A Skeleton-Based Movement Dataset for Autism Spectrum Disorder (ASD) Collected via an Augmented Reality Game

Iyas Qaddara
Department of Computer Science, Faculty of Information Technology, Al-Ahliyya Amman University, Jordan | Department of Computer Science, King Abdullah II School of Information Technology, University of Jordan, Jordan
i.qaddara@ammanu.edu.jo (corresponding author)

Mohammad O. Hiari
Department of Networks and Cybersecurity, Faculty of Information Technology, Al-Ahliyya Amman University, Jordan
m.hyari@ammanu.edu.jo

Talal. S. M. Haimur
Department of Allied Science, Faculty of Arts and Science, Al-Ahliyya Amman University, Jordan
thaimur@ammanu.edu.jo

Received: 28 January 2026 | Revised: 1 March 2026 | Accepted: 18 March 2026

 DOI: https://doi.org/10.48084/etasr.17815

ABSTRACT

This study recorded movement data for children with Autism Spectrum Disorder (ASD) and Typically Developing (TD) children on a skeleton-based serious game based on Augmented Reality (AR). The dataset contains full-body skeletal motions in interactive and task-oriented activities, unlike current datasets that use passive observation or detection of a part of the body. The AR environment was characterized as arousing clinically relevant movements of the upper limbs, lower limbs, balance, and body rotation. A Microsoft Kinect v2 sensor was used to capture data, and a series of steps were taken to process them, namely cleaning, segmentation, spatial normalization, and temporal normalization to create fixed-length sequences. The data consist of equal recordings of the ASD and TD groups and can be used in Machine Learning (ML) studies. Random Forest and CNN-LSTM baseline experiments show that the dataset can be used to learn discriminative movement patterns. This dataset can lead to the evolution of movement-based autism analysis and AR-based behavioral data collection.

Keywords-Autism Spectrum Disorder (ASD); skeleton-based movement analysis; augmented reality serious games; Machine Learning (ML)

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition commonly associated with challenges in social interaction, communication, and behavior. In recent years, research has increasingly reported that children with ASD also exhibit distinctive motor characteristics, including differences in coordination, balance, posture, and movement execution. These motor-related traits have drawn attention as potential objective indicators that may complement traditional ASD assessment methods [1]. Current ASD assessment practices are largely dependent on clinical observations and subjective assessments conducted by specialists. Although effective, such approaches can be time-consuming and influenced by individual judgment. Moreover, many existing computational studies focus on limited cues such as facial expressions, eye gaze, or upper-body motion, leaving full-body movement behavior relatively underexplored despite its relevance to motor development in autism [2].

Advances in motion sensing technologies have enabled non-invasive capture of human movement for behavioral analysis. In particular, skeleton-based representations obtained from depth sensors such as Microsoft Kinect provide structured and interpretable descriptions of full-body motion. However, there remains a lack of well-documented datasets that capture full-body skeletal movements of children with ASD during controlled but natural interaction scenarios [3]. Augmented Reality (AR) offers an effective solution for developing interactive environments that encourage predefined movements while maintaining engagement and comfort. AR-based serious games allow children to perform guided motor tasks in a motivating setting, which is especially important when collecting behavioral data from children with ASD [4].

This study introduces a skeleton-based movement dataset collected from children with ASD and Typically Developing (TD) peers using an AR game. The dataset was constructed through a structured AR interaction framework designed to generate a range of upper- and lower-body movements. Full-body skeletal data was captured using a Microsoft Kinect sensor and processed into standardized temporal sequences suitable for Machine Learning (ML) analysis. This dataset consists of synchronized RGB videos and 3D skeletal joint sequences, where each sample is represented as a normalized tensor of size (150×25×3). The dataset includes recordings from both ASD and TD children across five structured AR tasks, is currently available for research purposes upon request, and will be publicly released in future work.

The main contributions of this paper are as follows:

- A novel AR-based data collection framework to generate full-body movements, offering a balanced skeleton-based dataset for ASD research.
- A complete preprocessing pipeline for generating standardized movement sequences, and baseline ML and Deep Learning (DL) benchmarks.
- To support reproducibility and demonstrate dataset usability, a detailed description of the data collection protocol, preprocessing pipeline, and final data representation is provided. In addition, baseline classification experiments are conducted to validate the discriminative capability of the dataset for movement-based ASD research.

