

## Review

# Automatic Gesture and Movement Recognition for Child Behavioural Analysis: A Systematic Review of the Laboratory-to-Natural Setting Gap

Athifah Utami <sup>1,\*</sup> , David Mazoterias-Delgado <sup>2</sup> and Lucrezia Crescenzi-Lanna <sup>1,\*</sup> <sup>1</sup> Child Tech Lab, Department of Psychology and Education, Universitat Oberta de Catalunya, 08018 Barcelona, Spain<sup>2</sup> Department of Physiotherapy, Occupational Therapy, Rehabilitation and Physical Medicine, Universidad Rey Juan Carlos, 28922 Madrid, Spain; david.mazoterias@urjc.es

\* Correspondence: autami@uoc.edu (A.U.); lcrescenzi@uoc.edu (L.C.-L.)

## Abstract

Automatic gesture and movement recognition techniques are mainly used with adults for various purposes in public, clinical, and laboratory settings. Growing interest in this field has led to the increasing application of these methods in child behavioural analysis to serve different societal and educational functions. However, manual human annotation of behaviours remains the predominant method, and only a limited number of studies have explored the use of automatic recognition for children. This review aims to evaluate the rapidly developing techniques of automatic gesture and movement recognition that focus on child behaviour analysis across different settings and for different purposes. More specifically, it analyzes their purposes, target groups, settings, accuracy, and limitations, as well as the ethical issues and data privacy frameworks that should be considered in child-centred AI. Using a systematic review approach following the PRISMA guidelines, this study examines research published between 2021 and 2025 in four databases: Web of Science (WoS), Scopus, PubMed, and IEEE Xplore. From a total of 27 included studies, the findings reveal that automatic gesture and movement recognition is being applied across multiple fields, with consideration of children's developmental needs. However, a critical gap in technical reporting was identified: fewer than half of the included studies (44%) provided accuracy metrics or clinical validity. Furthermore, evidence of robust ethical safeguards remains limited. To support children's well-being, future studies must bridge the lab-to-field gap, prioritize natural research settings and enforce ethical and data protection measures.

**Keywords:** gesture and movement; automatic recognition; systematic review; child behaviour analysis



Academic Editors: Yang Gao and  
Yincheng Jin

Received: 19 May 2026

Revised: 8 June 2026

Accepted: 9 June 2026

Published: 12 June 2026

**Copyright:** © 2026 by the authors.  
Licensee MDPI, Basel, Switzerland.  
This article is an open access article  
distributed under the terms and  
conditions of the [Creative Commons  
Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

## 1. Introduction

Children communicate through their bodies from a very early age, even before language is fully developed. Pointing, imitation, and facial expressions reflect emotional, cognitive, and social states [1,2]. Analyzing gestures and movements from childhood to young adulthood is therefore essential for understanding developmental trajectories [3]. Technologies capable of automatically recognizing body gestures may contribute to improved learning experiences, clinical psychology, health-related assessment, and ethically responsible innovation for the early detection of children's social, emotional, language,

and motor capabilities [4–6]. Accordingly, the objective of this review is to identify which technologies are being used, for what purposes, what results have been obtained, and what barriers hinder their implementation.

Gesture and movement are related but distinct phenomena, including in anatomical, functional, and computational terms. These differences create specific challenges for automatic recognition systems: whereas movement primarily concerns physical displacement, gestures require the interpretation of semiotic intent, making the standardization of accuracy metrics—the direct measures used to evaluate a deep learning model’s precision in detection and classification. Gestures are actions that convey semantic content and pragmatic intent through semiotic signals, typically involving the hands and other parts of the body [7]. According to Novack and Goldin-Meadow [8], gestures are hand movements that accompany speech and play an essential role in human communication through joint attention and embodied symbolism [9]. In addition, joint attention and social relationships are among the factors that predict later linguistic outcomes through early gesture use [10]. By making internal mental processes visible, gestures can reinforce or illustrate verbal communication and function as cognitive tools that support concept acquisition and problem solving [8,9]. They are an integral part of communication because they encode spatial, temporal, and relational information and contribute to clearer and more efficient messages [11]. From a technical perspective, the challenge lies in moving from recognizing movement as kinematics to decoding gesture as semiotic action.

Gestures may take the form of either small or large actions. Although they often accompany speech [12], they may also occur independently or in combination with other systems such as sign language, which is particularly relevant for people with hearing impairments [13]. Studies linking the frequency of deictic gestures with progress in expressive language show that gestures are fundamental to social interaction and early communicative development. For example, Samanta et al. [14] developed an automatic system to identify prelinguistic gestures in child–caregiver interaction videos and found significant correlations with manual evaluations of children’s gestural behaviour.

Movement, unlike gesture, introduces the dimension of physical displacement, whether in relation to fine motor skills, such as writing and object manipulation, or gross motor skills, such as jumping and running [15]. In physics, movement is understood as a change in the position of a body in space over time. As Passmore et al. [16] show, the automated analysis of movement can support the assessment of motor development and the detection of non-typical developmental patterns. In their study, videos recorded by caregivers were used to train a deep learning model that identified non-typical infant movements with 94% accuracy.

Movement typically involves complete kinetic chains in which the limbs, trunk, and head participate, whereas gestures are more often produced in smaller body regions, such as the hands (74%) or the face (22%) [17,18]. From a neurophysiological perspective, movement is regulated by structures such as the basal ganglia and cerebellum, whereas gestures associated with linguistic planning are linked to activation in the left hemisphere, particularly Broca’s area [19]. Natural biological movements, such as defensive or flight responses, are broadly recognizable across cultures, while symbolic gestures for purposes such as greeting or insulting vary in both form and meaning from one culture to another [19,20]. Thus, phylogenetically rooted movements tend to be more universal, whereas gestures are strongly shaped by sociocultural conventions [21].

To distinguish movement from gesture, it is important to consider neural and physiological foundations, communicative intention, and sociocultural dimensions. From a biomechanical perspective, movement involves bodily action governed by both innate and learned motor patterns that emerge through the interaction between the nervous system

and the environment. In infancy, movement becomes meaningful when the child actively reproduces actions to relate to the physical and social world. Motor development and the acquisition of movement skills therefore depend not only on internal processes but also on participation in social contexts in which these movements are used [22].

By contrast, gesture is a semiotic act produced through specific movements and oriented toward conveying meaning; it is also culturally codified [23,24]. Gestures can serve as markers of self-identity and contribute to social cohesion. In teaching contexts, for example, gestures support spoken language and strengthen conceptual communication; one study suggests that 68% of teaching interactions in schools involve gestures as support for verbal explanation [23]. In contrast, general movements do not carry semantic meaning unless they are interpreted within a specific context. This communicative power stems from the symbolic, rapid, and economical nature of gesture: for instance, a thumbs-up is quickly understood as a sign of approval.

In some cases, gesture and movement coexist. Applause, for example, functions as a gesture of approval [23] while also requiring precise motor coordination, including rhythmic timing adjusted through auditory feedback. In this case, gesture as group intention guides physical execution as movement. This interaction operates at a neurosocial level and helps explain why spectators can synchronize clapping rapidly without prior training [23].

In children, gesture and movement are central to the development of symbolic and linguistic capacities through participation in joint attention with surrounding adults [2]. By around one year of age, children begin to produce different types of gestures to communicate specific motives, such as prompting adults to act in relation to an object or event (imperative motive) or sharing attention with them (declarative motive) [25]. From approximately 3 to 6 years of age, these abilities expand in complexity, as suggested by Nacher et al. [26]. Later, from 6 to 12 years of age, motor patterns tend to become more refined, with improvements in coordination and social skills. In this context, Lei et al. [27] developed and validated a method to assess motor coordination in children and found that automated systems could quantitatively assess gross motor skills and development in both educational and clinical settings.

The consolidation of basic motor skills and the acquisition of increasingly complex cognitive, social, and language competencies occur throughout childhood [28]. For this reason, the present review is primarily concerned with child-centred studies, with relevance to the preschool and primary-school years, while also capturing a broader developmental range represented in the included literature. This stage is especially important for cognitive, social, emotional, and physical development. Behavioural observation and analysis during childhood can help us understand the progression of gestural and motor abilities and explore their value as indicators of developmental markers, motivation, and potential difficulties.

Research on gestural expressiveness in children is increasing alongside new technological opportunities to detect, analyze, and transcribe gestures, behaviours, and reactions [29]. Recent studies have shown that prelinguistic gestures, such as intentional hand movements and pointing, are essential to children's communicative development. According to Samanta et al. [14], the emergence of these gestures predicts both linguistic ability and its development over time. Previous research has also suggested that adults' vocabulary size may be associated with the gestures they produced in infancy. This helps explain the current interest in analyzing gestures as early indicators of communicative development.

