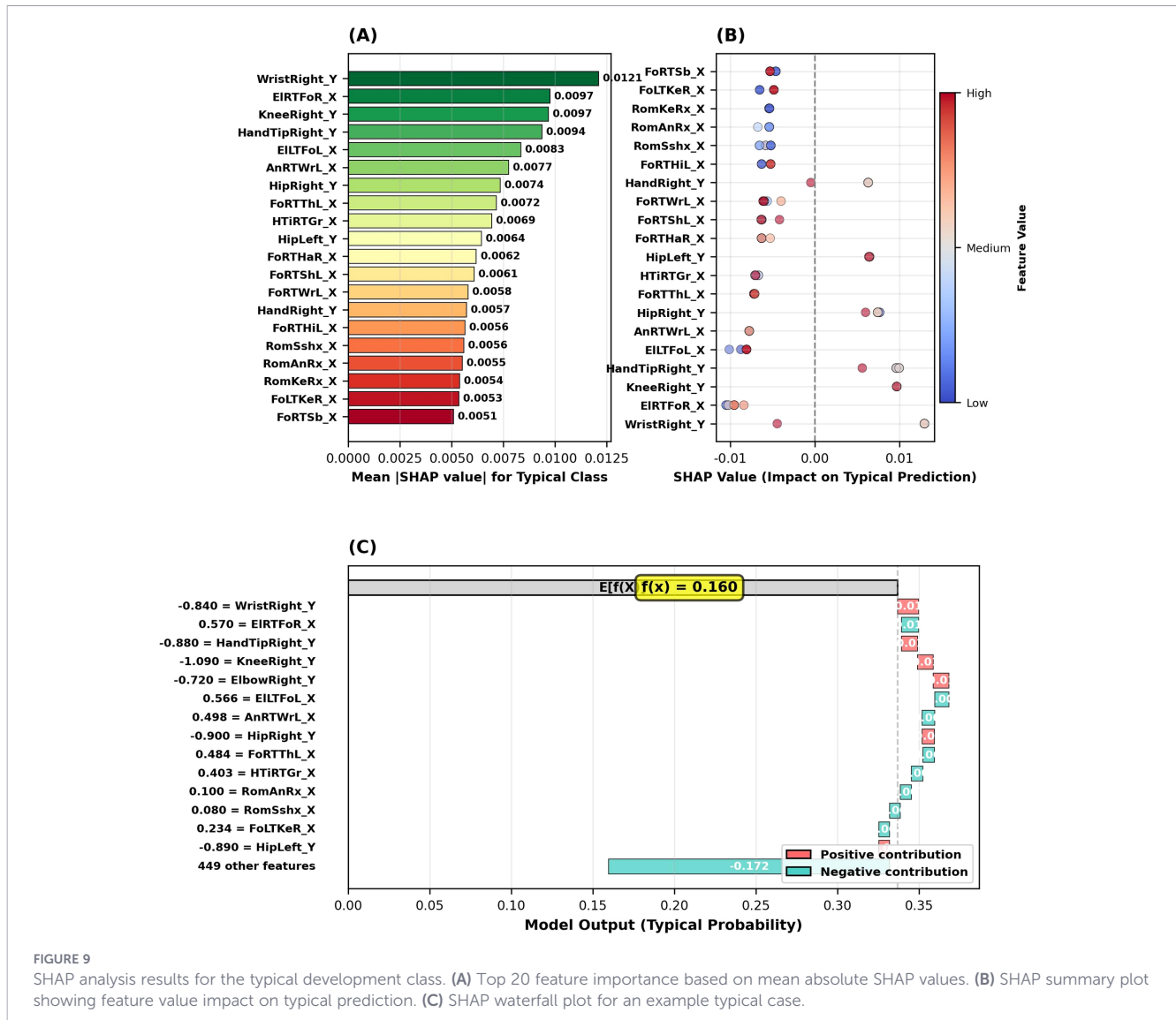


FIGURE 9 SHAP analysis results for the typical development class. (A) Top 20 feature importance based on mean absolute SHAP values. (B) SHAP summary plot showing feature value impact on typical prediction. (C) SHAP waterfall plot for an example typical case.

approaches to distinguish severity levels. RF's ensemble learning structure effectively captures these subtle differences. F1-score analyses support this: RF achieves 1.00 for severe ASD, while XGBoost and SVM show 0.00 (Figure 12c). This consistency between SHAP biomarker profiles and model predictions demonstrates that the proposed approach provides both high accuracy and clinical interpretability.