II. BACKGROUND

Research on ASD has traditionally focused on behavioral observation, clinical interviews, and questionnaire-based assessments. With the advancement of computational methods, recent studies have explored automatic ASD analysis using visual and sensor-based data, including facial expressions, eye gaze, speech patterns, and gesture analysis. Although these approaches have shown promise, many of them rely on partial behavioral cues and do not fully capture whole-body motor behavior.

Research on ASD increasingly incorporates objective, data-driven methods to characterize behaviors that extend beyond traditional clinical observation [5]. Several datasets have been introduced that include 3D behavioral measurements of children with ASD. For example, the DREAM dataset [6] contains motion information from body position, head orientation, and eye gaze during robot-enhanced therapy, providing structured 3D data that support a comprehensive investigation of ASD movement behavior.

Multimodal datasets such as MMASD [7] have also recently emerged to offer synchronized skeleton data and other modalities derived from play therapy interventions, contributing to research on joint movement and behavioral engagement in ASD from a multimodal perspective. Despite these advantages, there are still relatively few datasets focused specifically on full-body skeletal movement captured in interactive or enriched environments that are designed to elicit motor tasks relevant to analysis. The three-dimensional gait and full-body movement dataset collected via Kinect V2 [8] represents one of the earlier comprehensive efforts to provide 3D skeletal joint data of children with ASD in controlled tasks.

Nevertheless, the research community continues to highlight the lack of publicly accessible, well-documented skeletal motion datasets collected in engaging or naturalistic interaction contexts, particularly those incorporating interactive elements like AR that may better support richer motor elicitation across diverse movement profiles. This work proposes a dataset that addresses this gap by capturing full-body movement sequences through AR, with the aim of supporting ASD-oriented motion analysis and ML tasks.

Compared to current datasets, including DREAM and MMASD, as shown in Table I, which are mostly only obtained in a therapeutic or observational context, the suggested data is founded on an interactive setting of AR that actively activates structured full-body movements. Furthermore, datasets such as [9], which rely on controlled laboratory settings using skeleton and RGB data, remain limited to partially constrained movements and lack interactive task-driven engagement. Similarly, the study in [10] focuses on markerless skeleton data in a lab-based setup, but is restricted to gait analysis and does not capture diverse movement patterns. In addition, gait datasets that are based on Kinect v2 sensors are primarily concerned with lower-body movements, whereas the proposed data set involves a variety of multi-joint movement patterns that include both upper and lower limbs. This renders the data better suited to ASD analysis based on movement.

TABLE I. DATASETS' COMPARISON

Dataset	Modality	Interaction type	Full body	AR-based	Tasks	Size
DREAM [6]	Multimodal	Robot therapy	Partial	✗	Limited	61
MMASD [7]	Multimodal	Therapy/play	Partial	✗	Limited	32
[9]	Skeleton + RGB	Controlled lab	Partial	✗	Limited	100
[10]	Skeleton (markerless)	Lab-based	✗	✗	Gait only	53
Kinect Gait [8]	Skeleton	Lab gait	Lower-body	✗	Walking only	Varies
Proposed	Skeleton + RGB	AR interactive game	Full body	✓	Multi-task	30

III. AR GAME DESIGN & DATA COLLECTION

This section describes the design of the AR game and the data collection procedure used to construct the proposed skeleton-based movement dataset. The primary objective of the AR game is to provide a structured yet engaging environment that encourages children to perform predefined full-body movements in a natural and motivating manner. Unlike passive recording setups, the proposed AR-based approach actively elicits specific motor actions while maintaining child comfort and engagement, which is particularly important when working with children with ASD.

A. AR Game Design

The design aimed at offering an interactive and organized space where children will be motivated to perform specified physical activities as they watch themselves on the screen in real time. The system shows the live video stream of the child as recorded by the Microsoft Kinect V2 camera and superimposes game-related objects and indicators, colored objects and targets, directly onto the scene. The design enables children to play with the virtual objects with their bodies. Thus, the game is intuitive and fun, and ensures that meaningful full-body movements are elicited and controlled.

The main aim of the game is to make movements of the upper and lower limbs, trunk rotations, and postures such that a large number of Kinect joints are engaged in the various tasks. The system is organized in the form of a series of little mini tasks (or levels) during which a child does not spend more than several seconds, and in the form of which there is a clear visual and auditory feedback to keep the child focused and motivated. Before initiating the activities, a warm-up procedure is performed to verify that the child is in the tracking zone, and the Microsoft Kinect v2 sensor is properly tracking the entire skeleton. After calibration, the game proceeds automatically through a series of set tasks. The core interaction tasks encompass the following to cover the main Kinect joints:

1) Task 1 – Reach and Touch Floating Objects (Upper Limb Focus)

Colored virtual spheres or cubes appear at different positions around the child's upper body (e.g., above the head, to the left/right of the shoulders, and in front of the torso). The child is instructed to "touch the red ball" or "reach the yellow cube" by extending one or both arms. This task predominantly activates the shoulder, elbow, wrist, and hand joints, while also involving the head and upper spine joints as the child visually tracks the targets. Figure 1 shows an example.