Tools capable of analyzing children's movements and gestures for educational, clinical, and health-related purposes constitute an important and rapidly developing line of research [14]. Automatic recognition offers a way to overcome the limitations of manual observation, which is resource-intensive, costly, and difficult to scale. It can provide

non-invasive and more precise information about children's movements, gestures, modes of action, and learning processes [1]. However, although technical accuracy can be high in controlled settings, the ecological validity of these systems in the complex settings of children's natural environments remains underexplored.

Automatic gesture recognition is a complex computational process through which computer systems autonomously identify, classify, and interpret specific patterns, such as images or sounds, in digital data by means of learning algorithms and mathematical models that transform complex signals into simpler and more meaningful representations [30]. For example, systems used to recognize physical activity in schools often combine accelerometers with machine learning. Ahmadi and Trost [31] show that these systems can achieve more than 80% accuracy, suggesting that wearable devices can provide reliable and objective monitoring adapted to daily activities for educational and health purposes. Similarly, Cao et al. [32] demonstrated that tools such as OpenPose are reliable for the automatic transcription of children's gestures, thereby expanding possibilities for research and intervention in real-world contexts.

It is also important that these technologies adapt to the changing cognitive and motor demands associated with different developmental stages. Studies by Anzulewicz et al. [33] and Javed et al. [34], both reporting accuracies above 90%, show that automated systems can distinguish between typically developing children and children with autism by analyzing gesture-related patterns on tablets.

### *1.1. State of the Art: Clinical and Educational Applications*

Most existing studies have focused on analyzing movement in clinical settings, particularly for the rehabilitation or assessment of patients with specific conditions. For example, Groos et al. [35] showed that an algorithm trained on smartphone videos could identify atypical infant movements with an assessment accuracy comparable to that of experts, suggesting its potential value for population screening programmes. By contrast, research in educational contexts remains limited. Nevertheless, automatic recognition in classrooms, play environments, or homes offers considerable potential, for example, for detecting patterns of social interaction or personalizing educational activities [35].

New and promising applications are continuing to emerge. For example, Samanta et al. [14] developed a system based on convolutional neural networks to analyze gestures in videos of child-caregiver interactions, thereby facilitating large-scale studies of communicative development. In addition, studies such as Liu et al. [36] show how movement-related data can be used to examine children's engagement and physical activity in applied settings. These findings suggest that approaches tailored to educational environments could support more objective analyses of children's motor development and generate useful feedback for professionals, caregivers, and intelligent systems designed to recognize gestural patterns and make predictions from these data.

### *1.2. Knowledge Gaps and Technical Challenges*

The recent literature shows that, despite important advances, significant gaps remain. Previous systematic studies on automatic gesture analysis only focus on hand gestures, and not specifically those of children [37,38]. Performance differences across age groups suggest the need for adaptive models, as algorithmic accuracy may vary with development [39]. In addition, most systems have been validated using small and relatively homogeneous samples, which limits their generalizability. Many studies also rely on a single source of analysis, providing limited evidence for performance in uncontrolled environments. By contrast, Nacher et al. [26] used a combination of devices, tactile sensors, and cameras to analyze gestures in preschool children.

### 1.3. Ethical Considerations and Future Perspectives

The integration of emerging technologies into educational and everyday contexts requires particularly careful attention to ethical issues, especially in relation to data privacy and informed consent. In this regard, Crescenzi-Lanna and Grané Oro [40] highlight the risks associated with the use of cameras in school settings without clear and robust protocols, as well as the urgent need to establish strict guidelines and ethical procedures to safeguard children's rights.

Gesture recognition can deepen understanding of how children express themselves and learn while also supporting the study of early communication skills and their potential educational applications [14]. This systematic review aims to map recent research in this area, with particular attention to non-clinical contexts, such as formal education and informal learning. Specifically, it examines the current approaches used for the automatic recognition of children's gestures and movements, the main domains of application in child behavioural analysis, and the principal technical barriers and ethical considerations associated with the implementation of these technologies. Therefore, from both an ethical and technical standpoint, AI systems designed for children should be developed within robust privacy and security frameworks to protect sensitive data [41].

Although the included studies focus on different behavioural signals (for example, whole-body movement, hand gestures, eye movements, or facial expressions), we consider them together because they all rely on automatic detection of children's observable behaviour. From a child-centred perspective, these modalities offer complementary information about motor, attentional, emotional, and social processes, so we treat them as converging approaches within the broader framework of automatic gesture and movement recognition.

The associated research questions (RQs) are as follows:

RQ1. What are the specific target groups (1a), purposes (1b), and settings (1c) of automatic gesture and movement recognition in research involving children?

RQ2. What performance results, including accuracy (2a), have been reported, and what limitations (2b) characterize the methods used to automatically recognize children's gestures or movements?

RQ3. What ethical aspects have been considered in the use of automatic gesture and movement recognition with children?

## 2. Materials and Methods

This systematic review follows the PRISMA framework (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [38] (see Figure 1) and draws on four databases: Web of Science (WoS), Scopus, PubMed, and IEEE Xplore. The search covered a five-year publication period, from 2021 to 2025. These databases were selected to reflect both the interdisciplinary scope of the topic (WoS and Scopus) and its domain-specific relevance to health and technology (PubMed and IEEE Xplore). The publication period was chosen to capture the rapid development of gesture and movement recognition techniques in recent research.

### 2.1. Search Strategy

The search strategy generated different sets of keyword strings (see Table 1). The Boolean operators "OR", "AND", and "NOT" were used to combine terms. The search was limited to studies published in English and was designed to cover the 2021–2025 period. Study selection was conducted in two phases. First, the titles and abstracts of all identified records were screened for relevance. Second, the full texts of potentially eligible studies were assessed according to the inclusion and exclusion criteria (see Table 2). To minimize

selection bias and ensure methodological rigour, both phases were performed independently by two researchers. Any disagreements regarding study eligibility were resolved through discussion and consensus or, when necessary, by consulting a third reviewer. The final set of keyword strings is presented in Table 1. The table below summarizes the search strings used in “All Fields” for Web of Science and in “All” for Scopus, PubMed, and IEEE Xplore. Three stages of keyword refinement were carried out across all databases. This process led to the final inclusion of 25 studies for qualitative synthesis. Although the review was designed around the 2021–2025 search window, a small number of included studies fall just outside this range because they remained relevant to the final screened corpus and were retained for consistency with the database retrieval and eligibility process.

