

Characterization of Motor Signatures in Autism Using Interpretable Learning on 3D Motion Data

Lara Pereira

University of Coimbra

Teresa Sousa

University of Coimbra

Ricardo Vigário

NOVA University of Lisbon

Miguel Castelo-Branco

University of Coimbra

João Ruivo Paulo

`jpaulo@isr.uc.pt`

University of Coimbra

Research Article

Keywords: Autism, Motor impairments, Motion capture, Classification, Interpretability, Social motor learning, Kinematic analysis

Posted Date: September 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10816089/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Characterization of Motor Signatures in Autism Using Interpretable Learning on 3D Motion Data

Lara Pereira^{1,3}, Teresa Sousa², Ricardo Vigário³,
Miguel Castelo-Branco², João Ruivo Paulo^{1*}

¹Institute of Systems and Robotics, University of Coimbra, Coimbra,
Portugal.

²Coimbra Institute for Biomedical Imaging and Translational Research
(CIBIT), University of Coimbra, Coimbra, Portugal.

²Intelligent Systems Associate Laboratory (LASI), Portugal.

²Institute of Physiology, Faculty of Medicine, University of Coimbra,
Coimbra, Portugal.

³Physics Department, NOVA School of Science and Technology, NOVA
University of Lisbon, Caparica, Portugal.

*Corresponding author(s). E-mail(s): jpaulo@isr.uc.pt;
Contributing authors: lara.pereira@isr.uc.pt; tsousa@uc.pt;
mcbranco@fmed.uc.pt;

Abstract

Background: Autism is typically characterized by deficits in social communication and interaction, yet the mechanisms underlying associated motor impairments and difficulties in imitation abilities remain poorly explored.

Methods: This study leveraged the Move4AS dataset, comprising motion capture data from 34 participants (14 autistic and 20 neurotypical) performing emotional walking and social dancing tasks. Data were analyzed using machine learning classification with interpretability methods, examining discriminant movement patterns during the imitation of the naturalistic tasks across Support Vector Machine and Convolutional Neural Network models.

Results: Classification accuracy increased with task complexity on both models, indicating that emotional and social demands amplify group discriminability. Interpretability analyses revealed that discrimination between groups is systematically associated with distinct key body regions per task, demonstrating that whole-body movement contains discriminative information that characterizes and differentiates autism.

Conclusions: Our interpretable machine learning framework offers a generalizable approach for characterizing complex motor patterns in clinical populations, highlighting the importance of motor assessments that incorporate expressive, socially learned tasks and multiple kinematic dimensions.

Keywords: Autism, Motor impairments, Motion capture, Classification, Interpretability, Social motor learning, Kinematic analysis

1 Introduction

Autism is a complex neurodevelopmental condition marked by manifestations that emerge in the early years of life, and often with lifelong effects, that influence the way individuals perceive, interpret, and engage with the world around them [1]. These vary from individual to individual, in terms of nature and severity. Clinically, the diagnosis is based on two core domains: social communication challenges and restricted and repetitive behaviors and interests [2, 3]. However, autism encompasses a broader range of characteristics beyond these diagnostic criteria. In particular, differences in emotional [4, 5] and motor functioning are frequently observed [2, 3]. Motor differences such as coordination challenges, weak imitation skills, and altered movement patterns constitute some of the most frequently observed motor signs across different ages and levels of functioning [2]. These motor characteristics carry profound implications beyond movement itself, influencing social interaction, emotional expression, and adaptive behaviors throughout development [2, 6].

Despite this recognition of the involvement of motor components in autism, research has primarily focused on basic motor skills, balance, gait stability, and fine motor control [2], rather than emotionally charged or socially embedded movement. The question of how autistic individuals express emotions through whole-body kinematics or coordinate their movements during social interactions remains inadequately addressed. This gap is particularly striking, given that movements in everyday life inherently convey emotional states, social intentions, and interpersonal meaning [7]. Imitation provides a key context for examining how this information is processed and expressed. As a cornerstone of social learning and emotional understanding, imitation requires not only perceiving others' actions, but also mapping those observations onto one's own motor system while encoding the context of the movement [8, 9].

Although imitation and the processes it depends on play a significant role in embodied social cognition [10], they have received limited attention within computational research on expressive behavior. Over the past decade, research has expanded beyond facial and vocal cues to include full-body motion as a channel of emotional expression [4, 5, 11–14]. However, within autism research specifically, this broadening in modality has not been matched by a broadening in focus: most studies continue to concentrate on emotion recognition, particularly of facial expressions, rather than on how emotion is produced or expressed [1, 15]. Whether the same recognition/production imbalance observed for facial expression also holds for whole-body, embodied emotional communication remains largely unexamined, and, within autism research, body movement

itself still accounts for only a small fraction of emotion-related studies relative to other modalities [1], leaving this aspect of emotional communication underexplored. Additionally, despite growing interest in 3D motion data, the explainability of computational models applied to it is often limited [16]. This limitation makes it difficult to relate model outputs to interpretable motor or biomechanical features. In the context of neurodevelopmental conditions such as autism, interpretability is essential not only for scientific transparency but also for clinical translation and trustworthiness.



Fig. 1: Overview of the proposed interpretable classification pipeline. Using the motion capture data from the Move4AS dataset [17], the raw data first undergoes preprocessing to ensure signal consistency. The architecture then splits into two parallel streams for analysis: a feature-based classification explained via SHAP, and a data-driven classification interpreted using LRP.

To address these challenges, this work investigates how emotional and social contexts modulate motor execution in autism, through an interpretable analysis of 3D motion capture data from the Move4AS dataset [17], following the pipeline in Figure 1. We examine naturalistic imitation tasks charged with expressive cues where participants perform emotionally valenced actions (walking confidently vs. sadly) and socially configured movements (dancing solo vs. duo). To extract meaningful insights from this high-dimensional spatiotemporal data, we employ two complementary approaches: (i) Support Vector Machine (SVM) classification with SHapley Additive exPlanations (SHAP) to identify which kinematic features most strongly differentiate groups, and (ii) Convolutional Neural Network (CNN) with Layer-wise Relevance Propagation (LRP) to reveal which spatiotemporal patterns drive classification decisions. This dual strategy provides both feature-level interpretability (identifying important movement characteristics) and data-level interpretability (localizing where and when these differences manifest).

Hence, the contributions of this work are:

- Analysis of full-body motion using emotional and social imitation tasks designed to elicit expressive and ecological movements in autism.
- A comparison of explainable AI (XAI) techniques to provide comprehensive characterization of autistic motor function.
- Identification of body regions and kinematic features that distinguish autistic from neurotypical individuals during imitation of emotional (confident/sad) and socially-coordinated (solo/duo) movement patterns, revealing how emotional valence and movement complexity modulate motor execution differently between these groups.

2 Related Work

Imitation paradigms provide a rigorous and controlled means of assessing motor behavior [18], yet they have rarely been employed to systematically investigate the modulation of movement by emotional and social factors. This oversight is particularly significant, given that imitation impairments in autism are not homogeneous, but instead emerge from several distinct underlying pathways, suggesting that rigorous testing must account for broader contextual variables to fully understand these mechanisms [19]. McWeeny *et al.* substantiate this heterogeneity directly, establishing that motor execution processes are the strongest mediators of imitation accuracy differences between autistic and non-autistic children, with smaller but significant contributions from social motivation, action representation, and executive function. These findings suggest that motor execution differences constitute a foundational baseline shaping how actions are produced in autism. This is critical since group differences in kinematics may reflect underlying motor traits as much as, or more than, context-driven modulation.

Furthermore, imitation research in autism reveals task-dependent patterns, with some studies reporting impaired imitation abilities and others reporting intact performance [20]. Within this mixed body of evidence, a consistent dissociation emerges: autistic individuals reproduce action goals in a manner comparable to neurotypical individuals, but exhibit reduced fidelity in copying movement style and kinematics [20]. In addition, emotional and social contexts modulate motor performance, although their influences can manifest in different ways in autism. Emotional context appears to shape both the perception and production of biological motion, and autistic individuals have demonstrated poor performance in processing this higher-order information [21]. At the same time, the social context causes a flexible, top-down modulation of movement that goes beyond the basic principles of motor control. While actions in neutral environments ideally balance task completion with energy cost and fine motor control, social contexts systematically modulate the kinematics of movement [22].