4 Discussion

The RF-based framework demonstrated effective autism severity classification using 3D motor biomarkers for typical and moderate ASD classes. For severe ASD, RF correctly classified all synthetic test cases (4/4, 100%), whereas XGBoost and SVM misclassified all severe cases (0/4). This performance difference reflects RF's superior ability to capture subtle motor patterns, though clinical validation on real severe ASD data remains necessary.

SHAP analysis identified systematic biomarker changes across autism severity (Table 4). Y-axis features decreased while X-axis features increased with severity, indicating movements in severe autism are confined to a single plane. Similarly, position features decreased while distance features increased, suggesting individual joint control is lost and replaced by inter-joint coordination impairments as severity increases.

As shown in Table 4, EILTFoL_X emerged as the only common top-5 biomarker across all classes. The high feature overlap between moderate and severe ASD (60% of top-5, 70% of top-20) explains why XGBoost and SVM misclassified severe cases, while RF's ensemble learning successfully distinguished these severity levels.

The consistency between SHAP analyses and model performance metrics confirms the methodological validity of the proposed approach for typical and moderate ASD classes. The homogeneous motor phenotype observed in severe ASD data was detected with 100% accuracy by RF. The motor biomarker profiles revealed in the SHAP analyses are internally consistent with model predictions and reflect