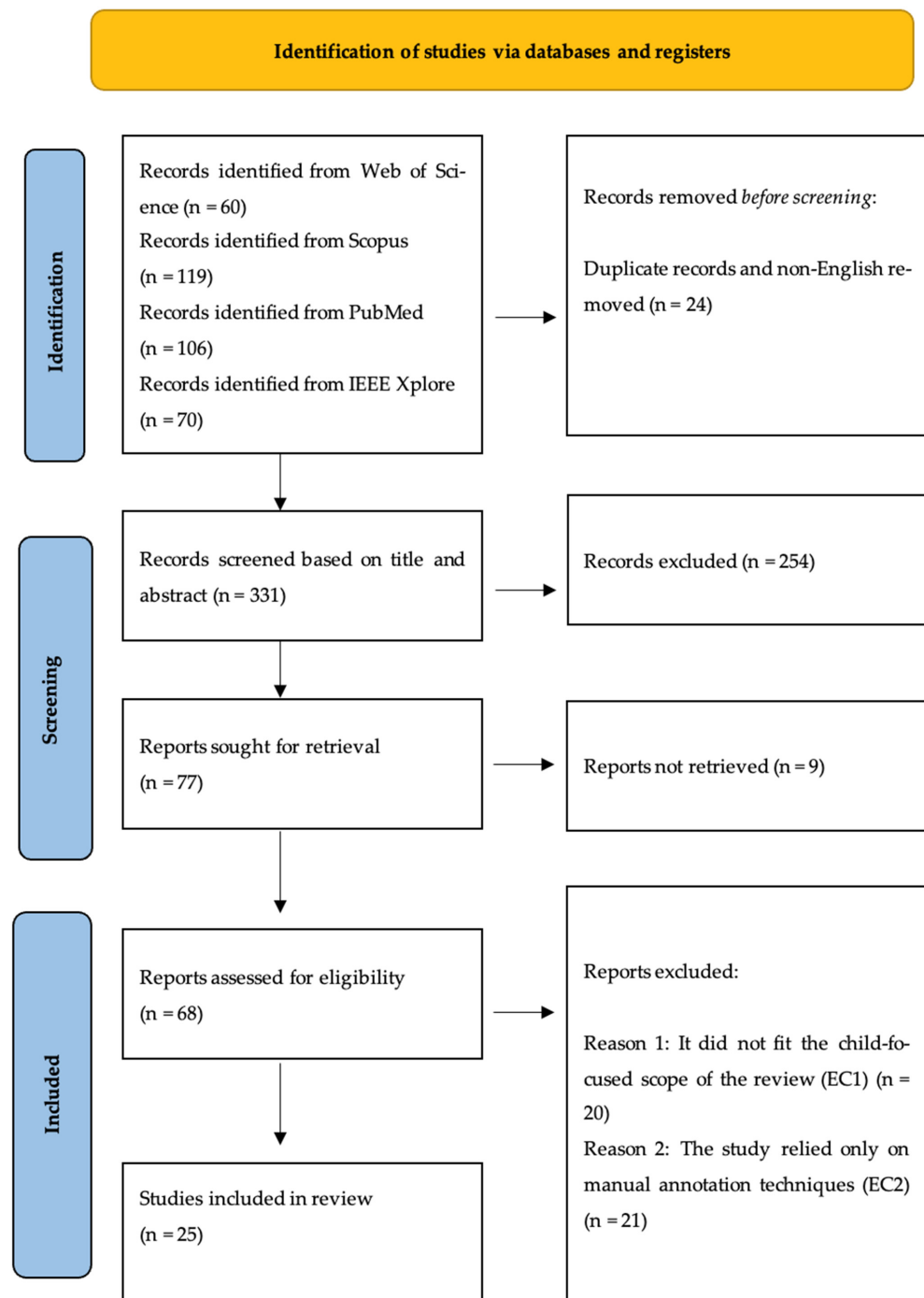


**Figure 1.** The PRISMA selection process with details of the samples.

**Table 1.** Databases, search strategy, and search strings.

| Database             | Word Strings                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Web of Science (WoS) | (ALL = ("child* movement") OR ALL = ("child* gesture") AND ALL = ("automatic detection") OR ALL = ("automatic recognition") AND ALL = ("behaviour analysis") NOT ALL = (marine) NOT ALL = (environment) NOT ALL = (medical) NOT ALL = (chronic OR traumatic) NOT ALL = (nature) NOT ALL = (sleep)) AND (PY == ("2025" OR "2024" OR "2023" OR "2022" OR "2021") AND DT == ("ARTICLE"))                                                                                                    |
| Scopus               | ALL (child* AND gesture) OR (child* AND movement) AND ("automatic detection") OR ("automatic recognition") AND (behaviour AND analysis) AND NOT (marine) AND NOT (environment) AND NOT (medical) AND NOT (chronic OR traumatic) AND NOT (nature) AND NOT (sleep) PUBYEAR > 2020 AND PUBYEAR < 2025 AND PUBYEAR > 2020 AND PUBYEAR < 2025 AND PUBYEAR > 2020 AND PUBYEAR < 2025 AND PUBYEAR > 2019 AND PUBYEAR < 2026 AND PUBYEAR > 2019 AND PUBYEAR < 2026 AND (LIMIT-TO (DOCTYPE,"ar")) |
| PubMed               | ("child" OR "children") AND ("movement" OR "gesture") AND ("automatic detection" OR "recognition" OR "tracking") AND ("pediatrics" OR "early childhood") AND ("2021"[PDAT] OR "2022"[PDAT] OR "2023"[PDAT] OR "2024"[PDAT] OR "2024"[PDAT])                                                                                                                                                                                                                                              |
| IEEE Xplore          | ("child" OR "children") AND ("movement" OR "gesture") AND ("automatic detection" OR "recognition OR "tracking") AND ("pediatrics" OR "early childhood") AND (2021 OR 2022 OR 2023 OR 2024 OR 2025)                                                                                                                                                                                                                                                                                       |

The inclusion criteria (IC) and exclusion criteria (EC) for the study are listed in Table 2.

**Table 2.** Inclusion and exclusion criteria.

| Inclusion Criteria                                                                       | Exclusion Criteria                                                                                      |
|------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| 1. The study and/or dataset includes children or child-relevant developmental stages.    | 1. The study does not include children, child-focused datasets, or child-relevant developmental stages. |
| 2. The study uses automatic recognition or detection technique/s of gesture or movement. | 2. The study uses manual annotations only.                                                              |
| 3. The form of publication is a scientific article.                                      | 3. The form of publication is not an article (e.g., conference proceedings, abstract, or book chapter). |
| 4. It is written in English.                                                             | 4. It is not written in English.                                                                        |

## 2.2. Selection Process

The general steps for identifying, selecting, and including studies are summarized in Figure 1, which presents the PRISMA flowchart for this review. In the identification stage, a total of 355 studies were retrieved from the selected databases. Before the screening stage, 24 records were excluded because they were duplicates or publications other than scientific articles. During screening, titles and abstracts were examined to assess their relevance to the review. Of the remaining 331 studies, 254 were excluded because they did not meet the inclusion criteria. Full-text versions of the remaining records were then retrieved and assessed in detail. A further 43 studies were excluded for three main reasons: they did not fit the child-focused scope of the review (EC1), the study relied only on manual annotation techniques (EC2), and/or the study was published outside of the search period (2021–2025) (EC5). Finally, 25 studies were selected for in-depth analysis. Each stage of the assessment was supervised and validated by an expert in child

psychology and child–computer interaction to ensure the quality of the studies included in the final analysis.

### 2.3. Risk-of-Bias Assessment

Quality appraisal was completed by two researchers and validated by an expert in child psychology. The first and second researchers conducted cross-checked screening of 77 studies before the eligibility assessment for full-text retrieval. All researchers reviewed the 25 included studies in strict accordance with the PRISMA guidelines and applied the inclusion and exclusion criteria. Although no specific inter-rater reliability statistic was calculated, any relevant disagreements were resolved through a transparent and systematic process based on discussion and consensus.

### 2.4. Data Extraction and Analysis

A standardized data extraction form was developed to ensure consistency across the included studies. The extracted variables were organized to address the research questions (RQs) directly.

To address RQ1, we extracted data on the target group (age range and population type), the purpose of the recognition system (e.g., social, educational, or clinical), and the physical setting in which the study was conducted (laboratory or naturalistic).

To address RQ2, we extracted performance results, specifically the reported accuracy of the models and other indicators of success (RQ2a), together with the technical and environmental limitations identified by the authors (RQ2b).

To address RQ3, we documented the presence or absence of ethical considerations, including informed consent, data privacy, and ethics committee approval. All details from the included studies were recorded in an Excel spreadsheet. The review involved a three-stage data extraction process to ensure the quality of the final selection: (1) all researchers contributed to developing the search strings, which were validated by two external experts in the field; (2) two researchers worked closely on the preliminary selection during the identification stage; and (3) an expert in child psychology supervised and validated the screening process before final selection. The recorded information covered descriptive characteristics of each study, sample and sampling information, data analysis techniques, and the specific gesture and movement variables examined.

Table 3 presents a descriptive summary of the included studies, detailing their citations, participant demographics—specifically sample size, age, and geographic location—the data types collected, the targeted gestures or movements, and the data analysis techniques or tools employed.