Within this, recent advances in motion-capture technology provide new opportunities to quantify human behavior with exceptional spatial and temporal precision. Marker-based 3D motion analysis offers the ability to track, measure, and evaluate body movements in 3D space, enabling the extraction of detailed features that capture the form, timing, and coordination of actions as they evolve in real time [23–25]. Against research focusing on movements from specific body parts, for example, directed only to the head, hands, upper limbs, or lower extremities [26–29], which limit the full understanding of motor function, some full-body imitation studies have demonstrated the feasibility of these automated, non-invasive tools, with classification accuracies often exceeding 80% where the classification refers to discrimination between autistic and typically developing participants based on the kinematic data [18, 30–32].

Key unsolved challenges remain in quantifying subtle motor differences in autism, despite advances in computational approaches. Notably, limited sample diversity and limited ecological validity represent persistent limitations throughout the literature. Much of the prior work has relied on highly controlled laboratory paradigms, such as simplified imitation tasks, constrained gait cycles, or balance assessments [2, 33, 34],

in which movement is treated as neutral or mechanically isolated. Although these designs ensure rigor, they capture only a narrow subset of real-world motor behavior, excluding the interpretability, expressiveness, and contextual responsiveness crucial to everyday interactions. Moreover, while evidence suggests that contextual modulations of movement in autism are distinct, the representation of these differences has rarely been scrutinized within a unified analytical framework.

To address these methodological limitations while also advancing the understanding of motor function in autism, the present study employs point-light displays to elicit two types of contextually enriched movement – confident or sad walking (functional + emotional) and solo or duo dancing (expressive + social) – and analyzes marker-based 3D motion to determine whether different contexts yield convergent or dissociable kinematic modulation profiles.

3 Data

3.1 Dataset Description

In this study, we used the Move4AS multimodal dataset [17]. It was collected at the Institute of Systems and Robotics, in partnership with the Portuguese Association for Development Disorders and Autism (APPDA), for the study of action-perception and motor function in autism during naturalistic imitation tasks. It contains electroencephalographic (EEG) and motion data. Based on this, the present work intends to focus on the systematization, treatment, and exploration of the 3D motion component, laying the groundwork for an autism-oriented biomechanical analysis.

The experimental setup consisted of a marker-based motion capture system coupled with a processing and display system. Motion data were captured using an OptiTrack Flex 3 system, configured with 10 cameras positioned around a 5×5 meter area, and recorded at 120 Hz. Participant movements were tracked using 37 reflective markers, enabling the automated reconstruction of a skeletal model. Visual instructions were displayed on a 49-inch LCD TV along with temporal references to guide participant performance throughout the experimental protocol. Data collection was managed through a dedicated workstation running OptiTrack’s Motive software. Temporal synchronization between the motion data and the experimental events was achieved by integrating trigger signals into the recorded data stream.

The resulting dataset contains full-body kinematic data from 34 participants, divided into 20 neurotypical controls (13 males and 7 females, with an average age of 25.9 ± 3.8 years) and 14 autistic individuals (12 males and 2 females, with an average age of 27 ± 6.8 years). Each participant performed two imitation paradigms, walking and dancing. These paradigms explore different social and emotional dimensions within motor tasks, integrating both into the experimental design. Each repetition of the imitation task is a trial, which followed a standardized temporal sequence divided into four stimulus:

1. Baseline: A fixation cross was presented;
2. Instruction: Point-light animation (skeletal figurine) with text legend that shows the movement to be imitated;

3. Execution: During this period, task performance occurred.
4. Return: The participants returned to the initial position to prepare for the next trial.

Throughout these repetitive trials, event triggers were embedded in the data to mark both the trial conditions (confident vs. sad or solo vs. duo) and temporal segments/stimulus (fixation, instruction, execution, return), enabling identification of movement phases and precise temporal segmentation for a specific trial-based analysis.

The data were organized as cell matrices, where rigid body data, used in this work, were stored as a $1 \times N_s$ cell array of a 7×21 matrix representing joint positions, where N_s represents the number of time samples. Each rigid body joint that models the skeleton was identified according to its respective anatomical location in the human body, as represented in Figure 2.

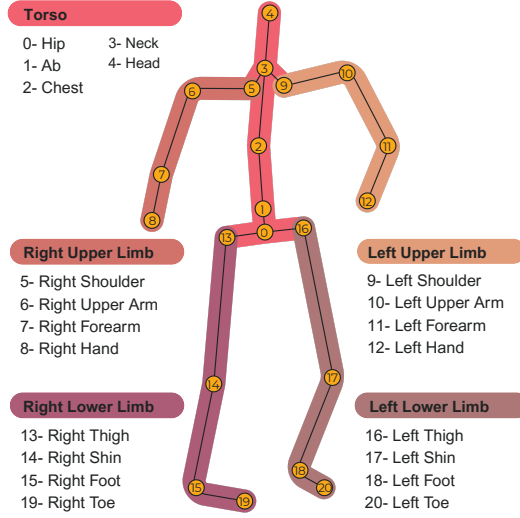


Fig. 2: Identification of the 21 rigid body joints.

3.2 Data Preprocessing

To ensure the quality and usability of the data, a comprehensive preprocessing was applied to prepare, correct, and optimize the data for subsequent analysis. To achieve this, it was essential to address factors such as noise or errors in the captured sequences, temporal variations, as well as anthropometric differences among subjects, and movement orientations.

3.2.1 Parsing and Selection

The first step was to access the motion files to extract the subset containing the Cartesian coordinates of the rigid body joint positions for analysis. Since each file

corresponds to a complete recording block of a participant’s task, each trial was segmented to retain only the period of interest for analysis, the execution phase. After that, the data were subjected to systematic visual inspection to identify and document artifacts. A trial-by-trial assessment was performed to detect issues such as uninterpretable skeletal shape, incorrect upper or lower limb tracking, and marker occlusions. Trials containing significant artifacts were excluded. In this context, prioritizing the broadest possible preservation of the dataset, for the walking data, it became essential to cut the period with the highest prevalence of errors, the first two seconds of the motion execution period, and exclude the upper limbs.

The distribution of usable trials across groups and conditions is shown in Figure 3. The imbalance among task conditions comes from multiple levels. First, the number of participants differs between groups, where each participant contributed varying numbers of trials (with four recording blocks for the control group and three recording blocks for the autistic group). Also, the preprocessing steps resulted in different numbers of valid trials per participant and condition. This multi-level imbalance reflects one naturalistic characteristic of the dataset, accounted for in our analysis.

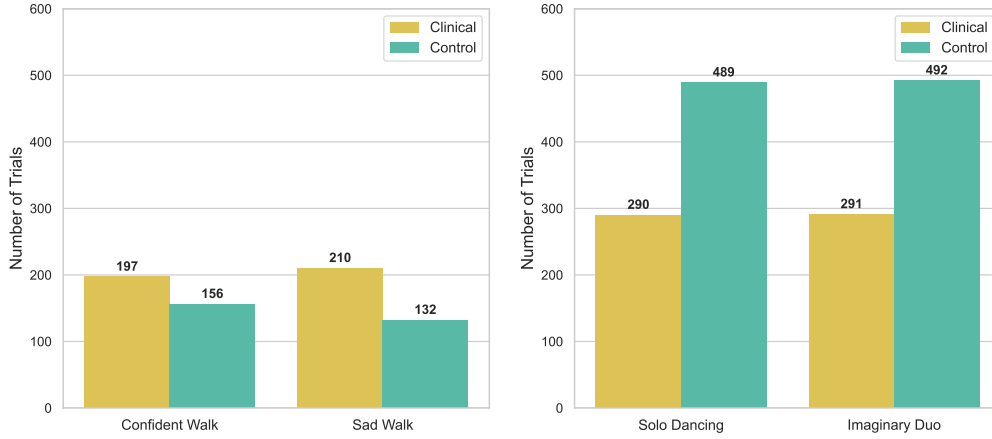


Fig. 3: Number of trials by group and task condition.

3.2.2 Temporal Interpolation

Due to the variable sampling rate inherent in the polling-based data acquisition process, motion samples exhibited non-uniform temporal intervals. To enable accurate comparisons, linear interpolation was applied to create uniformly sampled time series. For each execution segment within a trial, the interpolation was computed as:

$$S(t_{int}) = S(t_i) + \frac{(t_{int} - t_i)}{(t_{i+1} - t_i)} \cdot (S(t_{i+1}) - S(t_i)) \quad (1)$$

Where $S(t)$ is the 3D motion sample of a rigid body joint at timestamp t , t_{int} is the interpolated time point and t_i and t_{i+1} are the preceding and following recorded timestamps, respectively.