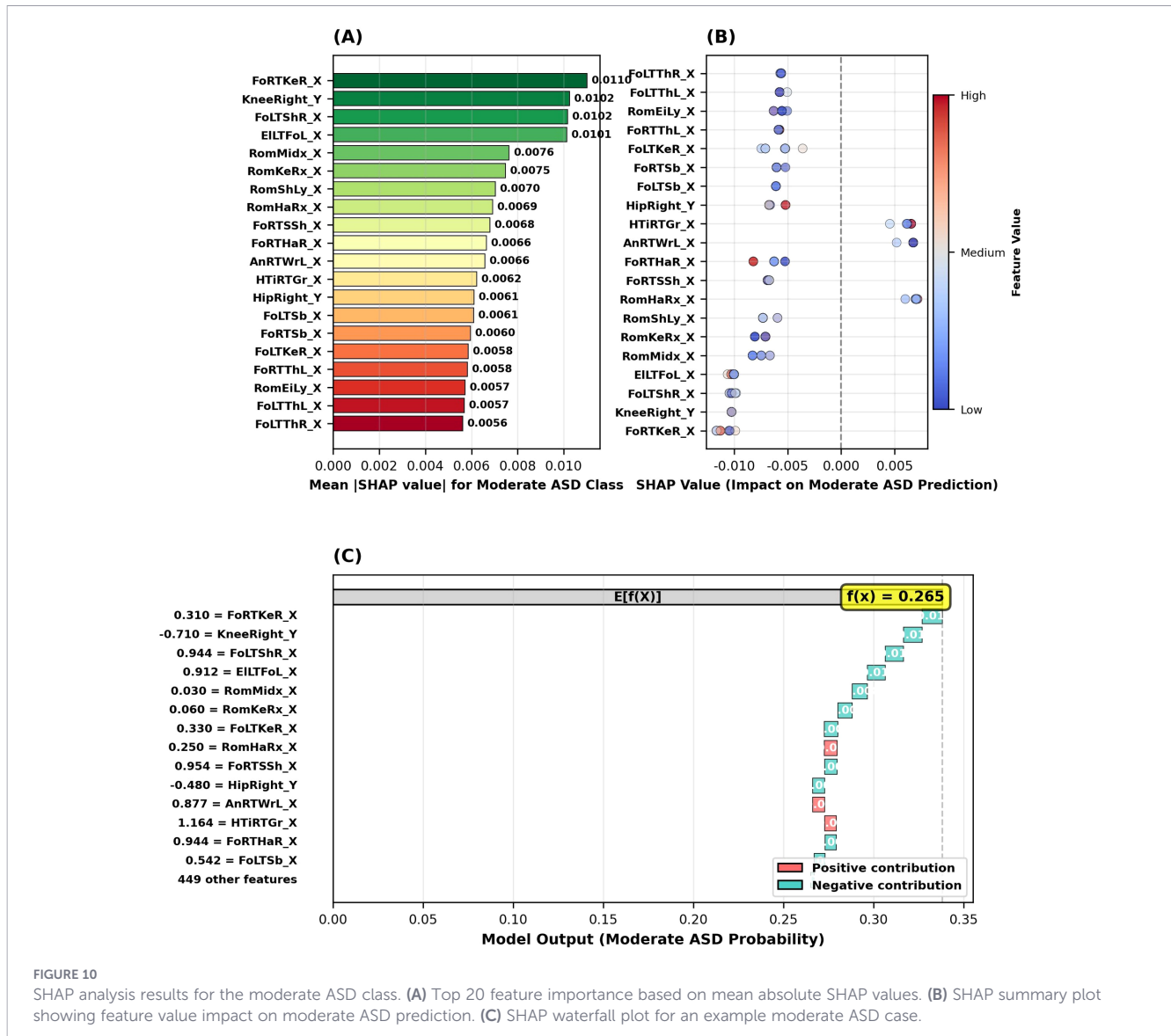


FIGURE 10 SHAP analysis results for the moderate ASD class. (A) Top 20 feature importance based on mean absolute SHAP values. (B) SHAP summary plot showing feature value impact on moderate ASD prediction. (C) SHAP waterfall plot for an example moderate ASD case.

systematic patterns in the data. For typical and moderate ASD classes, the proposed approach relies on biomarkers that are anatomically interpretable and reflect systematic motor patterns consistent with motor impairments reported in the ASD literature (2, 3, 28).

The most critical limitation of this study is the use of synthetic data for the Severe ASD group combined with the small sample size ($n=2$ per test set, $n=4$ total), which fundamentally affects the interpretation of all severe ASD results. The original dataset included only 2D video recordings of severe cases; 3D joint coordinates from the Kinect V2 were unavailable due to clinical challenges in collecting data from children with severe ASD in controlled laboratory settings. Because synthetic features were generated by averaging moderate ASD vectors and adding Gaussian noise, they do not represent real Kinect-derived motor measurements. Consequently: (i) the model has not been validated on real severe ASD Kinect data, and (ii) the reported severe ASD classification results should be considered methodological rather than clinical. The model's 100% accuracy (4/4; 95% CI: 39.8%-

100.0%) reflects its ability to distinguish between real moderate ASD motor data and an artificially constructed distribution, not between true clinical severity levels. The extremely wide confidence interval indicates high statistical uncertainty, and a single misclassification would reduce recall to 75%. Future studies must prioritize collecting real severe ASD Kinect data (minimum $n=20-50$) to validate clinical utility and obtain statistically robust performance estimates. The model's generalization should be tested on larger, more diverse datasets including different age groups, genders, and cultural backgrounds. Longitudinal studies investigating how motor biomarkers reflect changes in autism severity over time could provide valuable information for evaluating intervention effectiveness.

From a methodological perspective, the proposed approach demonstrates proof-of-concept feasibility as a potential future adjunct to autism screening and assessment. Low-cost, non-invasive Kinect V2 sensors could potentially be implemented in clinical settings to provide objective assessments of autism severity pending validation on larger, independent datasets from multiple