2) Task 2 – Cross Body Reach (Coordination and Trunk Rotation)

Virtual targets are displayed on the opposite side of the body (e.g., a target appears near the left side of the screen but must be reached with the right hand, and vice versa). The child is asked to "touch the star on the other side." This task encourages cross-body movements, trunk rotation, and lateral weight shifting, engaging shoulders, elbows, wrists, spine, and hip joints. Figure 2 shows an example.

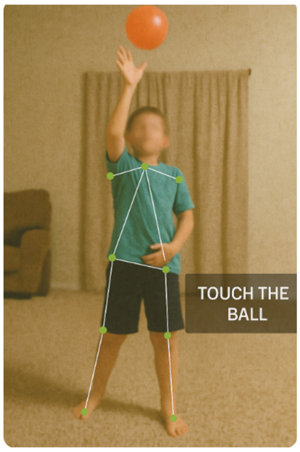

Fig. 1. Task 1 example.

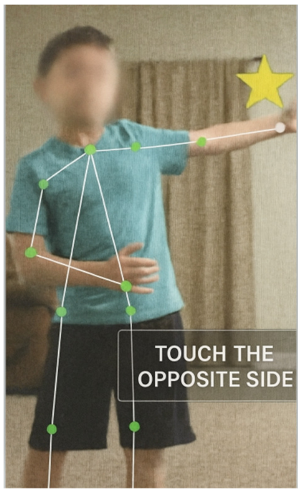

Fig. 2. Task 2 example.

3) Task 3 – Step Forward to Collect Ground Targets (Lower Limb and Balance)

A virtual marker (such as a star or coin) appears on the floor in front of the child. The child is instructed to "step on the star" or "walk to the coin." Completing the task requires taking one or more steps forward and then returning to the initial position, activating hip, knee, ankle, and foot joints, as well as balance-related postural adjustments. Figure 3 shows an example.

4) Task 4 – Kick the Balloon (Leg Movement)

There is a virtual balloon close to one of the child's feet. The child is requested to kick the balloon with the right or the left leg. When the balloon is kicked in the right way, it reacts with an animation. This activity focuses on hip, knee, and ankle movements and needs the support of one leg, which additionally promotes balance and coordination. Figure 4 shows an example.

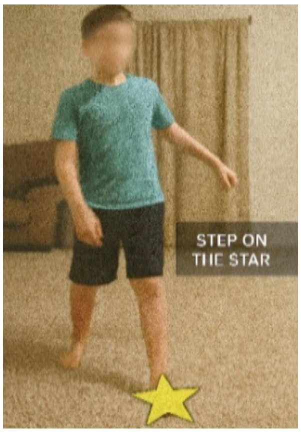

Fig. 3. Task 3 example.

Fig. 4. Task 4 example.

5) Task 5 – Turn and Follow the Target (Body Rotation)

A moving object or an arrow is shown either on the side or just behind the child, and the instruction is to turn around and follow the arrow or see the rocket. The target can also move to the other side of the child, which will involve a 90–180° body rotation. This task can engage the head, spine, hip, and lower limb joints and captures the pattern of rotation, which cannot be dealt with by only frontal movements. Figure 5 shows an example.

These tasks are arranged in such a manner that, during a single game session, all major Kinect joints (head, neck, spine, shoulders, elbows, wrists, hands, hips, knees, ankles, and feet) are engaged at least once under various conditions of movement. Tasks can be made harder and faster (e.g., by target size, distance, or time available) to the level of comfort and ability of the child so that the environment is both available and diagnostically informative. Considering the user experience, the game offers visual feedback immediately after a task is completed (such as the object exploding into small particles or turning colors) and a short positive sound. A simple score indicator or a progress bar that displays the number of tasks completed can be used, but without a sense of competition.

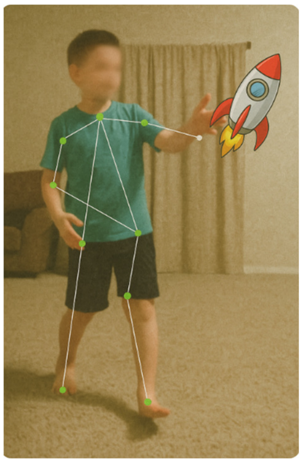

Fig. 5. Task 5 example.

