Received: 11 December 2023 | Revised: 25 March 2024 | Accepted: 7 April 2024

DOI: 10.1111/jcal.12990

ARTICLE

Journal of Computer Assisted Learning WILEY

RoboCA^{3}T: A Robot-Inspired Computer-Assisted adaptive autism therapy for improving joint attention and imitation skills through learning and computing innovations

Zunera Zahid[1] | Sara Ali[1,2,3] | Shehriyar Shariq[1] | Yasar Ayaz[1,4] | Noman Naseer[5] | Irum Yaseen[6]

[1]School of Mechanical and Manufacturing Engineering (SMME), National University of Sciences and Technology (NUST), Islamabad, Pakistan

[2]Intelligent Field Robotics Lab (IFRL), National Center of Artificial Intelligence (NCAI), National University of Sciences and Technology H-12, Islamabad, Pakistan

[3]NUST-COVENTRY Human Robot Interaction Lab, Department of Robotics and Artificial Intelligence (R&AI), School of Mechanical and Manufacturing Engineering (SMME), National University of Sciences and Technology H-12, Islamabad, Pakistan

[4]National Center of Artificial Intelligence (NCAI), National University of Sciences and Technology H-12, Islamabad, Pakistan

[5]Department of Mechatronics, Air University (AU), Islamabad, Pakistan

[6]Primal Support, E11, Islamabad, Pakistan

Correspondence
Sara Ali, NUST-COVENTRY Human Robot Interaction Lab, Department of Robotics and Artificial Intelligence (R&AI), School of Mechanical and Manufacturing Engineering (SMME), National University of Sciences and Technology (NUST), H-12, Islamabad 44000, Pakistan.
Email: sarababer@smme.nust.edu.pk

Abstract

Background: This study presents a Robot-Inspired Computer-Assisted Adaptive Autism Therapy (RoboCA^{3}T) focusing on improving joint attention and imitation skills of children with autism spectrum disorder (ASD). By harnessing the inherent affinity of children with ASD for robots and technology, RoboCA^{3}T offers a therapeutic environment designed to maximise engagement and facilitate effective skill development. It harnesses the advantages of Robot-Assisted Therapies (RATs) by employing robot avatars and integrating them with Computer-Assisted Therapies (CATs) within a web-based solution. The integration of automatic gaze and pose detection algorithms within RoboCA^{3}T addresses the challenge posed by potential human error and observation bias in assessing the child's progress, thereby ensuring accurate results. This research responds to the need for more effective, technology driven therapies for autism, filling gaps in existing methods.

Objectives: The primary goal of this research is to create a robot inspired computer assisted adaptive autism therapy that maximises engagement and enhances joint attention and imitation skills.

Methods: The study involves 11 ASD children with 30 sessions (divided into two halves) per module over eight months, comprising 660 experimental trials, 110 familiarizations, and 110 follow-up sessions. The joint attention module evaluates the subject's gaze pattern using WebGazer for gaze detection in response to four least-to-most robot-generated cues. The imitation module utilises robot-generated pose for comparing subjects' imitated actions using Tensorflow Lite for pose estimation.

Results and Conclusions: The effectiveness of therapy was substantiated by comparing Childhood Autism Rating Scale (CARS) scores before and after intervention. Significant improvements were noted between the first and second therapy halves, validated by Wilcoxon signed-rank tests ($p < 0.01$) and spearman's correlation analysis, reinforcing the observed improvements in joint attention and imitation skills.

KEYWORDS
Artificial Intelligence (AI), Computer Assisted Therapy (CAT), imitation, Joint Attention (JA), Robot Assisted Therapy (RAT), web-based therapy

1 | INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that manifests as impaired communication and social skills, unusual responses to sensory information, lack of attentiveness, and restricted repetitive sensory behaviour (American Psychiatric Association, 2013). The core impairments and early indicators of the disorder include impaired joint attention (JA) and imitation (Ingersoll, 2008; Stone et al., 1997; Winoto & Tang, 2019). Impaired JA refers to diminished tendency to establish visual coordination therefore displaying challenges in engaging in JA behaviours commonly observed in Typical Development (TD) child (Ames & Fletcher-Watson, 2010; Winoto & Tang, 2019). Imitation involves observing and replicating verbal or non-verbal actions naturally with varying accuracy. Individuals with ASD often face difficulties in imitating social cues and gestures, displaying delayed, atypical behaviours, and a limited repertoire(American Psychiatric Association, 2022; Barnhill, 2023; Grzadzinski et al., 2013; Ingersoll, 2008; Stone et al., 1997). These core deficits can have a negative impact on their social interaction and communication, thereby negatively impacting their development (Ali et al., 2022; David et al., 2020; Pop et al., 2013; Winoto & Tang, 2019).

Various treatment approaches have been developed for improving the underlying symptoms in children with ASD including pharmacological medicines and behavioural therapy (Aishworiya et al., 2022; Coury et al., 2012; Lovaas, 1987; Mannion & Leader, 2013; Odom et al., 2010; Rodrigues et al., 2021). Behavioural treatments have surpassed pharmacological therapies in treating the key symptoms of ASD (Alshawwa et al., 2023; Matson & Burns, 2019; Roane et al., 2016). These behavioural therapies include Applied Behaviour Analysis (ABA) therapy, Social Skills Training (SST), and Picture Exchange Communication System (PECS), which effectively improves social interaction and communication skills in children with ASD (Foxx, 2008). However, challenges such as limited access, inflexibility, training constraints, or high costs ($40,000 to $60,000 /individual/year) can hinder the effectiveness of these interventions (Bauer et al., 2022). Additionally, increasing prevalence rate of ASD estimated at about 1%–2% has surpassed availability of behavioural therapists resulting in waitlists of up to 18 months in the United States (Haber et al., 2020).

To address this challenge, technology-based treatments i.e., Computer-Assisted Therapies (CAT) and Robot-Assisted Therapies (RAT) are becoming increasingly popular for enhancing the abilities and quality of life (QoL) for individuals with ASD (Aresti-Bartolome & Garcia-Zapirain, 2014; Bauer et al., 2022; Boucenna et al., 2014; Chen et al., 2022; Eden & Oren, 2021; Nesterova et al., 2015; Rajendran, 2013). RAT integrates robotics, artificial intelligence, and machine learning to further enhance the well-being and autonomy of people with disabilities. RAT have emerged as an effective method for the development of crucial skills, such as JA and imitation in children with ASD cases (David et al., 2020; Pennisi et al., 2016; Richardson et al., 2019; Tleubayev et al., 2019). They offer a regulated, convenient, and predictable environment for learning, which can increase the learning motivation of children with autism (David et al., 2020; Grynszpan et al., 2014; Kapp, 2012). Children with ASD demonstrate notable levels of interest, and eye contact towards a robot's face during RAT (Ali et al., 2019; Alshammari et al., 2022; Marino et al., 2020; Martins et al., 2020; Salhi et al., 2022; Sigacheva et al., 2020). However, high cost of robots, availability of resources for intervention, time for training, and requirement for a developer's presence during intervention sessions are some common impeding factors that constrain RAT widespread adoption (Corcione et al., 2005). Rapid technological advancements may also render robots outdated quickly. Furthermore, emotional attachment to robots is a potential downside, which may cause distress when children recognises that robots are machines that can fail or become outdated (Ellery et al., 2008; Guntur et al., 2019; Wali & Sanfilippo, 2019).

Given the limitations of RAT, researchers have been prompted to explore alternative options, including CAT resulting in an increase in the research and development of such applications catering specifically to people with ASD (Den Bossche et al., 2019; Deng & Technology, 2022; Syriopoulou-Delli C, of EGIJ, 2022; Tsikinas & Xinogalos, 2019; Wali, 2019). CAT encompasses a range of technologies, including interactive websites, online learning modules, social skills training programmes, communication tools, and virtual therapy sessions. and are designed to create a safer, accessible, cost-effective, and repetitive environment to support individuals with ASD (Durkin, 2010; Bauer et al., 2022; Eden & Oren, 2021; Grynszpan et al., 2014; Winoto & Tang, 2019).They provide safe, secure, accessible, cost-effective, and repetitive environment for individuals with ASD to practice and develop various skills, including communication, social skills, and activities of daily living (ADL) (Wojciechowski & Al-Musawi, 2017), (Grynszpan et al., 2014). They also facilitate connections with others and enable individuals to track their progress over time (Ramos Aguiar et al., 2023; Winoto & Tang, 2019).

CAT based therapies are being developed in several areas including social skills training, cognitive behavioural therapy, sensory integration, educational, and behaviour management applications. Social skills training applications aim to provide a comprehensive communication system for individuals with autism. Examples include Autism iHelp-Play (Shrestha et al., 2021), Otismo, Proloquo2Go, Social Express, MyTalkTools Speech Assistant AAC, Card Talk, LetMeTalk Speech Therapy Articulation, ABtalk, NeuroGen Speech Play, 1 on 1: Communicate Easy, Social Skills for Autism, Autism Speech Diego-Says, Talking Pictures: Autism, CP, and CommBoards Lite (del PG et al., 2021). Cognitive Behavioural Therapy (CBT) Apps including MindShift CBT (Paul & Eubanks Fleming, 2019), MITA (Vyshedskiy & Dunn, 2015) are designed to help individuals with ASD manage anxiety and stress. Sensory integration applications, such as EduSense-AR (Sharma et al., 2022) and Autism & SPD and Sensory Timers & Tools (Sher, 2009) offer various activities designed to regulate sensory system specifically for individuals with ASD. Educational Apps including (Wojciechowski & Al-Musawi, 2017) and Knowledgemon hunter (Silva et al., 2017) help children with autism learn basic academic skills through interactive games. Behaviour management applications like Autism Tracker Pro, Autism & Me Behaviour Tracker, and AutiPlanner

(Inoue, 2019) help individuals with autism manage and track behaviours using data collection. Medical applications for ASD serve the purpose of assisting health care providers and therapists in identifying potential symptoms and facilitating further evaluation (Liu & Ma, 2022).

However, these applications have several limitations such as inability to deliver personalised content, track progress over time, and adapt to individual needs for targeted interventions. Several studies rely solely on researchers or parental observations as measurement methods, overlooking the potential of AI-based algorithms to analyse learners' responses and implement adaptive interventions. This introduces subjectivity and human error, potentially leading to less reliable results. Additionally, there is a noticeable lack of applications capable of predicting skill development in individuals with ASD (Rehman et al., 2021). Moreover, human interaction plays a vital role in social and emotional development for children with ASD. Existing applications fail to provide the necessary human interaction even as mediator, which is essential for effective therapy and lack rigorous validation procedures to ensure accuracy and reliability. Notably, none of the existing CAT solutions utilise robots as prompting agents. Our study investigates a novel perspective by capitalising on the affinity of children with ASD for robots while also incorporating humans as mediators, delivering a cost-effective adaptive solution suited to the learning, development, and therapy needs of children with ASD.

1.1 | Contributions

The aim of this study is to develop a Robot-Inspired Computer-Assisted Adaptive Autism Therapy (RoboCA^{3}T) focused on improving JA and imitation skills. To achieve this, the project combines the strengths of RATs through the utilisation of robot avatars and CATs by implementing low-cost web-assisted therapy. The integration of automatic gaze and pose detection algorithms within RoboCA^{3}T eliminates potential human error and observation bias in assessing the child's progress. This automation ensures objective and accurate results, enhancing the reliability of therapy outcomes. Additionally, by harnessing the inherent affinity of children with ASD for robots and technology, RoboCA^{3}T creates a unique therapeutic environment that maximises engagement and facilitates effective skill development.

1.2 | Architecture

RoboCA^{3}T utilises Next.js, React.js, and Tailwind CSS as front-end technologies to provide an intuitive and interactive user experience (Madurapperuma et al., 2022). On the back end, RoboCA^{3}T leverages Firebase authentication, Cloud Firestore, and storage. The database would allow us to monitor the effectiveness of the therapy and adjust the activities and exercises based on the child's progress. The front-end and back-end communicate through the Firebase Client-Side SDK. The architecture is shown in Figure 1.