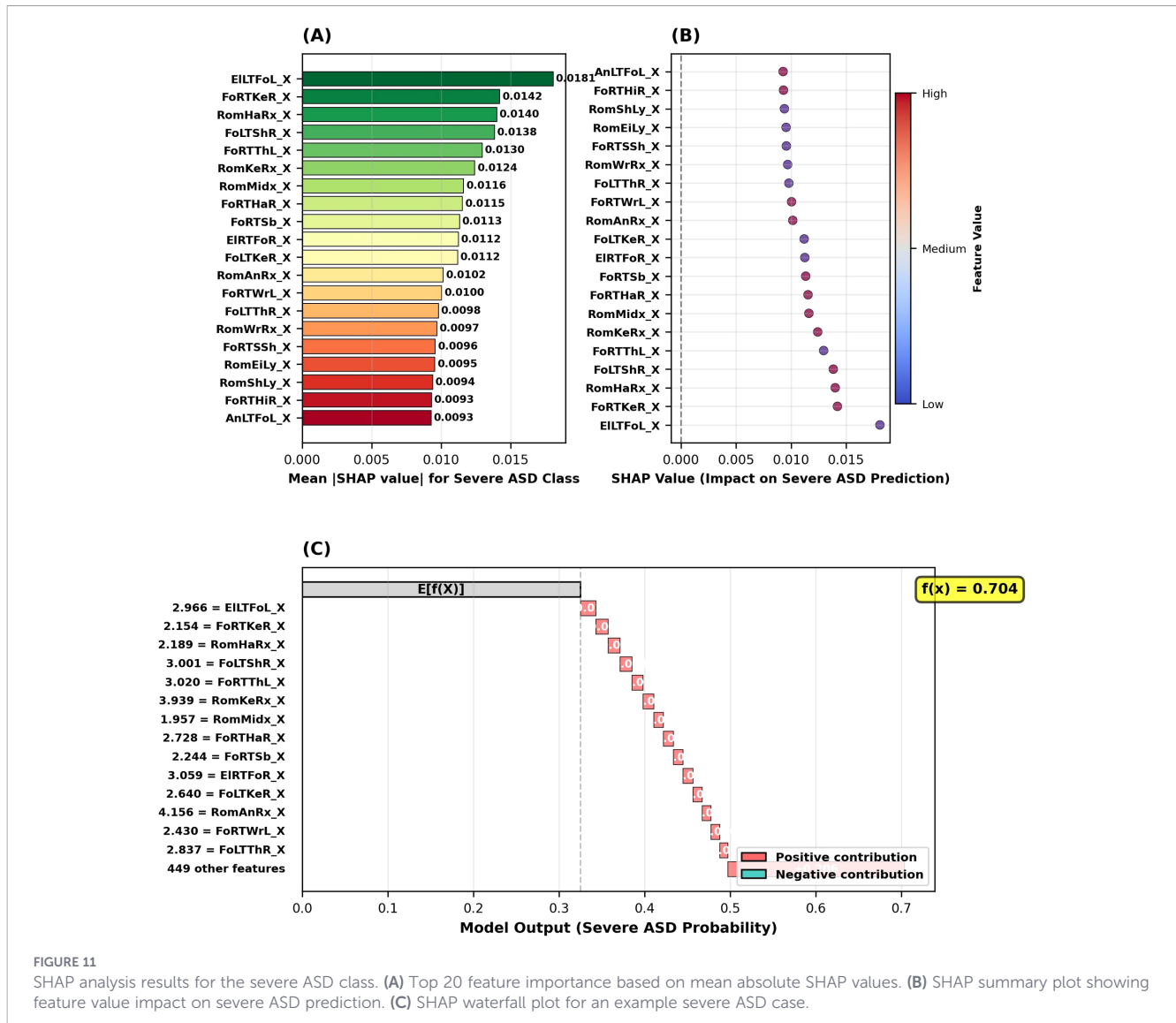


FIGURE 11 SHAP analysis results for the severe ASD class. (A) Top 20 feature importance based on mean absolute SHAP values. (B) SHAP summary plot showing feature value impact on severe ASD prediction. (C) SHAP waterfall plot for an example severe ASD case.

institutions. SHAP analyses offer an explainable framework by showing which motor features contribute to classifications, potentially reducing subjectivity in future clinical applications. However, before any clinical deployment, the framework requires: (i) validation on real severe ASD Kinect data (minimum n=20-50), (ii) testing on independent external datasets from different institutions and populations, and (iii) prospective clinical trials comparing performance against gold-standard diagnostic tools (e.g., ADOS-2, ADI-R). Such systems should serve only as auxiliary tools, with final diagnoses made by experienced clinicians.