B. Data Collection Procedure

The data collection process was conducted in a controlled indoor environment to ensure consistent lighting conditions and optimal Kinect tracking performance. At the beginning of each session, the child was positioned approximately 2 to 2.5 m in front of the Kinect sensor, standing within the predefined tracking area marked on the floor. A brief calibration step was performed to verify full-body skeleton detection, ensuring that all major joints were accurately captured before the game started.

After calibration, the AR game automatically progressed through the predefined task sequence described in Section III-A. The duration was between 5 and 10 s, and during this time, the child played with the virtual items on the screen and observed their personal live video feed with AR components on top. The Microsoft Kinect v2 sensor was able to capture the RGB video stream and all body joint coordinates in real-time at 30 fps, providing good alignment between hand and body movements.

During the session, auditory and visual feedback were used to keep the session active and motivate the participants to perform the tasks correctly. The researcher was also available to give clarification where necessary without directly interfering with the natural behavior of the child. The sessions were between 2 and 3 minutes long, depending on the responsiveness and level of comfort of the child. At the end of the recording, the obtained data were stored in a safe place and anonymized by removing the information on the faces and issuing individual session numbers.

C. Participants and Ethical Considerations

In this study, 30 children were involved, 15 with ASD and the remaining 15 TD who acted as the control population. The respondents, aged 5 to 10 years, were recruited from special autism centers and regular schools. The validity of the group categorization was ensured by the fact that all children with ASD had an official clinical diagnosis issued by certified professionals. Each group consisted of 15 participants (ASD: n=15, TD: n=15). Inclusion criteria required participants to be able to stand independently and follow simple motor instructions. Exclusion criteria included severe motor impairments or medical conditions affecting movement. The participation criteria were that the children should be able to stand up and respond to simple instructions of movement so that they could accomplish the necessary tasks of the game. Children who had severe motor impairments or conditions that might have made them unsafe to participate in the study were not included. All sessions were facilitated one at a time in a quiet place to minimize the issue of sensory overload and to help children with ASD feel comfortable.

The tasks were presented in a fixed sequential order. The instructions were delivered verbally by the researcher and visually through the AR interface. Each task lasted approximately 5 to 10 s.

Before data collection, parents or other legal guardians gave their written informed consent and children gave their verbal consent where necessary. During the session, a parent or guardian was present to provide emotional support and ensure the well-being of the child. To maintain the privacy of the participants, the facial information of the taped videos was blurred, and a special identifier was given to each participant in order to anonymize the data set.

This process adhered to ethical standards for conducting research with minors and participants with developmental issues. The institutional research ethics committee reviewed and approved the data collection procedure and the data storage protocol, complying with the ethical standards of research and the confidentiality principles. Table II summarizes the demographic features of the participants in terms of age, gender, height, and group assignment, with 750 AVG frame/sessions and 2-3 minutes long. The dataset demonstrates a balanced distribution across participants, session durations, and frame counts. This balance reduces potential bias and ensures fair representation across both ASD and TD groups.

TABLE II. DATASETS COMPARISON

Group	**N**	**Age range**	**Mean Age ± SD**	**Gender**	**Height**
ASD	15	5–10	7.1 ± 1.3	10M / 5F	122 ± 8 cm
TD	15	5–10	7.3 ± 1.2	9M / 6F	124 ± 7 cm
Total	30	5–10	7.2 ± 1.4	19M / 11F	-

D. System Setup and Hardware Configuration

The effectiveness and reliability of the proposed data collection process rely heavily on the technical setup used to capture both visual and skeletal movement information. To ensure accurate full body tracking and high quality recordings, a carefully controlled hardware and software configuration was established before conducting the sessions. This setup was designed to provide consistent performance across all participants, minimize tracking errors, and maintain synchronized acquisition of RGB video and joint data throughout gameplay. The Kinect v2 sensor tracked 25 skeletal joints in a 3D coordinate system relative to the sensor and RGB video at 30 fps. The Kinect v2 sensor was used to capture both RGB video and full-body skeletal joint data in real time, allowing children to interact naturally with the augmented reality game. The sensor is capable of tracking up to 25 skeletal joints per player, which is essential to analyze movements of both upper and lower limbs. Table III summarizes the main specifications of the Kinect v2 sensor used in the system.