RoboCA^{3}T utilises WebGazer for accurate real-time gaze detection (Papoutsaki et al., 2016) and Tensorflow Lite for precise pose estimation (Goldsborough, 2016). The prompting agent for both modules that is, joint attention and imitation is a robot avatar to replicate the movements and interaction that can be performed by a real robot. Both joint attention and imitation modules are based on the principles of Applied Behaviour Analysis (ABA) which utilises prompting, modelling, and reinforcement mechanism (Foxx, 2008). They theoretically

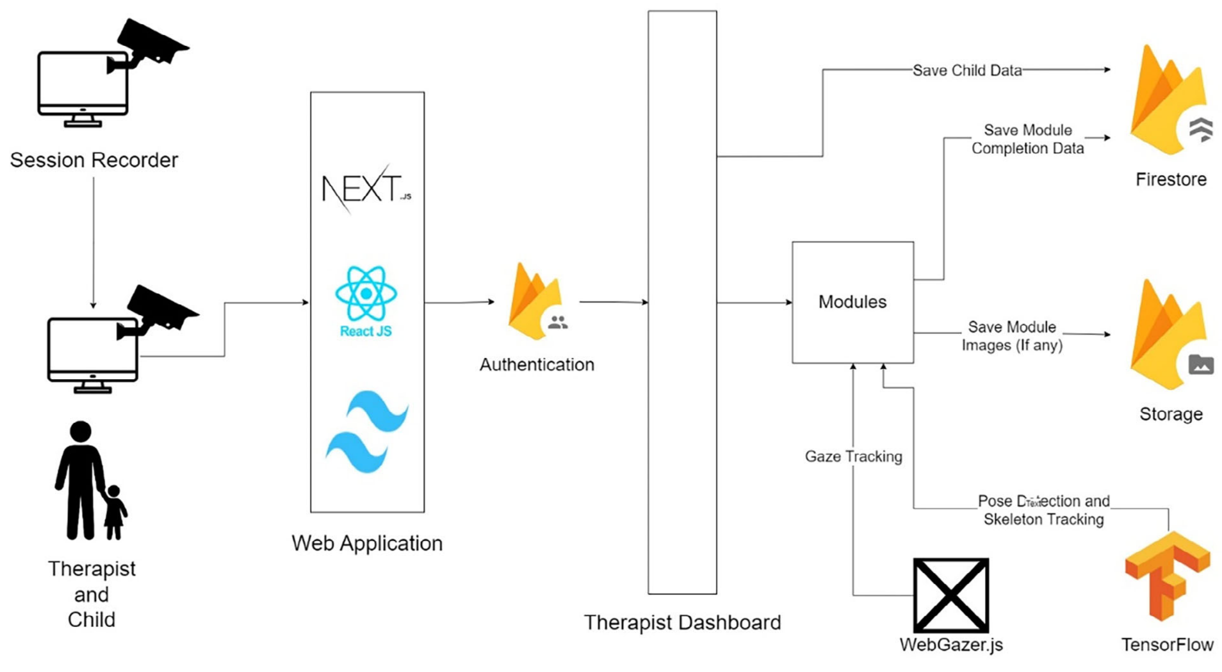

FIGURE 1 Diagram depicting RoboCA^{3}T architecture including backend and front-end services.

employ prompting using audio, visual and motor cues given by robot as stimulus. Modelling is implemented using robot as well as input from the therapist as mediator, and reinforcement mechanisms includes repeating cues and rewards.

1.3 | Adaptability and design inclusivity for children with ASD

RoboCA^{3}T prioritises inclusivity, by addressing the unique needs of children with autism spectrum disorder (ASD). Through consistency in design and interaction, the platform offers a uniform user experience. We prioritise clarity by using plain language, avoiding complex terms that might confuse users. Accessibility and usability are enhanced by verifying user-friendly attributes with tools like the WAVE Web Accessibility Evaluation Tool (Alsaeedi, 2020) and adhering to WCAG AA and AAA guidelines for contrast ratio and text size (Kumar et al., 2021). In line with the AutismGuide (Aguiar et al., 2022) principles, we strategically used colours to enhance visual cues, ensuring accessibility for users with ASD. This includes establishing a unified visual language using Tailwind CSS for styling and touch-friendly elements, while React components facilitate touch gestures, catering to varying motor skills. Progress indicators and delays between activities promote smoother transitions, aiding users in navigating through modules comfortably.

Personalised content delivery is another core feature that prioritises essential features and dynamically loads content tailored to each child's specific autism level. Additionally, it incorporates comprehensive visual, auditory, and motor support, including Robot images, videos, and audio cues, all aimed at enhancing understanding and engagement. Sensory sensitivities are addressed through adjustable sensory inputs, such as sound volume, brightness, and auditory cue pitch adjustment, including a Pakistani English accent, catering to individual preferences and sensitivities. To reduce cognitive load, tasks and activities are simplified and broken down into manageable subtasks, minimising distractions.

Additionally, human therapists act as mediators, providing valuable feedback and assisting ASD children in engaging with RoboCA^{3}T effectively. Integrated rewards provide incentives and positive reinforcement, encouraging ASD children to complete tasks and meet their goals, thereby making the learning process more enjoyable. Progress tracking is provided for therapists, parents, and caregivers, enabling them to monitor the child's development and achievements within the platform, facilitating assessment and support.

1.4 | Data privacy, security and robustness measures in RoboCA^{3}T

Before starting our study, explicit user consent was obtained. This involved presenting users with clear and concise information regarding the types of data collected, the purposes of data processing, and any third-party data sharing practices. Users were required to provide informed consent before proceeding with data collection, ensuring transparency and respect for user privacy preferences.

Moreover, RoboCA^{3}T follows data protection regulations, GDPR (General Data Protection Regulation). We adhere to strict data handling practices and ensure that user data is stored and processed securely in compliance with these regulations. A screenshot of GDPR policy page is shown in Figure 2.

RoboCA^{3}T ensures encryption in transit and at rest. Firebase services encrypt data during transmission using HTTPS (Hypertext Transfer Protocol Secure), ensuring that data exchanged between the Firebase servers and RoboCA^{3}T is protected from interception by unauthorised parties. Cloud Storage encrypts user data at rest using AES-256 encryption algorithm to prevent unauthorised access to stored information. Moreover, RoboCA^{3}T employs anonymization techniques to anonymize or pseudonymize personally identifiable information (PII) by assigning unique child IDs (cid) to subjects and removing identifiable information from stored data. For example, a child is saved with cid"1jSXCKxEGRup7qNgDIsB". This prevents the identification of individuals while preserving data integrity. The joint attention module extract gaze points from image data and stores x and y coordinates to prevent identification of individuals. Moreover, anonymization of the persons involved is ensured within the imitation module by saving skeletal points plotted against the black background as shown in Figure 3.

RoboCA^{3}T implements strict privacy protocols to protect user data. Authentication mechanisms enforced by Firebase restrict data access to authorised users only, reducing the likelihood of unauthorised data exposure or misuse as shown in Figure 4. Access controls further limit access to sensitive data. Moreover, comprehensive monitoring and auditing of access activities is also enabled to ensure accountability and traceability of data access, allowing for detection and investigation of any unauthorised access incidents.

Our system features comprehensive error handling mechanisms, including error logging, delayed responses between levels, and adjusted threshold values to accommodate non-standard movements or gaze patterns. In cases where the subject is not detected, an error message is promptly displayed on the screen, and the session is automatically paused, ensuring smooth interactions even in unpredictable scenarios. These error messages are implemented using interactive React JS components for action confirmation and cancellation. We have ensured consistent and informative error messages styled with Tailwind CSS for enhanced user experience. Additionally, our system employs understandable error codes to facilitate troubleshooting and diagnosis.

1.5 | Architecture for adaptive joint attention (JA) module

JA module consists of several modules that work together to achieve the desired functionality as illustrated in Figure 5. The modules below further describe Figure 5 in detail.

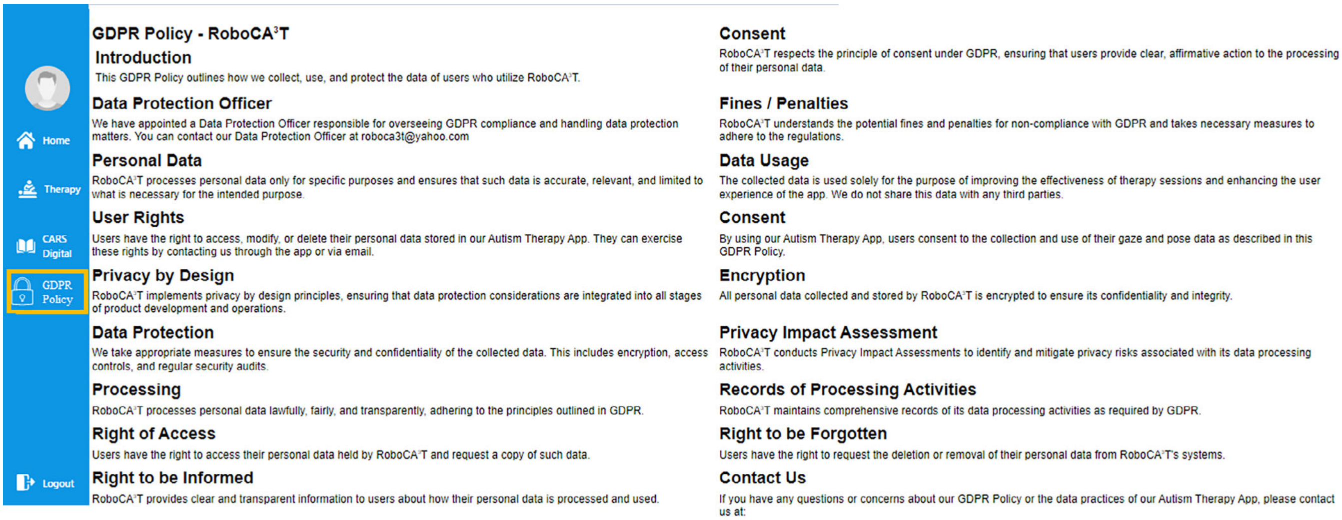

FIGURE 2 Diagram depicting RoboCA^{3}T General Data Protection Regulation (GDPR) Policy Page.

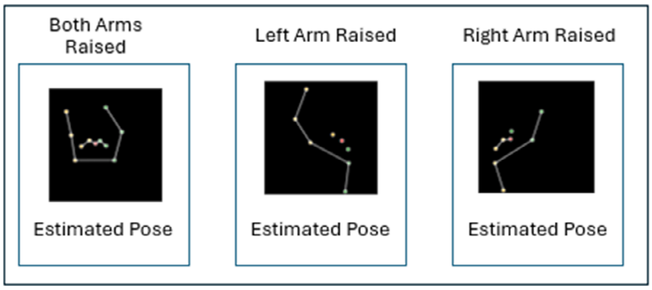

FIGURE 3 Skeletal points stored within imitation module to ensure subject anonymization.

1.5.1 | Selection of algorithm for eye tracking

JA module utilises eye tracking data and calculates joint attention score based on subject's response. When selecting eye tracking tools, various options were evaluated. OWLET (Werchan et al., 2023), while offering higher accuracy, yet it has limitations such as fixed four-point calibration, restricted processing capabilities (limited to 16:9 webcam videos at 30 frames per second or higher), and a lack of real-time gaze location prediction. Deep learning-based tools (Valliappan et al., 2020) were also explored but faced challenges as they were not openly available and may not generalise well to children due to their training data consisting solely of adults. Another deep learning-based tool, iCatcher+ (Erel et al., 2023), demonstrated high accuracy but lacked flexibility in calibration and real-time coordinate inference. Conversely, commercial platforms (Bánki et al., 2022; Finger et al., 2017; Heck et al., 2023; Lewandowska, 2019), offered convenience but lacked validation, were expensive, and lacked the transparency of open-source solutions. In contrast, WebGazer (Papoutsaki et al., 2016) emerged as a favourable choice—an open-source JavaScript-based framework designed for real-time eye movement monitoring. It features, including real-time tracking, compatibility with various devices, flexibility in setup, webcam formats, and calibration, alongside the benefits of open-source code, such as transparency, flexibility, and cost-effectiveness (Papoutsaki et al., 2016; Slim & Hartsuiker, 2022), make it an ideal fit for our research needs. While WebGazer may not be suitable for fine-grained spatial resolution, its spatial resolution is accurate enough to discriminate fixations across the quadrants of the screen (Steffan et al., 2024).