**Table 3.** Descriptive information of included entries.

| Study | Sample Size, Age and Location(s)                                 | Data Type                              | Gesture and/or Movement                 | Data Analysis Technique(s) or Tool(s)                   |
|-------|------------------------------------------------------------------|----------------------------------------|-----------------------------------------|---------------------------------------------------------|
| [42]  | 86 families<br>5 to 10 years old<br>Norway, Denmark, and Belgium | Video recording and body-tracking data | Spontaneous infant movements            | Automated body-tracking software and In-Motion-App      |
| [43]  | 10 children<br>2 to 5 years old<br>Sweden, Iraq, and Australia   | Video recording                        | Head movement and head control tracking | Sensor Fusion Algorithm (SFA) and Euler-based algorithm |
| [44]  | 3 to 5 years old<br>Sri Lanka                                    | Video recording                        | Motor skills detection                  | Computer vision techniques                              |

Table 3. Cont.

| Study | Sample Size, Age and Location(s)                                                      | Data Type                                                                | Gesture and/or Movement                                                                                         | Data Analysis Technique(s) or Tool(s)                                                                              |
|-------|---------------------------------------------------------------------------------------|--------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| [15]  | 405 children<br>2.5 to 4 years old<br>London (United Kingdom)<br>and Ontario (Canada) | Secondary dataset                                                        | Moderate-to-vigorous physical activity (MVPA), light-intensity physical activity (LPA), and sedentary behaviour | Statistical analysis tools: R software (v.4.1.3)                                                                   |
| [45]  | 80 children<br>18 to 59 months<br>Texas (United States of America/USA)                | Video recordings and baseline test                                       | Deictic, conventional, and iconic gestures                                                                      | Computerized Language Analysis software package                                                                    |
| [46]  | 281 children<br>3 to 5 years old<br>Virginia (USA)                                    | Video recordings                                                         | Deictic, iconic, metaphoric, symbolic, and beat gestures                                                        | Gesture coding protocol measuring frequency and type of gestures                                                   |
| [47]  | 141 mothers with their children<br>6 months to 2 years old<br>Arkansas (USA)          | Video recordings                                                         | Physical activity, sedentary behaviour, and sleep                                                               | Statistical tool: SAS software (version 9.4)                                                                       |
| [48]  | 100 children<br>3.5 to 11 years old<br>Switzerland and Canada                         | Video recordings                                                         | Hand and facial gestures                                                                                        | Eye tracking (The Observer XT software) and MATLAB R2019b (MathWorks)                                              |
| [27]  | Sample size and age are not specifically mentioned<br>Preschoolers<br>China           | Video dataset                                                            | Motor coordination                                                                                              | Pose estimation algorithm and key-points-based action quality assessment                                           |
| [49]  | 216 children<br>5 to 10 years old<br>China                                            | Behavioural tests data                                                   | Eye-movement patterns and gaze tracking                                                                         | Machine learning, 5-fold cross-validation                                                                          |
| [50]  | 335 infants<br>36 to 42 weeks old<br>USA                                              | Spatial and pressure data from a sensor mat, as well as video recordings | Cramped synchronized (CS) general movements—abnormal and spontaneous whole-body movements                       | Unsupervised machine learning (normalization, clustering) and supervised classification modelling (decision trees) |
| [51]  | Sample size is not specifically mentioned<br>3 to 12 years old<br>Philippines         | Eye-movement data (gaze tracking)                                        | Eye movement and fixation                                                                                       | CNN and K-means clustering                                                                                         |
| [52]  | Sample size and age are not specifically mentioned<br>China                           | Video recordings                                                         | Facial expression                                                                                               | Statistical analysis                                                                                               |
| [53]  | 35 children<br>4.5 years old<br>Michigan (USA)                                        | Video recordings                                                         | Locomotor skills, ball skills, and engagement                                                                   | Data Timer® app                                                                                                    |
| [54]  | Sample size and age are not specifically mentioned<br>India                           | Video recordings                                                         | Body pose estimation                                                                                            | Machine learning techniques                                                                                        |
| [55]  | 104 infants<br>3–17 months<br>Sweden                                                  | Video recordings of gaze behaviour and manual coding of gaze             | Head and eye movements in response                                                                              | Tablets (iPad) with front-facing camera + KiiTS App                                                                |
| [56]  | 40 children<br>10 to 14 years old<br>Aarhus (Denmark)                                 | Video recordings                                                         | Severe and extreme motions of head and chin movements                                                           | Statistical analysis                                                                                               |

Table 3. *Cont.*

| Study | Sample Size, Age and Location(s)                                           | Data Type                                | Gesture and/or Movement                                                  | Data Analysis Technique(s) or Tool(s)                                                                          |
|-------|----------------------------------------------------------------------------|------------------------------------------|--------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|
| [57]  | 95 children<br>4 to 12 years old<br>Texas (USA)                            | Video recordings                         | Ambulatory data                                                          | Statistical analysis                                                                                           |
| [58]  | Sample size is not<br>specifically mentioned<br>24 months<br>USA or Europe | Quantitative data<br>(tracking metrics)  | Eye movements                                                            | Measurements of latency (time to orient), fixation duration, and secondary fixation location with eye tracking |
| [59]  | 44 children<br>6 to 17 months<br>Stanford (USA)                            | Secondary dataset<br>(SEEDLings and CMU) | Body posture, touch and physical contact, and visibility engagement      | Computer vision and deep learning to extract 3D human meshes (HARMONI)                                         |
| [60]  | Seven children<br>5 to 21 years old<br>Japan                               | Video recordings                         | Matching-pair test using eye-gaze tracking                               | Eye-gaze system                                                                                                |
| [61]  | 178 children, age is not<br>specified<br>Osaka (Japan)                     | Video recordings                         | Emotional gait (anger, happiness, and sadness)                           | Statistical analysis                                                                                           |
| [62]  | 164 children<br>2 to 7 years old<br>Durham (United Kingdom)                | Quantitative movement and clinical tests | Free body movements within the center region of interest                 | Noldus EthoVision XT Version 14                                                                                |
| [63]  | 28 children<br>4 to 6 years old<br>China                                   | Video recordings and observational data  | Upright walking (UW), stoop walking (SW), & knee and hand crawling (KHC) | Statistical analysis and ErgoLab software                                                                      |
| [64]  | 330 children<br>1 to 3 years old Nanjing (China)                           | Quantitative data (tracking data)        | Eye-movement patterns classified as holistic and analytic                | Hidden Markov Modelling (HMM) with Variational Bayesian Expectation Maximization (EMHMM)                       |

### 3. Results

#### 3.1. RQ1. What Are the Specific Target Group (1a), Purpose (1b), and Setting (1c) of Automatic Gesture and Movement Recognition in Research with Children?

This review describes recent studies that use automatic gesture and movement recognition for child behavioural analysis. The developmental characteristics of children differ substantially from those of adults, which creates a need for techniques that are sensitive to child-specific developmental markers. Children use gestures well before they begin producing their first words, and their fine and gross motor skills present distinctive characteristics that require tailored analytical approaches. The results are organized to examine the relationships among specific target groups, research settings, and the emerging fields of application, together with their broader societal implications.

##### (1a) The specific target group

The classification of target age groups (see Figure 2) follows children's developmental markers across the ranges of 0–3 (infants and toddlers), 3–6 (preschools), and 6–12 (primary) (*x*-axis) and shows the frequency of each developmental group included in the review. However, the 25 included studies show mixed age groups because each study defined its age range differently. Organizing the studies in this way provides a more holistic overview of target participants in relation to developmental stage and helps identify possible links with the specific purposes and settings of the studies.

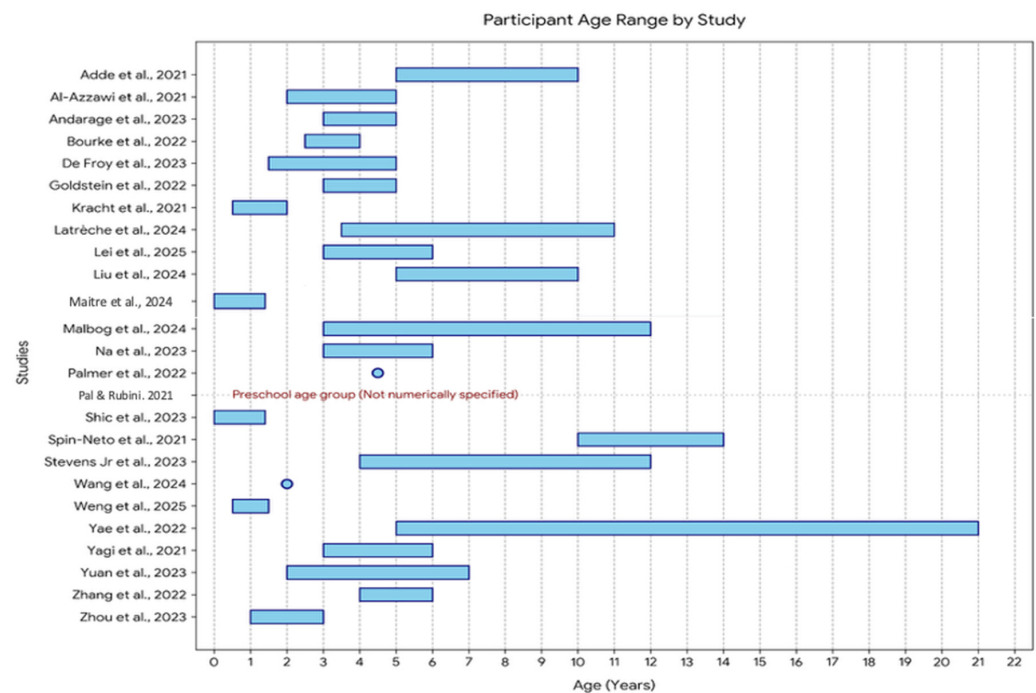


**Figure 2.** Distribution of studies by age category [15,27,42–64].

#### (1b) The purposes of using automatic gesture and movement recognition

After classifying the target groups, this section synthesizes the specific purposes of the studies that included children either as part of the dataset or as target participants in experimental research. Table 4 groups the studies by field and by specific purpose of gesture and motion recognition techniques. The findings reveal several key themes, which can be summarized as follows:

**Table 4.** Emerging research fields and specific purposes for using gesture and movement recognition techniques.

| Field                    | Study   | Specific Purposes                                                                                                                                                                                |
|--------------------------|---------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Clinical psychology      | [43]    | Measure head control and cervical range of motion (CROM) to evaluate accuracy and reliability in children with cerebral palsy using an inertial measurement unit (IMU).                          |
|                          | [48]    | Examine gesture imitation in children with ASD using eye tracking.                                                                                                                               |
|                          | [49]    | Support the auxiliary diagnosis of attention-deficit/hyperactivity disorder (ADHD) using eye tracking and machine learning based on visual attention patterns.                                   |
|                          | [51,54] | Assess attentive behaviour in children with ASD through a visual tracking model of eye movement.                                                                                                 |
|                          | [53]    | Measure children’s engagement during mastery motivational climate motor skill interventions.                                                                                                     |
|                          | [63]    | Identify gestures in children with ASD using person pose estimation. Evaluate the physiological burden associated with children’s movements during fire evacuation using wearable-based sensors. |
| Education and psychology | [42]    | Validate the use of a mobile app for remote infant assessment using movement-pattern detection.                                                                                                  |
|                          | [44]    | Assess children’s motor skills using the Early Childhood Action Monitoring and Analytics System (ECAMS), a video-based motion analysis approach.                                                 |
|                          | [15]    | Examine associations between children’s movement behaviour compositions, body mass index, and physical and psychosocial functioning.                                                             |

Table 4. Cont.

| Field                                       | Study   | Specific Purposes                                                                                                                                                                                    |
|---------------------------------------------|---------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Education and psychology                    | [45]    | Examine gesture rates in racially and ethnically diverse autistic children and their parents.                                                                                                        |
|                                             | [46]    | Explore preschool children’s gestures during pretend play with puppets and other embodied conditions.                                                                                                |
|                                             | [47]    | Examine movement behaviours and their associations with the home environment in early childhood.                                                                                                     |
|                                             | [57]    | Measure children’s ambulatory activity and heart rate during unstructured play in an inclusive playground.                                                                                           |
|                                             | [62]    | Explore associations between autistic children’s movement and location during free play and their social, communication, cognitive, adaptive, and language skills using automated movement tracking. |
|                                             | [64]    | Explore the relationship between eye-movement patterns and children’s cognitive ability.                                                                                                             |
| Healthcare                                  | [50,52] | Detect abnormal general movements to help predict neuromotor disorders.                                                                                                                              |
|                                             | [56]    | Explore facial-expression recognition for outpatient clinical contexts. Assess the frequency and characteristics of discomfort during cone beam CT examination simulation.                           |
|                                             | [65]    | Evaluate treatment-related performance in patients with spinal muscular atrophy using an eye-tracking system.                                                                                        |
| Research, development, and innovation (RDI) | [27]    | Assess the use of machine vision for preschoolers’ motor coordination and the improvement of objectivity and efficiency in assessment.                                                               |
|                                             | [55]    | Validate remote gaze-following assessment in infant twins.                                                                                                                                           |
|                                             | [59]    | Develop a three-dimensional computer vision and audio-processing method for caregiver–child behaviour and interaction analysis.                                                                      |
|                                             | [58]    | Investigate atypical face-scanning mechanisms in children with ASD (autism spectrum disorder) using eye tracking.                                                                                    |
|                                             | [61]    | Develop <i>ibuki</i> , a mobile robot with gait-induced emotional upper-body motion.                                                                                                                 |

As defined in the previous section, movement refers to physical displacement, whether in terms of fine motor skills (e.g., writing and object manipulation) or gross motor skills (e.g., jumping and running). Gestures are an integral part of communication because they encode spatial, temporal, and relational information and contribute to clearer and more efficient messages. The results show that five studies focus on movement, specifically eye movements, movement patterns, movement behaviour, and ambulatory activity, while only two studies focus on gestures.

Education and psychology are the predominant fields in the use of gesture and movement recognition, and they address a wide range of questions related to children’s physical activity and behaviour. The practical implications of studies in education and psychology are mainly directed towards improving children’s well-being and learning experiences in schools, homes, and public spaces.

The findings indicate that there is a need to bridge the laboratory-to-natural setting gap since there are overlapping cases between educational and clinical settings, as both may share similar social implications. In clinical psychology most studies focus on therapeutic or diagnostic issues related to specific conditions, such as autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), and cerebral palsy. Although ASD and ADHD are also common concerns in school-based interventions, the findings provide

insights from different types of data, including both multimodal and single-source data. Children's behaviours can be measured through distinct approaches, such as whole-body engagement or eye-movement tracking [33,39].

Classifying some studies under research, development, and innovation (RDI) is somewhat challenging, particularly when it is difficult to determine whether the focus is on assessing behavioural change after an intervention or on training a specific model or technique to achieve a certain level of accuracy or technical advancement. For this reason, the RDI category emerged to accommodate studies primarily concerned with advancing automatic gesture and movement recognition techniques. Six studies focused mainly on training with secondary data, such as video datasets, multimodal datasets, motion-tracking data, and behavioural datasets.

Finally, the findings show that healthcare includes the smallest number of studies using gesture or movement recognition with children. In this review, this category was defined narrowly to include specific health problems that go beyond general clinical settings and to highlight hospital-based treatment supported by innovative techniques.

#### (1c) The setting of the analyzed studies

Despite the predominance of studies in education and psychology (see Figure 3), a clear misalignment remains between research aims and research settings. There is a gap between laboratory and natural settings; research is rarely conducted in natural settings such as schools or playgrounds. Instead, laboratories remain the dominant context for data collection, even for educational research [33,39], which limits the ecological validity of findings intended to inform on children's real-world behaviour. This limitation is particularly relevant for preschool-aged children, as their behaviour may be strongly shaped by the conditions of a laboratory setting.

#### Research settings

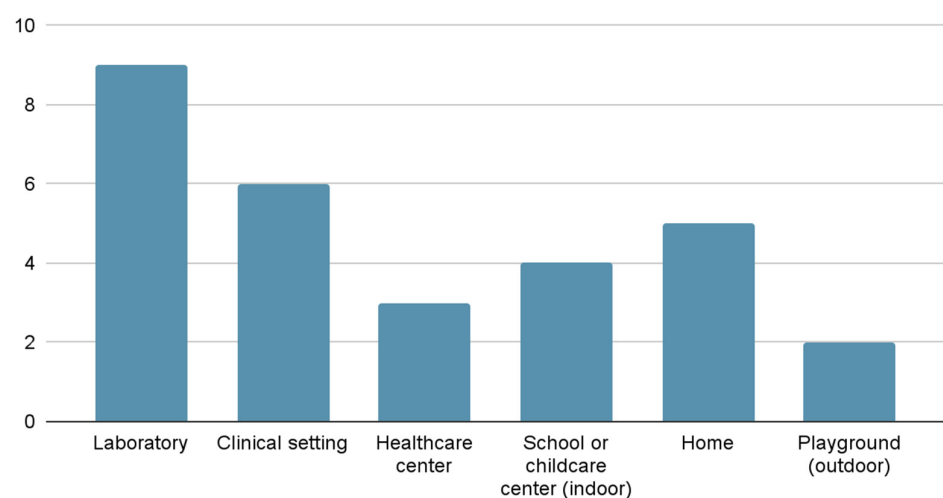


**Figure 3.** The diversification of research settings for automatic gesture and movement.

As discussed in Part 1b, the purposes of the studies were analyzed in relation to their implications for children's lives. From this standpoint, clinical studies show a clearer alignment between research purpose and research setting, as they are typically conducted in healthcare-related settings that correspond directly to their diagnostic or therapeutic aims.

