Tikrit Journal of Engineering Sciences (2025); 32(Sp1): 2831
DOI: http://doi.org/10.25130/tjes.sp1.2025.31

IRAQI Academic Scientific Journals

ISSN: 1813-162X (Print); 2312-7589 (Online)
Tikrit Journal of Engineering Sciences
available online at: http://www.tj-es.com
TJES Tikrit Journal of Engineering Sciences

A Systematic Survey of Vision Transformer-Based Autism Detection Using Full-Body Movement Data from Augmented Reality Gameplay

Iyas Qaddara *[a], Ahmad Sharieh [a], Huda Karajeh [b]
a Department of Computer Science, University of Jordan, Amman, Jordan.
b Department of Computer Information Systems, University of Jordan, Amman, Jordan.

Keywords:
Autism Spectrum Disorder; Augmented Reality; Full-Body Movement Analysis; Skeleton-Based Modelling; Spatiotemporal Transformers; Vision Transformers.

Highlights:

- AR-based standardized movement capture for ASD.
- Unified transformer taxonomy.
- End-to-end analysis pipeline.
- Addresses data scarcity in ASD research.

ARTICLE INFO
Article history:

Received	23 Nov.	2025
Received in revised form	02 Dec.	2025
Accepted	07 Dec.	2025
Final Proofreading	25 Dec.	2025
Available online	26 Dec.	2025

© THIS IS AN OPEN ACCESS ARTICLE UNDER THE CC BY LICENSE. http://creativecommons.org/licenses/by/4.0/

Citation: Qaddara I, Sharieh A, Karajeh H. **A Systematic Survey of Vision Transformer-Based Autism Detection Using Full-Body Movement Data from Augmented Reality Gameplay**. *Tikrit Journal of Engineering Sciences* 2025; **32**(Sp1): 2831.
http://doi.org/10.25130/tjes.sp1.2025.31

*Corresponding author:
Iyas Qaddara
Department of Computer Science, University of Jordan, Amman, Jordan.

Abstract: Early detection of autism spectrum disorder (ASD) increasingly relies on objective movement-based biomarkers. This Systematic survey explores the confluence of three promising areas: Augmented Reality (AR) gameplay for eliciting naturalistic full-body movement, skeleton tracking for data extraction, and Vision Transformer (ViT) models for analysing spatiotemporal movement patterns to detect ASD. This survey aims to summarize the current research landscape by considering available full-body skeleton datasets and transformer-based approaches, such as pure ViTs, spatiotemporal transformers, graph transformer variants, hybrid CNN-ViT models, and physics-informed approaches for AR systems that elicit ASD-relevant motor and social behaviors, and to organize them into a taxonomy. Our results reveal that models from the transformer learning family are better than traditional deep learning methods at all aspects of modelling coordination, time irregularity, and joint-dependency patterns related to ASD; however, this is limited due to a lack of AR gameplay data sets, pose-estimation accuracy, heterogeneity in task design, and a lack of clinically validated benchmarks. Building upon these results, we present a reference end-to-end pipeline for AR-collected movement analysis taking into account standardized task design, robust skeleton preprocessing, motion tokenization, and classifier through transformers and targeted suggestions for future research such as the development of shared AR-based datasets, multimodal fusion approaches (e.g., gaze + skeleton), self-supervised vital transformer pretraining with large motion datasets, physics-informed modeling, and clinically aligned evaluation protocols to ensure the development of scalable, robust and clinically meaningful vit-based autism detection systems.

1.INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neuro-developmental disorder in which there are chronic differences in social communication, sensory integration, and motor control. Over the past 20 years, clinical researchers have increasingly recognised that motor behavior, particularly full-body movement patterns, is a valuable indicator of the underlying mechanisms of ASD. It has been consistently shown that autistic individuals exhibit atypicalities in balance, gesture production, timing, coordination, postural stability, and whole-body motor planning. These differences are evident early in childhood and are highly informative for both screening and assessment, making motor behaviour one of the most promising objective indicators in the early identification of ASD [1]. Traditional methods for measuring motor behaviour in clinical settings rely on human observation, structured tasks, or standardised rating scales. While these approaches are valuable, they are subject to inter-rater variability. At the same time, children, especially those with ASD, may respond differently in clinical settings due to anxiety, unfamiliarity, or reduced engagement. These limitations have motivated a shift towards more naturalistic, technology-assisted assessment environments in which behaviour can be elicited more spontaneously, recorded more consistently, and analysed more objectively [2]. Augmented Reality (AR) gameplay has become a highly effective medium for this purpose. AR Tasks can be designed to elicit ASD-relevant behaviours, such as imitation, joint attention, gesture following, social orienting, and whole-body motor coordination, in an engaging, gamified, and ecologically valid environment [3]. Despite the potential of AR gaming, skeleton tracking, and transformer models, research at the intersection of these three remains in its infancy and is dispersed. Existing AR systems vary significantly in task design, sensing modalities, gameplay duration, and behavioural targets. Similarly, available datasets are limited in quantity, size, diversity, and clinical labelling [4]. Given this lack of fragmentation, there is a clear need for a structured, systematic synthesis that unifies the literature on AR-based movement elicitation, skeleton data representation, and transformer-based modeling approaches for ASD detection. A good mapping of this new domain, however, can help reveal methodological weaknesses, identify gaps in existing datasets, highlight interesting transformer architectures, and, indeed, unravel how AR gameplay can be standardised to enable more consistent behavioural evaluation. Such synthesis is of special significance as the field moves towards clinically meaningful, scalable, and privacy-preserving digital assessment tools [4]. To help fill this gap, this systematic survey provides an exhaustive review of AR systems deployed for the evaluation of ASD, complete-body skeleton datasets, movement representations, preprocessing approaches, and models in the transformer family, with a focus on motion analysis. It establishes a coherent taxonomy of transformer-based architectures, synthesises their advantages and limitations, and presents a consolidated end-to-end pipeline, with collected movement data in mind, developed by AR. The survey also highlights several challenges, including a dearth of datasets, the diversity of pose-tracking methods, the lack of clinical validation, and privacy concerns, and concludes with specific recommendations for future research directions.

2.BACKGROUND

Understanding the technological and behavioural basis of autism detection based on full-body movement requires understanding the elements underlying this field of study. This section contains the basic background necessary to understand the context of the rest of the survey, including evaluation of three basic dimensions: (i) the nature of behavioral and motor characteristics typically associated with ASD, (ii) augmented reality (AR) environments as naturalistic platforms for evoking some measurable whole-body actions, and (iii) the computational models and pose-estimation techniques used to translate raw movement to analyzable skeletal representations. Together, these components have provided the conceptual basis for the modern transformer-based ASD detection systems.

2.1.Autism and Motor Behaviour Characteristics

Autism Spectrum Disorder (ASD) is gaining widespread attention as a condition that includes not only a difference in social communication, but characteristic motor and coordination patterns that occur early in development. A growing body of literature indicates that autistic individuals exhibit measurable differences in full-body movement, including postural stability, gesture execution, bilateral coordination, gait regularity, and the timing and smoothness of limb trajectories. These movement atypicalities can be observed in both structured tasks and spontaneous play and often precede symptoms in verbal and social behaviours, thus motor behaviour appears as a good channel for screening at an early stage [5]. Full-body skeleton tracking enables these. Computational behavioral patterns are to be characterised as joint trajectories (2D or 3D) over time, and computational models can be used to analyse joint velocities, accelerations, relative joint

angles, symmetry differences, and temporal coordination. These representation formats preserve essential information about kinematics relevant to the distinction between ASD-Typical and Neurotypical movement dynamics. Since the motor patterns are less susceptible to cultural, linguistic, and environmental variability, they provide the foundation for a stable and objective basis for two-way automated detection of ASD, mainly when elicited naturally, through interactive and engaging AR experiences [6].

2.2.Augmented Reality for ASD Assessment

Augmented Reality (AR) has become an increasingly exciting platform for behavioral assessment when investigating autism due to its capacity to combine elements of the digital world with aspects of the real environment, while still preserving the naturalistic interaction. Unlike traditional clinical settings, with their tasks that can seem rigid, artificial, and anxiety-inducing, AR is an engaging, play-based medium that encourages spontaneous, full-body movement. This makes it particularly useful for eliciting behaviors relevant to ASD, including imitation, gesture following, social orienting, joint attention, response inhibition, and whole-body coordination [7]. AR gameplay has actions. These tasks can be standardized for timing, difficulty, and structure and are reproducible, but have ecological validity. Importantly, because AR-based environments allow for free movement on the part of the children, barriers to compliance are mitigated, and opportunities for attaining movement patterns translatable to in-the-world behavior are increased [8]. Overall, AR offers a robust, child-friendly, and high ecological validity platform to capture complete body movement data that will be required to foster the automated detection of ASD, which is an essential bridging platform between naturalistic behavior.

2.3.Skeleton Tracking and Pose Estimation

Skeleton tracking and pose estimation are the backbone of any movement-based system for ASD detection. These techniques range from converting the raw visual input, usually RGB, depth, or RGB-D stream, into structured representations at the joint level and therefore enabling the analysis of full body motion accurately and consistently. At the heart of each of these processes lies the extraction of a set of anatomical keypoints (e.g., head, shoulders, elbows, wrists, hips, knees, ankles) and their 2D or 3D trajectories over time. As discussed, these trajectories preserve identifiable facial details, making the skeleton-based representations significantly more private than the raw video recording [9]. Modern pose estimation frameworks such as OpenPose, MediaPipe, VNect, and the Kinect SDK employ deep neural networks that can detect joint positions despite occlusions, varying lighting conditions, and fast movement. Depth sensors (e.g., Kinect Azure, Intel RealSense) provide geometric information that improves the accuracy and temporal stability of 3D joints. These technologies enable the recording of complex movement patterns important for ASD, such as irregularities in rhythm, asymmetries, hesitations, repetitive movements, and deviations in velocity or joint coordination [10].