To address this limitation the Joint attention module was designed to capture fixation in response to images shown in two quadrants of the screen. WebGazer consists of the front-end code, machine learning models, and the backend server. The front-end code captures video streams from the user's webcam and processes them to track the position of the user's eyes. The machine learning models are trained on large datasets of eye-tracking data utilising deep learning algorithms to perform accurate eye-tracking. These models predict the position of the user's eyes based on the video stream. The back-end server of WebGazer is responsible for storing and retrieving eye-tracking data for individual users.

1.5.2 | Calibration module

The JA module starts with calibration module as shown in Figure 6.

The calibration interface of WebGazer was redesigned and robot head marker was used for calibration. The robot head marker appears nine times on screen, with users performing eye calibration via 4 clicks per position within a 3 × 3 grid as shown in Figure 6. Further refinement occurs as users gaze at the screen center for 5 s, extracting image data to map stimulus location with gaze direction for correctness measurement. Calibration gestures were added to incorporate both touch and click actions. A new feature of setting a threshold level of 80% before proceeding to Real time Tracking and Evaluation Module was set to ensure reliable eye tracking throughout the session within the calibration process. Pseudocode for calibration phase is below:

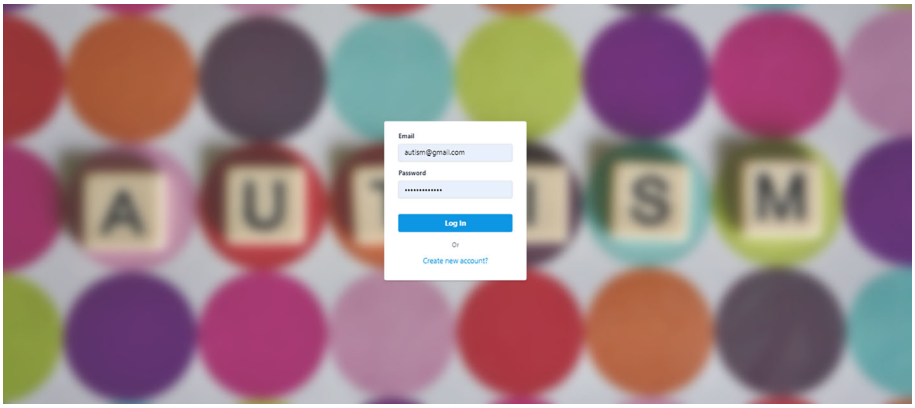

FIGURE 4 Diagram depicting RoboCA^{3}T Login Page, where users are prompted to enter their username and password for authentication.

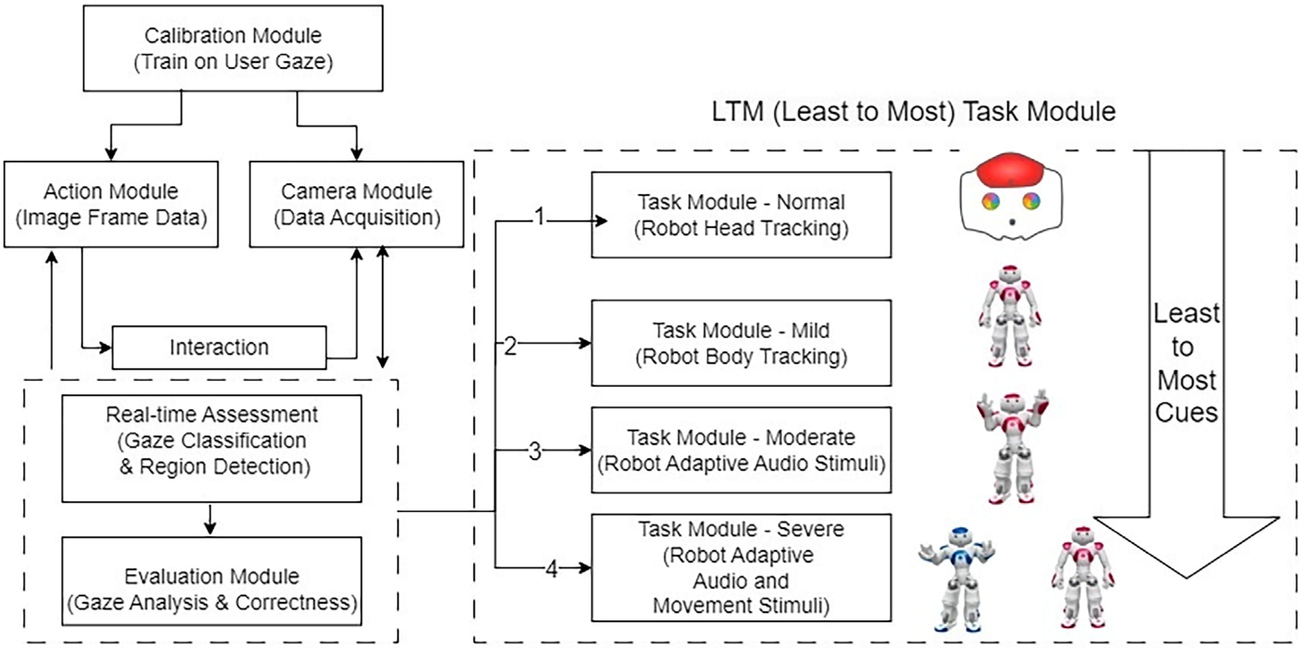

FIGURE 5 Diagram depicting joint attention module architecture.

WHILE camera == True DO.
IF child_detected() == True THEN.
IF eye_contours_detected() == True THEN.
ShowFirstCalibrationPoint().
FOR point IN CALIBRATION_POINTS DO.
IF point clicked 4 times THEN.
ShowNextCalibrationPoint().
END IF.
END FOR.
CalculateCalibrationAccuracy().
IF accuracy > = 80% THEN.
ShowTask().
ELSE.
Recalibrate().
END IF.
END IF.
END IF.
END WHILE.

1.5.3 | Real time tracking and evaluation module

This module displays the prompting agent five times across two screen regions. It tracks, records, and assesses gaze accuracy in comparison to the agent's location, classifying responses as correct if they match. The module concludes therapy for JA if over 75% of tracked points align with the prompting agent's region, displaying the score. If not, the next LTM-based task is loaded. The equations for calculating the total score S_t can be expressed as:

$$S_t = \Sigma_{n=1}^{4} \left(n.[T_n \geq 0.75 \times 5]\right) + 4 \cdot [T_4 < 0.75 \times 5] \quad (1)$$

where T_n is the number of correct responses in task n. n takes values from 1 to 4. [] is the inversion bracket notation, which evaluates to

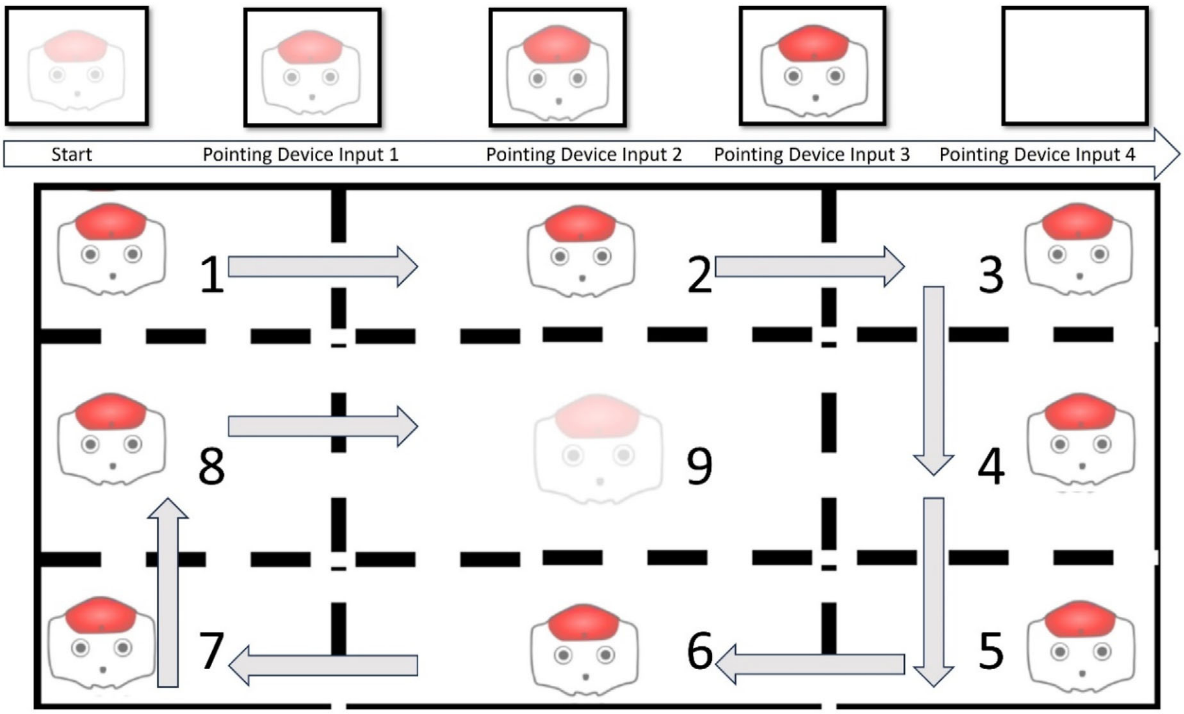

FIGURE 6 Diagram depicting marker placement sequence in 3 × 3 grid across screen and eye calibration with gaze and pointing device input within joint attention module.

1 if the condition is true, and 0 if it's false. T_4 is the number of correct responses in task 4. Figure 7 indicates the flow of activities within the Joint Attention module.

1.5.4 | Adaptive least to Most (LTM) module

Tasks within JA module vary in stimuli and prompting agents. Task 1 employs minimal stimuli (visual), using a robot head as prompting agent with rotating eyes to attract the child's gaze. Task 2 involves visual and motor cues with a full-body robot as prompting agent. Task 3 increases stimuli, utilising a full-body robot as prompting agent, and introduces adaptive auditory cues along with visual, and motor cues to guide the child's gaze. Lastly, task 4 distinguishes itself with multiple prompting agents (two full-body robots), each with a distinct appearance (red and blue), employing a combination of adaptive auditory, visual (colour), and motor cues to direct the child's gaze as both robots move across the screen. Auditory cues are tailored for adaptability, greeting the child by name. Furthermore, the module's adaptability extends to task selection, dynamically loading tasks based on the child's autism level using application data. For existing users, scores are determined by their performance in the previous therapy session.