TABLE III. BASIC SPECIFICATIONS OF THE KINECT V2 SENSOR

Specification	**Value / Range**
Color camera	1920×1080 @ 30 FPS
Depth camera	512×424
Depth distance	0.5–8.0 m
Horizontal field of view	70°
Vertical field of view	60°
Full skeletons tracked	6
Skeleton joints defined	25 joints per player

The AR game was developed using the Unity game engine. Unity is a cross-platform development tool that finds wide applications in the gaming, simulation, education, medical, architecture, and interactive training systems. Complex processes, such as graphics rendering, physics simulation, and compilation, are managed by the engine, allowing game developers to work on game design and interaction logic. The environment and the game mechanics were developed with the Unity editor, whereas the tracking and data acquisition functionality was developed in C# with the help of the Microsoft Kinect v2 SDK. C# was selected because it has an object-oriented and multi-paradigm implementation that enables it to integrate effectively with Unity and the Kinect libraries. The last system logs synchronized RGB video and skeleton data, which creates a powerful platform for movement analysis and follow-up ML analysis.

E. Sensor Placement and Environment Setup

The sensor was installed in a controlled indoor setting that did not have significant changes in light and background disturbance factors. The sensor was mounted about 1 m in the air and tilted at a low angle to ensure that the full body of the participant, including the lower limbs and feet, which is very important in the study of movement, is visible to the sensor. The child was told to move 2 to 2.5 m from the sensor and placed in the center of the tracking space to have optimal depth detection and minimize the chances of having joint occlusion. The recording space was sufficiently spaced to allow the children to perform the necessary tasks without any difficulties, such as stepping forward, reaching, kicking, and turning around. The background colors were kept soft and neutral to avoid impacting the depth sensor, whereas there was a minimization of reflective surfaces, as well as direct sunlight; the latter was not wanted because depth noise and tracking instability would occur.

A display was placed in front of the participant, where real-time visualization of the AR elements could be seen on the live video feed. This arrangement was useful in keeping the children engaged and also ensured that they could clearly view the virtual objects with which they were directed to interact. The same Kinect sensor, display, and tracking space settings were used throughout the sessions to ensure consistency and data of similar quality. Table IV shows the setup parameters.

TABLE IV. SENSOR PLACEMENT AND ENVIRONMENT SETUP PARAMETERS

Parameter	Value / Description
Sensor height	~1.0 m
Distance to child	2.0 – 2.5 m
Sensor tilt angle	~15° downward
Tracking area width	~2.5 m
Room type	Indoor, controlled
Lighting	Stable, non-reflective
Display position	In front of the participant
Floor surface	Non-slippery, open space

Throughout each game, the system captured RGB video and skeletal joint data in real-time. The data streams were recorded with an identical frame rate of 30 fps to ensure temporal consistency between the visual data and the position of the body joints. Each frame was timed by the Kinect SDK, offering a high level of accuracy in terms of synchronization between the RGB frames and the 3D joint coordinates.

The recording process stored the RGB video as MP4 files, while the skeletal data was saved in structured formats containing the 3D coordinates of 25 joints for every frame. This structure enabled the accurate reconstruction of movement sequences and facilitated subsequent processing for ML analysis. Each recording session generated a synchronized pair of video and skeleton files, representing the full interaction of the participant with the AR tasks.

IV. DATASET DESCRIPTION

The data utilized in this study were specifically created to study ASD-based movement and classification, as there is no publicly available data that involves full-body motion capture, real-time interaction, and controlled motor tasks in an AR setting. All recording sessions were conducted individually, where a child interacting with the AR system and the Kinect v2 sensor was used to record the RGB video and skeletal joints in sync. Kinect sensor monitored 25 anatomical joints in 3D space, providing a detailed spatial and time representation of the movement of the participant during the session. The RGB recordings were used to record the visual context of the task performance, whereas the skeletal data presented a structured and noise-reduced representation of body movement that could be processed effectively by a computer.

To achieve diversity and representativeness, the dataset comprises the recording of both ASD-diagnosed and TD children. All participants performed the same set of AR tasks, making it possible to compare the movement peculiarities of the two groups directly. The recording protocol was standardized with respect to the environment, sensor placement, sequence of tasks, and flow of interaction, allowing consistency across the sessions and minimizing variation based on external factors and not on the behavior of the participants.

The participants were required to perform five different tasks in the AR game, oriented to a particular movement pattern, including reaching, lifting, stepping, kicking, or jumping. Such activities were chosen according to the suggestions of the occupational and behavioral therapy practices, which often implement the same movements to measure motor abilities and coordination in children with autism. Thus, the resulting dataset is not only useful in computational classification but also corresponds to clinically relevant movement assessment principles.