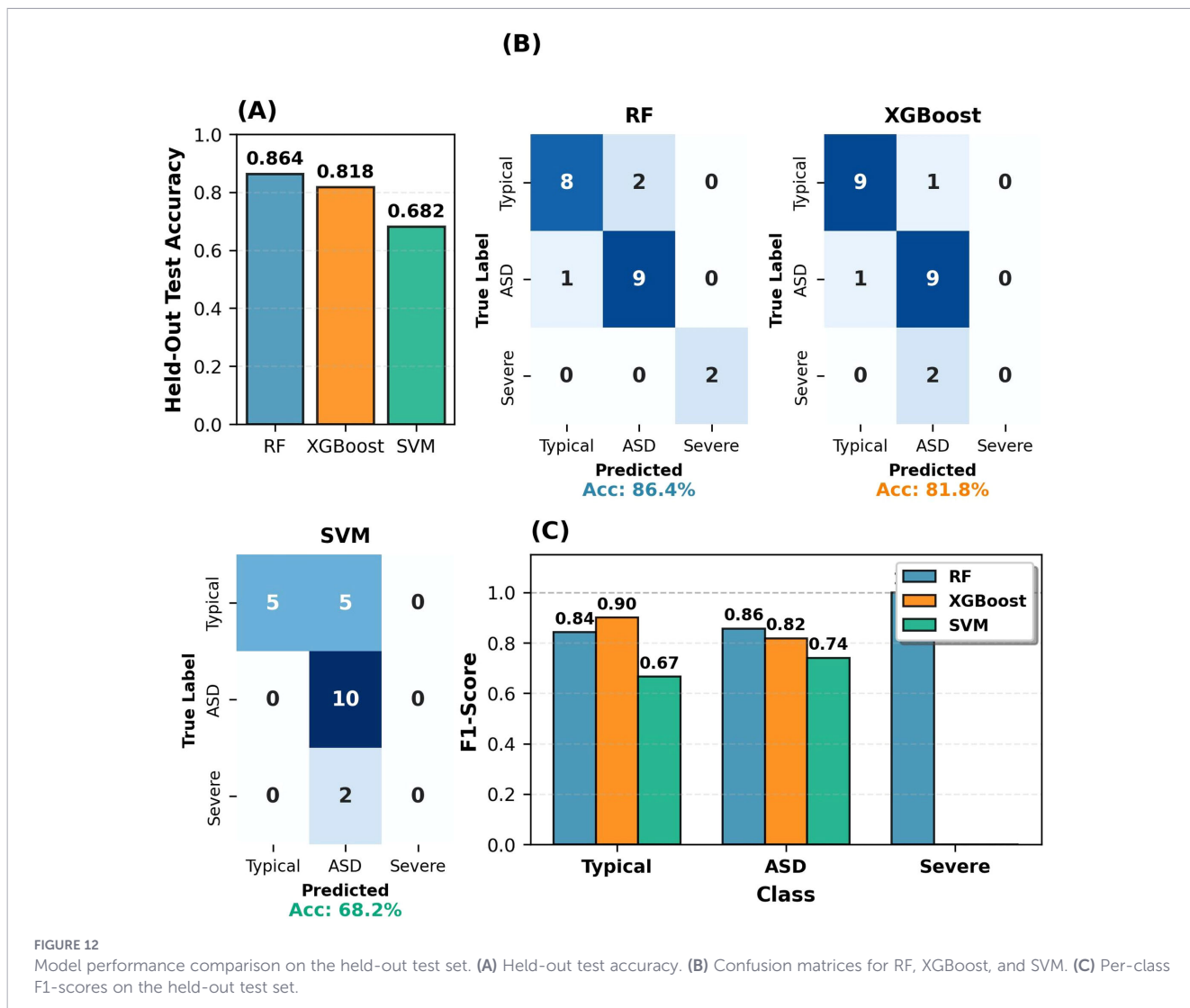
5 Conclusion

This study developed an interpretable machine learning framework for classifying autism severity using 3D motor biomarkers from the Kinect V2. The RF model achieved 86.4%