The inclusion of home and playground settings further suggests that educationally relevant topics extend beyond formal learning environments. For example, a study conducted in inclusive outdoor community playgrounds found a strong positive correlation between playground time, understood as unstructured play, and total steps, ambulatory

time, and moderate-to-vigorous physical activity (MVPA). These findings were then used to examine the physical, social, and psychosocial benefits associated with children's play in naturalistic environments [40].

Overall, the findings indicate that research settings are becoming more diverse, but this diversification remains limited. Expanding automatic gesture and movement recognition into natural settings such as homes, classrooms, and playgrounds is essential to provide more representative evidence of children's everyday experiences and developmental contexts.

### *3.2. RQ2. What Are the Accuracy (2a) and Limitations (2b) of the Methods Used to Automatically Recognize Children's Gestures or Movement?*

Body expression and movement are fundamental to communication, socialization, and the early detection of possible alterations in children's neurodevelopment. Recent advances in computer vision, machine learning, and gesture analysis have led to increasingly accurate automatic methods for the recognition and quantification of gestures and movements in early childhood.

However, the literature also shows substantial methodological heterogeneity and, at the same time, a limited number of studies reporting accuracy metrics and clinical validity. Although high performance is reported for specific tasks, such as 96% accuracy for attentive behaviour through K-means clustering or 94% accuracy for atypical movement detection, tracking precision appears to depend strongly on the anatomical focus of analysis. For example, central body points such as the thorax (93%) appear to be tracked more reliably than distal extremities such as the wrists (74.4%), which are more prone to displacement.

Machine learning performance also varies considerably by model; comparative results suggest that Random Forest often outperforms convolutional neural networks and Support Vector Classification. In addition, the findings point to a clear ecological validity gap: systems that achieve over 84% consistency in controlled laboratory settings may decline to around 64% when compared with manual annotations in more varied or naturalistic contexts. Factors such as high child mobility and the need for continuous calibration can therefore limit the applicability and reliability of these techniques outside the laboratory.

Only 44% of the included studies reported specific accuracy metrics or clinical validity scores. Accuracy metrics are considered direct measurements of deep learning models performance to precisely detect and classify gestures. As shown in Table 5, there is also considerable variation in how researchers quantify the performance of automatic recognition systems. This diversity reflects a broader challenge in the field: the difficulty of establishing a universal ground truth for children's spontaneous behaviours, especially in natural environments. The results indicate that researchers assess technological performance through three main methodological lenses: percentage-based, which quantify the frequency of correct classifications, such as the 96% accuracy reported for attentive behaviour analysis; statistical significance measures ( $p$ -values), which determine whether a model's performance particularly in gesture and gait analysis is significantly better than chance, rather than providing a direct success rate; and reliability coefficients, such as Cohen's kappa, which measure the level of agreement between an automated system and human expert coding. Considering that these approaches capture different dimensions of performance, from raw computational precision to human-like interpretative consistency, they are not directly comparable, which partly obscures the objective technical standing of one system relative to another.

**Table 5.** Accuracy (2a) and limitations (2b) of the methods used.

| Study | Method and Its Accuracy Percentage                                                                                                                                                                                                                                                                                                                     | Limitations Identified by the Authors                                                                                                                                                                                                                                                |
|-------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [42]  | Method: Automated body-tracking software, general movement assessment, and In-Motion-App.<br>Accuracy: It consists of 7 body points, resulting in 83.15% prediction accuracy from 5493 videos. Percentage (%) for each point: nose (93.3), thorax (93), right wrist (74.4), left wrist (83.1), pelvis (63), right ankle (79.5), and left ankle (79.3). | Further analysis is needed for the selection of body points, tracked body point accuracy and object movement.                                                                                                                                                                        |
| [43]  | Method: Sensor Fusion Algorithm (SFA) and Euler-based algorithm.<br>Accuracy: It uses an IMU (an inertial sensor to detect children's head movements) with 81% accuracy.                                                                                                                                                                               | It lacks qualitative metrics, has no real-time feedback, and is insufficiently validated in homes or outdoor settings.                                                                                                                                                               |
| [44]  | Method: Computer vision techniques, machine learning techniques, and TGM2D (Test of Gross Motor Development-2nd Edition) toolkit.<br>Accuracy: It is based on the Euclidean distance as a proportion of 0.05 diagonal of the human skeleton bounding box.                                                                                              | It only works for explicitly supervised motor activities due to the small dataset used to train the model.                                                                                                                                                                           |
| [15]  | Method: Accelerometer and statistical analysis tools: R software (v.4.1.3).<br>Accuracy: The optimal movement behaviour is based on moderate-to-vigorous physical activity (MVPA), which results in 11.7 min/h, with sedentary time at 40.28 min/h and low levels of light-intensity physical activity (LPA) at 9.02 min/h.                            | It is necessary to measure children's movements outside the research setting (childcare), which may give an insightful comparison with the results.                                                                                                                                  |
| [45]  | Method: Computerized Language Analysis software package.<br>Accuracy: The children's gestures were measured with reliability (Cohen's Kappa), $M = 0.92$ .                                                                                                                                                                                             | The authors mention the lack of statistical power for their pairwise comparisons of parent-child gestures across race/ethnicity groups.                                                                                                                                              |
| [46]  | Method: Gesture coding protocol for frequency and type of gestures.<br>Accuracy: The frequencies across gesture types and embodiment conditions were measured using one-way ANOVA, which reveals that children gesture more with puppets and costumes ( $p < 0.001$ ) than with no embodiment at all.                                                  | It lacks measurement of fully physical embodied actions.                                                                                                                                                                                                                             |
| [47]  | Method: Pediatric Review of Children's Environmental Support and Stimulation (PROCESS) and statistical tool: SAS software (version 9.4).<br>Accuracy: The variables were measured using Correlation Coefficients at 6 months and 2 years, which showed $p = 0.65$ for physical activity and $p = 0.86$ for sleep guideline.                            | It lacks variation in settings, which are solely at home, and predominantly in the homes of white and middle- or low-income families.                                                                                                                                                |
| [48]  | Method: Eye tracking and behavioural tests of imitation.<br>Accuracy: Cohen's kappa was 0.944 for frequency and 0.841 for accuracy.                                                                                                                                                                                                                    | The small group of typically developed (TD) participants implies a lack of imitation during the experiment.                                                                                                                                                                          |
| [27]  | Method: A pose estimation algorithm and key-points-based action quality assessment.<br>Accuracy: It is based on the measurement of the predicted score and true score, with a mean error value of 0.071 and a standard deviation of 0.074.                                                                                                             | It lacks age-specific tasks and task sensitivity and underutilizes eye-tracking data.                                                                                                                                                                                                |
| [49]  | Method: Machine learning and eye tracking.<br>Accuracy: It reports an accuracy of almost 91% and an ROC (Receiver Operating Characteristic) of 0.96 in machine learning and eye tracking.                                                                                                                                                              | It has limitations due to excessive head movements or lack of collaboration with individuals, and its applicability is also limited to the laboratory. In addition, it is highly dependent on good calibration and technical supervision to ensure the quality of the data obtained. |
| [50]  | Method: Unsupervised machine learning and pressure sensors.<br>Accuracy: The authors developed a deep learning model for the detection of certain abnormal movements in neonates using videos from smartphones. This showed an AUC of 0.80 and a sensitivity of 76%.                                                                                   | It is limited to hospitals, requires that the individual be lying down, is dependent on the placement of the sensors and does not work for activities of daily living.                                                                                                               |
| [51]  | Method: CNN and K-means clustering.<br>Accuracy: The K-means model resulted in 96% accuracy and 94% recall.                                                                                                                                                                                                                                            | It can be improved by expanding the number of ASD participants and age groups with more diverse backgrounds.                                                                                                                                                                         |

Table 5. Cont.