A. Data Modalities

The data used in this study were collected using two complementary data modalities, namely RGB video records and skeletal joint data obtained using the Microsoft Kinect v2 sensor. All modalities represent various parts of the motor performance of the children in the AR tasks, offering a more detailed description of movement patterns with reference to behavioral analysis in ASD. The RGB modality gives a visual representation of how individual members engage with the AR environment. These videos include body position, gesture performance, task interaction, and the surrounding environment. Computer vision researchers widely use RGB video because it is familiar and intuitively easy to use, and can also preserve natural appearance and motion information as part of ASD assessment.

As can be seen in Table V, there were 150 recorded sessions, and the number of videos per task was the same (30). This balanced representation ensures that all types of movements are equally represented so that one can make equal comparisons across tasks and thereby eliminate the risk of having a bias about certain motion patterns. The constant time in all recordings also promotes uniformity in the course of preprocessing and model training.

TABLE V. DATASET RECORDINGS PER AR TASK

Task ID	Task description	Number of videos	Average duration (s)
Task 1	Touch the ball	30	25
Task 2	Raise both hands	30	25
Task 3	Kick the object	30	25
Task 4	Side step movement	30	25
Task 5	Jumping action	30	25
Total		**150**	

The dataset, as shown in Table VI, is balanced in terms of ASD and TD children, with each group providing 75 recordings. Such a balanced distribution reduces the chances of classification bias and also ensures that the ML models do not overlearn patterns of one group at the expense of the other. The similarity between the number of frames in the two groups also facilitates equitable comparison of the two in training and assessment, which allows for evaluating more reliably the differences in movements relating to autism.

TABLE VI. PARTICIPANT AND RECORDING DISTRIBUTION ACROSS GROUPS

Group	Number of participants	Number of videos	Approx. frames
ASD	15	75	56,250
TD	15	75	56,250
Total	**30**	**150**	**112,500**

B. Data Labeling

Children diagnosed with ASD and TD were grouped separately. All participants with ASD were previously diagnosed by licensed experts in the field of ASD assessment. The diagnoses were made according to standardized clinical assessments that were performed before participation. In the case of TD participants, the inclusion criterion was the absence of developmental, neurological, or motor disorders, which was ensured by parental reports and school records. This served the purpose of having the TD group as a stable baseline and helped to eliminate the possibility of misclassification based on unreported developmental conditions.

Since each child performed all five AR tasks, all recordings related to any particular participant were labeled with the same label. This approach eliminated the fragmentation of the identity of participants between tasks and ensured the consistency of model training and testing. It also helped to split datasets in a participant-independent manner, where the data of one participant was confined to training, validation, or testing sets, preventing information leakage between subsets.

In order to keep the labeling intact, two verification steps were conducted. To begin with, names were given according to the documentation of the participants. Second, the dataset was audited to ensure that all recordings were properly stored and labeled with respect to the defined folder hierarchy and nomenclature. This validation guaranteed proper matching of recordings and clinical categories. Table VII indicates that labels were created at the participant level so that all records created by the same child had a similar label.

TABLE VII. LABEL ASSIGNMENT SUMMARY

Group	Label type	Number of participants	Number of recordings	Label source
ASD	Clinical diagnosis	15	75	Licensed specialist evaluation
TD	Non-clinical	15	75	Parental report & institutional records
Total		**30**	**150**	

C. Data Preprocessing

This section describes the preprocessing pipeline applied to the collected skeletal data prior to model training. Due to natural variations in task execution speed, tracking stability, and recording duration, raw skeletal sequences require several preprocessing steps to ensure consistency, reliability, and suitability for ML-based analysis. This pipeline was designed to preserve meaningful movement characteristics while eliminating noise, irrelevant variability, and subject-dependent differences.

1) Data Cleaning

The initial preprocessing stage involved the elimination of incomplete or poor-quality recordings that occurred due to loss of tracking, extreme joint occlusion, or participants who moved beyond the specification tracking region. Recordings with long gaps or unstable skeletal data were not used, and short gaps were rectified where possible. Non-informative frames were also cut out before the onset of the task or after the completion of the task, leaving behind only the active movement segments on the respective task.

2) Synchronization and Temporal Alignment

The data of skeletal joints and RGB video streams were recorded simultaneously when the game was played, but in different locations. Skeletal frames were synchronized to RGB frames based on depth sensor timestamp information to ensure that the bones did not move during that time interval. Buffering led to minor temporal discrepancies that were removed by a nearest-frame matching strategy.