2.4.Vision Transformers and Spatiotemporal Modeling

Vision Transformers (ViTs) have become one of the most important architectural innovations in the fields of computer vision and sequence modeling, and possess powerful capabilities for analysing full-body movement data. Unlike convolutional neural networks (CNNs), which model dependencies among meaningful elements in space using fixed receptive fields, transformers rely on mechanisms of global self-attention between joints and across the entire temporal sequence. This is a property that makes them particularly suitable for the detection of ASD, in which there may be slight abnormalities in timing, coordination, or cross-limb interactions with diagnostic value [11]. More advanced madiapproaches include Graph Transformers, which incorporate skeletal connectivity as a constraint in the attention mechanism; Hybrid CNN-ViT, which uses a CNN to extract low-level motion cues and then a transformer to refine the results; and Physics-informed Transformers, which incorporate biomechanical constraints to improve interpretability and robustness. Additionally, the self-supervised version of ViTs on large-scale motion datasets has demonstrated excellent generalisation ability—important given the limited availability of labelled data for ASD [12]. Vision Transformers (ViTs) have several benefits over a more traditional deep-learning architecture when modeling skeleton-based ASD motor patterns. CNN-based approaches are good at learning local spatial structures of joints. Still, they cannot learn long-range interactions between distal body parts, which are critical for comprehending full-body coordination. RNNs and LSTM variations can capture the dynamics of time series. Still, they are more susceptible to timing anomalies and differences in movement velocity, as well as to the presence of noise and missing frames, which are typical of the motor behavior of autistic children. Conversely, since ViTs have a self-attention mechanism, the model can simultaneously analyze all joints and time steps and extract global spatiotemporal correlations, which are not limited by locality or the order of the sequence. This renders ViTs especially appropriate to ASD movement analysis, in

which abnormal patterns tend to encompass distributed postural asymmetries, retarded or irregular time relations among limbs, and long-range joint interactions, which grow over the temporal span of a movement sequence.

2.5.Research Questions (RQ)

To steer this systematic survey and organize the analysis in the categories of AR systems, skeleton-based representations, and transformer-family models, the following research questions were formulated:

RQ1. AR Systems: How are augmented reality (AR) environments being utilized to induce clinically relevant individuals with ASD to exhibit full-body movement behavior, and what tasks, sensors, and design paradigms predominate current investigation approaches using AR-based assessments?

RQ2. Skelton-Based Movement Analysis: What are the types of full-body skeleton datasets, movement tasks, and preprocessing strategies for collecting ASD relevant motor patterns, and what are the limitations of the current collected data sources?

RQ3. Transformer Models: How have Vision Transformers, spatiotemporal transformers, graph transformer architectures, and CNN-ViT hybrid modeling approaches been employed to tackle challenges in object recognition of full-body movement data in the scope of ASD detection, with the modeling strategy that shows the most promising success?

RQ4. Gaps and Future Needs: What are the major challenges, methodological in nature, and open research opportunities at the intersection of AR gameplay, modeling based on skeletons, and ASD detection via transformers?

3.SURVEY METHODOLOGY

Conducting a systematic survey across the domains of AR-based behavioral elicitation, skeleton-tracking technologies, and transform-based ASD detection requires a structured, transparent methodology. This chapter describes the process of identifying, selecting, and analyzing the relevant literature, including the databases searched, the search strategies used, the inclusion and exclusion criteria, and the multi-stage screening procedure. The forging of a rigorous methodological base in this section ensures that the synthesis presented in the following chapters is comprehensive, reproducible, and consistent with the formulation of research questions earlier in this study.

3.1.Databases and Sources

To ensure coverage of as much literature as possible, several academic databases were used, including IEEE Xplore, ACM Digital Library, PubMed, Scopus, ScienceDirect, and arXiv. Searches focused on peer-reviewed journals, conferences, and preprints that have content related to AR systems, full-body movement analysis, skeleton tracking, and transformer-based modeling used for the detection of ASD. Backwards and forward reference tracing was also conducted to identify studies that could not be obtained through the direct search process. By using a variety of databases, both the technical and clinical domains are covered. IEEE and ACM for engineering, robotics, AR, and computer vision studies, while PubMed and ScienceDirect have medically oriented ASD research. Scopus offers cross-disciplinary indexing with an emphasis on identifying the studies that use AR-based behavioral tasks combined with computational modeling. The inclusion of arXiv is critical given how fast-paced transformer architectures are changing, with many seemingly published first as preprints. This use of many sources addresses biases and helps identify emerging applications for AR-skeleton-transformer intersections.

3.2.Search Strategy

A defined keyword approach was applied. Searches for combinations of terms related to ASD, augmented reality, movement analysis, skeleton tracking, and transformer architectures. Examples of representative search phrases were:

Autism" OR "ASD" AND "augmented reality". Autism motor behavior" AND "skeleton tracking". Vision Transformer" OR "spatiotemporal transformer" AND "movement analysis". AR gameplay" AND "pose estimation".

Search strings were progressively refined to maximise recall while excluding irrelevant literature.

Keyword combinations were modified repeatedly by reference to initial search results. Adding terms such as "pose estimation", "skeleton data", "Kinect", and "joint trajectories" increased the retrieval of movement analysis papers relevant to ASD to a large extent. Likewise, terms like "transformer encoder, "attention mechanism, and "graph transformer" appeared in recent studies on specific models. Boolean operators and symbols for wild cards were used to represent different terminology embraced by different communities (e.g., "motion capture", "body tracking", "movement disorder", "ViT-based. This became a refinement that ensured maximum precision without sacrificing recall.

3.3.Inclusion and Exclusion Criteria

Defining specific inclusion and exclusion criteria is key to ensuring that the selected studies are meaningful to the goals of the survey and the research questions that have been developed earlier. Because the intersection of AR gaming, skeleton-based analysis of the movement, and transformer-based detection of ASD lies across multiple disciplines, it is

important to filter the literature in a systematic way to retain only such studies that offer empirical, methodological, or technical relevance to this field. The criteria below set a high bar, distinguishing between ASA research based on movement and unrelated to AR, clinical experiments, or machine learning research, and therefore, such a synthesis is focused, coherent, and scientifically valid. Inclusion criteria: Studies either focused on ASD assessment, behavioural modelling, or the movement-based detection, works that involve AR environments, motion-eliciting tasks, or interactive gameplay, Studies on skeleton-based representations (skeleton-based or pose-estimation-based pipelines), Machine learning or deep learning models such as ViTs or the like, and Papers that provide some empirical data, evaluation metrics, or some methodological contributions. Exclusion criteria: Studies not related to ASD or movement behavior, AR research without complete body analysis of motion, on-ASD relevant motion-analysis papers, and Non-technical work, reviews without data or opinion work. Together, these inclusion and exclusion criteria provide a focused, methodologically sound basis for the survey. By limiting the research sample to the space between AR-based behavior elicitation, skeleton-driven movement analysis, and transformer family modelling, the selected literature aligns directly with the survey's research questions. This helps to make sure that the resulting synthesis is both coherent and relevant to helping to ensure the coherence of the next stage - the structured screening and selection of eligible studies - to proceed with clarity and coherence.

3.4.Study Selection and Screening Process

A structured process with multiple screening stages was used to ensure that studies directly relevant to AR-based behavioural elicitation, skeleton-tracking methodologies, and transformer-driven ASD detection were included. Given the heterogeneity of research on the topics of interest, spanning clinical, computational, and interaction design disciplines, a stepwise selection workflow was needed to screen out studies that were unrelated or methodologically weak. This process helps make the survey as scientific as possible and ensures that the final collection of papers is informative and meaningful in relation to the research questions.

The three successive phases of the selection pipeline included:

a) **Title and Abstract Screening**: Initial filtering to identify studies mentioning ASD, AR interaction, movement analysis, pose estimation, or transformer-based modeling. Papers that were obviously irrelevant were discarded at this stage.
b) **Full-Text Evaluation**: Detailed study of methodology, such as movement elicitation paradigm, data collection setup, skeleton extraction method, and machine learning methods. Studies with inadequate methodological information or lacking information on movement were excluded.
c) **Final Eligibility of Final Filtering:** Papers that met all the inclusion criteria were kept for review. Ambiguous cases (e.g., partial AR tasks, limited joint-tracking detail, hybrid behavioral setups) were double-checked through secondary review to ensure consistency and minimize subjectivity in the measures.

The multi-stage screening procedure ensures that the final body of literature, which is part of this survey, is methodologically reliable and strictly relevant to the intersection of AR-based behavioral elicitation, skeleton-driven motion analysis, and transformer-family modeling. By narrowing selection increasingly from the keyword filtering to the full-text evaluation and eligibility checking, the selection process minimizes the bias and prevents inclusion of studies that don't have the empirical rigour or that don't align with the research questions. This approach to structuring a set of studies provides a clean, validated set on which the synthesis in the next section can be constructed.

3.5.Synthesis Approach

A well-organized synthesis approach was required to combine the wide variety of research in the fields of AR-based behavior tasks, skeleton-based movement representations, and transformer-based modeling approaches. Because the included studies span disciplines such as clinical science, computer vision, AR interaction design, and deep learning research, a common analytical framework is needed to derive meaningful patterns and answer the survey's research questions. This section describes the methodology employed to categorise, compare, and interpret the chosen literature, to allow a coherent understanding of the intersections between these domains that support the development of a coherent understanding of the detection of ASD. To achieve this, each of the studies was mapped onto one of three analytical dimensions relating to the survey RQ's (Research Questions) 1-4. First, AR-based ASD systems were analyzed in reference to their task structures (e.g., gesture imitation, object tracking, spatial navigation), hardware platforms (e.g., Kinect, ARCore/ARKit, depth cameras), and behavioral goals. Second, data characteristics, joint representations, preprocessing strategies, and types of motor features extracted were reviewed in the skeleton-based ASD studies. Third, models of the transformer family were synthesized taking into consideration the architecture type (Vision

Transformers, spatiotemporal transformers, graph transformers, or hybrid CNN-ViT structures), tokenization formats, attention mechanisms, and evaluation methodology. To specify the synthesis process, the studies included were then classified into three fields of analysis with respect to their primary contribution to methodology: (i) AR-based behavioral elicitation systems, (ii) skeleton-based movement data and pose-extraction pipelines, and (iii) transformer-based modeling methods. All studies were grouped into one or more categories based on the introduction of a new AR task paradigm, the presentation of structured skeletal motion data, or the application of a modeling method to the analysis of movement in ASD. In this categorization, a structured comparison was possible according to the research questions of the survey. In this way, it was able to support the necessity to guarantee the integration of insights from heterogeneous fields in a coherent manner. An additional flow diagram is presented to depict this classification process and visualize the idea of how the studies moved in the direction of first identification and further division into domain-specific grouping into AR systems, skeleton data, and transformer models. Through this framework of a structured synthesis, the survey brings together the learnings of different fields of research into a single perspective, paving the way for a better understanding of the concerted effort coming from limb skeletal representations, transformer-driven models, and gameplay with AR games to tackle the problem of ASD detection. This integrative approach not only organizes the current findings but also lays the groundwork for the in-depth analysis presented in the following chapters.