The pseudocode for real time tracking and evaluation module is given below:

```
WHILE camera == True DO.
IF child_detected() == True THEN.
WHILE stimuli_count <= MAX_STIMULI DO.
ShowStimuli().
WHILE is_tracking DO.
IF gaze inside allowed_region THEN.
region_count = region_count +1.
END IF.
IF region_count/total_points_tracked <0.75 THEN.
IncreaseScore().
END IF.
END WHILE.
END IF.
END WHILE.
```

1.6 | Architecture for real time imitation module

The imitation module combines advanced machine learning algorithms, computer vision, and auditory and motor stimuli to enhance the imitation skills of children with autism spectrum disorder. The pseudocode for imitation module is given below:

```
WHILE camera == True DO.
IF child_detected() == True THEN.
IF post_detected() == True THEN.
WHILE stimuli_count <= MAX_STIMULI DO.
ShowStimuli().
IF is_child_imitating == True THEN.
Increase Score().
END IF.
END WHILE.
```

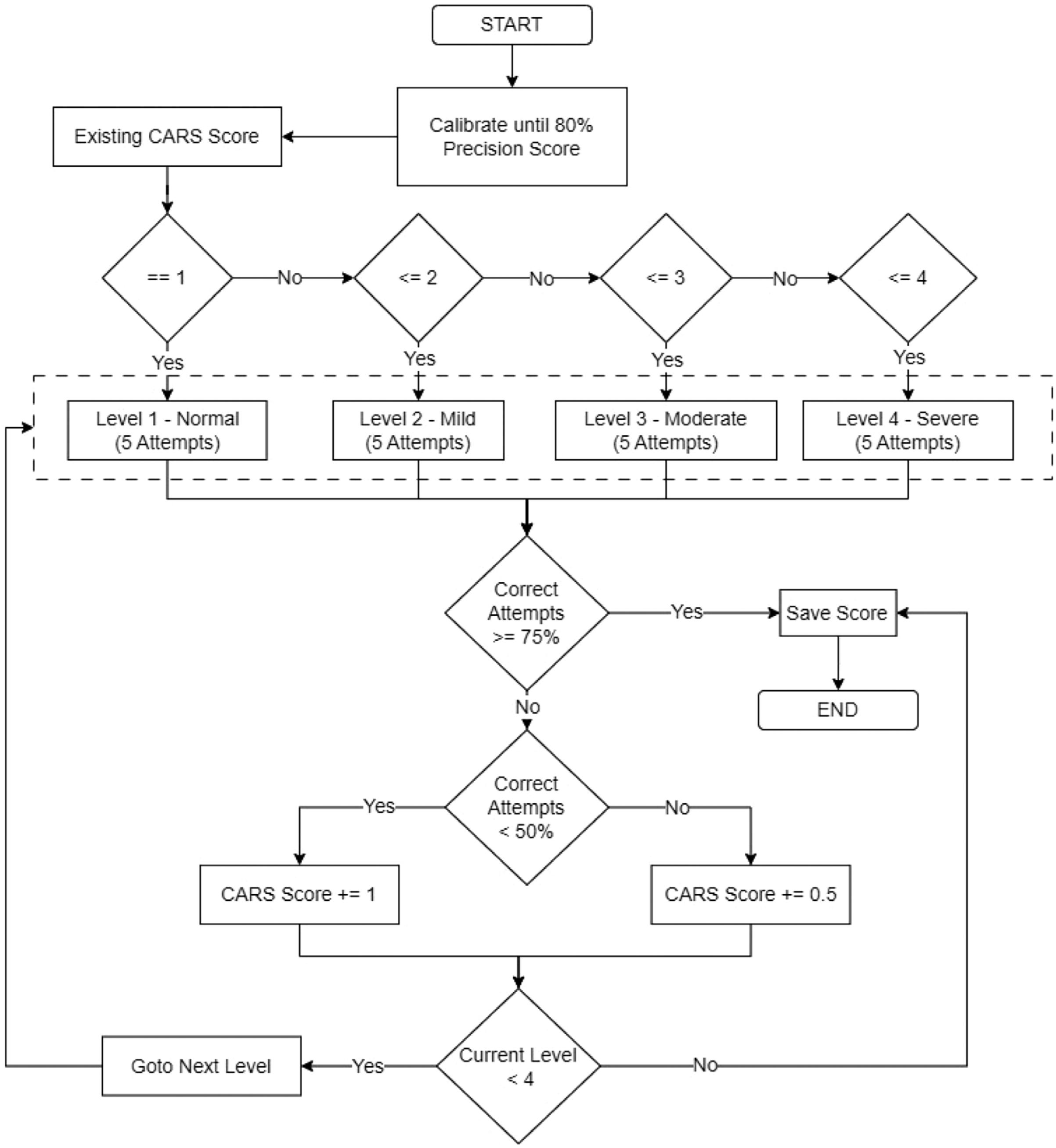

FIGURE 7 Flow of activities within joint attention module.

SaveScore().
EndModule().
END IF.
END IF.
END WHILE.

1.6.1 | Selection of algorithm for pose estimation

The imitation module requires a Pose Estimation (PE) algorithm to capture the child's body movements and joint points in real-time video (Liu et al., 2015). Various Image Processing (IP) based PE algorithms, including those utilising hand-crafted features and Deep Learning (DL) based approaches, were considered. However, hand-crafted feature methods often lack robustness (Lotti et al., 2023; Sharma et al., 2018). Despite the effectiveness of deep-learning-based PE methods to address this limitation, their computational intensity and long latency (Sun et al., 2019; Zhang et al., 2022) pose challenges for real-time interactions required by RoboCA^{3}T. Tensorflow Lite (TFLite) emerged as the optimal choice for RoboCA^{3}T due to its low latency, accelerated inference, and lightweight cross-platform compatibility, addressing the platform's requirements for reduced dimensionality

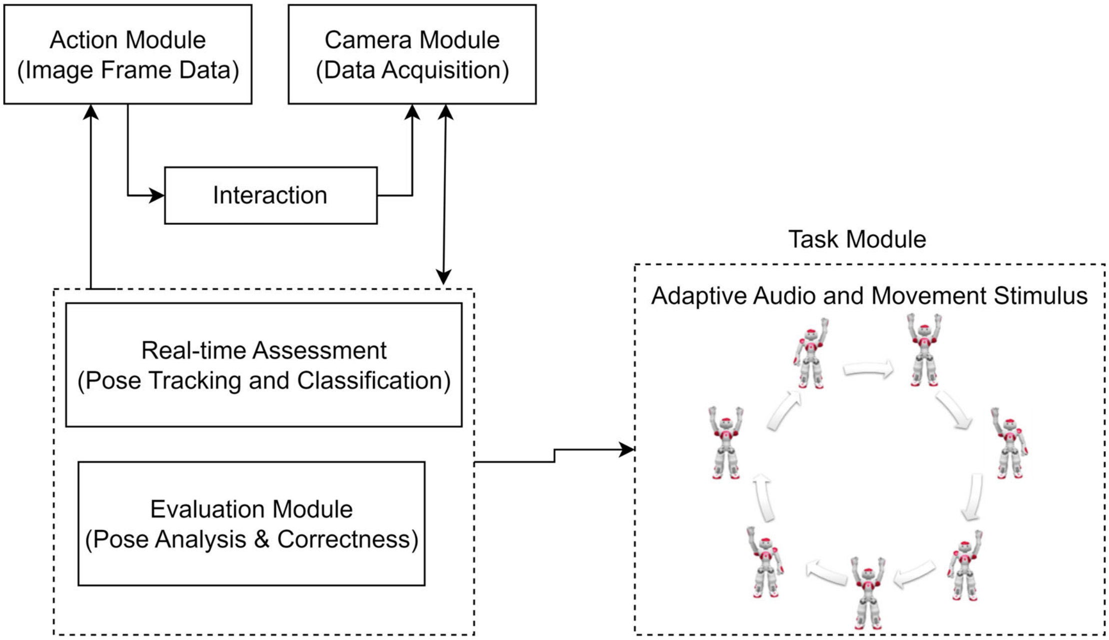

FIGURE 8 Diagram depicting imitation module architecture.

and anonymity of individuals involved (Goldsborough, 2016; Zhang et al., 2020).

1.6.2 | Action and camera module

Within imitation module, the action module presents gestures to the child and records the image frame data. This module is responsible for presenting the imitation poses to the child and recording the child's movements through a camera. The image frame data is then fed to the camera module for further processing. The camera module is responsible for acquiring relevant data from the camera. It captures the child's movements during the imitation task and provides the relevant data to other modules.

1.6.3 | Real time assessment module

Real time assessment module is designed to evaluate the child's imitation skills in real-time. This module includes pose classification, pose detection, and regions detection. It is responsible for detecting and classifying the child's body posture and joint movements according to the targeted pose within the camera region.

1.6.4 | Task module

The task module works on task determination, dynamic pose tracking, adaptive auditory stimuli, and motor stimuli. Task Determination decides which imitation task the child should perform. Dynamic pose tracking tracks the child's movements. Adaptive auditory stimuli provide the child with child specific auditory instructions for imitating the presented pose. Motor Stimuli provides the child with visual feedback, reinforcing the correct imitation movements. The prompting agent is a single robot that presents multiple poses for the child to imitate. It accurately and precisely demonstrates seven robot gestures, with three being unique and four repeated, as depicted in Figure 8.

1.6.5 | Evaluation module

The evaluation module is responsible for conducting an initial assessment of the child's imitation skills. This module involves engagement, pose analysis, and correctness measurement. Engagement component is designed to keep the child interested and motivated throughout the imitation task. Pose analysis is responsible for analysing the child's body posture and joint movements before and during the task. Additionally, a feedback Module is designed to provide immediate feedback to the child and therapist using a small overlay window that shows the child real time video and highlights joints and poses, enabling them to adjust and correct the child's movement in real-time as shown in Figure 9.

Correctness measurement assesses the accuracy of the child's imitation performance. The child is expected to imitate Robot pose, and the accuracy of imitation is measured using the laptop's camera. A pre-trained pose estimation model using TFLite was deployed. This model will detect key points on the subject, including the positions of the shoulder, elbow, and wrist. Next, the positions of the shoulder, elbow, and wrist key points are extracted from the output of the pose estimation model. The next step involves calculating the vertical distance between the wrist and the shoulder key points. The threshold

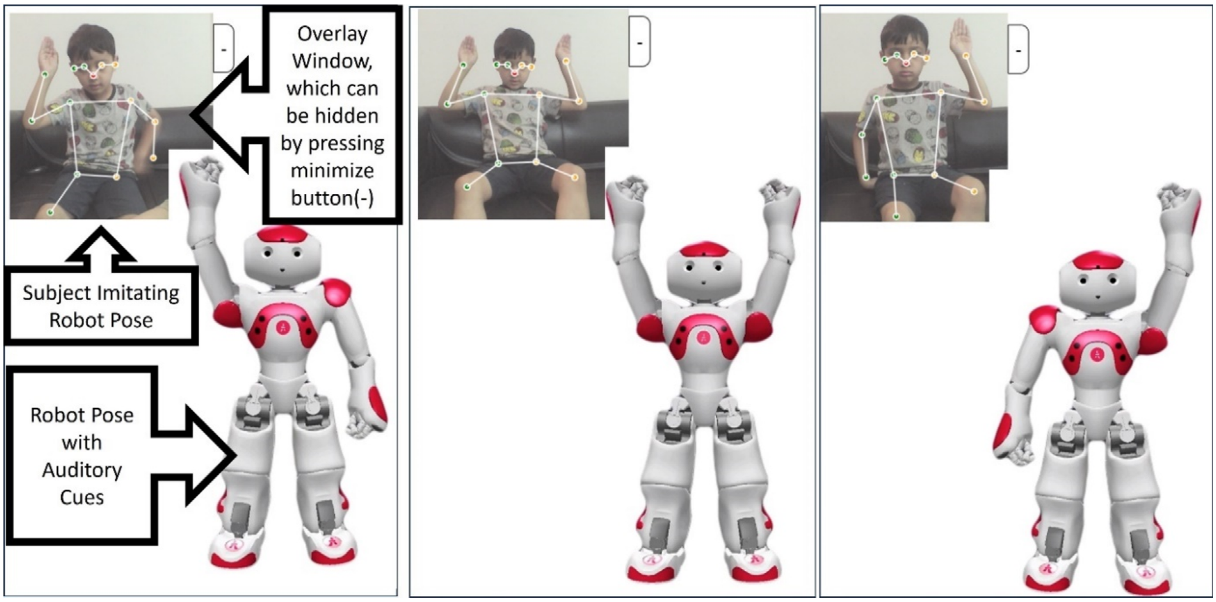

FIGURE 9 Diagram showing overlay window, pose detection and robot pose.

value for wrist's y-coordinate was set to above 40% relative to the shoulder's y-coordinate indicating a raised arm. If the wrist height conditions are met, it is inferred that the arm is raised. The system compared a child's pose with robot's model, adjusting scoring based on incorrect poses to cater to all severity levels.

Ensure that the.