3) Task Segmentation

The recording sessions were divided into tasks, and the main execution part was isolated. Joint motion magnitude was identified as a signal of task initiation and termination, with dominant joints (wrists, elbows, knees, and ankles) being targeted.

4) Skeleton Extraction

In each segmented sequence, 25 anatomical joint 3D coordinates were extracted at a rate of 30 fps. Frames whose joint tracking confidence was of low reliability were either interpolated or dropped. The resulting skeleton data was arranged as time sequences.

5) Spatial Normalization

Spatial normalization was applied to eliminate variations caused by participant height and distance from the sensor. Joint coordinates were translated so that the spinal base joint served as the origin and scaled relative to torso length.

6) Temporal Normalization

All sequences were resampled to a fixed length of 150 frames using linear interpolation to ensure consistent input length.

7) Final Data Representation

After preprocessing, each movement sequence was represented as a normalized tensor of shape $150 \times 25 \times 3$, labeled according to the clinical diagnosis.

V. BASELINE EXPERIMENTS

Baseline experiments were carried out to determine the applicability and discriminative ability of the proposed skeleton-based dataset. These experiments are not intended to present new classification models, but a set of reference performance results to support the appropriateness of the dataset to analyze movement-based ASD.

A. Experimental Task

The formulation of the baseline task is that of a binary classification problem where each skeletal movement sequence is either ASD or TD. A subject-independent evaluation protocol was adopted to ensure that the data of the same subject is not repeated in the training and testing sets.

B. Input Representation

The representation of each input sample is that of a normalized skeleton-tensor $150 \times 25 \times 3$, involving 150 timeframes, 25 skeletal joints, and 3 spatial coordinates (x, y, z). Full-body spatial arrangement and time-change dynamics are retained in this representation.

C. Baseline Models

Two baseline models were tested to give representative points of reference for both classical ML and DL methods:

- Random Forest (RF): A traditional ML classifier, trained on hand-made statistical features obtained based on the skeletal joint trajectories: the mean and standard deviation of joint displacements [11].
- A lightweight DL model that bypasses the widely-used convolutional layers by extracting spatial features but uses a Long Short-Term Memory (LSTM) layer to learn time-dependent interactions between sequences of movement data (CNN-LSTM) [12].

D. Evaluation Protocol and Metrics

Five-fold cross-validation was used on the participant level. In each sample, all data belonging to one participant were limited to one fold, avoiding leakage of subjects and providing a reasonable assessment of the generalization performance [13]. Accuracy, precision, recall, and F1-score were used to assess classification performance, providing a full evaluation on the ASD vs TD binary classification task [12, 14].

E. Results

Table VIII shows the results of the baseline classification averaged over all cross-validation folds. The findings show that the proposed dataset has discriminative movement patterns that can effectively be learnt by both models.

TABLE VIII. BASELINE RESULTS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Random Forest	78.4	77.9	79.1	78.5
CNN-LSTM	85.6	84.8	86.2	85.5

The CNN-LSTM model achieved higher performance across all metrics compared to the Random Forest classifier, highlighting the importance of modeling temporal dynamics in skeleton-based movement analysis. These baseline results confirm the suitability of the proposed dataset for ML and DL research. The baseline experiments demonstrate that movement sequences captured through the AR-based data collection framework contain meaningful patterns capable of distinguishing between ASD and TD participants. Although DL models benefit from temporal modeling, the competitive performance of the Random Forest classifier also indicates that the dataset supports traditional ML approaches. These results provide reliable reference benchmarks for future studies utilizing the proposed dataset.

VI. CONCLUSION

This paper presented a skeleton-based movement dataset that was recorded with children with ASD and TD peers using an AR serious game. The dataset addresses a key research gap, as full-body skeletal motion data related to ASD remain relatively scarce in the existing literature, and the interactions are focused not on observation or incomplete behavioral indicators. The purpose behind the development of the AR game was to stimulate clinically relevant motor behaviors that involve upper limbs, lower limbs, balance, and body rotation in an exciting and child-friendly setting, providing data quality and comfort to participants. The protocol and data collection pipeline were thoroughly described, as well as the cleaning of data, their division, spatial normalization, and time normalization, so that they have fixed-length skeleton sequences and could undergo analysis by an ML framework. In addition, baseline classification experiments were carried out to show the applicability and discriminative power of the dataset. The baseline results indicated that both classical ML and DL models can answer meaningful calculations of movement patterns, providing reference performance benchmarks rather than optimized classification outcomes.