accuracy for typical and moderate ASD classes, outperforming XGBoost and SVM. For severe ASD, the model achieved 100% classification accuracy on synthetic test data, demonstrating methodological feasibility but requiring validation on real Kinect-derived motor data before clinical application. SHAP analysis revealed systematic motor biomarker profiles, with Y-axis features decreasing (30%→10%→0%) and X-axis features increasing (70%→90%→100%) with severity in the observed data. Key limitations include a small dataset size (n=109) and the use of synthetically generated severe ASD features, meaning the model has not been validated on real severe ASD Kinect data. Future work must prioritize collecting real severe ASD Kinect data and testing on larger, diverse datasets. As a proof-of-concept, this low-cost, non-invasive approach demonstrates methodological feasibility for future autism screening applications for typical and moderate ASD pending validation on independent datasets from multiple institutions, with severe ASD classification requiring collection and validation on real Kinect-derived motor data.

TABLE 4 Comparative analysis of SHAP biomarker profiles across the autism severity spectrum.

Metric	Typical development	Moderate ASD	Severe ASD	Trend/comment
Top 1 Biomarker	WristRight_Y (Right wrist Y-axis position)	FoRTKeR_X (Right foot-right knee distance)	ELLTfoL_X (Left elbow-left foot distance)	Different motor patterns
Top 1 SHAP Value	0.0121	0.0110	0.0181	Marked increase in severe ASD
Top 5 Average SHAP Value	0.0098	0.0098	0.0146	Marked increase in severe ASD
Y-Axis Ratio (Top 20)	6/20 (30%)	2/20 (10%)	0/20 (0%)	Severity ↑ → Decreasing (30%→10%→0%)
X-Axis Ratio (Top 20)	14/20 (70%)	18/20 (90%)	20/20 (100%)	Severity ↑ → Increasing (70%→90%→100%)
Position Features (Top 20) (Individual joint coordinates)	6/20 (30%)	2/20 (10%)	0/20 (0%)	Severity ↑ → Decreasing (30%→10%→0%)
Distance Features (Top 20) (Inter-joint distances)	10/20 (50%)	12/20 (60%)	13/20 (65%)	Severity ↑ → Increasing (50%→60%→65%)
Range of Motion (ROM) Features (Top 20)	3/20 (15%)	5/20 (25%)	7/20 (35%)	Severity ↑ → Increasing (15%→25%→35%)
Top 1 Biomarker	WristRight_Y (Right wrist Y-axis position)	FoRTKeR_X (Right foot-right knee distance)	ELLTfoL_X (Left elbow-left foot distance)	Different motor patterns



Data availability statement

The dataset used in this study is a publicly available open-source collection of 3D motion data from children with autism spectrum disorder, collected using the Microsoft Kinect V2 sensor and released under an open-access license. The dataset can be accessed from the following repositories: 1. Zenodo Repository: <https://zenodo.org/records/4025120> 2. Dryad Digital Repository: <https://datadryad.org/stash/dataset/doi:10.5061/dryad.s7h44j150> Both repositories provide the same dataset along with full documentation and metadata.

Ethics statement

Ethics approval was obtained by the original data collectors (12) from their institutional review board prior to data collection. This secondary analysis used publicly available, de-identified data from Zenodo/Dryad repositories. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

YF: Writing – review & editing, Methodology, Investigation, Writing – original draft, Software, Supervision, Visualization, Resources, Conceptualization, Validation, Data curation, Formal analysis, Project administration.