4.AR SYSTEMS FOR AUTISM ASSESSMENT

Augmented Reality (AR) has very quickly become one of the most promising platforms for providing naturalistic, measurable full-body movement behaviors to people with autism spectrum disorder (ASD). Unlike the conventional assessment environment, which is based on structured clinical tasks or lab-style environments, AR allows immersive, game-like interactions that encourage children to move around freely while keeping a task structure consistent. This puts AR in a unique position to facilitate behavioral assessment and motor analysis, as well as providing the early screening - especially when paired with modern skeleton tracking frameworks and learning-based models [13]. This chapter provides a review of the existing AR systems that have been developed for ASD research, with special emphasis on the type of interactions each can incorporate, the types of sensors employed to capture movement in the world, the behavioral objectives each system aims to measure, and the methodological gaps that have motivated the development of AR-based ViT-driven assessment pipelines.

4.1.AR Interaction Paradigms Used in ASD Research

AR systems dedicated to autism assessment usually use interaction paradigms that encourage children to complete actions involving whole-body responses to virtual cues, objects, or characters [14,15]. These paradigms fall into several recurring categories:

- **Gesture-Imitation Tasks**: Where the child is asked to mimic poses, arm movements, or whole-body gestures that an avatar performs.
- **Tasks of Object-Based Interaction**: AR elements (balls, shapes, targets) pop up around the child.
- **Navigational/spatial-complexity Exploration Tasks**: Children do physical movements in a specific space in front of AR overlays.
- **Social-Response AR Tasks**: Some systems use AR avatars that call the child's name, point to objects, and give instructions.
- **Reward-Based AR Gameplay:** Gamified AR tasks that play animations if the child makes the correct movement.

These paradigms show, in total, how AR can efficiently induce reliable motor patterns required for automated detection of ASD [16]. In Figure 1, the five primary AR interaction categories above, identified for use in ASD research, are summarized.

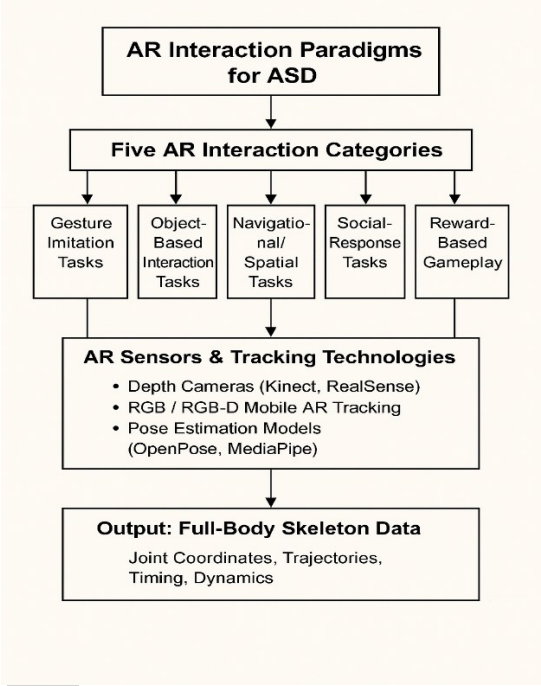

Fig. 1 Overview of AR interaction Paradigms for ASD.

4.2.Sensors Used in AR-Based ASD Systems

AR systems for autism assessment make good use of sensing technologies to capture full-body movement with sufficient accuracy, temporal resolution, and robustness for clinical or computational analysis. The selection of the sensor directly influences the quality of the extracted skeleton data, the description of details in joint trajectories, and the generic suitability of the system overall for ASD-related tasks, not least because autistic children may be in unpredictable motion, speed, and even irregular motion profiles. This section provides an overview of the most common sensing platforms for AR-based ASD-related research and notes the strengths and limitations of each [17].

Depth Cameras (as mentioned above, e.g., Microsoft Kinect, Intel RealSense) : Depth cameras are the most popular type of sensor for AR-for-ASD systems because they provide high-quality 3D joint information, robust skeletal tracking, and good resistance to lighting differences. Specific examples of the applications of Kinect-based systems include: gesture imitation, movement copying, and balance tasks[18,19].

RGB -D Sensor Incorporated as AR Glasses / Mobile AR Platform: Modern AR systems (ARKit, ARCore) use body trackers, which are based on camera data, to estimate the 2D or pseudo-3D skeletons in real-time.

RGB Cameras and Models for Pose Estimation using OpenPose and MediaPipe: Some AR systems combine simple RGB cameras with pose estimation software to extract joint trajectories.

Multi-Sensor Fusion: Few advanced systems use a combination of sensors (RGB + depth) to enhance the accuracy of the joints further.

Wearable Sensors (less ASD-focused AR): Some research uses IMUs or other wearable motion sensors, but they aren't as suitable for children with ASD because autistic children often don't like to wear devices due to tactile discomfort and sensory sensitivities.

4.3.ASD-Related AR Tasks

AR systems intended for autism evaluation are usually embedded systems that incorporate structured activities within interactive, game-like environments that motivate children to make specific full-body movements in response to a virtual cue. These tasks are designed to intentionally elicit the motor patterns known to be different in individuals with autism than in those who are neurotypical to measure behavior in a naturalistic, but standardized, way[20]. Below are the main types of tasks used in the AR field of ASD research aimed at different types of motor or cognitive behavior [21]:

Gesture-Imitation Tasks: In these tasks, a virtual avatar or AR character performs a movement—such as raising an arm, stepping to the side, or assuming a full-body pose—and the child must imitate it. Measures: Motor imitation ability, Bilateral coordination, Upper/lower limb timing.

Pointing, Reaching, and Locomotor Interaction: Here, AR objects appear at different spatial locations (front/back, left/right, high/low), requiring the child to reach, tap, move toward, and swipe virtual objects. Measures: Reaction time, spatial orientation, and range of motion.

Target-Following and Path-Tracking Tasks: AR cues (arrows, footsteps, glowing markers) appear in the physical environment, instructing the child to walk, step, or follow a path. Measures: Gait stability, Dynamic balance, and step timing.

Multi-Step Action Tasks: Some AR setups require the child to combine multiple actions—for example, raise a hand, touch a virtual object, and step back. Measures motor planning, sequencing skills, and transition between sub-actions.

Socially AR Tasks: AR characters may: point to cues, ask the child to perform actions, and display emotional expressions. Measures: joint attention and social response time.

Reward-Based AR Gameplay: Many systems embed simple rewards, such as animations, sound, stars/tokens, and character reactions.

4.4.Summary of AR-Based ASD Studies

To provide a structured understanding of how augmented reality has been implemented in autism research, this section will provide an overview of the key studies that implemented AR environments to elicit full-body movement behaviors in individuals with ASD. Unlike the conceptual overview in the previous section, this section is concerned with concrete research evidence - namely, the details of the various studies about their specific objectives, characteristics of participating individuals, registration style of AR interaction, sensing, movement tasks, and main findings. By placing these studies into the types of tasks and sensors discussed in Sections 4.1-4.3, this summary provides a solid ground where one can gather some immutable trends of the methods under consideration, examine those systems' effectiveness, and recognize gaps that help to spur the development of far more advanced AR-based assessment pipelines.

Gesture-Imitation Tasks: Gesture-imitation studies consistently demonstrate that AR environments enhance children's ability to reproduce whole-body actions. For example, Pérez-Fuster [26] showed improved imitation accuracy and joint attention in full-body pictogram-guided actions. At the same time, Amara [27] reported increased engagement and

better hand–eye coordination during AR-based hand-gesture tasks. Similarly, Lee [28] observed notable gains in role-play imitation using a Kinect-based AR avatar. These findings collectively indicate that AR-guided imitation tasks can reliably elicit repeatable motor patterns and expose core ASD-related difficulties such as delayed motor copying and reduced body-schema awareness—patterns clearly reflected across the Gesture-Imitation studies summarized in Table 1.

Object-Based Interaction Tasks: Object-based AR tasks converge on a common outcome: improved attentional engagement and interaction with physical–virtual objects. Tang [31,32] demonstrated that AR-supported object recognition improves vocabulary acquisition and object learning, while Wedyan [33] highlighted the need for standardization in multi-sensor AR object systems. Koumpouros [34] further showed cognitive accessibility gains when autistic learners interacted with AR-enhanced objects. Together, these studies (listed in Table 1) underscore that object-based AR tasks are effective at drawing and sustaining attention while eliciting fine-motor planning sequences, making them particularly useful for structured ASD motor assessments.

Social-Response AR Tasks: Social-response AR systems integrate motor actions with social cues, and the studies listed in Table 1 consistently show improvements in joint attention and social-motor alignment. Cheng & Bololia [38] synthesized multiple AR social-interaction tasks and found consistent improvements in following social cues, while Liu [39] demonstrated enhanced social communication and gaze stability using AR smart glasses. These findings align with the tabled results, highlighting that embedding social agents or avatars within AR environments generates more ecologically valid motor responses—responses that better reflect real-world social–motor challenges experienced by autistic children.

Reward-Based AR Gameplay: visual and auditory rewards increase participation, consistency, and the quality of motor data collected. The study by Pérez-Fuster (listed under multiple categories due to cross-task design) demonstrates that immediate reinforcement promotes more stable whole-body imitation. In contrast, others indicate that reward-triggered movement repetition improves data density and reduces behavioral variability. As reflected in Table 1, reward-driven systems achieve higher engagement levels and produce cleaner motor sequences, making them ideal for collecting standardized full-body data for ASD motion analysis.