The scoring system is based on the number of incorrect poses imitated by the child (C). If the child gets 6 or 7 poses wrong (C ≥ 6), they receive a score of 4, indicating poor imitation performance. The score decreases by 0.5 for each additional incorrect pose imitated, starting from 3.5 for 5 incorrect poses (C = 5) to 1 for no incorrect poses (C = 0). If the child correctly imitates all 7 poses (C = 0), they receive a score of 1, indicating excellent imitation performance.

$$S = \{4, if\, C \geq 6\ 3.5, if\, C = 5\ 3, if\, C = 4\ 2.5, if\, C = 3\ 2, if\, C = 2\ 1.5, if\, C = 1\ 1, if\, C = 0\} \quad (2)$$

where S is Score obtained by the child based on their imitation performance and C is Number of incorrect poses imitated by the child.

2 | METHODOLOGY

2.1 | Experiment design

The study was conducted in Primal Support located in Islamabad, Pakistan over a period of 8 months (34 weeks). The study was approved by the autism specialist and director board of PSC and written parental consent was obtained before the intervention.

The iterative design process for RoboCA^{3}T includes requirement gathering, analysis, design, pilot testing, familiarisation, experimentation, follow-up, and future enhancements/scalability phases. RoboCA^{3}T was developed collaboratively with input from diverse stakeholders within the autism community, including therapists, parents, educators, and individuals with ASD. These unique perspectives were integral in shaping RoboCA^{3}T's design and functionality. Compliance with the established UNICEF Policy guidance on AI for children (Lemaignan et al., 2021) ensured the establishment of a transparent framework that prioritises principles of privacy, safety, inclusivity, and responsible AI practices, particularly for children. Additionally, the integration of key principles from the Autism-Guide (Aguiar et al., 2022) into the developmental framework of RoboCA^{3}T ensured a user-centric design tailored specifically for individuals diagnosed with ASD. Table 1 outlines the various phases of the project lifecycle, including key activities, associated artefacts, and involved stakeholders at each stage. The experiments started with familiarisation phase in which each subject was subjected to 5 sessions of joint attention and 5 sessions of imitation module. The familiarisation phase was followed by 30 experimental trails per module per subject. In total, 60 (30 × 2) experimental trials were conducted for each subject. In this phase, the children went through 5 follow- up sessions per module accounting for a total of 10 sessions after a break of two months.

To enhance the impact of our findings across different ASD severity levels and demographics, strategies involving active participation of parents and online accessibility will be incorporated in Future Enhancements/Scalability phase, facilitating broader applicability and impact. Parents will play a pivotal role in facilitating online therapy sessions and providing valuable feedback on their child's progress and experience with the updated platform. Additionally. being accessible online, RoboCA^{3}T's can reach individuals from various demographic backgrounds and geographical locations, ensuring a more representative participant pool.

TABLE 1 Inter-disciplinary approach for evolution and iterative development process highlighting theoretical foundation of RoboCA^{3}T.

Project phase	Description	Artefacts	Stakeholders
Requirement Gathering	Collaborative sessions with therapists, educators, parents, and researchers to define study objectives, scope, and requirements.	Project Charter, SRS Requirements Document	Therapists, Educators, Parents, Researchers
Analysis	Identification and documentation of specific therapeutic goals and target areas of intervention through use case analysis.	Use Case Document, User Stories	Therapists, Researchers
Design	Creation of a functional prototype of the therapy platform based on identified use cases, followed by iterative refinement. Integration of key principles from the AutismGuide (Aguiar et al., 2022) into the developmental framework of RoboCA^{3}T	Prototype Design, Wireframes, Mockups, Evaluation Matrix to ensure compliance with AutismGuide (Aguiar et al., 2022)	Therapists, Developers, ASD children, Researchers
Pilot Testing	Small-scale study to assess feasibility, usability, and efficacy of therapy interventions in a real-world setting.	Pilot Study Plan, Participant Consent Forms, Data Collection Forms	Subjects, Therapists, Researchers
Familiarisation	Introduction of the therapy platform to therapists and educators through training sessions and onboarding processes.	Training Materials, Onboarding Guides	Subjects, Therapists, Researchers
Experimentation	Deployment of therapy interventions with a larger group of participants, followed by comprehensive evaluation.	Experiment Protocol, Evaluation Metrics, Therapy Notes, Data Analysis Plan	Subjects, Therapists, Researchers
Follow-up	Assessment of long-term effects of therapy interventions and monitoring of progress through periodic check-ins.	Follow-up Plan, Progress Reports, Participant Feedback Forms	Subjects, Therapists, Researchers
Future Enhancements/ Scalability	Incorporation of additional core impairments (Emotional Response, Body Use, Adaptation to Change, Visual Response, Listening Response, Verbal Communication, Non-verbal Communication, Activity Level, Intellectual Response) into the therapy platform. Implementing further updates and improvements based on feedback and emerging needs. Exploring feasibility of parent facilitated online therapy sessions.	Updated Therapy Platform, Revised Recruitment Criteria	Subjects, Therapists, Researchers, Parents

Moreover, this online accessibility offers a cost-effective solution, eliminating barriers related to travel and accommodation often associated with traditional therapy settings.

2.2 | Participants

To conduct this research, a total of 12 children (9 males and 3 females) with autism spectrum disorder, aged between 3 and 11 years old, participated in the study. However, the consent of one of the children (male) was withdrawn, reducing the number to 11. The recruited participants were already evaluated clinically based on Childhood Autism Rating Scale Schedule (CARS) criteria by the experts. The details of participants are presented in Table 2.

2.3 | Research hypothesis

H_0 is the null hypothesis, while H_A is the alternative hypothesis for our current study. H_0 states that mean score of child joint attention skills in the first half of therapy sessions is greater than the mean score in the second half of therapy sessions. Whereas H_A states that the mean score of child joint attention skills in the first half of therapy sessions is equal to or greater than the mean score in the second half

TABLE 2 Summary of subjects.

Subject ID	Age	Gender	CARS ASD category	CARS
1	4	Male	Mild	31
2	5	Male	Mild	32.5
3	5	Female	Mild	34.5
4	11	Male	Moderate	37
5	5	Male	Mild	31.5
6	4	Male	Mild	34
7	4	Male	Mild	33
8	4	Male	Moderate	37
9	3	Female	Moderate	36.5
10	5	Male	Mild	30.5
11	4	Female	Mild	35.5

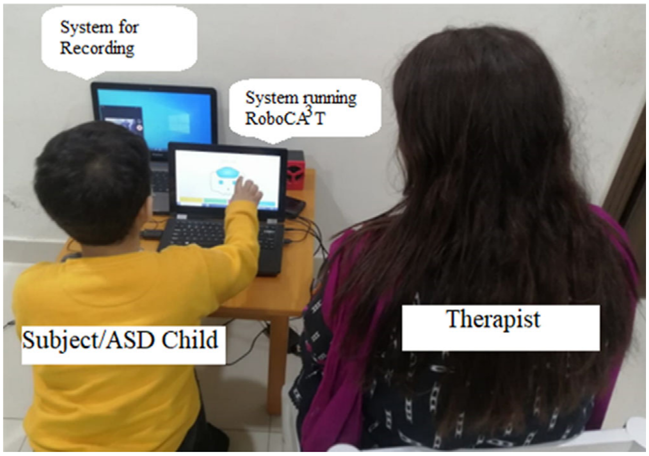

FIGURE 10 Visual Representation of Experimental Setup, depicting the roles of the child, therapist, and system within the intervention context.

of therapy sessions for each subject. We will reject H_0 if we find sufficient evidence to support H_A. The scores represent the level of improvement after therapy, with reduced scores indicating improvement after each session.

2.4 | Experimental setup

An overview of the experimental setup is shown in Figure 10. The child sits in front of a laptop running RoboCA^{3}T, while a therapist would be positioned beside the child to aid. The child was positioned in a way that aligned their eye level with the camera of the first laptop, facilitating precise recording of gaze patterns. A second laptop equipped with a camera was positioned to record the child's interaction with the laptop in real-time, capturing both visual and auditory information for subsequent analysis and assessment.

3 | RESULTS

Our study included an assessment of the Childhood Autism Rating Scale (CARS) scores, which were recorded before and after the intervention to gauge improvements in the participants' conditions as shown in Table 3.

This analysis of CARS scores recorded before and after the intervention, indicated a marked reduction in Childhood Autism Rating Scale (CARS) scores for each child, highlighting significant improvements, particularly in the JA (Joint Attention) and imitation modules. The average mean scores for JA (Joint Attention) and imitation showed a notable reduction from the first half (sessions 1–15) to the second half of therapy, further indicating significant improvement (Table 3).

Moreover, descriptive analysis was applied to study the trend in scores over time for each subject. Heatmaps represented in Figure 11 and Figure 12 were generated for both modules, showing how scores varied across sessions and subjects.

The visual analysis of heatmaps (Figure 11 and Figure 12) presented a compelling trend in the scores across therapy sessions. The leftmost columns, representing the initial sessions, exhibited higher colour intensity, gradually transitioning to lower colour intensity as the sessions advanced towards the rightmost columns (final sessions). This visual observation strongly reinforces the therapeutic intervention's potential efficacy in enhancing the imitation and joint attention skills of individuals with ASD.

To enable a comparison of the score distributions for each subject in the two session groups, the therapy sessions were divided into two equal halves: 'Sessions 1–15' (first half) and 'Sessions 16–30' (second half). Point plots (Figures 13 and 14) were created for each group of session for both modules.

Each subject has its own representation within the point plot, allowing to compare how the scores vary between the two session groups for individual subjects. These point plots also provide valuable insights into the distribution of scores, including the range of scores for each subject, both in the first half and the second half of the therapy sessions.

Both the imitation and joint attention modules revealed a noteworthy pattern in the data. The mean scores for 'Sessions 1–15' were consistently higher than those for 'Sessions 16–30', as evident from point plots (Figures 13 and 14) thus rejecting our null hypothesis. This signifies a significant decrease in scores as therapy sessions progressed (higher session numbers), indicating marked improvement in both joint attention and imitation skills. Overall, the scores for both modules exhibited significant decreases from the initial therapy sessions to the later therapy sessions, providing further support for the alternative hypothesis that RoboCA^{3}T positively influences Joint Attention and Imitation abilities. The results consistently indicate that as the session number increases, the subjects' scores decrease, implying a reduction in the severity level and, consequently, an improvement in their skills.

TABLE 3 Summary of results.

Subject ID	*Age*	*CARS pre intervention*	*CARS post intervention*	*Average joint attention score 1ST half*	*Average joint attention score 2nd half*	*Average imitation score 1ST half*	*Average imitation score 2nd half*
1	4	31	28	3.5	2.5	3.5	2.5
2	5	32.5	31	2	1.5	3	1.5
3	5	34.5	30	3.5	2	3	1.5
4	11	37	34	3.5	3	3	1.5
5	5	31.5	29	3.5	2	3.5	1.5
6	4	34	30	3	2.5	3.5	1.5
7	4	33	29	3.5	2	3	1
8	4	37	33	4	3	2.5	1.5
9	3	36.5	32	3.5	2.5	3	1
10	5	30.5	27	3	1.5	2	1.5
11	4	35.5	30	3	2.5	3	1

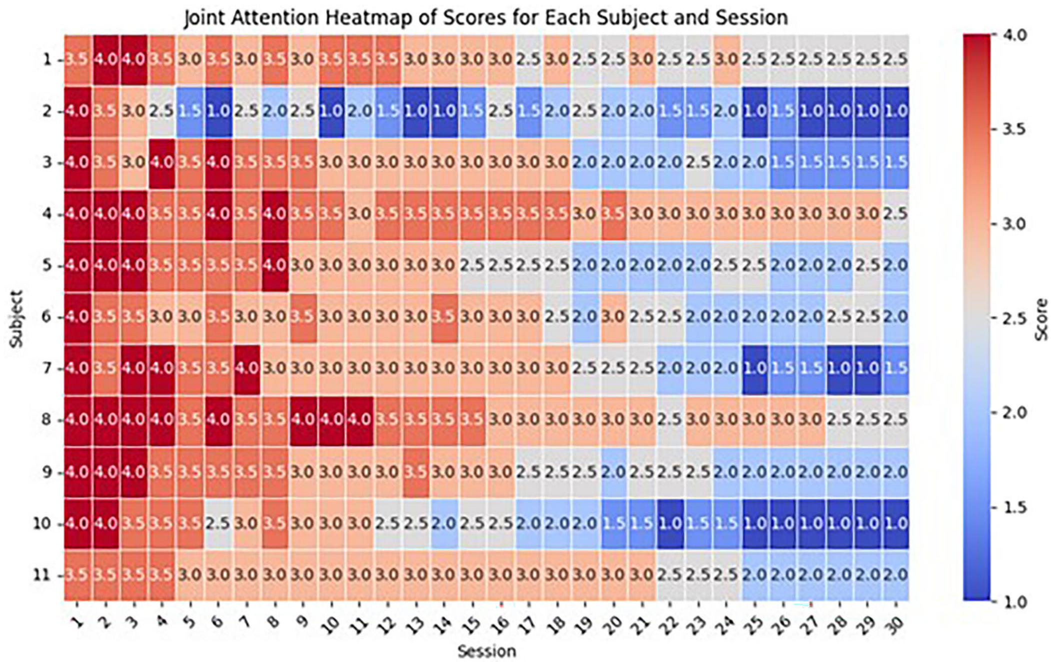

FIGURE 11 Joint Attention Module Heatmap of Scores for Each Subject and Session.