This procedure also ensured that all execution segments contained the same duration, with the same number of samples.

3.2.3 Normalization

Scale

Considering the inter-participant variability in body dimensions and to enable valid comparisons of movement patterns across individuals, segment-wise skeleton normalization was performed to create an equally sized skeleton across subjects. For each direct joint-to-joint segment (e.g., shoulder to forearm) the segment length was normalized to the average length of that segment computed across all participants in the dataset [35]. This normalization approach was applied to each skeletal segment while preserving the hierarchical structure and kinematic relationships between joints.

Position and Orientation

Establishing a consistent reference across all motion sequences was necessary to mitigate the impact of different starting positions, which varied depending on the locations of participants within the capture volume. Drawing on the methodologies presented in [36], two distinct positional normalization strategies were applied. These are dependent on the below-presented analysis methodology, and are as follows:

- i) *First-Pose Position Normalization*: This involves setting the same origin for all motion sequences, while preserving the natural translational dynamics throughout the remainder of the sequence. Taking the skeleton’s root joint (hip) position from the first frame as the origin $(0, 0, 0)$, a per-trial coordinate translation was made, according to:

$$S'(t, J) = S(t, J) - S(0, 0) \quad (2)$$

- ii) *All-Poses Position Normalization*: This involves constraining the hip joint to remain at the global coordinate origin $(0, 0, 0)$ throughout the entire motion sequence. This transformation converts the motion from world-space to root-relative space, isolating the intrinsic joints kinematics from extrinsic translational displacement of the body. For this, each trial sample was processed individually following:

$$S'(t, J) = S(t, J) - S(t, 0) \quad (3)$$

$S'(t, J)$ denotes the normalized position of joint J at time t , obtained by subtracting the hip joint position either (i) from the first frame $S(0, 0)$ or (ii) from every frame $S(t, 0)$ in each joint motion sequence.

In the preprocessing pipeline, position normalization was performed by default using the approach described in (i).

Orientation normalization addresses rotational misalignment arising from variable

initial facing directions relative to the motion capture coordinate system. For this, a first-frame orientation normalization was applied based on the methodology described in [36], to standardize each sequence’s initial heading while preserving the relative rotational dynamics of the movement.

The procedure began by selecting the first frame of each sequence as the reference pose. From this pose, the initial direction of the skeleton was estimated by defining a body vector, in this case the pelvic vector, by extracting the 3D coordinates of the right and left thighs, as expressed by:

$$\mathbf{v}_{\text{pelvis}} = S(0, J_{RThigh}) - S(0, J_{LThigh}) \quad (4)$$

This vector represents the transverse axis of the pelvis and points from left to right. Then, within the horizontal plane (XY plane), the orientation of the pelvic vector relative to a fixed reference axis was calculated following:

$$\theta = \arctan 2(\mathbf{v}_y, \mathbf{v}_x) \quad (5)$$

Let $\mathbf{v}_y$ and $\mathbf{v}_x$ denote the y - and x -components of the vector $\mathbf{v}_{\text{pelvis}}$, respectively. The $\arctan 2$ function was used to determine the signed angular displacement of the reference vector to the positive x -axis of the coordinate system, represented by θ .

With this, it was possible to compute the corrective rotation that was needed to align the reference vector with a standard orientation, as given by:

$$\phi = -\theta \quad (6)$$

This angle ϕ was defined to put the pelvic vector exactly along the x -axis, with the skeleton facing the positive y -direction. For this, a planar rotation of the whole skeleton was performed in XY plane, rotating around the z -axis (up/down axis) by the angle ϕ .

The definition of the rotation matrix R_z is provided in the equation that follows, and the corresponding rotational transformation (shown in the subsequent equation) was applied uniformly to every joint position throughout the entire motion sequence.

$$R_z(\phi) = \begin{bmatrix} \cos(\phi) & -\sin(\phi) & 0 \\ \sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

$$S'(t, J) = S(t, J) \cdot R_z(\phi) \quad (8)$$

4 Methodology

4.1 Pipeline

The proposed pipeline, illustrated in Figure 1, processes 3D motion capture data through two complementary pathways: 1) a classical machine learning path that extracts handcrafted kinematic features for SVM classification with SHAP-based explanations, and 2) a deep learning path that processes spatio-temporal tensors through a CNN with LRP for relevance mapping. These parallel paths are integrated to provide insights into how emotional and social states influence motion patterns in autism.

4.2 Motion Classification Framework

Within this motion analysis framework, we adopted a classification strategy to identify autistic or neurotypical subjects at the trial level.

4.2.1 Feature-Based Analysis

To enable discrimination between autistic and neurotypical movement patterns, a comprehensive set of handcrafted kinematic features was extracted from the preprocessed 3D motion data [37]. These features were designed to capture multiple aspects of motor execution, including displacement patterns, speed characteristics, angular dynamics, and body coordination. An overview of the resulting kinematic feature set is reported in Table 1.

Table 1: Motion Features

Feature Category	Description
<i>Displacement-Based</i>	Total displacement: Sum of the overall displacement of the joints. Individual joint’s displacement ratio: Ratio of each joint’s displacement and the total displacement. Upper-lower body displacement ratio: Ratio between upper body’s joints’ displacement (above hip) and the lower body’s joints’ displacement.
<i>Velocity-Based</i>	Average joint’s speed: Total joint’s displacement divided by the execution’s time interval.
<i>Task-Specific Spatial</i>	Individual joint’s position differences (Walking): Difference between initial and final position. Floor area coverage (Dancing): Area covered by foot movements. Floor circularity (Dancing): Geometric profile of the movement path, expressed in terms of circularity and elongation.
<i>Angular</i>	Mean angle: Average angle of joint triplets. Angle range: Difference between maximum and minimum angles of joint triplets. Total angle change: Sum of angular variations within a trial for joint triplets. Min/max angles: Minimum and maximum angles of joint triplets. Left-right symmetry: Comparison of overall angular variation between corresponding left and right joint triplets.

Each feature was computed from the execution period considered for each trial, with joint positions referenced to the hip joint coordinate at the trial’s first sample. Angle-based features were computed for the following anatomical locations: torso, neck, left and right hips and knees, and for left and right shoulders and elbows (dancing only).

This process resulted in 79-dimensional feature vectors for the walking tasks and 100-dimensional feature vectors for dancing. The higher dimensionality of

dance reflects the inclusion of features of the upper limbs and additional spatial characteristics relevant to dance movements.

Support Vector Machine Classification

Binary classification based on the extracted kinematic features was performed using a SVM. The classifier was implemented in Python using scikit-learn, with probability estimation enabled, and default hyperparameters.

Leave-One-Group-Out (LOGO) cross-validation was employed for model evaluation. In each iteration, all trials from a participant of each group were held out as the test set, while the model was trained on the trials from all remaining participants. The performance metrics were then averaged across iterations.

4.2.2 Deep Data-Driven Analysis

To explore whether additional discriminative information exists in the sequential structure of motion data, we employed a CNN that processes motion trajectories as spatiotemporal tensors.

Model Architecture

To learn spatiotemporal representations from skeletal data, this work proposes a dual-input fusion extension CNN that processes complementary data through parallel branches. The fusion approach, which includes a mid-level concatenation strategy and dual-branch processing of coordinates and vector norms, represents an original methodological contribution designed to enhance the discriminative power between autistic and neurotypical individuals. The representation of the architecture is shown in Figure 4, followed by its explanation.

Before model training, the skeletal data underwent the preprocessing steps previously described, where the position normalization approaches were applied one for each input representation. The All-Poses Position Normalization and the First-Pose Position Normalization methods were applied separately, in this order, one for each of the following input branches:

- **Coordinates Branch:** 3D skeletal coordinates. Since motion is attached to the global origin, this root-centric representation enables the analysis of the motion patterns, focusing purely on the morphological and kinematic properties of the movement itself.
- **Norms Branch:** Vector norms of joint movements. As a magnitude-based measure, it provides information on spatial occupancy and relates to movement intensity or joint activity.