Table 1 summarises the studies reviewed in the context of the five AR interaction paradigms. Each entry focuses on the essential elements within these areas, required for systematic comparison of outcomes, such as the type of task, the type of sensing modality, target skills, and key outcomes. This table is the basis of the answer to RQ1, and the lack of information in the AR-based assessment of ASD.

Table 1 Summarize Studies in AR Interaction Paradigms.

Category	Study / Citation	AR Task Type	Sensors Used	Targeted Skills / Behaviors	Key Findings
1. Gesture-Imitation Tasks	Pérez-Fuster et al. (2025) [26]	Whole-body imitation, pictogram-guided actions	Depth camera / AR room	Imitation, joint attention, body knowledge	AR improved imitation accuracy and social attention
	Amara et al. (2023) [27]	Hand-gesture + voice AR tasks	RGB camera + gesture tracking	Gesture recognition, hand–eye coordination	AR increased engagement + gesture learning
	Lee (2021) [28]	Kinect-based role-play with AR avatar	Kinect depth sensor	Imitation, role-play skills	Notable improvement in motor copying
	Alcaniz Raya et al. (2020) [29]	VR/AR body-movement classification	Kinect	Full-body movement patterns	ML classified ASD vs NT with good accuracy
	Hu et al. (2025) [30]	AR emotion-driven imitation	Mobile AR	Emotional imitation, expression	Better recognition of emotional cues
2. Object-Based Interaction Tasks	Tang et al. (2019a) [31]	AR vocabulary with object recognition	Mobile RGB	Object learning, attention	AR improved engagement with objects
	Tang et al. (2019b) [32]	Lightweight AR object detection	Mobile camera	Object recognition	Faster learning of new vocabulary
	Wedyan et al. (2020) [33]	AR system review + prototype	Mixed sensors	Object-interaction, recognition	AR is promising, but needs standardization
	Koumpouros (2025) [34]	AR tools for autism	Mobile devices	Object interaction, basic motor coordination	AR enhanced cognitive accessibility
3. Navigational / Spatial AR Tasks	Lee & Huang (2025) [35]	AR + MR map-based navigation	Kinect / MR	Social cues, spatial sequencing	Improved sequencing and cues following

	McMahon et al. (2015) [36]	AR navigation for employment	Mobile AR	Real-world navigation	Increased independence and planning
	Fridhi et al. (2020) [37]	Geospatial AR exploration	Mixed VR/AR	Spatial–motor integration	Better exploration and direction following
4. Social-Response AR Tasks	Cheng & Bololia (2024) [38]	Systematic review of AR social tasks	Various	Joint attention, social cues	AR improves social interaction consistency
	Liu et al. (2017) [39]	AR smartglasses system	AR glasses	Social communication, gaze	High feasibility + child compliance
	Syahputra et al. (2018) [40]	AR social stories	Mobile AR	Social rules, social understanding	Improved story comprehension
	Nekar et al. (2022) [41]	Multiplayer AR dual task	AR tablet + sensors	Social + cognitive motor	Significant social-cognitive gains
	Rega et al. (2018) [42]	AR motivation enhancer	Mobile AR	Motivation, engagement	AR increased willingness to participate
5. Reward-Based AR Gameplay	Williams & Chandramouli (2025) [43]	VR/AR reward-training	VR + sensors	Positive reinforcement, communication	Increased engagement + response stability
	Nekar et al. (2022) [44]	Cognitive-motor AR game	AR game	Cognitive control, RRBs	Reduction in repetitive behaviors
	Brandão et al. (2015) [45]	AR gamebook	Mobile AR	Motivation, reading/action coupling	Improved task initiation
	Bhatt et al. (2014) [46]	AR game therapy	Mixed sensors	Engagement, basic motion	AR games increased participation

As shown in Table 1, the five categories of AR interaction exhibit clear patterns in the approaches used across studies to elicit motor and social behavior for the assessment of ASD. Tasks of gesture imitation demonstrate a good consistency across the literature, with most studies focusing on coordination, timing, and joint synchronization, key signs of the ASD-related motor atypicalities. Object-based interaction and locomotor navigation tasks tend to utilise a wider range of whole-body motions and are suitable for analysing vasomotor integration and dynamic balance.

5.SKELETON-BASED ASD DETECTION

The skeleton-based method for ASD detection has become one of the most promising signals for the acquisition of objective and quantitative indicators of autistic motor behavior. Instead of relying on subjective observation or manual rating scales, skeleton analysis takes the full-body movement. It converts it into a structured and joint-level data - in other words, it tracks how each limb is moving, aligning, accelerating, and coordinating with time. Because autistic people tend to have measurable differences in timing, fluidity, balance, gait, postural control, and the synchronization of the joints, the trajectories of the skeleton are also a rich source of behavioral information that is both standardized and unobtrusive [47,48]. This chapter provides an overview of the current state of the art in skeleton-based ASD detection, covering available datasets, movement tasks for motor elicitation, feature-representation methods, preprocessing pipelines, and the limitations of these datasets. Together, these components make up the structural backbone of the transformer-based movement analysis investigated in the next chapter. Figure 2 shows the Skeleton-Based ASD Detection Pipeline.

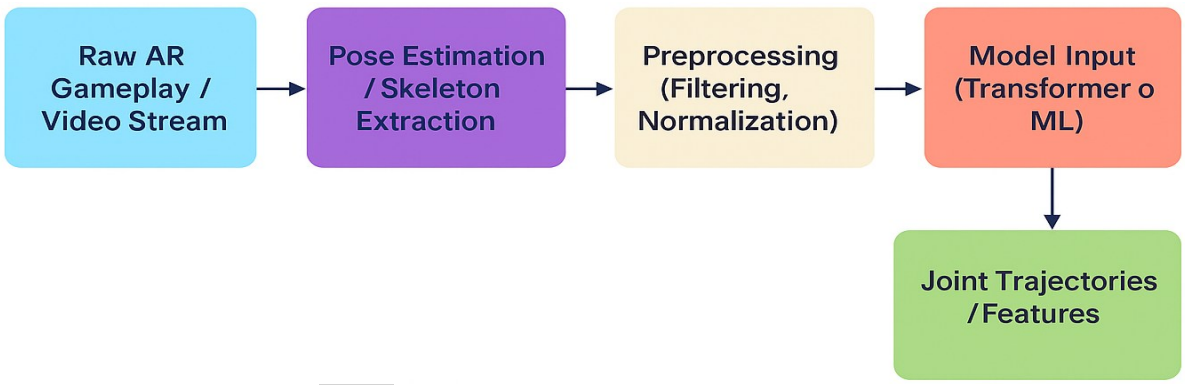

Fig. 2 Skeleton-Based ASD Detection Pipeline.

5.1.Datasets

Skeleton-based ASD detection requires full-body movement data to be available, of high quality, and of a specific structure. Due to the subtlety, heterogeneity, and context-dependence of autistic motor behavior, high-resolution joint trajectories across a range of tasks, participants, and movement conditions are important for data sets. Yet the field is plagued by having far too few publicly available datasets of autistic people, and most were originally designed for more general human-action recognition, which are not intended for clinical assessment. This section reviews the datasets that are currently used for the skeletons in a way related to ASD. It makes a distinction between (i) datasets that are converted to use specifically for ASD, (ii) general human movement data sets that are frequently adopted for model pretraining, and (iii) custom-built in-house data sets that are designed for small-scale clinical usage.

5.1.1.ASD-Specific Skeleton Datasets

ASD-specific skeleton datasets directly capture the full-body motion of autistic children during controlled and naturalistic tasks, yielding joint-level trajectories suitable for computational movement analysis. Table 2 presents the main datasets available in the literature for ASD research.

5.1.2.RRB-Based Skeleton Datasets

Repetitive and stereotypical behaviors (RRBs), e.g., hand flapping, rocking, and cyclic movements of a particular limb, are among the most typical motor patterns observed in ASD. Skeleton-based datasets targeted at RRBs yield useful information about the temporal regularity, amplitude, and spatial repetition of these behaviors such that machine learning models can recognize stimming patterns directly from the joint trajectories. The following studies are primary efforts made to capture and analyze RRB movements by using pose estimation and skeletal tracking.

5.1.3.Gait & Full-Body Movement ASD Datasets

Gait-based locomotor patterns for the whole body are another major source of motor signatures associated with ASD. Datasets in this category are concerned with walking cycles, balance control, stride variability, and global coordination; in many cases, obtained via depth cameras or a motion capture system. These datasets offer rich temporal-spatial joint information for the quantitative analysis of the postural stability and movement variability of autistic individuals.

5.1.4.Markerless Pose-Estimation Datasets (Open Pose/Validation)

Markerless pose estimation algorithms extract joint lesions in the form of skeletal trajectories directly from RGB no-deep-sense video. These datasets are important because they closely reflect the real-world conditions of AR-based assessment, where children move freely, and only a single camera may be available. The following studies are among the markerless datasets and validation efforts that are key to an ASD motor behavior analysis.

5.2.Movement Tasks

Skeleton-based ASD studies use a wide range of movement tasks to elicit a motor pattern that can be analyzed from joint-level movement trajectories. Each task category focuses on specific aspects of motor coordination, timing, posture, and repetitive behavior. Understanding these types of tasks is crucial before the features and preprocessing steps of ASD movement classification can be defined. Figure 3 shows the Movement Task Categories Used in Skeleton-Based studies of ASD.

Fig. 3 Movement Tasks.

5.3.Skeleton Features

Skeleton-based ASD studies use a variety of motion features derived from joint trajectories to describe motor coordination, temporal stability, and spatial consistency. These features provide quantitative representations of gait, gesture execution, upper-limb control, and repetitive behavior. The following categories summarize the most commonly used feature types in the literature on skeleton analysis in ASD, as shown in Figure 4.