After conducting descriptive analysis, subsequent statistical analyses were performed to determine the significance of any changes observed.

3.1 | Statistical analysis

To test the data normality, Shapiro–Wilk was applied on the data, which indicated that the scores of the subjects were not normally distributed ($p < 0.05$). To particularly determine the significance of changes in scores, a non-parametric Wilcoxon signed-rank test was applied on the data. For each subject, the therapy sessions were divided into two equal halves: 'Sessions 1–15' (first half) and 'Sessions 16–30' (second half) for both modules. Scores from the first half were compared with the second half using Wilcoxon signed-rank test. The Spearman correlation coefficient was also performed on the two groups to measure the strength and relationship between 'session number' and 'score' for each subject for both modules.

Statistical analyses, including the Wilcoxon signed-rank test, consistently reported exceptionally low p-values (<0.01) for all subjects. This statistical significance implies that the observed strong negative correlations between 'session' and 'score' are highly unlikely to occur by chance. Consequently, these results suggest that with the progression of therapy sessions (higher 'session' values), the 'score' tends to decrease significantly, implying potential enhancements in both Joint Attention and Imitation skills.

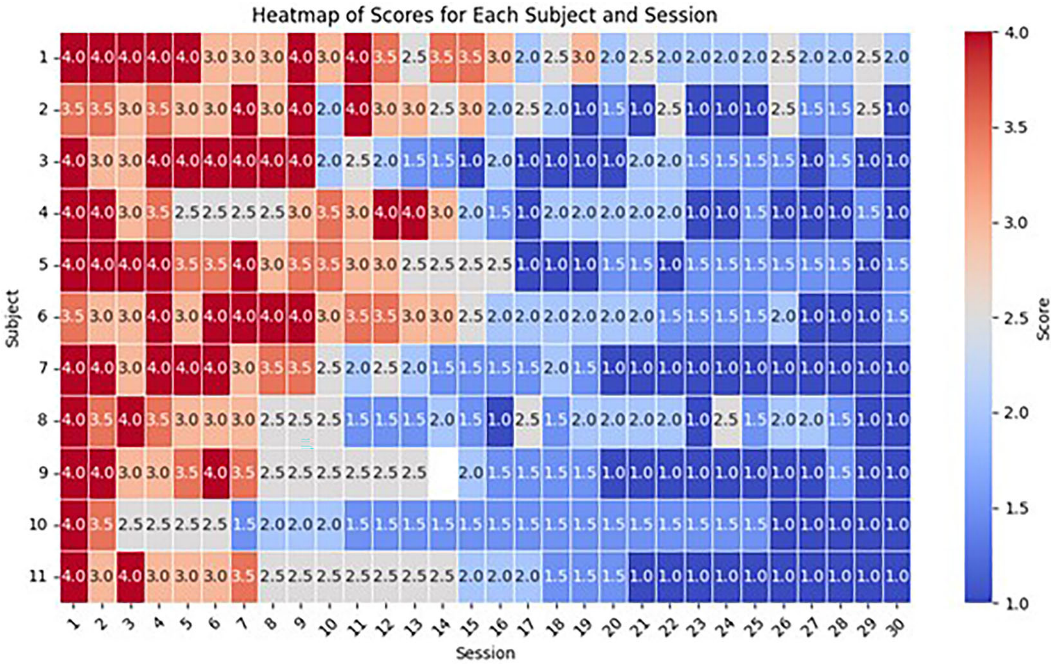

FIGURE 12 Imitation Module Heatmap of Scores for Each Subject and Session.

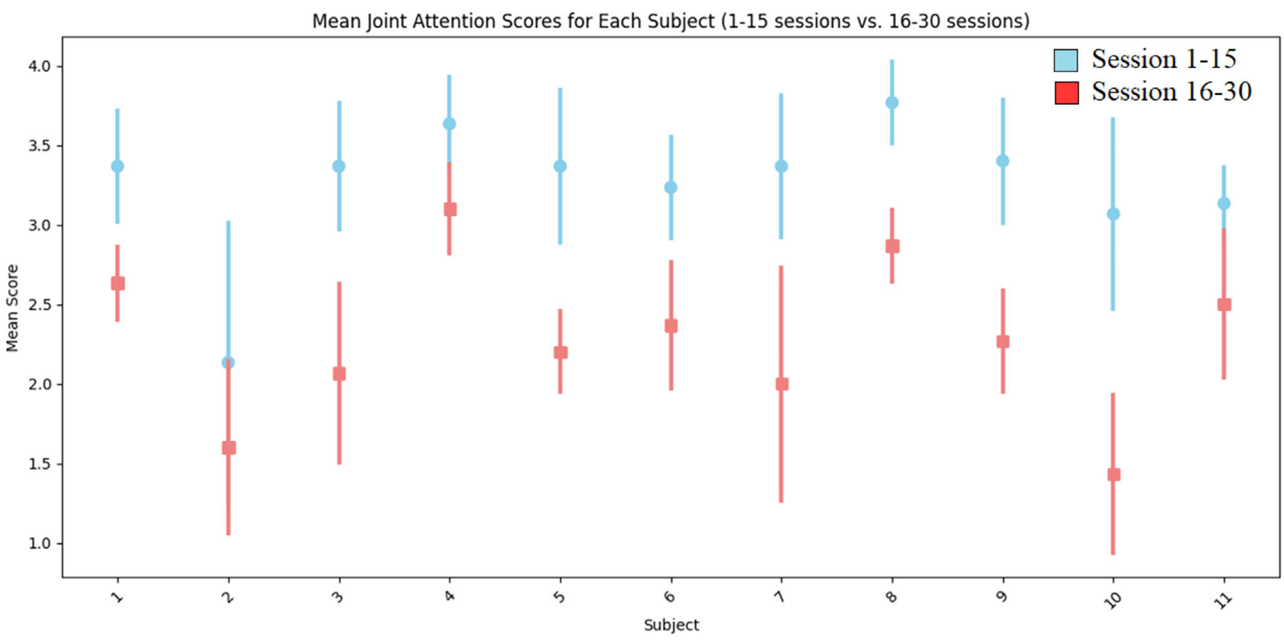

FIGURE 13 Joint attention point plot: scores in sessions 1–15 versus sessions 16–30 for each subject.

Furthermore, the Spearman correlation coefficient outcomes supported these findings, revealing robust negative correlations between 'session' and 'score' for all subjects ($p < 0.01$). These correlations indicate a substantial improvement in Joint Attention skills as sessions advanced and a similar noteworthy improvement in Imitation skills for the Imitation module.

3.2 | Assessment of system accuracy and performance

To validate how accurately the system detects the child's gaze and pose in real-time a comparison was made to ground truth/manual annotations derived from the therapist notes. This sheet allowed

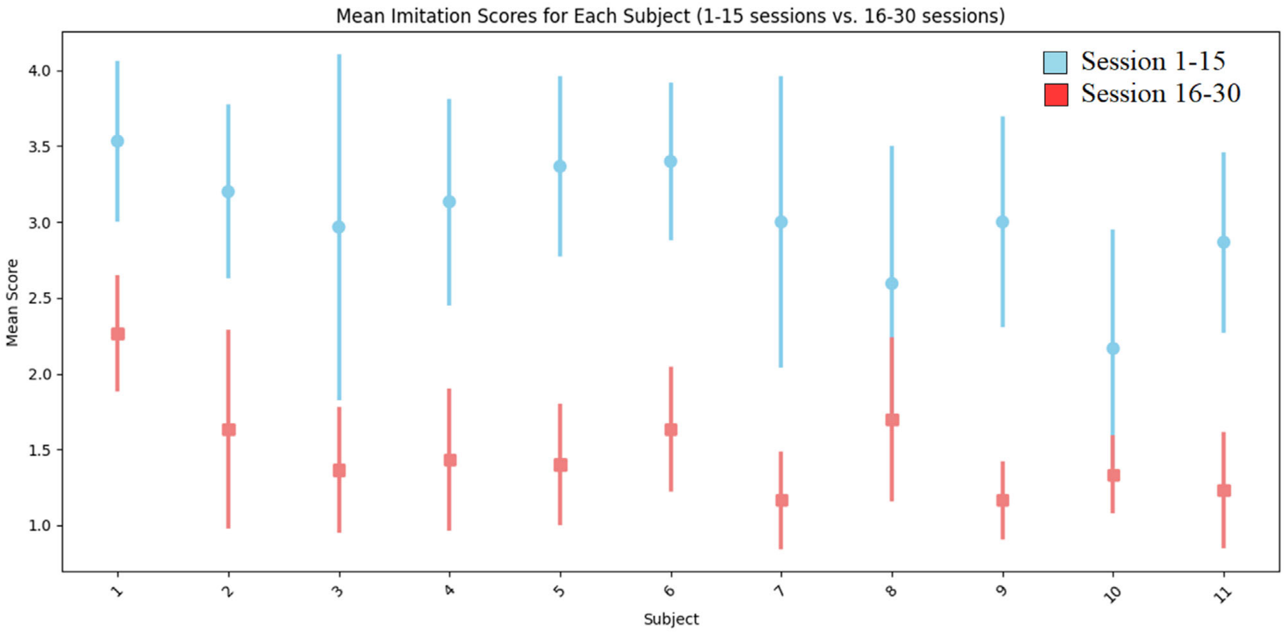

FIGURE 14 Imitation module point plot: scores in sessions 1–15 versus sessions 16–30 for each subject.

TABLE 4 Summary of evaluation matrix.

Metric	Joint attention	Imitation
Accuracy	92.0%	92.4%
Precision	0.957	0.958
Recall	0.878	0.893
F1 Score	0.916	0.924

therapists to document any observed discrepancies between the child's actual gaze and pose behaviour and the scores provided by the system. In assessing the system's performance, four key metrics were utilised: TP (True Positive), FP (False Positive), TN (True Negative), and FN (False Negative). TP denotes instances where the therapist observed a correct response, and the system also identified it as correct. Conversely, FP represents cases where the therapist noted an incorrect response, yet the system classified it as correct. TN reflects situations where both the therapist and the system identified a response as incorrect. Finally, FN indicates instances where the therapist recognised a correct response, but the system categorised it as incorrect. These metrics collectively provide insights into the system's accuracy as shown in Table 4.