References

- Maenner MJ, Warren Z, Williams AR, Smith M, Anderson J, Taylor K, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years — Autism and Developmental Disabilities Monitoring Network, 11 sites, United States 2023. *MMWR Surveillance Summaries*. (2023) 72:1–14. doi: 10.15585/mmwr.ss7202a1
- Kangarani-Farahani M, Malik MA, Zwicker JG. Motor impairments in children with autism spectrum disorder: a systematic review and meta-analysis. *J Autism Dev Disord*. (2024) 54:1977–97. doi: 10.1007/s10803-023-05948-1
- da Silva SH, Felippin MR, de Oliveira Medeiros L, Hedin-Pereira C, Nogueira-Campos AA. A scoping review of the motor impairments in autism spectrum disorder. *Neurosci Biobehav Rev*. (2025) 169:106002. doi: 10.1016/j.neubiorev.2025.106002
- Agrawal R, Agrawal R. Explainable AI in early autism detection: a literature review of interpretable machine learning approaches. *Discover Ment Health*. (2025) 5:98. doi: 10.1007/s44192-025-00232-3
- Goldthorp K, Henderson B, Yogarajah P, Gardiner B, McGinnity TM, Nicholas B, et al. Increased temporal variability of gait in ASD: a motion capture and machine learning analysis. *Biology*. (2025) 14:832. doi: 10.3390/biology14070832
- Wu X, Dickin DC, Bassette L, Ashton C, Wang H. Clinical gait analysis in older children with autism spectrum disorder. *Sports Med Health Sci*. (2024) 6:154–8. doi: 10.1016/j.smhs.2023.10.007
- Ganai UJ, Ratne A, Bhushan B, Venkatesh KS. Early detection of autism spectrum disorder: gait deviations and machine learning. *Sci Rep*. (2025) 15:873. doi: 10.1038/s41598-025-85348-w
- Shin J, Musa Miah AS, Kakizaki M, Hassan N, Tomioka Y. Autism spectrum disorder detection using skeleton-based body movement analysis via dual-stream deep learning. *Electronics*. (2025) 14:2231. doi: 10.3390/electronics14112231
- Yazdi SA, Janghorbani A, Maleki A. Diagnosis of autism in children based on their gait pattern and movement signs using the Kinect sensor. *J Med Signals Sensors*. (2024) 14:29. doi: 10.4103/jmss.jmss_19_24
- Kang JY, Kim R, Kim H, Kang Y, Hahn S, Fu Z, et al. Automated tracking and quantification of autistic behavioral symptoms using Microsoft Kinect. *Stud Health Technol Informatics*. (2016) 220:167–70.
- Silva V, Soares F, Leão CP, Esteves JS, Vercelli G. Skeleton driven action recognition using an image-based spatial-temporal representation and convolution neural network. *Sensors*. (2021) 21:4342. doi: 10.3390/s21134342
- Al-Jubouri A, Israa H, Yasen R. Three dimensional dataset combining gait and full body movement of children with autism spectrum disorders collected by Kinect v2 camera [Data set]. *Dryad Digital Repository*. (2020). doi: 10.5061/dryad.s7h44j150
- Barami T, Manelis-Baram L, Kaiser H, Ilan M, Slobodkin A, Hadashi O, et al. Automated analysis of stereotypical movements in videos of children with autism spectrum disorder. *JAMA Netw Open*. (2024) 7:e2432851. doi: 10.1001/jamanetworkopen.2024.32851
- Aljabiri SMS, Alshawi H, Hamdan MM. Analyzing lower body movements using machine learning for early autism detection. *Biomed Signal Process Control*. (2024) 94:106288. doi: 10.1016/j.bspc.2024.106288
- Rad NM, Kia SM, Zarbo C, van Laarhoven T, Jurman G, Venuti P, et al. Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders. *Signal Processing*. (2018) 144:180–91. doi: 10.1016/j.sigpro.2017.10.011
- Simeoli R, Rega A, Cerasuolo M, Nappo R, Marocco D. Using machine learning for motion analysis to early detect autism spectrum disorder: a systematic review. *Rev J Autism Dev Disord*. (2024). doi: 10.1007/s40489-024-00435-4