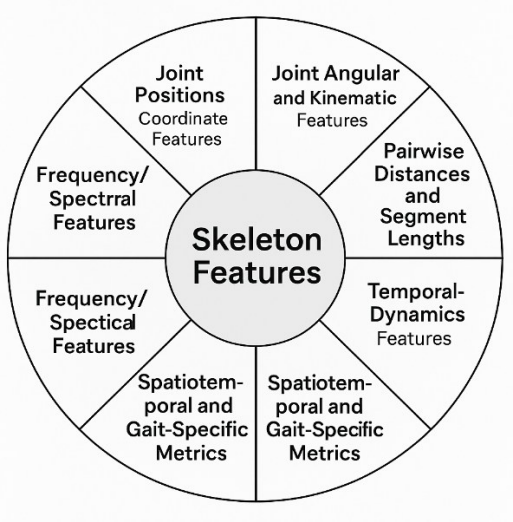

Fig. 4 Skeleton Features.

In summary, the range of features of skeleton use goes from the raw implementation of joint coordinates to the advanced implementation of relational or frequency-domain features. Each feature category involves social and unique aspects of how kids with ASD move, which form the basis for the preprocessing and feature engineering steps outlined in the next section.

5.4.Preprocessing

Raw skeleton data obtained from RGB cameras, depth sensors, or markerless pose estimation models often contain noise, missing joints, variable frame rates, and variable sequence lengths [49]. Preprocessing is thus admitted to be crucial for the purpose of transforming the heterogeneous motion recordings into clean, standardized input that is ready to be fed into the feature extraction and the transformer-based modeling. Figure 4 provides an overview of the (typical) preprocessing steps applied to skeleton-based ASD datasets, including missing-joint handling, smoothing and normalization, temporal alignment, sequence segmentation, and data augmentation. The following steps summarize the most common preprocessing strategies that are used across the skeleton data sets with a connection to the study of ASD:

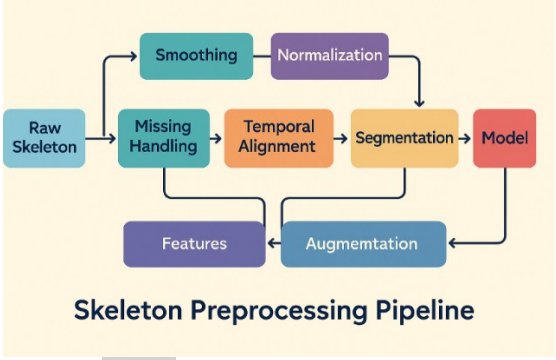

Fig. 5 Preprocessing Pipeline.

Before discussing dataset limitations, here is an integrated summary of all the datasets of skeletal-based ASDs included in this survey. Table 2 summarizes the comprehensive coverage of available datasets by bringing together ASD-specific datasets, datasets specifically about repetitive behaviours, datasets specifically about gait, and datasets on markerless pose estimation. This unified tabulation highlights the diversity of sensing modalities, participant experience, and motor tasks across the existing literature and serves as a useful reference point for understanding methodological differences before exploring their limitations.

Table 2 Unified Skeleton-Based ASD Dataset Summary.

Dataset Type	Study	Year	Sensor / Extraction	Participants	Task Type	Notes	Publicly Available
Type 1 — ASD-Specific	Li et al. [50]	2023	RGB + Depth + Skeleton	ASD children	Free-play, therapy	Multimodal dataset	YES
	Zhang et al. [51]	2021	Pose Estimation (2D)	ASD children	Gesture imitation	LSTM-based	YES
	Al-Jubouri et al. [52]	2020	Kinect v2 (3D)	ASD children	Gait	Structured dataset	YES
	Muty & Azizul [53]	2016	Pose Estimation	ASD children	Arm-flapping	Early RRB	YES
	Shin et al. [54]	2025	Pose Estimation	ASD children	Motor tasks	Dual-stream DL	YES
	Zahan et al. [55]	2023	Pose Estimation	ASD/Non-ASD	Gait + gesture	Large dataset	YES
	Yazdi et al. [56]	2024	Kinect	ASD children	Gait	Spatial temporal	NO
	Paulo et al. [57]	2025	RGB+Depth+Skeleton	ASD individuals	Motor tests	Clinical	NO
Type 2 — RRB	Lemler et al. [58]	2025	Post-hoc Skeleton	ASD children	Mannerisms	Multi-label	NO
	Muty & Azizul [53]	2016	Pose Estimation	ASD children	Arm-flapping	RRB	YES
	Lin et al. [59]	2025	Robot vision	ASD children	Body language	Robot-assisted	YES
Type 3 — Gait	Goldthorp et al. [60]	2025	Motion-capture	ASD/Non-ASD	Gait	Variability	NO
	Wu et al. [61]	2024	Markerless tracking	ASD children	Gait	Clinical	YES
Type 4 — Markerless	Kalam et al. [62]	2024	OpenPose	ASD/Non-ASD	General movement	Video-based	NO
	Anderson et al. [63]	2025	Markerless gait	Toddlers	Gait	Validation	NO
	Barahona[64]	2025	Open-source pose	Infants	Arm motion	Tracking	NO

5.5.Dataset Limitations

Although this new and greater access to skeleton-based ASD datasets provides a significant improvement, it is still limited in several ways, which have not allowed for generalisation of the models or diagnostic validity. To begin with, the available datasets tend to be very small, including fewer than 50 people with autism, leading the a loss of statistical power and the failure of deep learning models to trigger learning across the full range of motor variability. Second,

numerous datasets are biased in terms of demographics, particularly with respect to age, gender, and levels of ASD severity. As a result, biased representations of motion and, consequently, may not be generalizable in other subpopulations. The other great limit is that the composition of senses one to themselves is very different. Data obtained using Kinect, OpenPose, motion capture laboratories, and custom-engineered depth cameras differ enormously in the definition of joints, sampling rate, noise level, and coordinate system. These discrepancies make the comparison of various studies and the training of single models without a significant amount of preprocessing more difficult. A few are also task-specific (e.g., gait-only, arm-flapping-only), which constrains the range of motor behaviors that can be studied with them and the creation of holistic classifiers of ASD movements. Lastly, each dataset has a different level of annotation quality. Others are founded on the manual tagging of motion episodes or stimming incidences, whereas others are linked with feeble or crude tags when reducing the accuracy of supervised study. The combination of these constraints demonstrates the need not only to use standardized acquisition procedures, but also to use larger and more varied samples and rich, multi-task skeleton datasets in future studies of ASD movement.

6.TRANSFORMER MODEL TAXONOMY

Among the latest developments in transformer-based architectures has come a fundamental restructuring of the skeleton-based movement analysis field, providing the skeleton-based motor behavior researchers with potent ways to model whole-body motor behavior with autism spectrum disorder (ASD). A transformation of local spatial patterns or short-term time relations, as is the case with deep-learning models such as CNNs, RNNs, or graph convolutional networks, transformers use self-attention mechanisms that are capable of jointly modelling long-range dependencies and long-range time dynamics on a global scale. The latter is especially applicable in the case of ASD detection, where subtle anomalies in coordination, timing, balance, bilateral synchronization, and posture are observed in long sequences of movements. To offer the systematic overview of this new domain, this chapter offers a taxonomy of the transformer-based methods employed to analyze the ASD-related skeleton, grouping the literature in five broad categories: Vision Transformers, Spatiotemporal Transformers, Graph Transformers, Hybrid CNN-Transformer models, and Self-Supervised or Pertained Transformer models as shown in Figure 6. We summarize representative studies in each category in Table 3. Altogether, the literature review has shown that transformer-based models have great potential in ASD analysis, especially in cases when both temporal and structural data are considered. Table 4 Comparative Summary of AR, Skeleton, and Transformer Approaches in ASD Movement Assessment.

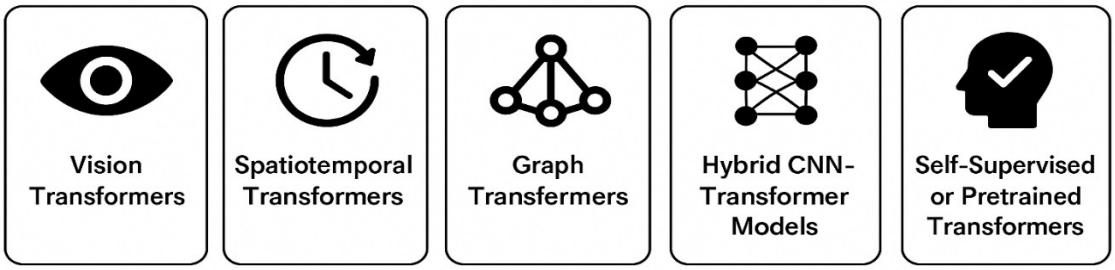

Fig. 6 Transformer Model Taxonomy.

Table 3 Summary of Transformer-Based Approaches for ASD Analysis.

Category	Study	Data Type	Task	Key Outcome
Vision Transformer	Shin,et al. (2025)[65]	Facial images	ASD classification	ViT outperformed CNNs in spatial feature extraction
Spatiotemporal Transformer	Gupta et al. (2025)[66]	Multi-modal behavioral	ASD severity assessment	Improved robustness through spatiotemporal attention
Graph Transformer	Zhang et al. (2023)[67]	Skeleton action data	Action recognition (relevant to ASD)	Graph-aware attention captured joint dependencies
Hybrid (CNN–Transformer)	Anand & Kini (2024)[68]	Clinical + Neuroimaging	ASD classification	The hybrid model outperformed the standalone CNN/Transformer
Multi-Task Transformer	Gao et al. (2024)[69]	Behavioral/clinical	ASD detection across multiple tasks	Multi-task learning improved generalization

Table 4 Comparative Summary.