Each branch comprised identical architectural components: an initial 1D convolutional layer incorporating batch normalization, ReLU activation (enabling non-linear feature learning), and dropout regularization (rate = 0.3) to establish foundational feature extraction while preventing early overfitting. The core processing was performed through configurable residual blocks that implemented adaptive skip connections to

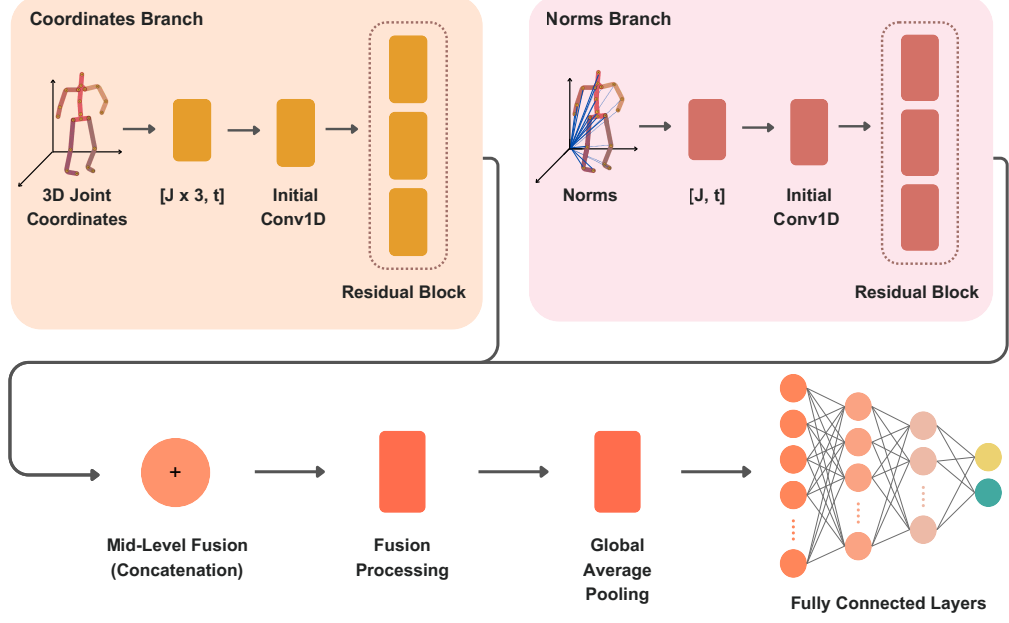


Fig. 4: Dual-Input CNN architecture for motion classification. The coordinates branch (left) processes 3D joint positions, while the norms branch (right) processes movement magnitudes. Both branches employ initial Conv1D layers and residual blocks. Features are integrated via mid-level fusion (concatenation), followed by fusion processing, global average pooling, and fully connected layers for binary classification.

facilitate gradient flow and enable deeper network training. Each residual block contained three sequential convolutional layers with ReLU activation following each batch normalization, using kernel sizes of 5, 3, and 1, which allowed the network to capture features at multiple scales. Batch normalization and dropout (rate = 0.5) were applied after each convolutional layer to enhance regularization and training stability. The final ReLU activation was applied after the element-wise addition of the skip connection, following standard residual learning principles.

Following this, feature integration was accomplished through mid-level fusion, where feature maps from both branches were concatenated. This concatenated representation was subsequently processed through an additional convolutional layer that reduced dimensionality, incorporating batch normalization, ReLU activation, and dropout before global average pooling, to generate compact feature vectors for classification.

The classification head consisted of a three-layer fully connected network with progressive dimensionality reduction, and ReLU activation between layers, transitioning from the initial channel dimension to channels/2, then channels/4, and finally to 2 classes for binary classification. Dropout was incorporated between each fully connected layer to mitigate overfitting risks.

Training Methodology

To address class imbalance at the trial level, weighted cross-entropy loss was employed with class weights calculated using scikit-learn’s balanced approach, which inversely weights classes proportional to their frequency in the training data. This ensured that the model maintained sensitivity to minority class examples during optimization.

The training process employed a conservative learning strategy with a batch size of 8 and an initial learning rate of 1×10^{-5} . Learning rate scheduling was managed using `ReduceLROnPlateau`, which adaptively reduced the learning rate when validation performance plateaued. Regularization was applied through multiple mechanisms: weight decay (1×10^{-3}) and L1 regularization ($\lambda = 1 \times 10^{-5}$) to limit parameter overgrowth, as well as gradient clipping with a maximum norm of 1.0 to ensure training stability.

Training was configured for a maximum of 200 epochs with enhanced early stopping criteria to mitigate overfitting. The early stopping mechanism monitored validation loss with a patience of 15 epochs and required a minimum improvement threshold of 0.001 to continue training. Upon early stopping, the model weights corresponding to the best validation performance were restored to ensure optimal generalization.

Model selection was based on validation balanced accuracy across all cross-validation folds, with the best-performing model selected for final evaluation on the holdout test set. This procedure ensured robust model selection while maintaining strict separation between validation and test data.

Data Configuration

In this CNN implementation, data splitting at the participant level was maintained and fixed throughout the analysis. To ensure robust generalization and prevent data leakage, while also considering temporal and computational constraints, a two-stage framework was implemented:

1. **Holdout Split:** The participant pool was randomly divided into development and holdout test sets (with an approximately 80/20 proportion, participant-based), maintaining equal representation of clinical and control participants in the test set.
2. **K-Fold Cross-Validation:** The development set underwent stratified k-fold cross-validation ($k=3$) at the participant level, ensuring complete separation between training and validation participants within each fold.

4.3 Explainable AI Methods

To interpret model decisions and identify discriminative motion patterns, XAI was integrated into both classification pathways.

4.3.1 SHapley Additive exPlanations

SHAP was used to provide an interpretable analysis of kinematic features and their directional influences, within the classical SVM model.

SHAP values were computed using `LinearExplainer` to provide precise feature attributions for the linear SVM classifier. In each cross-validation fold, the trained model and corresponding training data served as background to calculate SHAP values for

the test samples. This enabled quantification of each feature’s contribution to individual classification decisions. Finally, feature importance was determined by aggregating the mean absolute SHAP values across all test samples and folds, which yielded robust and interpretable estimates of feature relevance in the decision boundary.

4.3.2 Layer-wise Relevance Propagation

LRP was applied to identify anatomical regions of broader importance, indicating where discriminative information, temporal consistency, and the magnitude of relevance were located, from an autistic against neurotypical group perspective.

To apply this explanation technique to motion classification tasks, the LRP- ε rule was specifically implemented [38]. The implementation involved registering forward and backward hooks within the model architecture to capture activations and gradients during both forward and backward passes. This enabled the propagation of relevance scores backward from the output layer to the input layer, using an epsilon term to stabilize the division of relevance by weighted activations, according to:

$$R_j = \sum_k \frac{a_j w_{jk}}{\varepsilon + \sum_{0,j} a_j w_{jk}} R_k \quad (9)$$

R_j and R_k denote the relevance scores of neurons in two consecutive layers (lower and upper, respectively), a_j represents the activation of the lower-layer neuron j , and w_{jk} is the weight connecting neuron j to neuron k . The term ε is a small positive constant that ensures numerical stability and acts as a noise filter.

This generated relevance scores for coordinate inputs, representing the x , y , and z positions of each joint, and norm inputs. To derive joint-level relevance scores from the coordinate data, the relevance scores for each direction of each joint were aggregated through a process that took the absolute value of each direction’s relevance, summed these values, and divided by three. These resulting joint-level scores were then normalized to a range of $[-1, 1]$. This normalization ensures consistent visualization across trials with varying magnitude ranges and enables meaningful comparison of relevance patterns, as it reflects relative importance within each motion sequence rather than absolute magnitude.

The analysis extended to class-specific examination by computing relevance scores for all trials within the holdout test set and grouping them according to their classification as either autistic or neurotypical. Average relevance scores were calculated for each class to reveal characteristic patterns between groups. To further emphasize the differences between autistic or neurotypical populations, difference heatmaps were generated by subtracting the average relevance scores of the neurotypical group from those of the autistic group for both joint coordinates and norms. These difference maps were normalized to $[-1, 1]$ to maintain consistency. The resulting zero-centered representation provides immediate insight into discriminative attention: positive values indicate regions where the network focuses more heavily when classifying autistic trials, negative values indicate greater attention for neurotypical trials, and near-zero values represent comparable attention across groups.

5 Experimental Results

This section presents the classification performance of feature-based learning (SVM) and deep learning (CNN) approaches, as well as the explainable analysis inherent in distinguishing between autistic and neurotypical participants in the motion tasks.

5.1 Classification Performance

Table 2 summarizes the classification metrics results for the two methods used for classification.

Table 2: Classification results across tasks and conditions.