| Study | Method and Its Accuracy Percentage                                                                                                                                                                                                                                                                                                                                                                        | Limitations Identified by the Authors                                                                                                                                                                               |
|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [52]  | Method: Statistical analysis.<br>Accuracy: Children’s facial expressions are classified into four descriptive levels (i.e., happy, sad, surprised, and scared).                                                                                                                                                                                                                                           | It lacks quantitative data to show the accuracy levels of facial recognition.                                                                                                                                       |
| [53]  | Method: Data Timer® app.<br>Accuracy: Children engage in appropriate behaviours for around 36% (motor skills intervention).                                                                                                                                                                                                                                                                               | It does not address individual factors, initial skills, sex or the quality of engagement.                                                                                                                           |
| [54]  | Method: Machine learning techniques and pose estimation (OpenPose).<br>Accuracy: The model with the Random Forest technique resulted in 83% accuracy and 71% recall.                                                                                                                                                                                                                                      | It can be improved by adding more gesture classes and including speech and video to merge with real-time gestures.                                                                                                  |
| [55]  | Method: Tablets (iPad) with front-facing camera and KiiTS App.<br>Accuracy: It performs remote assessment of gaze tracking through tablets. It presents good consistency with eye-tracking systems of traditional clinical methods and validates them for large-scale studies. In uncontrolled environments, it presents high success, which makes it possible to process eye-tracking data in real time. | It relies on child cooperation and is not validated in natural settings; video quality is dependent on uncontrolled factors; the recruitment rate is low; and automated metrics are lacking.                        |
| [56]  | Method: Statistical analysis.<br>Accuracy: Video observation provides 80% movement detection.                                                                                                                                                                                                                                                                                                             | It can be improved by considering efficient head immobilization devices and data correction algorithms.                                                                                                             |
| [57]  | Method: Statistical analysis.<br>Accuracy: The ambulatory activity is classified into easy or moderate with a percentage between age groups.                                                                                                                                                                                                                                                              | The playground is designed for children aged 2–10 years; hence, older children are rarely present.                                                                                                                  |
| [58]  | Method: Eye-tracking technology with a Tobii Pro X3-120 device.<br>Accuracy: The Welch two-sample <i>t</i> -test revealed a similar number of valid trials across groups (Expression × Guiding condition) with $p = 0.206$ .                                                                                                                                                                              | Limitations include data exclusion, exclusive laboratory validation, calibration errors, high sensitivity to head movements, sample biases and static stimuli.                                                      |
| [59]  | Method: Computer vision and deep learning to extract 3D human meshes (HARMONI).<br>Accuracy: It achieves 63.8 to 80% consistency with manual annotations and 84% to 93% consistency with laboratory setting videos (on touch and visibility). Lastly, the audio analysis achieves 72.8% prediction for adults’ word count using LENA.                                                                     | The dataset has limitations in terms of diversity, sample size, and range of child development.                                                                                                                     |
| [60]  | Method: Eye-tracking system and statistical analysis.<br>Accuracy: The matching-pair tests resulted in 45.8% in week 24 for seven patients from baseline.                                                                                                                                                                                                                                                 | A small number of patients participated in the study; a wide age range, a short observation period, and the mood swings of participants are also cited as limitations.                                              |
| [61]  | Method: Mobile Android (Ibuki) and statistical analysis.<br>Accuracy: The recognition rate was 56% for anger, 77.7% for happiness, and 97% for sadness. The human gait motions resulted in recognition rates of 97% for anger, 93.2% for happiness, and 96.2% for sadness.                                                                                                                                | It only uses online video surveys with a predesigned response range of questionnaires.                                                                                                                              |
| [62]  | Method: Noldus EthoVision XT Version 14.<br>Accuracy: The first video tracking revealed higher expressive vocabulary ( $p = 0.026$ ) and fewer autism symptoms ( $p = 0.033$ ) in the centre of ROI (region of interest).                                                                                                                                                                                 | It lacks a comparison sample of neurotypical controls and the measurement of children’s skills during observation.                                                                                                  |
| [63]  | Method: Statistical analysis and ErgoLab software.<br>Accuracy: Children’s evacuation velocity and physiological responses are classified into knee and hand crawling (KHC), stoop walking (SW), and upright walking (UW), which are all increased from baseline values (i.e., 78.88%, 88.03%, and 105.72% for KHC).                                                                                      | It only focuses on electrodermal activity (EDA) and skin temperature (SKT) without considering other physiological factors (i.e., respiratory exchange and oxygen consumption).                                     |
| [64]  | Method: Eye tracking and HMM.<br>Accuracy: It uses machine learning and relates eye patterns to cognitive development in children. It has a high accuracy of 83–90% and identifies specific patterns that are linked to memory or attention. Accuracy is high but varies with the age of the child.                                                                                                       | This does not consider body gestures, only eye tracking (gaze patterns); its application is limited if children have a lot of motor activity. It was not evaluated in natural environments, only in the laboratory. |

For this part of the review, the included studies were examined in terms of the performance information they reported for automatic recognition methods. As Table 5 shows, the evidence base is heterogeneous: some studies provide direct accuracy, whereas others report related indicators, such as reliability coefficients, consistency with manual annotation, descriptive performance outcomes, or inferential statistics. This variation reflects the current state of the field and helps explain why technical comparison across studies remains difficult.

This approach helps ensure that the analysis and conclusions of this article remain grounded in the range of performance evidence currently available on automatic gesture and movement recognition methods for children, while also making visible the substantial inconsistencies in reporting practices across the literature. In turn, this better aligns the review with its objectives and research questions.

### *3.3. RQ3. What Ethical Aspects Have Been Considered in Using Automatic Gesture and Movement Recognition for Children?*

The ethical implications of using automatic gesture and movement recognition are discussed in some of the studies involving direct human participation. Only 10 of the 25 studies reported their ethical procedures and data protection measures to ensure privacy and safety throughout the research process [42,43,48–50,61,63]. Privacy-preserving deep learning approaches, such as differential privacy-based frameworks for video- and sensor-based systems, offer concrete strategies for mitigating these risks and should be considered in future research on gesture and motion recognition focused on children [65]. The ethical concerns addressed, either explicitly or implicitly, can be grouped into four main areas: (1) personal data protection and safety, (2) justification of societal contribution, (3) bias and fairness, and (4) privacy-preserving methods within the secure AI framework.

All studies involving children under 18 required parental or guardian consent before data collection, which is a common safeguard in research with children. At the same time, children should be given explanations that are appropriate to their level of understanding so that they can better comprehend their rights regarding the data collected during a study. Research on explainable artificial intelligence (AI) for children may help support their more active participation in understanding data protection and informed involvement in research processes [66].

The second ethical concern relates to the justification of societal contribution. Some studies are more strongly oriented towards technological advancement, which can make their direct significance for children's well-being less clear. The development of new technologies often requires large training datasets and diverse participant groups to address concerns about bias and fairness. However, the value of technological innovation also depends on its contextual relevance and practical usefulness for children's lives.

Third, bias and fairness are recurring concerns when datasets include children. Underrepresentation of communities, languages, and age groups may reduce developmental appropriateness and introduce bias into classification systems. Even within primary research, age ranges are often divided differently across studies, which can create measurement bias when behavioural cues are matched to specific psychological or developmental conditions. In this respect, combining primary data with secondary data for training purposes may help mitigate some of the biases in automatic gesture and movement recognition systems for children.

Finally, privacy-preserving strategies within a secure AI framework should be implemented both before and throughout the study, given the sensitivity of using video recordings of children's data to train deep learning models. In line with Wu et al. [67], this should include privacy-risk assessment and prevention-control mechanisms to strengthen data protection. Similarly, Wang et al. [65] emphasize that privacy-preserving deep learning

methods are particularly crucial for automatic gesture and movement recognition systems that depend on video-based, sensor-based, or deep learning technologies in schools, homes, and other naturalistic child-centred environments.

#### 4. Discussion

This systematic review identifies a substantial misalignment between current research on the automatic recognition of children's gestures and movements, established principles of psychological adjustment, and the laboratory-to-natural setting gap, understood here as children's adaptive functioning within naturalistic environments. Although 78% of the analyzed studies were conducted primarily in laboratories, many focused on educational, psychological, or clinical applications. This concentration limits the transferability of findings to real-world contexts, where behavioural adjustment unfolds dynamically. This concern is consistent with recent research on emotion recognition using computer vision [68,69]. The gap is further compounded by the frequent use of small and relatively homogeneous samples [70].

A central finding of this review is the technical disparity between tracking physical displacement (movement) and interpreting semantic meaning (gesture). The systems analyzed generally achieve high accuracy when monitoring stable central anatomical points, such as the thorax, but their precision decreases substantially when they attempt to infer the semiotic intent of gestures [42]. This pattern supports the conceptual distinction proposed by Tomasello and Camaioni [20] between biological movement as kinematics and communicative intent as meaning-bearing action. Translating this distinction into computational systems remains a major challenge.

Several technical limitations help explain why systems that perform well in controlled settings often become less reliable in natural settings. For example, Passmore et al. [16] reported 94% accuracy for detecting atypical movements in controlled clinical conditions, yet such performance is difficult to maintain in everyday contexts. Three factors appear particularly important. First, developmental variability reduces algorithmic stability in younger populations, as systems must adapt to rapid changes in motor and attentional control; Zhou et al. [64] identified specific challenges related to developmental differences in oculomotor control and attention. Second, methodological constraints remain substantial. As shown in Table 5, error patterns frequently reflect failures in naturalistic settings, where excessive participant mobility and the need for continuous calibration reduce performance; this is evident in studies such as Liu et al. [49] and Al-Azzawi et al. [43]. Third, ethical requirements also shape what can be studied and how. Collecting data in natural environments demands rigorous privacy protections, particularly in settings such as schools, where obtaining informed consent for camera-based monitoring can be especially challenging [71].