Dimension	AR-Based Systems	Skeleton-Based Analysis	Transformer-Based Models
Engagement Level	Very high; interactive, motivating, suitable for children with ASD	Moderate; passive capture without active interaction	Low by itself; it depends on the upstream data source
Data Standardization	Low; tasks vary widely, sensors differ, no consistent protocols	Medium; structured joint coordinates but heterogeneous sensors (Kinect, OpenPose, MoCap)	High requirement; needs clean, uniformly structured sequences
Modeling Complexity	Low to moderate; mostly rule-based or traditional ML	Moderate; CNN/RNN/LSTM/GCN pipelines; limited long-range modeling	High; self-attention enables global temporal–spatial reasoning
Clinical Readiness	Promising but inconsistent; limited formal clinical validation	Growing; used in gait and motor-control studies but not standardized	Emerging; strong potential but still exploratory in ASD research
Key Limitations	Lack of standard tasks; variable environments; inconsistent recording quality	Small datasets; demographic bias; different joint definitions and frame rates; task-specific	Requires large, clean datasets; sensitive to noise; high compute; few clinical benchmarks available

7.PROPOSED PIPELINE

This paper suggests the development of an integrated pipeline that incorporates the augmented reality (AR) gameplay, movement capture using skeletons, and classification based on transformers to aid in autism detection. The pipeline will fill in the gaps in the literature, especially the gap in the unstandardized data on motor tasks and the paucity of more advanced temporal and structural modeling approaches to movement analysis in relation to ASD. By shaping the participants to perform the structured AR tasks and converting the recorded movement sequences to the representation that transformers can accept, the proposed system aims at the creation of credible behavioral indicators and at preserving the engaging and child-friendly environment. The suggested pipeline combines the AR-motor task with whole-body skeleton tracking and transformer analysis, offering a stepwise workflow for evaluating motor movement in ASD. During the gameplay, participant actions are captured as indicated in Figure 7, and their sequence is converted into skeletal joint sequences and processed, which is then classified using a transformer model. The design enables standardized data collection in an engaging environment and the analysis of movement patterns.

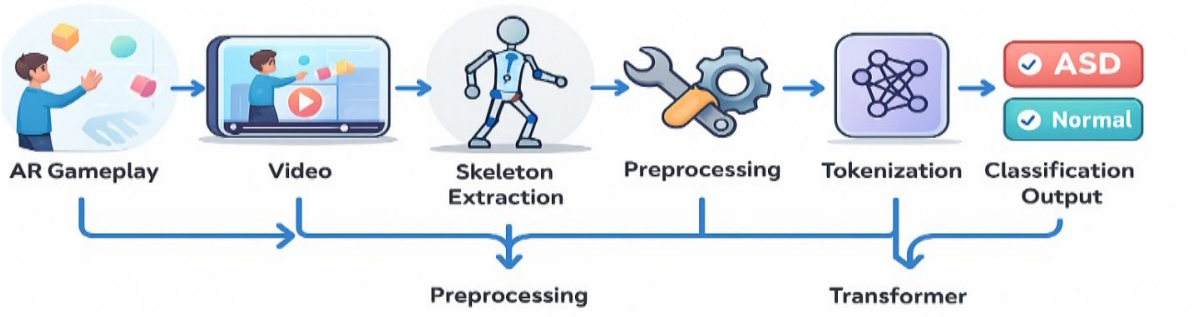

Fig. 7 Proposed Pipeline.

7.1.AR Gameplay Design

It is suggested that the AR game be created to encourage children to perform certain upper- and lower-limb movements in a controlled yet naturalistic environment. The gameplay features will encourage the users to reach, point, step, imitate gestures, and follow movement patterns, allowing a gathering of consistent motor data among the participants. Engagement and anxiety will be ensured through visual and auditory feedback [70]. In contrast, the difficulty of tasks will be altered dynamically based on the performance of the user to support people of different levels of ability. The AR setting tries to provide equilibrium between structure and playfulness, where they can ensure meaningful motor patterns are captured without affecting the comfort of the users.

7.2.Data Collection

Video streams will be captured during the gameplay sessions of ASD and typically developing participants by the system. The protocols to be used in each session will remain constant to maintain uniformity across individuals, including the task sequence, time duration, and distance. The sample will comprise 30 participants (15 with ASD, 15 controls), consistent with the recent literature. Data collection will occur before acquiring ethical approval and parental consent. All videos will be safely stored and anonymized to ensure the privacy of participants.

7.3.Skeleton Extraction

The body joints will be tracked to provide upper- and lower-limb and trunk movement with the help of skeleton data being extracted at the level of recorded videos based on either a Microsoft Kinect camera or a similar pose estimation system. The skeleton data will be used to provide full-body movement at the frame rate, using the 25-joint Kinet model. Such representation allows organizing the analysis of coordination patterns and mobility dynamics.

7.4.Preprocessing

Noise filtering, joint coordinates normalization, and sequence-to-sequence temporal alignment will be part of the preprocessing. Missing or fluctuating joint estimates will be interpolated, and segments with low tracking quality will be removed. The purpose of these steps is to enhance the reliability of the data as well as minimize sensor noise.

7.5.Tokenization

The skeleton sequences obtained after processing are converted into a form that could be read directly by the transformer model. The movement records are initially broken into small time clusters, which in turn record a short continuous part of the movement of the child. In each of the segments, a video frame is an individual token. The 3D joint positions of that frame are flattened into a solitary feature vector and are fed through a learnable projection layer that transforms it into a small numerical embedding that is processed by transformers. Positional information is then added to make the model aware of the sequential arrangement of frames, and an optional classification token can be added at the start of the sequence to provide an overview of the entire segment. The resulting design, based on one token per frame and partially overlapping segments, will ensure that the transformer can capture a wider range of patterns over time, relationships in the world between joints that are distant to one another, and low-level timing anomalies. The latter features of the tokenization strategy render it particularly suitable for the analysis of ASD motor behavior, where the lack of proper coordination, weak time consistency, and distributed joint dependency is frequent over large parts of the movement routine.

7.6.Transformer Model

The array of proposed classifications will utilize a transformer-based framework that is developed to learn long-range temporal dynamics and the interactions among limbs. The self-attention systems will help the model to detect the lack of coordination, timing disparities, and unusual movement organization, which is commonly linked to ASD. It will train the model to differentiate between the ASD and control groups using extracted movement features, leveraging transformer capabilities described in recent literature.

7.7.Evaluation Plan

The system will be tested based on the normal performance measures, including accuracy, precision, recall, and F1-score. The cross-validation will be done to determine the generalization, and it will be compared to the baseline models that include CNN or LSTM architectures. The objective of the evaluation is to know whether the transformer-based representations can be used to achieve quantifiable gains in the ASD classification.

7.8.Summary

Overall, the suggested pipeline is a combination of AR-driven motor activity and transformer-enhanced movement tracking, which will be a systematic and innovative method of assessing autism. The system should produce significant behavioral understanding and help create more useful computational means for ASD assessment.

8.COMPARATIVE ANALYSIS

In this chapter, the author discusses the three main directions of research considered to take place in the context of this survey: AR-based systems, skeleton-based, and transformer-based models. It aims to determine their respective strengths and weaknesses and to emphasize how their results are applied to the development of the proposed pipeline. This section summarizes the lessons learned in the preceding chapters, explains the methodological gap the present research should address, and provides the rationale for combining the AR-guided tasks with the full-body skeleton capture and a transformer architecture.

8.1.Research Directions Comparison

AR-based systems have shown considerable potential to raise the engagement and motivation levels of children with ASD, especially by using interactive and gamified systems. Nonetheless, most AR solutions have no standardized data acquisition, which leads to the inconsistency of movement records and the inability to apply them to objective evaluation. Conversely, skeleton-based techniques offer positional descriptions of motor behaviors and can be used to make quantitative assessments of coordination, posture, and movement performance. However, most of these approaches are based on conventional classifiers or sequence models like CNNs or LSTMs, which do not conceptualize long-range temporal interactions and complicated inter-limb interactions. Transformer models overcome these limitations in that they provide better performance in the context of modeling the temporal patterns and structural dependencies between joints. The literature consulted in Chapter 6 indicates that the spatiotemporal

transformer and the graph-based transformer are superior to traditional models for detecting abnormal movement patterns associated with ASD. Nonetheless, these models require high-quality, consistent input data and are not often implemented in controlled task environments, which limits their usability.

8.2.Integrated Insights

The comparative results have shown that there is no single direction of research that can offer a comprehensive solution to ASD movement assessment. AR-based solutions are the most effective for engagement; skeleton-based solutions provide quantitative analysis of muscular movements but have less developed modeling; and transformer-based models have excellent representational capability but heavily rely on high-quality inputs. Such strengths can be used to overcome some of their weaknesses, suggesting that a synthesised framework may yield more valid and apparent assessment results.

8.3.Recommendations to the Proposed Pipeline

As a result of this investigation, the pipeline proposed in Chapter 7 is directly aligned with the most promising research direction. The system guarantees a controlled and standardized data collection and keeps the participants motivated using AR-guided tasks. Whole-body skeleton capture offers structured data on movement that is useful in computational modeling, and transformer-based structures allow modeling of both temporal and structural dependencies. Such a combination addresses gaps identified in the current literature and provides a unified method that balances usability and analytical rigor. In general, this course of comparison shows that current available methods offer quality but incomplete solutions to the assessment of ASD movement. AR systems enhance interactivity without structured data, skeleton-based methods have a quantitative representation of movements but only limited modeling techniques, and the transformer-based models allow more complex analysis but demand steady inputs. All these restrictions underscore the necessity of a combined framework, which is at the foundation of the troubles in the following chapter.

9.CHALLENGES AND LIMITATIONS

Although the integration of AR-guided motor tasks, fully capturing the skeleton, and transformer-based analysis takes a promising direction, there are several challenges and limitations. First, not many high-quality records of movement among children with ASD exist, because recruiting, seeking ethical approval, and control of data gathering take much time and need organization with clinical and educational facilities. Such a lack of standardised data can affect the generalisability of models and limit large-scale validation. Second, skeleton-tracking systems like Kinetect or pose-estimation systems may introduce noise, occlusions, or missing joints, especially when children behave chaotically or lose attention during tasks. Such tracking anomalies can decrease the accuracy of movement features being extracted and can affect the performance of transformers, which also depends on a reliable temporal and structural input. Third, behavioral heterogeneity among the ASD population is a given challenge. Variations in cognitive capacity, motor dexterity, attention span, and sensory sensibilities might also affect the performance of the tasks, and it may not be easy to determine a common pattern of movement among the participants. Due to these factors, models may not fully represent the full range of ASD-related motor characteristics. Lastly, the suggested pipeline is yet to be proven empirically, and the future application and appraisal will determine its efficiency. The assessment of real-world performance, usability, and clinical relevance, especially across a variety of settings, needs to be further conducted.