Paradigm	Condition	Method	Balanced Accuracy	Precision	Recall	F1-Score
Walking	Confident	SVM	0.66	0.70	0.64	0.63
		CNN	0.64	0.79	0.36	0.49
	Sad	SVM	0.71	0.78	0.69	0.69
		CNN	0.74	0.68	0.84	0.75
Dancing	Solo	SVM	0.63	0.66	0.46	0.49
		CNN	0.75	0.84	0.61	0.71
	Duo	SVM	0.57	0.60	0.36	0.39
		CNN	0.79	0.84	0.72	0.78

For walking tasks, both methods achieved moderate to good classification performance. In the confident walking condition, the SVM marginally outperformed the CNN in balanced accuracy (66% vs. 64%). However, the CNN exhibited substantially higher precision (79%) but considerably lower recall (36%), suggesting a more conservative classification strategy that minimized false positives at the expense of missing true cases. In the walking sad condition, the CNN achieved a balanced accuracy of 74%, outperforming the SVM (balanced accuracy: 71%). The CNN showed notably higher recall, correctly identifying 84% of true positive cases, though with slightly lower precision.

Classification performance was notably higher in the dancing paradigm, particularly for the CNN model. For solo dancing, the CNN outperformed the SVM (balanced accuracy: 75% vs. 63%), with particularly strong precision (84%). In duo dancing, the CNN achieved the highest overall performance across all conditions. The SVM showed limited performance in both dancing conditions, with F1-Score below 50%, indicating difficulty in capturing the discriminative patterns in expressive movement.

5.2 Explainable Analysis

To interpret the decision logic of the classification architectures, two complementary explainable AI (XAI) strategies were employed.

Figure 5 presents the SHAP values for the top features contributing to SVM classification decisions. Features are ranked by their mean absolute SHAP value, indicating their overall importance to the model’s predictions.

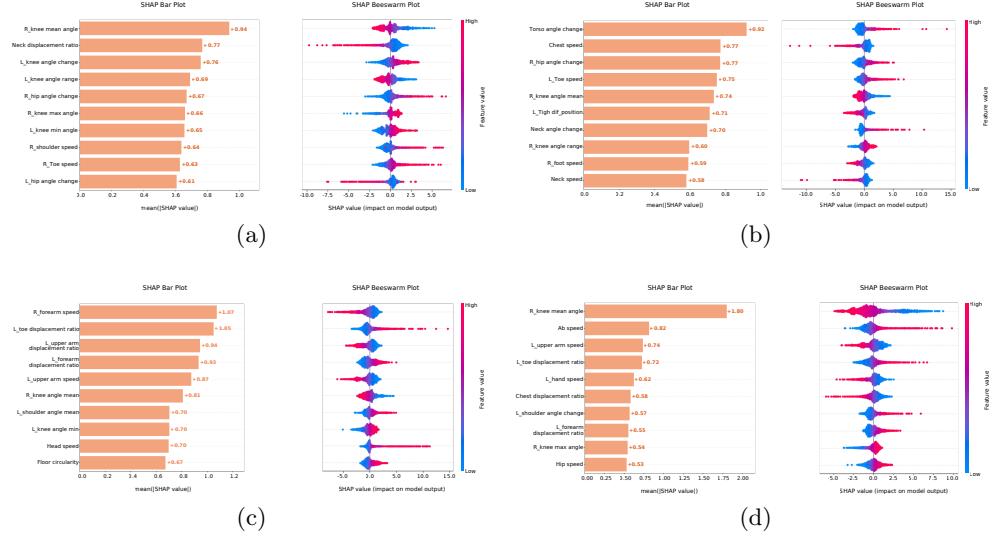


Fig. 5: SHAP results for the walking (top) and dancing (bottom) tasks. Each panel contains a bar plot (left) showing the mean absolute SHAP values for the top 10 most important features, and a beeswarm plot (right) showing how feature values relate to model predictions. In the beeswarm plots, each point represents a sample; the x-axis indicates the feature’s impact on the prediction (positive values: autism; negative values: neurotypicality) and colour encodes feature value (red: high; blue: low). (a) Confident Walk. (b) Sad Walk. (c) Solo Dance. (d) Duo Dance.

In the confident walk condition (Figure 5a), lower-body kinematics dominated the feature importance ranking. Knee-related features were particularly influential, with right knee mean angle showing the highest importance (mean $|\text{SHAP}| = 0.94$). Hip angle dynamics exhibited lateralized patterns: right hip angle change showed predominantly positive SHAP values, indicating that larger angular changes were associated with autism classification, while left hip angle change demonstrated the opposite relationship. Neck displacement ratio showed a strong inverse relationship (mean $|\text{SHAP}| = 0.77$), with lower values contributing to autism classification. Velocity-based features demonstrated highly positive relationships, indicating faster movements associated with a higher likelihood of autism.

The sad walk condition (Figure 5b) demonstrated a fundamental shift in discriminative features compared to confident walking. Upper-body and axial kinematics emerge in the feature importance ranking, with torso angle change showing the highest importance (mean $|\text{SHAP}| = 0.92$), followed by chest speed (0.77) and neck angle change (0.70). This reorientation toward trunk and postural features suggests

that sad states amplify group differences in axial motor control and postural stability. Speed parameters revealed region-specific patterns. Upper-body velocity features (chest, neck) exhibited complex relationships with classification, with high values contributing to neurotypical predictions depending on other kinematic factors.

Solo dancing (Figure 5c) demonstrated strong upper-limb dominance in feature importance, with right forearm speed showing the highest importance (mean $|\text{SHAP}| = 1.07$), followed by upper arm displacement ratios (left: 0.94, right via forearm: 0.93) and upper arm speed (0.87). Head speed (0.70) and shoulder kinematics (0.70) further emphasized the prominence of upper-body dynamics in expressive movement classification. Notably, displacement ratio features, quantifying the spatial extent of movement, emerged alongside velocity parameters, suggesting that both the amplitude and speed of arm movements distinguish groups during solo dance. The mixed associations suggest that discriminative patterns lie in the temporal dynamics or coordination of movements rather than simple velocity or position magnitude. The presence of floor circularity (0.67) among the top features indicates that spatial movement patterns, beyond pure joint kinematics, contribute to classification.

Duo dancing exhibited the highest overall feature importance values across all conditions, with right knee mean angle achieving exceptional discriminative power (mean $|\text{SHAP}| = 1.80$), nearly double the highest values observed in other tasks. The feature profile was more distributed across body regions compared to solo dancing. While upper-limb features remained important, trunk and core kinematics gained prominence: ab speed (0.82), chest displacement ratio (0.58), and hip speed (0.53). Also lower-limb features remained influential, indicating that full-body coordination contributes to group discrimination, potentially revealing more comprehensive motor control differences.

Looking at the LRP values, we identify other relevance patterns to distinguish between groups. For confident walking (Figure 6a), the toe joints exhibited a strong negative relevance (≈ -0.75 to -1.00), sustained throughout the movement sequence, indicating that the network relies predominantly on distal lower-limb kinematics to identify neurotypical participants. Proximal regions showed near-zero to weakly positive relevance (hip/abdomen: $\approx +0.25$). The convergence of this weak positive signal across both coordinate and norm representations suggests it may reflect genuine, though minor, hip kinematic differences between groups, with these features contributing to autism classification. However, the concentration of discriminative power in the feet, coupled with near-zero relevance differences in the upper body and most mid-body regions, indicates a focused classification strategy centered on fine gait mechanics rather than gross postural or upper-limb control. The sustained temporal profile essentially demonstrates that these kinematic differences persist across all gait phases rather than emerging at specific movement events.

In sad walking (Figure 6b), lower-limb patterns persisted, with bilateral toes maintaining neurotypical-associated relevance throughout the sequence. With more distinct patterns in the norm relevance map, a temporal gradient suggesting differences between groups becomes more pronounced as sad walking continues. Within this, the chest and head regions exhibited late-phase emergence of positive relevance.