4 | DISCUSSION

Robot-Assisted Therapies (RATs) have demonstrated effectiveness in developing crucial skills such as JA and imitation (Ali et al., 2019; Alshammari et al., 2022; Marino et al., 2020; Martins et al., 2020; Salhi et al., 2022; Sigacheva et al., 2020). However, challenges such as cost, resource availability, and emotional attachment to robots persist (Ellery et al., 2008; Guntur et al., 2019; Wali & Sanfilippo, 2019). Given these challenges, researchers have explored alternative options, including Computer-Assisted Therapies (CATs). CAT encompasses a range of technologies designed to create a safe, accessible, and cost-effective environment for individuals with ASD to practice and develop various skills, including communication and social skills (Bauer et al., 2022; Eden & Oren, 2021; Grynszpan et al., 2014; Winoto & Tang, 2019). These applications facilitate connections with others, enable progress tracking, and support skill development across various domains (Ramos Aguiar et al., 2023; Winoto & Tang, 2019). However, existing CAT applications have limitations, including the inability to deliver personalised content, track progress over time, and adapt to individual needs. Additionally, there is a noticeable lack of applications capable of predicting skill development in individuals with ASD. Human interaction remains vital for effective therapy, yet existing applications often lack rigorous validation procedures and fail to provide necessary human interaction as mediators. Our study addresses these limitations by developing a novel intervention, Robot-Inspired Computer-Assisted Adaptive Autism Therapy (RoboCA^{3}T), focused on improving JA and imitation skills. By integrating RAT and CAT approaches, RoboCA^{3}T utilises the strengths of both methodologies while mitigating their respective limitations. Automatic gaze and pose detection algorithms within RoboCA^{3}T eliminate potential human error and observation bias, ensuring objective and accurate assessment of progress. Moreover, by capitalising on children with ASD's affinity for robots and technology, RoboCA^{3}T creates a unique therapeutic environment that maximises engagement and facilitates effective skill development. It offers a cost-effective and adaptive solution tailored to the learning, development, and therapy needs of children with ASD, ultimately aiming to improve their JA and imitation skills and enhance their overall quality of life.

5 | CONCLUSION

In this study, we aimed to develop Robot-Inspired Computer-Assisted Adaptive Autism Therapy (RoboCA^{3}T) to improve Joint Attention (JA) and imitation skills in children with ASD. RoboCA^{3}T integrated Robot-Assisted Therapies (RATs) and Computer-Assisted Therapies (CATs), utilising robot avatars and web-based solutions. It prioritised personalised content delivery and design inclusivity. The technology stack included Next.js, React.js, Tailwind CSS, Firebase, WebGazer, and Tensorflow Lite. By leveraging AI-based gaze and pose detection algorithms, the therapy mitigated potential human errors and observation biases, thereby providing a more reliable assessment of the child's development. Throughout 660 experimental trials involving 11 ASD children, significant reductions in CARS scores, notably in JA and imitation modules, were observed. Descriptive analysis revealed a consistent decreasing trend in mean scores across therapy sessions for both modules, visually supported by heatmaps and point plots. This improvement remained consistent in both the initial (sessions 1–15) and later phases (sessions 16–30), as confirmed by Wilcoxon signed-rank tests ($p < 0.01$) and negative correlations between session number and scores for all subjects. Overall, RoboCA^{3}T exhibited positive outcomes, potentially transforming autism therapy, and expanding accessible support for individuals and caregivers. These findings have important implications for the development of effective therapies for individuals with joint attention and imitation deficits.

AUTHOR CONTRIBUTIONS

Sara Ali: Supervision; conceptualization; visualization; validation; writing – review and editing. **Zunera Zahid:** Conceptualization; investigation; methodology; validation; software; formal analysis; data curation; writing – review and editing; writing – original draft; visualization. **Shehriyar Shariq:** Conceptualization; formal analysis; data curation. **Yasar Ayaz:** Conceptualization; supervision. **Noman Naseer:** Validation; methodology; conceptualization. **Irum Yaseen:** Conceptualization; methodology; formal analysis; data curation.

ACKNOWLEDGEMENTS

We extend our gratitude to the Primal Support Center for their invaluable feedback and for providing both subjects and space, which were instrumental in conducting our experiments.

FUNDING INFORMATION

This research received no external funding.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest and the study is carried out in the absence of any commercial or financial relationships.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Sara Ali https://orcid.org/0000-0002-5100-9430

REFERENCES

Aguiar, Y. P. C., Galy, E., Godde, A., Trémaud, M., & Tardif, C. (2022). AutismGuide: A usability guidelines to design software solutions for users with autism spectrum disorder. *Behaviour & Information Technology*, *41*(6), 1132–1150.

Aishworiya, R., Valica, T., Hagerman, R., & Restrepo, B. (2022). An update on psychopharmacological treatment of autism Spectrum disorder. *Neurotherapeutics*, *19*(1), 248–262. https://doi.org/10.1007/s13311-022-01183-1

Ali, S., Mehmood, F., Ayaz, Y., Sajid, M., Sadia, H., & Nawaz, R. (2022). An experimental trial: Multi-robot therapy for categorization of autism level using hidden Markov model. *Journal of Educational Computing Research*, *60*(3), 722–741.

Ali, S., Mehmood, F., Dancey, D., Ayaz, Y., Khan, M. J., Naseer, N., Amadeu, R. D. C., Sadia, H., & Nawaz, R. (2019). An adaptive multi-robot therapy for improving joint attention and imitation of ASD children. *IEEE Access*, *7*, 81808–81825.

Alsaeedi, A. (2020). Comparing web accessibility evaluation tools and evaluating the accessibility of webpages: Proposed frameworks. *Information*, *11*(1), 40.

Alshammari, R. F. N., Arshad, H., Abd Rahman, A. H., & Albahri, O. S. (2022). Robotics utilisation in automatic vision-based assessment systems from artificial intelligence perspective: A systematic review. *IEEE Access*, *10*, 77537–77570.

Alshawwa, S., Alarnous, R., & Albasalah, A. (2023). The most common treatment methods through comparing the studies results of children with autism spectrum syndrome. *Periodicals of Engineering and Natural Sciences*, *11*(2), 155–165. http://pen.ius.edu.ba/index.php/pen/article/view/3506

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders. *Diagnostic and Statistical Manual of Mental Disorders*. (5th ed.), 1–947.

American Psychiatric Association. (2022). *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition Text Revision DSM-5-TR*. (*5th ed.*) 1–1120.

Ames, C., & Fletcher-Watson, S. (2010). A review of methods in the study of attention in autism. *Developmental Review*, *30*(1), 52–73.

Aresti-Bartolome, N., & Garcia-Zapirain, B. (2014). Technologies as support tools for persons with autistic spectrum disorder: A systematic review. *International Journal of Environmental Research and Public Health*, *11*(8), 7767–7802.

Bánki, A., de Eccher, M., Falschlehner, L., Hoehl, S., & Markova, G. (2022). Comparing online webcam-and laboratory-based eye-tracking for the assessment of infants' audio-visual synchrony perception. *Frontiers in Psychology*, *12*, 733933.

Barnhill, J. W. (2023). DSM-5-TR® Clinical Cases. https://doi.org/10.1176/appi.books.9781615375295

Bauer, V., Bouchara, T., & Bourdot, P. (2022). Extended reality guidelines for supporting autism interventions based on stakeholders' needs. *Journal of Autism and Developmental Disorders*, *1–34*, 2078–2111.

Boucenna, S., Narzisi, A., Tilmont, E., Muratori, F., Pioggia, G., Cohen, D., & Chetouani, M. (2014). Interactive technologies for autistic children: A review. *Cognitive Computation*, *6*, 722–740.

Chen, Y., Zhou, Z., Cao, M., Liu, M., Lin, Z., Yang, W., Yang, X., Dhaidhai, D., & Xiong, P. (2022). Extended reality (XR) and telehealth interventions for children or adolescents with autism spectrum disorder: Systematic review of qualitative and quantitative studies. *Neuroscience and Biobehavioral Reviews*, *1*(138), 104683.

Corcione, F., Esposito, C., Cuccurullo, D., Settembre, A., Miranda, N., Amato, F., Pirozzi, F., & Caiazzo, P. (2005). Advantages and limits of robot-assisted laparoscopic surgery: Preliminary experience. *Surgical Endoscopy and Other Interventional Techniques*, *19*, 117–119.

Coury, D. L., Anagnostou, E., Manning-Courtney, P., Reynolds, A., Cole, L., McCoy, R., Whitaker, A., & Perrin, J. M. (2012). Use of psychotropic medication in children and adolescents with autism spectrum disorders. *Pediatrics*, *130*(suppl 2), S69–S76.

David, D. O., Costescu, C. A., Matu, S., Szentagotai, A., & Dobrean, A. (2020). Effects of a robot-enhanced intervention for children with ASD on teaching turn-taking skills. *Journal of Educational Computing Research*, *58*(1), 29–62.

del Pilar Gallardo Montes, C., Fuentes, A. R., & Cara, M. J. C. (2021). Apps for people with autism: Assessment, classification and ranking of the best. *Technology in Society*, *64*, 101474.

Den Bossche, V., Abdallah, W., Vella, F., Vigouroux, N., Van den Bossche, A., & Val, T. (2019). A collaborative talking assistive technology for people with autism spectrum disorders. *Springer*, *26*, 3–12. https://doi.org/10.1007/978-3-030-23563-5_1

Deng, L., & Technology, P. R. A. (2022). The need for and barriers to using assistive technologies among individuals with Autism Spectrum Disorders in China. *Taylor & Francis*, *34*(2), 242–253. https://doi.org/10.1080/10400435.2020.1757787

Durkin, K. (2010). Videogames and young people with developmental disorders. *Review of General Psychology*, *14*(2), 122–140.

Eden, S., & Oren, A. (2021). Computer-mediated intervention to foster prosocial ability among children with autism. *Journal of Computer Assisted Learning*, *37*(1), 275–286.

Ellery, A., Kreisel, J., & Sommer, B. (2008). The case for robotic on-orbit servicing of spacecraft: Spacecraft reliability is a myth. *Acta Astronautica*, *63*(5–6), 632–648.

Erel, Y., Shannon, K. A., Chu, J., Scott, K., Kline Struhl, M., Cao, P., Tan, X., Hart, P., Raz, G., Piccolo, S., Mei, C., Potter, C., Jaffe-Dax, S., Lew-Williams, C., Tenenbaum, J., Fairchild, K., Bermano, A., & Liu, S. (2023). iCatcher+: Robust and automated annotation of Infants' and young Children's gaze behavior from videos collected in laboratory, field, and online studies. *Advances in Methods and Practices in Psychological Science*, *6*(2), 25152459221147250.

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: A unified JavaScript framework for online studies. In *International conference on computational social science (cologne)* (pp. 1–3). University of Osnabrück Cologne.

Foxx, R. M. (2008). Applied behavior analysis treatment of autism: The state of the art. *Child and Adolescent Psychiatric Clinics of North America*, *17*(4), 821–834.

Goldsborough, P. (2016). A tour of tensorflow. *arXiv Preprint arXiv:161001178*. *1*, 1–16.

Grynszpan, O., Weiss, P. L., Perez-Diaz, F., & Gal, E. (2014). Innovative technology-based interventions for autism spectrum disorders: A meta-analysis. *Autism*, *18*(4), 346–361.

Grzadzinski, R., Huerta, M., & Lord, C. (2013). DSM-5 and autism spectrum disorders (ASDs): An opportunity for identifying ASD subtypes. *Molecular Autism*, *4*(1), 1–6. https://doi.org/10.1186/2040-2392-4-12

Guntur, S. R., Gorrepati, R. R., & Dirisala, V. R. (2019). Robotics in healthcare: An internet of medical robotic things (IoMRT) perspective. In *Machine learning in bio-signal analysis and diagnostic imaging* (pp. 293–318). Elsevier.

Haber, N., Voss, C., Daniels, J., Washington, P., Fazel, A., & Kline, A. (2020). A wearable social interaction aid for children with autism. arXiv Preprint arXiv:200414281.

Heck, M., Becker, C., & Deutscher, V. (2023). *Webcam eye tracking for desktop and Mobile devices: A systematic review*. University of Hawaii.

Ingersoll, B. (2008). The social role of imitation in autism: Implications for the treatment of imitation deficits. *Infants and Young Children*, *21*(2), 107–119.