10.FUTURE DIRECTIONS

The work will be conducted in the future to apply and verify the proposed AR-skeleton-transformer pipeline and implement it in the real-world environment. Sensory modalities can be extended by including audio or facial expression analysis to increase the capability of the system to detect multimodal behavioral signals that are linked to ASD. The deepening of the sample of participants and longitudinal research may also enhance the generalizability of the model and help to determine the developmental patterns across the years. The metrics used in clinically aligned evaluation protocols must enable the comparative, meaningful representation of the computational model relative to pre-existing diagnostic practices. This involves correlation of model output with standardized clinical scales like ADOS, ADI-R, or SRS-2 so that atypicality of motor manifestation at the predicted level can be said to be in line with severity scores that are clinically sound. Also, ASD motor subtypes should be evaluated sensitively in that they determine patterns between bilateral coordination problems, gait abnormalities, the lack of postural stability, and repetitive motor behaviours, and not give a single global prediction. Longitudinal validation should also be included: it is necessary to test the models with movement data measured on several sessions to confirm their stability over time and be able to observe the changes in development or the effect of the intervention. More measures can be tested, including retest reliability, agreement with clinician-rated video measures, and the ability

to identify small within-child differences associated with familiar clinical indicators. The combination of these elements guarantees the statistical accuracy of computational predictions, as well as their clinical interpretability, clinical reproducibility, and true relevance to real-world ASD assessment procedures.

11.CONCLUSION

This survey has reviewed three significant research directions in the field of assessing the movement of ASD, which are AR-based systems, skeleton-based analysis, and transformer-based models. AR solutions are highly interactive but lack standardized data-acquisition procedures. In contrast, skeleton-based methods offer organized motor data at the expense of traditional models with limited ability to model the complex temporal and inter-joint relationships. A transformer-based strategy presents better modeling abilities, especially long-range temporal-spatial relationships, though it needs clean and consistently structured input data, which most present ASD datasets lack. Together, these results indicate the obvious methodological gap: there is no current framework that can combine the control of task conditions, full-body behavior in an ecologically valid way, measure standard bone motions, and use the advanced transformer architectures to understand the movement deeply. The proposed AR-Skeleton-Transformer pipeline directly fills this gap by providing a well-controlled and clinically equivalent behavioral elicitation with the use of AR tasks, generating high-quality and standardized skeleton data that can be used in further modeling, and using transformer models to obtain rich and temporal-structural biomarkers that are not accessible in other modalities. This combined workflow is not only able to overcome the space between the existing studies of ASD movements but also offers a scalable, reproducible, and clinically promising basis for next-generation computational ASD diagnostics, making the pipeline a formidable and visionary future research choice.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Iyas Qaddara: Writing original draft, survey methodology, background, proposed pipeline, formal analysis, and comparison analysis. **Ahmad Sharieh**: Supervision, review, and editing. **Huda Karajeh:** Supervision.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

ACKNOWLEDGEMENTS

The authors would like to thank the administration of Jordan University for providing all forms of support to the university's Students, especially in scientific research.

REFERENCES

[1] Hirota T, King BH. **Autism Spectrum Disorder: A Review**. *JAMA* 2023; **329**(2):157–168.

[2] Su Q, Wong OW, Lu W, Wan Y, Zhang L, Xu W, Ng SC. **Multikingdom and Functional Gut Microbiota Markers for Autism Spectrum Disorder**. *Nature Microbiology* 2024; **9**(9):2344–2355.

[3] Barta S, Gurrea R, Flavián C. **Augmented Reality Experiences: Consumer-Centered Augmented Reality Framework and Research Agenda**. *Psychology & Marketing* 2025; **42**(2):634–650.

[4] Khowaja K, Banire B, Al-Thani D, Sqalli MT, Aqle A, Shah A, Salim SS. **Augmented Reality for Learning of Children and Adolescents with Autism Spectrum Disorder (ASD): A Systematic Review**. *IEEE Access* 2020; **8**:78779–78807.

[5] Taha BA, Addie AJ, Kadhim AC, Azzahran AS, Haider AJ, Chaudhary V, Arsad N. **Photonics-Powered Augmented Reality Skin Electronics for Proactive Healthcare: Multifaceted Opportunities**. *Microchimica Acta* 2024; **191**(5):250.

[6] Smadi TA, Al-Maitah M. **Artificial Intelligent Technology for Safe Driver Assistance System**. *International Journal of Computer Aided Engineering and Technology* 2020; **13**(1-2):183–191.

[7] Omarov N, Omarov B, Azhibekova Z, Omarov B. **Applying an Augmented Reality Game-Based Learning Environment in Physical Education Classes to Enhance Sports Motivation**. *Retos* 2024; **60**:269–278.

[8] Stalheim OR, Somby HM. **An Embodied Perspective on an Augmented Reality Game in School: Pupil's Bodily Experience Toward Learning**. *Smart Learning Environments* 2024; **11**(1):24.

[9] Roggio F, Trovato B, Sortino M, Musumeci G. **A Comprehensive Analysis of the Machine Learning Pose Estimation Models Used in Human Movement and Posture Analyses: A Narrative Review**. *Heliyon* 2024; **10**(21): e39977.

[10] Phalke DA, Kotipalli V, Ranjan P, Pawar Y, Bharat P. **Artificial Intelligence in Fitness: Pose Estimation and**

Movement Correction. *Cureus Journals* 2025; **2**(1): 1-12.

[11] Kim JW, Khan AU, Banerjee I. **Systematic Review of Hybrid Vision Transformer Architectures for Radiological Image Analysis**. *Journal of Imaging Informatics in Medicine* 2025; **37**(4): 1–15.

[12] Ibadi H, Lakizadeh A. **ASDvit: Enhancing Autism Spectrum Disorder Classification Using Vision Transformer Models Based on Static Features of Facial Images**. *Intelligence-Based Medicine* 2025; **11**: 100226.

[13] Astafeva D, Syunyakov T, Shapievskii D, Malashonkova E, Vlasov A, Shport S, Smirnova D. **Virtual Reality/Augmented Reality (VR/AR) Approach to Develop Social and Communication Skills in Children**. *Source Not Fully Specified* 2024.

[14] Habboush A, Elzaghmouri B. **Methodological Integration of Machine Learning and Metaheuristics for Seismic Risk Assessment of Irregular Buildings**. *Asian Journal of Civil Engineering* 2025:1–16.

[15] Yang YM, Chang KC, Luo JN. **Hybrid Neural Network-Based Intrusion Detection System: Leveraging LightGBM and MobileNetV2 for IoT Security**. *Symmetry* 2025; **17**(3):314.

[16] Alnabhan M, El-Qasass A, Atoum M, Al-Haija QA, Habboush A. **A Lightweight Cryptographic Solution for Enhanced Image Security**. *Engineering, Technology & Applied Science Research* 2025; **15**(5):27052–27059.

[17] Al Smadi T. **Application of Fuzzy Logic to Cognitive Wireless Communications**. *Journal of Advanced Sciences and Engineering Technologies* 2019; **2**(03): 2228-2234.

[18] Zhou Y, Rashid FAN, Mat Daud M, Hasan MK, Chen W. **Machine Learning-Based Computer Vision for Depth Camera-Based Physiotherapy Movement Assessment: A Systematic Review**. *Sensors* 2025; **25**(5):1586.

[19] Qaddara I, Alraba'nah Y. **Enhancing Requirements Classification Using Machine Learning Techniques**. *SN Computer Science* 2025; **6**(6):649.

[20] Khowaja K, Banire B, Al-Thani D, Sqalli MT, Aqle A, Shah A, Salim SS. **Augmented Reality for Learning of Children and Adolescents with Autism Spectrum Disorder (ASD): A Systematic Review**. *IEEE Access* 2020; **8**:78779–78807.

[21] Cihak DF, Moore EJ, Wright RE, McMahon DD, Gibbons MM, Smith C. **Evaluating Augmented Reality to Complete a Chain Task for Elementary Students with Autism**. *Journal of Special Education Technology* 2016; **31**(2):99–108.

[22] Latrèche K, Kojovic N, Pittet I, Natraj S, Franchini M, Smith IM, Schaer M. **Gesture Imitation Performance and Visual Exploration in Young Children with Autism Spectrum Disorder**. *Journal of Autism and Developmental Disorders* 2024:1–14.

[23] Van der Hallen R, Evers K, de-Wit L, Steyaert J, Noens I, Wagemans J. **Multiple Object Tracking Reveals Object-Based Grouping Interference in Children with ASD**. *Journal of Autism and Developmental Disorders* 2018; **48**(4):1341–1349.

[24] Gilabert-Cerdá A, Lledó GL, Lorenzo-Lledó A, Carreres AL, Pérez-Vázquez E. **Augmented Reality Promotes Social Responses in Autism Spectrum Disorder Subjects**. Education and New Developments 2023: 424-428.

[25] Razhkou I. **In-Game Reward Systems and their Effect on the Player**. 2024.

[26] Pérez-Fuster P, Herrera G, Vera L, Nadel J, Tijus C, López-Fernández A, Leppink J. **Pictogram Room Augmented Reality Technology Games Improve Body Knowledge, Imitation, and Joint Attention Skills in Autistic Children with Intellectual Disability**. *Scientific Reports* 2025; **15**(1):34966.

[27] Amara K, Boudjemila C, Zenati N, Djekoune O, Aklil D, Kenoui M. **AR Computer-Assisted Learning for Children with ASD Based on Hand Gesture and Voice Interaction**. *IETE Journal of Research* 2023; **69**(12):8659–8675.

[28] Lee IJ. **Kinect-for-Windows with Augmented Reality in an Interactive Roleplay System for Children with an Autism Spectrum Disorder**. *Interactive Learning Environments* 2021; **29**(4):688–704.