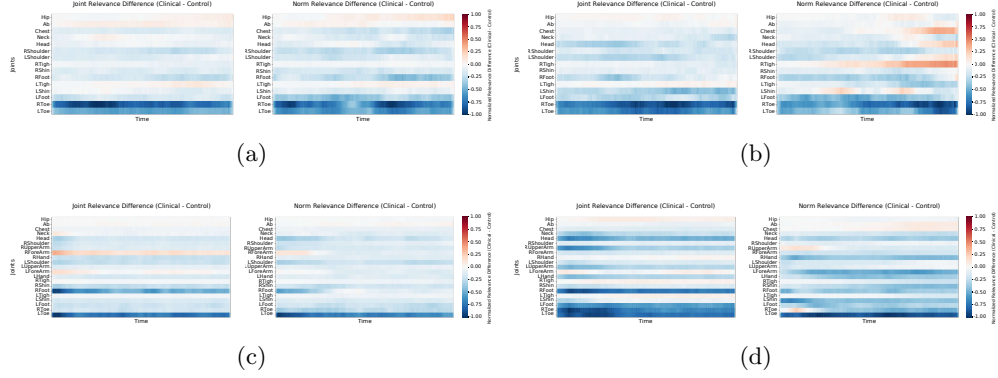


Fig. 6: LRP results for the walking and dancing tasks. The heatmaps represent the difference in joint relevance scores between clinical and control groups across time and body joints, considering coordinates (left) and norms data (right). All heatmaps use a normalized scale ranging from -1.00 to 1.00 , where positive values (warmer colours) indicate higher relevance when analysing the autistic group, negative values (cooler colours) indicate lower relevance, and values near zero represent minimal difference between groups. (a) Confident Walk. (b) Sad Walk. (c) Solo Dance. (d) Duo Dance.

This temporal specificity indicates that upper-body movement magnitude differences manifest progressively rather than immediately. Importantly, this upper-body discrimination occurred essentially in norm-based relevance (movement magnitude), not in joint-coordinate relevance (spatial positioning). This dissociation indicates that while the spatial trajectories are more similar between groups, the amplitude of movement differs systematically as the trial progresses.

Solo dancing (Figure 6c) demonstrated a dual-region discrimination strategy with pronounced temporal dynamics. The right forearm showed positive relevance, strongest during early phases ($\approx +0.50$) in coordinate space, fading to weaker signals during sustained execution ($\approx +0.20$ to 0.30). Bilateral toe kinematics exhibited the strongest discrimination, also higher in the early phases, with robust neurotypical-associated relevance, which diminished during sustained execution. The temporal specificity, with maximum discrimination during initiation rather than sustained throughout, indicates that this expressive movement reveals group differences most clearly when motor planning or postural establishment is involved. As the dance becomes established, discrimination diminishes, suggesting either that groups converge toward more similar execution patterns, that initial planning is more informative than sustained performance, or that compensatory strategies become effective once movement is underway.

Duo dancing (Figure 6d) produced the most distributed discrimination pattern across all experimental conditions, with substantially intensified signals. In the joint coordinate space, discrimination was stronger across head, upper-limb, and lower-limb regions with predominantly neurotypical-associated patterns. The head exhibited strong negative relevance sustained throughout the sequence with minimal signal in

norm space, indicating that head spatial positioning distinguishes neurotypical participants while movement magnitude remains similar between groups. Upper arm positioning showed strong relevance with sustained temporal profiles, as well as distal lower-limb features, which exhibited very strong neurotypical-associated patterns (≈ -1.00), extending robust discrimination throughout the lower kinematic chain. The sustained temporal profile indicates that discrimination patterns remain approximately consistent throughout movement reproduction, but with maximum discrimination during initiation in joint coordinates, revealing higher distinction during early stages of the movement. Critically, the substantially stronger discrimination in both coordinate and norm representations compared to other conditions indicates that this condition amplifies group differences.

6 Discussion

This study reveals that motor signatures in autism are fundamentally context-dependent and movement demands systematically modulate both the spatial distribution and the temporal dynamics of kinematic differences. Rather than representing a fixed deficit, autistic motor function appears to involve adaptive strategies that respond to task complexity, emotional content, and social demands, following distinct patterns compared to those of neurotypical individuals. These findings align with the movement-sensing perspective in autism [39, 40], which reconceptualizes motor differences as context-sensitive sensory-motor variations that contribute to adaptive and compensatory mechanisms, rather than static impairments.

The persistence of distal lower-limb discrimination across all conditions provides evidence for a hierarchical organization of motor differences. Anterior foot-based signatures emerged as a fundamental, context-independent marker, suggesting that distal motor control differences constitute a baseline characteristic that persists regardless of task demands. Consistent with prior research identifying altered foot mechanics and distal lower extremity strategies as characteristic features of autism gait [41, 42], our results extend these observations by demonstrating that distal motor differences also emerge in a dance context. This signature then interacts with context-dependent features that emerge under specific movement conditions, creating a layered architecture where fundamental execution differences are modulated by task-specific control processes.

In fact, the nature of this modulation depends critically on the type of demand imposed. The emotional context introduces dynamic temporal evolution: while confident walking maintains stable anterior foot-based discrimination throughout gait cycles, sad walking exhibits progressive intensification of distal features alongside late-phase emergence of upper-body magnitude differences. This temporal progression suggests that emotional-motor integration operates through active, evolving processes rather than static postural or kinematic deviations. Importantly, upper-body differences manifest in movement amplitude rather than spatial trajectories, indicating preserved gross motor planning but altered execution dynamics. This dissociation between planning and execution is consistent with theoretical frameworks in autism research that attribute differences in motor behavior to disruptions in prospective

sensory-motor coordination, where high-level intentions may be preserved but their translation into coordinated, timed action is altered [39].

Social demands produce an even more pronounced reorganization. The transition from solo to duo dance transforms discrimination from localized to comprehensive full-body patterns, engaging head, upper-limb, and lower-limb kinematics simultaneously. This distributed engagement suggests that reproducing socially structured movements requires integrated multi-system coordination. The presence of both magnitude and positional differences across body segments in the duo dance condition is consistent with challenges across the observation–representation–execution cascade, potentially reflecting differences in movement encoding, internal motor representation, or execution fidelity. These may pose a greater challenge for the mirror neuron system in individuals with autism. The temporal analysis adds another dimension to understanding this context effects. Dance, analyzed from movement onset, produces maximum discrimination during initiation that diminishes as execution progresses. Motor planning difficulties manifest most strongly during movement initiation and early execution phases, when demands for forming and implementing internal motor representations are highest. While previous studies suggest that basic, cue-based action planning may be relatively preserved in autism [43], our findings indicate that socially structured and internally driven movements, such as duo dance, place substantially greater demands on motor planning processes, revealing difficulties during movement initiation and early execution. This view is further supported by longitudinal evidence of atypical development in visuomotor planning mechanisms underlying precise motor output [44].

Moreover, the divergence between SVM and CNN performance across tasks provides methodological insight into these distinctions. CNN superiority in dancing contexts reflects the value of hierarchical representation learning for capturing complex spatiotemporal patterns in expressive movement, while comparable performance between methods in walking suggests that explicit kinematic features adequately capture the relevant variance in stereotyped locomotion. This pattern reinforces that different movement contexts engage qualitatively different motor processes with distinct discriminative structures. Taking the classification results, despite the variability driven by context, the analysis also demonstrated robust intrapersonal consistency. This points to the fact that while motor strategies adapt to external demands, each individual retains a strong, idiosyncratic motor signature.

Collectively, these findings contribute to a more nuanced understanding of the autistic phenotype, indicating that motor strategies are continuously reconfigured to meet the varying demands of real-world social interactions and emotional contexts.

All the findings evidenced in this work were possible through the use of the Move4AS dataset, which is the only publicly available dataset for motor function studies in autism, comprising naturalistic imitation tasks. This dataset contains a representative, although limited number of participants, which is compensated by the substantial number of collected trials for each participant, which is in line with current practices.

In addition, in this work, the choice was made to analyze the movement patterns cross-sectionally, which essentially assumes that motor organization remains constant,

ignoring learning contexts. This may not align with dynamic adaptation processes, and the repetitive nature of paradigms could introduce effects of fatigue or disengagement that alter natural motor patterns. Although trial-wise analysis restricts insights into dynamic learning processes, and task repetition may influence the performance, these elements serve as a rigorous validation of the classifier’s robustness, ensuring it captures the most persistent and fundamental motor patterns. This resilience to variability is similarly observed regarding the sample heterogeneity (varying in age, sex, and full-scale IQ), which significantly enhances the generalizability of the findings. The model’s ability to successfully distinguish the autistic group despite such demographic noise indicates that the learned features are not artifacts of a specific age or cognitive profile, but rather represent a fundamental kinematic phenotype of autism that transcends individual variability.

7 Conclusions

This research identified distinctive movement patterns that characterize motor function in autism across different contexts, through interpretable machine learning approaches.