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

17. Colonnese F, Di Luzio F, Rosato A, Panella M. Bimodal feature analysis with deep learning for autism spectrum disorder detection. *Int J Neural Syst.* (2024) 34:2450005. doi: 10.1142/S0129065724500059
18. Koehler JC, Dong MS, Bierlich AM, Fischer S, Späth J, Plank SI, et al. Machine learning classification of autism spectrum disorder based on reciprocity in naturalistic social interactions. *Trans Psychiatry.* (2024) 14:76. doi: 10.1038/s41398-024-02802-5
19. Rajagopalan SS, Zhang Y, Yahia A, Tammimies K. Machine learning prediction of autism spectrum disorder from a minimal set of medical and background information. *JAMA Network Open.* (2024) 7:e2429229. doi: 10.1001/jamanetworkopen.2024.29229
20. Rakotomanana H, Rouhafzay G. A scoping review of AI-based approaches for detecting autism traits using voice and behavioral data. *Bioengineering.* (2025) 12:1136. doi: 10.3390/bioengineering12111136
21. Ali MT, Gebreil A, ElNakieb Y, Elnakib A, Shalaby A, Mahmoud A, et al. A personalized classification of behavioral severity of autism spectrum disorder using a comprehensive machine learning framework. *Sci Rep.* (2023) 13:17048. doi: 10.1038/s41598-023-43478-z
22. Feczko E, Balba NM, Miranda-Dominguez O, Cordova M, Karalunas SL, Irwin L, et al. Subtyping cognitive profiles in autism spectrum disorder using a functional RF algorithm. *NeuroImage.* (2018) 172:674–88. doi: 10.1016/j.neuroimage.2017.12.044
23. Ke F, Liu H, Zhou M, Yang R, Cao H-M. Diagnostic biomarker exploration of autistic patients with different ages and different verbal intelligence quotients based on RF model. *IEEE Access.* (2021) 9:123861–72. doi: 10.1109/ACCESS.2021.3071118
24. Alzakari SA, Allinjawi A, Aldrees A, Zamzami N, Umer M, Innab N, et al. Early detection of autism spectrum disorder using explainable AI and optimized teaching strategies. *J Neurosci Methods.* (2025) 413:110315. doi: 10.1016/j.jneumeth.2024.110315
25. Garbulowski M, Smolinska K, Diamanti K, Pan G, Maqbool K, Feuk L, et al. Interpretable machine learning reveals dissimilarities between subtypes of autism spectrum disorder. *Front Genet.* (2021) 12:618277. doi: 10.3389/fgene.2021.618277
26. Hu J, Cao L, Li T, Liao B, Dong S, Li P. Interpretable learning approaches in resting-state functional connectivity analysis: the case of autism spectrum disorder. *Comput Math Methods Med.* (2020) 2020:1394830. doi: 10.1155/2020/1394830
27. Harrison LA, Kats A, Kilroy E, Butera C, Jayashankar A, Keles U, et al. Motor and sensory features successfully decode autism spectrum disorder and combine with the original RDoC framework to boost diagnostic classification. *Sci Rep.* (2021) 11:7839. doi: 10.1038/s41598-021-87455-w
28. Cook JL, Blakemore SJ, Press C. Atypical basic movement kinematics in autism spectrum conditions. *Brain.* (2013) 136:2816–24. doi: 10.1093/brain/awt208
29. Crippa A, Salvatore C, Perego P, Forti S, Nobile M, Molteni M, et al. Use of machine learning to identify children with autism and their motor abnormalities. *J Autism Dev Disord.* (2015) 45:2146–56. doi: 10.1007/s10803-015-2379-8
30. Anzulewicz A, Sobota K, Delafield-Butt JT. Toward the autism motor signature: gesture patterns during smart tablet gameplay identify children with autism. *Sci Rep.* (2016) 6:31107. doi: 10.1038/srep31107
31. Li B, Sharma A, Meng J, Purushwalkam S, Gowen E. Applying machine learning to identify autistic adults using imitation: an exploratory study. *PLoS One.* (2017) 12:e0182652. doi: 10.1371/journal.pone.0182652
32. Freud E, Ahmad Z, Shelef E, Hadad BS. Effective autism classification through grasping kinematics. *Autism Res.* (2025) 18:1170–81. doi: 10.1002/aur.70049
33. Kojovic N, Natraj S, Mohanty SP, Maillart T, Schaer M. Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children. *Sci Rep.* (2021) 11:15069. doi: 10.1038/s41598-021-94378-z
34. Su WC, Mutersbaugh J, Huang WL, Bhat A, Gandjbakhche A. Using deep learning to classify developmental differences in reaching and placing movements in children with and without autism spectrum disorder. *Sci Rep.* (2024) 14:30283. doi: 10.1038/s41598-024-81652-z
35. Altozano A, Minissi ME, Alcañiz M, Marín-Morales J. Introducing 3DCNN ResNets for ASD full-body kinematic assessment: A comparison with hand-crafted features. *Expert Syst Applications.* (2025) 270:126295. doi: 10.1016/j.eswa.2024.126295
36. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874.* (2017). doi: 10.48550/arXiv.1705.07874