[29] Alcaniz Raya M, Marín-Morales J, Minissi ME, Teruel Garcia G, Abad L, Chicchi Giglioli IA. **Machine Learning and Virtual Reality on Body Movements' Behaviors to Classify Children with Autism Spectrum Disorder**. *Journal of Clinical Medicine* 2020; **9**(5):1260.

[30] Hu X, Jiang Y, Sun Y, Xu Z, Zheng F, Hu X. **Exploring the Effects of Augmented Reality on the Emotion Recognition Skills of Autistic**

Children. *Journal of Special Education Technology* 2025; **40**(3):384–399.

[31] Tang TY, Xu J, Winoto P. **An Augmented Reality-Based Word-Learning Mobile Application for Children with Autism to Support Learning Anywhere and Anytime: Object Recognition Based on Deep Learning**. *International Conference on Human-Computer Interaction* 2019:182–192.

[32] Abbas AK, Ayop R, Tan CW, Al Mashhadany Y, Takialddin AS. **Advanced Energy-Management and Sizing Techniques for Renewable Microgrids with Electric-Vehicle Integration: A Review**. *Results in Engineering* 2025; **27**:106252.

[33] Wedyan M, Al-Jumaily A, Dorgham O. **The Use of Augmented Reality in the Diagnosis and Treatment of Autistic Children: A Review and a New System**. *Multimedia Tools and Applications* 2020; **79**(25):18245–18291.

[34] Koumpouros Y. **Digital Horizons: Enhancing Autism Support with Augmented Reality**. *Journal of Autism and Developmental Disorders* 2025:1–17.

[35] Lee IJ, Huang YC. **Improving Social Skills in Children with Autism Spectrum Disorder Using Augmented Reality and Mixed Reality Technology Combined with Concept Maps**. *Universal Access in the Information Society* 2025; **24**(2):1255–1281.

[36] McMahon D, Cihak DF, Wright R. **Augmented Reality as a Navigation Tool to Employment Opportunities for Postsecondary Education Students with Intellectual Disabilities and Autism**. *Journal of Research on Technology in Education* 2015; **47**(3):157–172.

[37] Fridhi A, Bali N, Rebai N, Kouki R. **Geospatial Virtual/Augmented Environment: Applications for Children with Pervasive Developmental Disorders**. *Neurophysiology* 2020; **52**(3):239–246.

[38] Cheng Y, Bololia L. **The Effects of Augmented Reality on Social Skills in Children with an Autism Diagnosis: A Preliminary Systematic Review**. *Journal of Autism and Developmental Disorders* 2024; **54**(4):1317–1331.

[39] Liu R, Salisbury JP, Vahabzadeh A, Sahin NT. **Feasibility of an Autism-Focused Augmented Reality Smartglasses System for Social Communication and Behavioral Coaching**. *Frontiers in Pediatrics* 2017; **5**:145.

[40] Syahputra MF, Arisandi D, Lumbanbatu AF, Kemit LF, Nababan EB, Sheta O. **Augmented Reality Social Story for Autism Spectrum Disorder**. *Journal of Physics: Conference Series* 2018; **978**(1):012040.

[41] Nekar DM, Kang H, Alao H, Yu J. **Feasibility of Using Multiplayer Game-Based Dual-Task Training with Augmented Reality and Personal Health Record on Social Skills and Cognitive Function in Children with Autism**. *Children* 2022; **9**(9):1398.

[42] Rega A, Mennitto A, Vita S, Iovino L. **New Technologies and Autism: Can Augmented Reality (AR) Increase the Motivation in Children with Autism?** *INTED2018 Proceedings* 2018:4904–4910.

[43] Williams A, Chandramouli M. **Virtual Reality (VR)-Based Training Tool for Positive Reinforcement & Communication in Autistic Children**. *2025 International Conference on Intelligent Computing and Virtual & Augmented Reality Simulations (ICVARS)* 2025:78–82.

[44] Nekar DM, Lee DY, Hong JH, Kim JS, Kim SG, Seo YG, Yu JH. **Effects of Augmented Reality Game-Based Cognitive–Motor Training on Restricted and Repetitive Behaviors and Executive Function in Patients with Autism Spectrum Disorder**. *Healthcare* 2022; **10**(10):1981.

[45] Brandão J, Cunha P, Vasconcelos J, Carvalho V, Soares F. **An Augmented Reality Gamebook for Children with Autism Spectrum Disorders**. *The International Conference on E-learning in the Workplace 2015* 2015:1–6.

[46] Bhatt SK, De Leon NI, Al-Jumaily A. **Augmented Reality Game Therapy for Children with Autism Spectrum Disorder**. *International Journal on Smart Sensing & Intelligent Systems* 2014; **7**(2): 519-536.

[47] Qaddara I, Alraba'nah Y, Hiari MO. **Evaluation of SQL and NoSQL Databases on Parallel Processing**. *Engineering, Technology & Applied Science Research* 2025; **15**(4):24298–24304.

[48] Lu H, Chen J, Zhang Z, Liu R, Zeng R, Hu X. **Emotion Recognition from Skeleton Data: A Comprehensive Survey**. *arXiv preprint* 2025:arXiv :2507.18026.

[49] Qaddara I, Naffar E, Hyassat A, Kenanah A, Alraban'Ah Y, Abualhaj MM, Alzubaidi R. **Arabic News Text Summarization: An Extractive**

Technique. *2025 12th International Conference on Information Technology (ICIT)* 2025:571–576.

[50] Li J, Chheang V, Kullu P, Brignac E, Guo Z, Bhat A, Barmaki RL. **Mmasd: A Multimodal Dataset for Autism Intervention Analysis**. *Proceedings of the 25th International Conference on Multimodal Interaction* 2023:397–405.

[51] Zhang Y, Tian Y, Wu P, Chen D. **Application of Skeleton Data and Long Short-Term Memory in Action Recognition of Children with Autism Spectrum Disorder**. *Sensors* 2021; **21**(2):411.

[52] Al-Jubouri AA, Ali IH, Rajihy Y. **Generating 3D Dataset of Gait and Full Body Movement of Children with Autism Spectrum Disorders Collected by Kinect v2 Camera**. *Compusoft* 2020; **9**(8):3791–3797.

[53] Muty N, Azizul Z. **Detecting Arm Flapping in Children with Autism Spectrum Disorder Using Human Pose Estimation and Skeletal Representation Algorithms**. *2016 International Conference on Advanced Informatics: Concepts, Theory and Application (ICAICTA)* 2016:1–6.

[54] Shin J, Miah ASM, Kakizaki M, Hassan N, Tomioka Y. **Autism Spectrum Disorder Detection Using Skeleton-Based Body Movement Analysis via Dual-Stream Deep Learning**. *Electronics* 2025; **14**(11):2231.

[55] Zahan S, Gilani Z, Hassan GM, Mian A. **Human Gesture and Gait Analysis for Autism Detection**. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 2023:3328–3337.

[56] Yazdi SA, 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 & Sensors* 2024; **14**(10):29.

[57] Paulo JR, Sousa T, Perdiz J, Pereira L, Vasen M, Mouga S, Castelo-Branco M. **A Multimodal Dataset Addressing Motor Function in Autism**. *Scientific Data* 2025; **12**(1):959.

[58] Lemler C, Kleber SK, Polzer L, Raji N, Kitzerow-Cleven J, Kim Z, Bast N. **Semi-Automated Multi-Label Classification of Autistic Mannerisms by Machine Learning on Post Hoc Skeletal Tracking**. *Autism Research* 2025; **18**(4):833–844.

[59] Lin WS, Peng XY, Cheng YF. **Using Intelligent Robots to Detect Body Language and Improve Social Development in Children with Autism Spectrum Disorder**. *International Journal of Intelligent Robotics and Applications* 2025; **9**(2):592–607.

[60] Goldthorp K, Henderson B, Yogarajah P, Gardiner B, McGinnity TM, Nicholas B, Wimpory DC. **Increased Temporal Variability of Gait in ASD: A Motion Capture and Machine Learning Analysis**. *Biology* 2025; **14**(7):832.

[61] Wu X, Dickin DC, Bassette L, Ashton C, Wang H. **Clinical Gait Analysis in Older Children with Autism Spectrum Disorder**. *Sports Medicine and Health Science* 2024; **6**(2):154–158.

[62] Kalam SA, Prome TR, Ullah MA. **Machine Learning Approach for Identification of Autism Spectrum Disorder from Video Using OpenPose**. *2024 27th International Conference on Computer and Information Technology (ICCIT)* 2024:1720–1725.

[63] Anderson JT, Stenum J, Roemmich RT, Wilson RB. **Validation of Markerless Video-Based Gait Analysis Using Pose Estimation in Toddlers with and without Neurodevelopmental Disorders**. *Frontiers in Digital Health* 2025; **7**:1542012.

[64] Barahona M. **Evaluation of an Open-Source Pose Estimation Tool to Track Infant Arm Motion**. 2025.

[65] Shin J, Miah ASM, Kakizaki M, Hassan N, Tomioka Y. **Autism Spectrum Disorder Detection Using Skeleton-Based Body Movement Analysis via Dual-Stream Deep Learning**. *Electronics* 2025; **14**(11):2231.

[66] Gupta K, Aly A, Ifeachor E. **Multi-Modal Framework for Autism Severity Assessment Using Spatio-Temporal Graph Transformers**. *18th International Conference on Health Informatics (HEALTHINF)* 2025.

[67] Zhang J, Xie W, Wang C, Tu R, Tu Z. **Graph-Aware Transformer for Skeleton-Based Action Recognition**. *The Visual Computer* 2023; **39**(10):4501–4512.

[68] Anand, Kini. **Hybrid Transformer-CNN Models for Enhanced ASD Classification**. 2024.

[69] Gao L, Wang Z, Long Y, Zhang X, Su H, Yu Y, Hong J. **Autism Spectrum Disorders Detection Based on Multi-Task Transformer Neural Network**. *BMC Neuroscience* 2024; **25**(1):27.

[70] Hiari M, Alraba'nah Y, Qaddara I. **A Deep Learning-Based Intrusion Detection System Using Refined LSTM for DoS Attack Detection**. *Engineering,*

Technology & Applied Science Research
2025; **15**(4):25627–25633.