Together, the findings challenge unitary models of motor function in autism, revealing that motor differences emerge from interactions between foundational execution characteristics and context-specific control demands, depending on the situational context. The persistence of distal motor differences as a baseline characteristic, combined with context-dependent modulation under cognitive, social, and emotional demands, suggests a hierarchical organization where fundamental sensorimotor features interface with higher-order control processes that become increasingly challenged as task complexity increases. The capacity for context-dependent modulation, alongside the persistence of individual motor signatures, suggests that motor function in autism involves adaptive strategies rather than fixed deficits, though these strategies produce systematic deviations from neurotypical patterns under increasing cognitive, social, and planning demands. A better understanding of these layered, context-sensitive motor signatures, grounded in both group-level patterns and individual variation, could open pathways for more nuanced assessment approaches and targeted interventions that address specific aspects of motor control rather than treating motor differences as homogeneous impairment. Even more, future work can integrate multimodal data, particularly by combining these motor findings with the EEG data available in the Move4AS dataset. Such approach would offer critical insights into the neural mechanisms that drive motor differences and action–perception coupling in autism. By linking kinematic and neural signatures, research can move beyond descriptive behavioral patterns toward a mechanistic understanding of the neurobiological processes underlying context-dependent motor differences. This multimodal framework holds promise for identifying objective biomarkers that could enhance autism assessment and ultimately inform more targeted intervention strategies.

Declarations

Ethics approval and consent to participate

Data collection and study procedures for the Move4AS dataset used in this study were approved by the Ethical Committee of The Faculty of Medicine of the University of Coimbra (Ref. CE-084/2022). All procedures were carried out in accordance with the relevant guidelines and regulations, including the Declaration of Helsinki and Regulation (EU) 2016/679 (General Data Protection Regulation). Informed consent was obtained from all participants prior to participation, including consent for the dataset to be shared for research purposes. This study constitutes a secondary analysis of the publicly available, de-identified Move4AS dataset.

Consent for publication

Not applicable.

Availability of data and materials

The Move4AS dataset supporting the findings of this study is publicly available in the Figshare repository at <https://doi.org/10.6084/m9.figshare.28296518>, and is described in full in the accompanying data descriptor [17]. This work used only the 3D motion capture modality of Move4AS (the EEG modality was not used), specifically the rigid-body joint data for the 21 anatomical joints shown in Figure 2. From each recording block, only the execution segment of each trial was retained; for the walking paradigm, the first two seconds of the execution window were used and upper-limb joints were excluded, as described in Section 3.2. Block-level selection and exclusion decisions (participant, session, number of trials available, selected trial indices, and exclusion reasons where applicable) are documented in the 3D motion participant information file provided with the dataset. Code implementing select preprocessing steps described in this study is also included in the repository;

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by national funds through FCT - Foundation for Science and Technology, I.P., under the grants UID/00048/2025 (DOI: 10.54499/UIDB/00048/2025) and LA/P/0112/2020 (DOI: 10.54499/LA/P/0112/2020).

Authors' contributions

LP and JRP conceived and designed the study. LP performed data preprocessing and implemented both feature-based (SVM/SHAP) and deep learning (CNN/LRP) analyses, and analyzed and interpreted the results. LP drafted the manuscript. JRP

supervised the study closely, contributing ideas throughout its development, and critically revised the manuscript. RV contributed to the overall supervision of the study. TS and MCB provided clinical expertise on autism, contributed to the interpretation of the findings, and reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the participants who took part in this study, and the Portuguese Association for Development Disorders and Autism (APPDA) for their support in participant recruitment.

References

- [1] Banos, O., Comas-González, Z., Medina, J., Polo-Rodríguez, A., Gil, D., Peral, J., Amador, S., Villalonga, C.: Sensing technologies and machine learning methods for emotion recognition in autism: Systematic review. *International Journal of Medical Informatics* **187** (2024) <https://doi.org/10.1016/j.ijmedinf.2024.105469>
- [2] McCleery, J.P., Elliott, N.A., Sampanis, D.S., Stefanidou, C.A.: Motor development and motor resonance difficulties in autism: Relevance to early intervention for language and communication skills. *Frontiers in Integrative Neuroscience* **7** (2013) <https://doi.org/10.3389/fnint.2013.00030>
- [3] 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. *Review Journal of Autism and Developmental Disorders* (2024) <https://doi.org/10.1007/s40489-024-00435-4>
- [4] Simões, M., Monteiro, R., Andrade, J., Mouga, S., França, F., Oliveira, G., Carvalho, P., Castelo-Branco, M.: A novel biomarker of compensatory recruitment of face emotional imagery networks in autism spectrum disorder. *Frontiers in Neuroscience* **12**, 418353 (2018) <https://doi.org/10.3389/FNINS.2018.00791>
- [5] Monteiro, R., Simões, M., Andrade, J., Branco, M.C.: Processing of facial expressions in autism: a systematic review of eeg/erp evidence. *Review Journal of Autism and Developmental Disorders* 2017 4:4 **4**, 255–276 (2017) <https://doi.org/10.1007/S40489-017-0112-6>
- [6] Nebel, M.B., Eloyan, A., Nettles, C.A., Sweeney, K.L., Ament, K., Ward, R.E., Choe, A.S., Barber, A.D., Pekar, J.J., Mostofsky, S.H.: Intrinsic visual-motor synchrony correlates with social deficits in autism. *Biological Psychiatry* **79**, 633–641 (2016) <https://doi.org/10.1016/J.BIOPSYCH.2015.08.029>
- [7] Bachmann, J., Munzert, J., Krüger, B.: Neural underpinnings of the perception of emotional states derived from biological human motion: A review of neuroimaging research. *Frontiers in Psychology* **9**, 368410 (2018) <https://doi.org/10.3389/>

- [8] McEllin, L., Knoblich, G., Sebanz, N.: Imitation from a joint action perspective. *Mind and Language* **33**, 342–354 (2018) <https://doi.org/10.1111/mila.12188>
- [9] Latrèche, K., Kojovic, N., Pittet, I., Natraj, S., Franchini, M., Smith, I.M., Schaer, M.: Gesture imitation performance and visual exploration in young children with autism spectrum disorder. *Journal of Autism and Developmental Disorders* (2024) <https://doi.org/10.1007/s10803-024-06595-w>
- [10] Ducret, M., Chabanat, E., Kambara, A., Rossetti, Y., Quesque, F.: Monkey do, monkey see? the effect of imitation strategies on visuospatial perspective-taking and self-reported social cognitive skills. *Behavioral sciences (Basel, Switzerland)* **15** (2025) <https://doi.org/10.3390/BS15081112>
- [11] Babu, P.R.K., Martino, J.M.D., Chang, Z., Perochon, S., Carpenter, K.L.H., Compton, S., Espinosa, S., Dawson, G., Sapiro, G.: Exploring complexity of facial dynamics in autism spectrum disorder. *IEEE Transactions on Affective Computing* **14**, 919–930 (2023) <https://doi.org/10.1109/TAFFC.2021.3113876>
- [12] Xia, C., Chen, H., Han, J., Zhang, D., Li, K.: Identifying children with autism spectrum disorder via transformer-based representation learning from dynamic facial cues. *IEEE Transactions on Affective Computing* **16**, 83–97 (2025) <https://doi.org/10.1109/TAFFC.2024.3412032>
- [13] Leung, F.Y.N., Stojanovik, V., Micai, M., Jiang, C., Liu, F.: Emotion recognition in autism spectrum disorder across age groups: A cross-sectional investigation of various visual and auditory communicative domains. *Autism Research* **16**, 783–801 (2023) <https://doi.org/10.1002/AUR.2896>
- [14] Mazzoni, N., Landi, I., Ricciardelli, P., Actis-Grosso, R., Venuti, P.: "motion or emotion? recognition of emotional bodily expressions in children with autism spectrum disorder with and without intellectual disability". *Frontiers in psychology* **11** (2020) <https://doi.org/10.3389/FPSYG.2020.00478>
- [15] Webster, P.J., Wang, S., Li, X.: Review: Posed vs. genuine facial emotion recognition and expression in autism and implications for intervention. *Frontiers in Psychology* **12**, 653112 (2021) <https://doi.org/10.3389/FPSYG.2021.653112/TEXT>
- [16] Chakraborty, A., Karagoz, G., Meratnia, N.: A cross-domain benchmark of intrinsic and post hoc explainability for 3d deep learning models. *Journal of Imaging* **12** (2026) <https://doi.org/10.3390/jimaging12020063>
- [17] Paulo, J.R., Sousa, T., Perdiz, J., Pereira, L., Vasen, M., Mouga, S., Pires, G., Castelo-Branco, M.: A multimodal dataset addressing motor function in autism. *Scientific Data* **12** (2025) <https://doi.org/10.1038/s41597-025-05313-0>