DECLARATION OF COMPETING INTERESTS

The authors declare no conflicts of interest.

ACKNOWLEDGMENT

Not applicable to this work.

DATA AVAILABILITY

The dataset generated and analyzed during this study is not publicly available due to ethical and privacy constraints involving human participants, but is available from the corresponding author upon reasonable request.

REFERENCES

[1] M. Kangarani-Farahani, M. A. Malik, and J. G. Zwicker, "Motor Impairments in Children with Autism Spectrum Disorder: A Systematic Review and Meta-analysis," *Journal of Autism and Developmental Disorders*, vol. 54, no. 5, pp. 1977–1997, May 2024, https://doi.org/10.1007/s10803-023-05948-1.

[2] R. A. J. De Belen, T. Bednarz, A. Sowmya, and D. Del Favero, "Computer vision in autism spectrum disorder research: a systematic review of published studies from 2009 to 2019," *Translational Psychiatry*, vol. 10, no. 1, Sept. 2020, Art. no. 333, https://doi.org/10.1038/s41398-020-01015-w.

[3] B. Ren, M. Liu, R. Ding, and H. Liu, "A Survey on 3D Skeleton-Based Action Recognition Using Learning Method." arXiv, 2020, https://doi.org/10.48550/ARXIV.2002.05907.

[4] M. V. Toma, C. E. Turcu, C. O. Turcu, S. Vlad, D. E. Tiliute, and P. Pascu, "Extended Reality–Based Mobile App Solutions for the Therapy of Children With Autism Spectrum Disorders: Systematic Literature Review," *JMIR Serious Games*, vol. 12, Feb. 2024, Art. no. e49906, https://doi.org/10.2196/49906.

[5] T. Barami *et al.*, "Automated Analysis of Stereotypical Movements in Videos of Children With Autism Spectrum Disorder," *JAMA Network Open*, vol. 7, no. 9, Sept. 2024, Art. no. e2432851, https://doi.org/10.1001/jamanetworkopen.2024.32851.

[6] E. Billing *et al.*, "The DREAM Dataset: Supporting a data-driven study of autism spectrum disorder and robot enhanced therapy," *PLOS ONE*, vol. 15, no. 8, Aug. 2020, Art. no. e0236939, https://doi.org/10.1371/journal.pone.0236939.

[7] J. Li *et al.*, "MMASD: A Multimodal Dataset for Autism Intervention Analysis." arXiv, 2023, https://doi.org/10.48550/ARXIV.2306.08243.

[8] A. A. Al-Jubouri, I. Hadi, and Y. Rajihy, "Three dimensional dataset combining gait and full body movement of children with autism spectrum disorders collected by Kinect v2 camera." Dryad, Sept. 11, 2020, https://doi.org/10.5061/DRYAD.S7H44J150.

[9] A. A. Al-Jubouri, I. H. Ali, and Y. Rajihy, "Gait and Full Body Movement Dataset of Autistic Children Classified by Rough Set Classifier," *Journal of Physics: Conference Series*, vol. 1818, no. 1, Mar. 2021, Art. no. 012201, https://doi.org/10.1088/1742-6596/1818/1/012201.

[10] N. K. Zakaria, N. M. Tahir, and R. Jailani, "A Markerless-based Gait Analysis and Visualization Approach for ASD Children," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, 2021, https://doi.org/10.14569/IJACSA.2021.0120553.

[11] M. Alanazi, R. S. Aldahr, and M. Ilyas, "Human Activity Recognition through Smartphone Inertial Sensors with ML Approach," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12780–12787, Feb. 2024, https://doi.org/10.48084/etasr.6586.

[12] M. Hiari, Y. Alraba'nah, and I. Qaddara, "A Deep Learning-Based Intrusion Detection System using Refined LSTM for DoS Attack Detection," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 25627–25633, Aug. 2025, https://doi.org/10.48084/etasr.11499.

[13] H. A. Hatim, Z. A. A. Alyasseri, and N. Jamil, "A recent advances on autism spectrum disorders in diagnosing based on machine learning and deep learning," *Artificial Intelligence Review*, vol. 58, no. 10, July 2025, Art. no. 313, https://doi.org/10.1007/s10462-025-11302-x.

[14] I. Qaddara and Y. Alraba'nah, "Enhancing Requirements Classification Using Machine Learning Techniques," *SN Computer Science*, vol. 6, no. 6, July 2025, Art. no. 649, https://doi.org/10.1007/s42979-025-04158-z.