Inoue, M. (2019). Assessments and interventions to address challenging behavior in individuals with intellectual disability and autism spectrum disorder in Japan: A consolidated review. *Yonago Acta Medica*, *62*(2), 169–181.

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. John Wiley & Sons.

Kumar, S., Shree, D. V. J., & Biswas, P. (2021). Comparing ten WCAG tools for accessibility evaluation of websites. *Technology and Disability*, *33*(3), 163–185.

Lemaignan, S., Newbutt, N., Rice, L., Daly, J., & Charisi, V. (2021). UNICEF guidance on AI for children: Application to the design of a social robot for and with autistic children. *arXiv Preprint arXiv:210812166*. *1*, 1–7.

Lewandowska, B. (2019). *RealEye eye-tracking system technology whitepaper* (p. 2022). Warsaw University of Technology.

Liu, M., & Ma, Z. (2022). A systematic review of telehealth screening, assessment, and diagnosis of autism spectrum disorder. *Child and Adolescent Psychiatry and Mental Health*, *16*(1), 1–15.

Liu, Z., Zhu, J., Bu, J., & Chen, C. (2015). A survey of human pose estimation: The body parts parsing based methods. *Journal of Visual Communication and Image Representation*, *32*, 10–19.

Lotti, A., Modenini, D., Tortora, P., Saponara, M., & Perino, M. A. (2023). Deep learning for real-time satellite pose estimation on tensor processing units.*Journal of Spacecraft and Rockets*, *60*(3), 1034–1038.

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*, *55*(1), 3–9.

Madurapperuma, I. H., Shafana, M. S., & Sabani, M. J. A. (2022). *State-of-art frameworks for front-end and back-end web development*. ICST.

Mannion, A., & Leader, G. (2013). Comorbidity in autism spectrum disorder: A literature review. *Research in Autism Spectrum Disorder*, *7*(12), 1595–1616.

Marino, F., Chilà, P., Sfrazzetto, S. T., Carrozza, C., Crimi, I., Failla, C., Busà, M., Bernava, G., Tartarisco, G., Vagni, D., Ruta, L., & Pioggia, G. (2020). Outcomes of a robot-assisted social-emotional understanding intervention for young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *50*(6), 1973–1987.

Martins, P. S., Faria, G., & de Jesus Fiais Cerqueira, J. (2020). I2E: A cognitive architecture based on emotions for assistive robotics applications. *Electronics*, *9*(10), 1590.

Matson, J. L., & Burns, C. O. (2019). Pharmacological treatment of autism spectrum disorder. *APA Handbook of Psychopharmacology*, *18*, 373–396.

Nesterova, A. A., Aysina, R. M., & Suslova, T. F. (2015). Recent technologies to improving social and communication skills in children with asd: Systematization of approaches and methods. *Modern Applied Science*, *9*(11), 38.

Odom, S. L., Boyd, B. A., Hall, L. J., & Hume, K. (2010). Evaluation of comprehensive treatment models for individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *40*(4), 425–436.

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., Hays, J. (2016). Webgazer: scalable webcam eye tracking using user interactions. In: *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 3839–3845.

Paul, A. M., & Eubanks Fleming, C. J. (2019). Anxiety management on campus: An evaluation of a mobile health intervention. *Journal of Technology in Behavioral Science*, *4*, 58–61.

Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S., & Pioggia, G. (2016). Autism and social robotics: A systematic review. *Autism Research*, *9*(2), 165–183.

Pop, C. A., Simut, R. E., Pintea, S., Saldien, J., Rusu, A. S., Vanderfaeillie, J., David, D. O., Lefeber, D., & Vanderborght, B. (2013). Social robots vs. computer display: Does the way social stories are delivered make a difference for their effectiveness on ASD children. *Journal of Educational Computing Research*, *49*(3), 381–401.

Rajendran, G. (2013). Virtual environments and autism: A developmental psychopathological approach. *Journal of Computer Assisted Learning*, *29*(4), 334–347.

Ramos Aguiar, L. R., Álvarez Rodríguez, F. J., Madero Aguilar, J. R., Navarro Plascencia, V., Peña Mendoza, L. M., Quintero Valdez, J. R., Pech, J. R. V., Leon, A. M., & Ortiz, L. E. L. (2023). Implementing gamification for blind and autistic people with tangible interfaces, extended reality, and universal design for learning: two case studies. *Applied Sciences*, *13*(5), 1–17.

Rehman, I. U., Sobnath, D., Nasralla, M. M., Winnett, M., Anwar, A., & Asif, W. (2021). Features of Mobile apps for people with autism in a post COVID-19 scenario: Current status and recommendations for apps using AI. *Diagnostics*, *11*(10), 1923 https://www.mdpi.com/2075-4418/11/10/1923/htm

Richardson, K., Cao, H., Esteban, P. G., Bartlett, M., Belpaeme, T., Billing, E., Cai, H., Coeckelbergh, M., Pop, C. A., & David, D. (2019). Robot-enhanced therapy: Development and validation of supervised autonomous robotic system for autism spectrum disorders therapy. *IEEE Robotics & Automation Magazine*, *26*(2), 1–10.

Roane, H. S., Fisher, W. W., & Carr, J. E. (2016). Applied behavior analysis as treatment for autism Spectrum disorder. *Journal of Pediatrics*, *175*, 27–32. https://pubmed.ncbi.nlm.nih.gov/27179552/

Rodrigues, R., Lai, M. C., Beswick, A., Gorman, D. A., Anagnostou, E., Szatmari, P., Anderson, K. K., & Ameis, S. H. (2021). Practitioner review: Pharmacological treatment of attention-deficit/hyperactivity disorder symptoms in children and youth with autism spectrum disorder: A systematic review and meta-analysis. *Journal of Child Psychology and Psychiatry*, *62*(6), 680–700.

Salhi, I., Qbadou, M., Gouraguine, S., Mansouri, K., Lytridis, C., & Kaburlasos, V. (2022). Towards robot-assisted therapy for children with autism—The ontological knowledge models and reinforcement learning-based algorithms. *Frontiers in Robotics AI*, *6*(9), 90.

Sharma, B., Mantri, A., Singh, N. P., Sharma, D., Gupta, D., & Tuli, N. (2022). EduSense-AR: A Sensory Learning Solution for Autistic Children. In: 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO). IEEE, 1–4.

Sharma, S., Ventura, J., & D'Amico, S. (2018). Robust model-based monocular pose initialization for noncooperative spacecraft rendezvous. *Journal of Spacecraft and Rockets*, *55*(6), 1414–1429.

Sher, B. (2009). *Early intervention games: Fun, joyful ways to develop social and motor skills in children with autism Spectrum or sensory processing disorders*. John Wiley & Sons.

Shrestha, S., Shah, A., Dhakal, P., & Dhakal, N. (2021). Head Gesture and Voice Based Learning App for Children with Autism of Nepal. *2021 3rd International Conference on Sustainable Technologies for Industry 4.0 (STI)*. (pp. 1–6). IEEE. https://doi.org/10.1109/sti53101.2021.9732539

Sigacheva, N., Baranova, A. R., & Makaev, K. F. (2020). Communication problems: Advantages and disadvantages of teaching autistic children with humanoid robots. *ARPHA Proceedings*, *3*, 2387.

Silva, S. D., Neto, F. M. M., de Lima, R. M., de Macêdo, F. T., Santo, J. R. S., & Silva, W. L. N. (2017). Knowledgemon hunter: a serious game with geolocation to support learning of children with autism and learning difficulties. In 2017 19th symposium on virtual and augmented reality (SVR). *IEEE*, *2017*, 293–296.

Slim, M. S., & Hartsuiker, R. J. (2022). Moving visual world experiments online? A web-based replication of Dijkgraaf, Hartsuiker, and Duyck (2017) using PCIbex and WebGazer. js. Behavior Research Methods, 1–19.

Steffan, A., Zimmer, L., Arias-Trejo, N., Bohn, M., Dal Ben, R., Flores-Coronado, M. A., Franchin, L., Garbisch, I., Grosse Wiesmann, C., Hamlin, J. K., Havron, N., Hay, J. F., Hermansen, T. K., Jakobsen, K. V., Kalinke, S., Ko, E. S., Kulke, L., Mayor, J., Meristo, M., et al. (2024). Validation of an open source, remote web-based eye-tracking method (WebGazer) for research in early childhood. *Infancy*, *29*(1), 31–55.

Stone, W. L., Ousley, O. Y., & Littleford, C. D. (1997). Motor imitation in young children with autism: What's the object? *Journal of Abnormal Child Psychology*, *25*(6), 475–485. https://doi.org/10.1023/A:1022685731726

Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5693–5703). IEEE.

Syriopoulou-Delli C, of EGIJ. (2022). Review of assistive technology in the training of children with autism spectrum disorders. *Taylor & Francis*. *68*(2), 73–85. https://doi.org/10.1080/20473869.2019.1706333

Tleubayev, B., Zhexenova, Z., Zhakenova, A., & Sandygulova, A. (2019). Robot-assisted therapy for children with ADHD and ASD: a pilot study. In *Proceedings of the 2019 2nd International Conference on Service Robotics Technologies*, (pp. 58–62). Association for Computing Machinery.

Tsikinas, S., & Xinogalos, S. (2019). Studying the effects of computer serious games on people with intellectual disabilities or autism spectrum disorder: A systematic literature review. *Journal of Computer Assisted Learning*, *35*(1), 61–73.

Valliappan, N., Dai, N., Steinberg, E., He, J., Rogers, K., Ramachandran, V., Xu, P., Shojaeizadeh, M., Guo, L., Kohlhoff, K., & Navalpakkam, V. (2020). Accelerating eye movement research via accurate and affordable smartphone eye tracking. *Nature Communications*, *11*(1), 4553.

Vyshedskiy, A., & Dunn, R. (2015). Mental imagery therapy for autism (MITA)-an early intervention computerized brain training program for children with ASD. *Autism Open Access*, *5*(2), 1000153.

Wali, L. J., & Sanfilippo, F. (2019). A Review of the State-of-the-Art of Assistive Technology for People with ASD in the Workplace and in Everyday Life. In: Digital Transformation for a Sustainable Society in the 21st Century: 18th IFIP WG 611 Conference on e-Business, e-Services, and e-Society, I3E 2019, Trondheim, Norway, September 18–20, 2019, Proceedings 18. Springer, 520–32.

Werchan, D. M., Thomason, M. E., & Brito, N. H. (2023). OWLET: An automated, open-source method for infant gaze tracking using smartphone and webcam recordings. *Behavior Research Methods*, *55*(6), 3149–3163.

Winoto, P., & Tang, T. Y. (2019). Training joint attention skills and facilitating proactive interactions in children with autism Spectrum disorder: A loosely coupled collaborative tabletop-based application in a Chinese special education classroom. *Journal of Educational Computing Research*, *57*(1), 32–57. https://doi.org/10.1177/0735633117745160

Wojciechowski, A., & Al-Musawi, R. (2017). Assisstive technology application for enhancing social and language skills of young children with autism. *Multimedia Tools and Applications*, *76*, 5419–5439.

Zhang, J., Zhang, D., Xu, X., Jia, F., Liu, Y., & Liu, X. (2020). MobiPose: Real-time multi-person pose estimation on mobile devices. In: Proceedings of the 18th Conference on Embedded Networked Sensor Systems, 136–49.

Zhang, Q., Li, X., Che, X., Ma, X., Zhou, A., Xu, M., Wang, S., Ma, Y. & Liu, X. (2022). A comprehensive benchmark of deep learning libraries on mobile devices. In *Proceedings of the ACM Web Conference 2022* (pp. 3298–3307).

How to cite this article: Zahid, Z., Ali, S., Shariq, S., Ayaz, Y., Naseer, N., & Yaseen, I. (2024). RoboCA^{3}T: A Robot-Inspired Computer-Assisted adaptive autism therapy for improving joint attention and imitation skills through learning and computing innovations. *Journal of Computer Assisted Learning*, *40*(5), 2031–2048. https://doi.org/10.1111/jcal.12990