- [18] Zampella, C.J., Sariyanidi, E., Hutchinson, A.G., Bartley, G.K., Schultz, R.T., Tunç, B.: Computational measurement of motor imitation and imitative learning differences in autism spectrum disorder. In: ICMI 2021 Companion - Companion Publication of the 2021 International Conference on Multimodal Interaction, pp. 362–370. Association for Computing Machinery, Inc, New York, NY, USA (2021). <https://doi.org/10.1145/3461615.3485426>
- [19] McWeeny, S., Kolday, A., Wodka, E.L., Mostofsky, S.H., Caffo, B.S., Ewen, J.B.: Mechanisms of altered imitation in autism spectrum disorders. *Autism Research* **18**, 1397–1411 (2025) <https://doi.org/10.1002/aur.70046>
- [20] Forbes, P.A.G., Pan, X., Antonia, A.F.: Reduced mimicry to virtual reality avatars in autism spectrum disorder. *Journal of Autism and Developmental Disorders* **46**, 3788–3797 (2016) <https://doi.org/10.1007/s10803-016-2930-2>
- [21] Todorova, G.K., Hatton, R.E.M.B., Pollick, F.E.: Biological motion perception in autism spectrum disorder: A meta-analysis. *Molecular Autism* **10** (2019) <https://doi.org/10.1186/s13229-019-0299-8>
- [22] Trujillo, J.P., Simanova, I., Bekkering, H., Özyürek, A.: Communicative intent modulates production and comprehension of actions and gestures: A kinect study. *Cognition* **180**, 38–51 (2018) <https://doi.org/10.1016/J.COgnITION.2018.04.003>
- [23] Das, K., Paula Oliveira, T., Newell, J.: Comparison of markerless and marker-based motion capture systems using 95% functional limits of agreement in a linear mixed-effects modelling framework. *Scientific Reports* **13**(1), 22880 (2023) <https://doi.org/10.1038/s41598-023-49360-2>
- [24] Riglet, L., Delphin, C., Claquesin, L., Orliac, B., Ornetti, P., Laroche, D., Gueugnon, M.: 3d motion analysis dataset of healthy young adult volunteers walking and running on overground and treadmill. *Scientific Data* 2024 11:1 **11**, 556 (2024) <https://doi.org/10.1038/s41597-024-03420-y>
- [25] Kinfu, K.A., Pacheco, C., Sperry, A.D., Crocetti, D., Tuncgenc, B., Mostofsky, S.H., Vidal, R.: Computerized assessment of motor imitation for distinguishing autism in video (cami-2dnet). *IEEE Transactions on Biomedical Engineering* (2025) <https://doi.org/10.1109/TBME.2025.3637089>
- [26] Zhao, Z., Zhu, Z., Zhang, X., Tang, H., Xing, J., Hu, X., Lu, J., Peng, Q., Qu, X.: Atypical head movement during face-to-face interaction in children with autism spectrum disorder. *Autism Research* **14**, 1197–1208 (2021) <https://doi.org/10.1002/aur.2478>
- [27] Lakkapragada, A., Kline, A., Mutlu, O.C., Paskov, K., Chrisman, B., Stockham, N., Washington, P., Wall, D.P.: The classification of abnormal hand movement to aid in autism detection: Machine learning study. *JMIR Biomedical Engineering*

7 (2022) <https://doi.org/10.2196/33771>

- [28] Vabalas, A., Gowen, E., Poliakoff, E., Casson, A.J.: Applying machine learning to kinematic and eye movement features of a movement imitation task to predict autism diagnosis. *Scientific Reports* **10** (2020) <https://doi.org/10.1038/s41598-020-65384-4>
- [29] Eggleston, J.D., Harry, J.R., Cereceres, P.A., Olivas, A.N., Chavez, E.A., Boyle, J.B., Dufek, J.S.: Lesser magnitudes of lower extremity variability during terminal swing characterizes walking patterns in children with autism. *Clinical Biomechanics* **76** (2020) <https://doi.org/10.1016/j.clinbiomech.2020.105031>
- [30] Yazdi, S.A., Janghorbani, A., Maleki, A.: Diagnosis of autism in children based on their gait pattern and movement signs using the kinect sensor. *Journal of Medical Signals and Sensors* **14** (2024) https://doi.org/10.4103/jmss.jmss_19_24
- [31] Altozano, A., Minissi, M.E., Alcañiz, M., Marín-Morales, J.: Introducing 3dcnn resnets for asd full-body kinematic assessment: A comparison with hand-crafted features. *Expert Systems with Applications* **270** (2025) <https://doi.org/10.1016/j.eswa.2024.126295>
- [32] Ganai, U.J., Ratne, A., Bhushan, B., Venkatesh, K.S.: Early detection of autism spectrum disorder: gait deviations and machine learning. *Scientific Reports* **15** (2025) <https://doi.org/10.1038/s41598-025-85348-w>
- [33] Silva, S.H., Felippin, M.R., Oliveira Medeiros, L., Hedin-Pereira, C., Nogueira-Campos, A.A.: A scoping review of the motor impairments in autism spectrum disorder. *Neuroscience & Biobehavioral Reviews* **169**, 106002 (2025) <https://doi.org/10.1016/J.NEUBIOREV.2025.106002>
- [34] Zakaria, N.K.: Experimental approach in gait analysis and classification methods for autism spectrum disorder: A review. *International Journal of Advanced Trends in Computer Science and Engineering* **9**, 3995–4005 (2020) <https://doi.org/10.30534/IJATCSE/2020/224932020>
- [35] Zanfır, M., Leordeanu, M., Sminchisescu, C.: The moving pose: An efficient 3d kinematics descriptor for low-latency action recognition and detection. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2752–2759. Institute of Electrical and Electronics Engineers Inc., Sydney, NSW, Australia (2013). <https://doi.org/10.1109/ICCV.2013.342>
- [36] Sedmidubský, J., Elias, P., Zezula, P.: Effective and efficient similarity searching in motion capture data. *Multimedia Tools and Applications* **77**, 12073–12094 (2017) <https://doi.org/10.1007/s11042-017-4859-7>
- [37] Ahad, M.A.R., Ahmed, M., Antar, A.D., Makihara, Y., Yagi, Y.: Action recognition using kinematics posture feature on 3d skeleton joint locations. *Pattern*

- [38] Montavon, G., Binder, A., Lapuschkin, S., Samek, W., Müller, K.-R.: In: Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.-R. (eds.) *Layer-Wise Relevance Propagation: An Overview*, pp. 193–209. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-28954-6_10
- [39] Trevarthen, C., Delafield-Butt, J.T.: Autism as a developmental disorder in intentional movement and affective engagement. *Frontiers in Integrative Neuroscience* **7**, 1–31 (2013) <https://doi.org/10.3389/FNINT.2013.00049/FULL>
- [40] Torres, E.B., Travers, B.G., Delafield-Butt, J.T., Srinivasan, A.: Editorial: Autism: the movement (sensing) perspective a decade later. *Frontiers Media SA* (2025). <https://doi.org/10.3389/fnint.2025.1634265>
- [41] Pradhan, A., Chester, V., Padhiar, K.: Classification of autism and control gait in children using multisegment foot kinematic features. *Bioengineering* (Basel, Switzerland) **9** (2022) <https://doi.org/10.3390/BIOENGINEERING9100552>
- [42] Gong, L., Liu, Y., Yi, L., Fang, J., Yang, Y., Wei, K.: Abnormal gait patterns in autism spectrum disorder and their correlations with social impairments. *Autism research : official journal of the International Society for Autism Research* **13**, 1215–1226 (2020) <https://doi.org/10.1002/AUR.2302>
- [43] Stoit, A.M.B., Schie, H.T.V., Slaats-Willemse, D.I.E., Buitelaar, J.K.: Grasping motor impairments in autism: not action planning but movement execution is deficient. *Journal of autism and developmental disorders* **43**, 2793–2806 (2013) <https://doi.org/10.1007/S10803-013-1825-8>
- [44] Bäckström, A., Johansson, A.M., Rudolfsson, T., Rönqvist, L., Hofsten, C., Rosander, K., Domellöf, E.: Atypical development of sequential manual motor planning and visuomotor integration in children with autism at early school-age: A longitudinal kinematic study. *Autism : the international journal of research and practice* **29**, 1510–1523 (2025) <https://doi.org/10.1177/13623613241311333>