A further gap persists between research objectives and research settings. In educational research, for example, academic performance is often prioritized over the spontaneous emotional and social adjustment that occurs during learning interactions. As Table 5 indicates, technological adaptations are still concentrated in controlled educational and clinical contexts, and when these systems are deployed in less regulated environments, technical and methodological difficulties may introduce additional measurement bias. Future research should therefore prioritize the use of non-invasive technologies and context-sensitive protocols—such as accelerometers and RGB-D cameras—in naturalistic environments, including classrooms, playgrounds, and homes. Expanding empirical datasets involving children is essential for improving model adaptability and technical accuracy across developmental stages. Addressing these challenges will require a multidisciplinary approach that integrates expertise from education, child psychology, and computer science to ensure that automated systems are both developmentally appropriate and scientifically robust. At

the same time, the implementation of strong ethical frameworks remains essential to protect children's rights while promoting transparency and fairness in data use and algorithmic design [70]. Beyond basic consent and anonymization, future work on child-centred gesture and movement recognition should follow clearer data-governance practices, with policies that specify who can access the data, how long it is kept, and under what conditions it is shared. It is also important to monitor algorithmic fairness so that system performance does not systematically differ between age groups, developmental profiles, or sociocultural contexts. Finally, secure AI frameworks that integrate privacy-preserving methods, such as zero-trust architectures, dynamic security, or differential-privacy-based deep learning approaches [65,67], will be essential for protecting sensitive video and sensor data in real-world child-centred settings.

## 5. Conclusions

Automatic gesture and movement recognition has been widely applied to adult populations in public, clinical, and laboratory settings for purposes such as surveillance, health monitoring, and human–computer interaction. More recently, growing interest has turned toward adapting these technologies to the behavioural analysis of children because of their potential value in educational, developmental, healthcare, and social contexts. Despite this momentum, research involving children still relies heavily on manual observation and expert annotation, which remain labour-intensive and difficult to scale. As a result, only a limited body of research has specifically examined automatic gesture and movement recognition in child populations.

This review provides a systematic evaluation of the current state of automatic gesture and movement recognition in child behavioural analysis. Following the PRISMA guidelines, it examined studies retrieved through searches designed to cover the 2021–2025 period across four major databases: Web of Science (WoS), Scopus, PubMed, and IEEE Xplore. The review analyzed the purposes for which these technologies are being used, the age groups they target, the settings in which they are implemented, their reported accuracy and technical limitations, and the ethical considerations associated with their use. In total, 27 studies met the inclusion criteria and were included in the analysis. The final sample spans a broader developmental range than the core preschool-to-primary focus of the review, reflecting the diversity of the literature captured during screening.

The findings show that automatic behaviour recognition in children is being applied across multiple domains, including healthcare, clinical and developmental psychology, education, and child-centred technological innovation. However, the evidence base remains uneven. Fewer than half of the reviewed studies reported detailed accuracy percentages or clear information about methodological limitations, highlighting a substantial gap in technical reporting and practical evaluation. The review also identified a persistent tension between promising performance in controlled environments and the limited ecological validity of these systems in children's everyday settings.

Ethical considerations are equally central to the development of this field. The use of automated behaviour recognition with children raises important concerns regarding data protection, privacy, fairness, and the justification of societal benefit. Future research should therefore adopt a more purpose-driven and child-centred approach, with stronger ethical safeguards and greater attention to developmental appropriateness. It should also prioritize naturalistic settings such as classrooms, playgrounds, and homes while using non-invasive technologies capable of capturing authentic social interaction and behaviour. Advancing the field will depend not only on improving technical performance but also on ensuring that these systems are transparent, context-sensitive, and genuinely supportive of children's health, development, and learning.

**Author Contributions:** Conceptualization, A.U., D.M.-D. and L.C.-L.; methodology, A.U. and L.C.-L.; validation, L.C.-L.; formal analysis, A.U. and D.M.-D.; investigation, L.C.-L., A.U. and D.M.-D.; resources, L.C.-L.; data curation, A.U. and D.M.-D.; writing—original draft preparation, A.U. and D.M.-D.; writing—review and editing, A.U., D.M.-D. and L.C.-L.; visualization, A.U.; supervision, L.C.-L.; project administration, L.C.-L.; funding acquisition, L.C.-L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work has been partially supported by Internal Grant of the Predoctoral Program, Academic Year 2023/2024, funded by the Foundation of the Open University of Catalonia and grant RYC2021-031152-I, funded by MCIN/AEI/10.13039/501100011033.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Liszkowski, U.; Carpenter, M.; Henning, A.; Striano, T.; Tomasello, M. Twelve-month-olds point to share attention and interest. *Dev. Sci.* **2004**, *7*, 297–307. [\[CrossRef\]](#)
2. Tomasello, M. *Constructing a Language: A Usage-Based Theory of Language Acquisition*; Harvard University Press: Cambridge, MA, USA, 2003.
3. Pollux, P.M. Age-of-Actor Effects in Body Expression Recognition of Children. *Acta Psychol.* **2021**, *220*, 103421. [\[CrossRef\]](#)
4. Kushawaha, A. AI and Non-Verbal Communication: Enhancing Understanding of Emotional Cues for Hearing Impairment Children. In *Transforming Learning: The Power of Educational Technology*; Blue Rose Publishers: Noida, India, 2024. [\[CrossRef\]](#)
5. Rudenko, I.; Rudenko, A.; Lilienthal, A.J.; Arras, K.O.; Bruno, B. The Child Factor in Child–Robot Interaction: Discovering the Impact of Developmental Stage and Individual Characteristics. *Int. J. Soc. Robot.* **2024**, *16*, 1879–1900. [\[CrossRef\]](#)
6. Figueiredo, A.R.; Pereira, A.; Frias, F.; Rodrigues, L.M.D.; Diogo, P. Applications of Artificial Intelligence in Emotion Recognition in Pediatrics Health Care: Scoping Review. *J. Pediatr. Nurs.* **2025**, *85*, 593–606. [\[CrossRef\]](#)
7. Kendon, A. *Gesture Visible Action as Utterance*, 1st ed.; Cambridge University Press: Cambridge, UK, 2004.
8. Novack, M.A.; Goldin-Meadow, S. Gesture as Representational Action: A Paper about Function. *Psychon. Bull. Rev.* **2017**, *24*, 652–665. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Radman, Z. *The Hand, an Organ of the Mind: What the Manual Tells the Mental*; MIT Press: Cambridge, MA, USA, 2013.
10. Shin, S.Y.; Rowe, M.L.; Lee, H.S. Early Gesture Use Predicts Children’s Language Development in South Korea: New Evidence Supporting the Cross-cultural Importance of Pointing. *Infancy* **2024**, *29*, 327–354. [\[CrossRef\]](#)
11. McNeill, D. *Hand and Mind: What Gestures Reveal About Thought*; University of Chicago Press: Chicago, IL, USA, 1992.
12. Bozkurt, E.; Yemez, Y.; Erzin, E. Affective Synthesis and Animation of Arm Gestures from Speech Prosody. *Speech Commun.* **2020**, *119*, 1–11. [\[CrossRef\]](#)
13. Lopez-Ozieblo, R. Proposing a Revised Functional Classification of Pragmatic Gestures. *Lingua* **2020**, *247*, 102870. [\[CrossRef\]](#)
14. Samanta, S.; Bannard, C.; Pine, J. The Language05 Team. Can Automated Gesture Recognition Support the Study of Child Language Development? In Proceedings of the 42nd Annual Meeting of the Cognitive Science Society, Austin, TX, USA, 29 July–1 August 2020; pp. 343–349.
15. Bourke, M.; Vanderloo, L.M.; Irwin, J.D.; Burke, S.M.; Johnson, A.M.; Driediger, M.; Timmons, B.W.; Tucker, P. Association between Childcare Movement Behaviour Compositions with Health and Development among Preschoolers: Finding the Optimal Combinations of Physical Activities and Sedentary Time. *J. Sports Sci.* **2022**, *40*, 2085–2094. [\[CrossRef\]](#)
16. Passmore, E.; Kwong, A.L.; Greenstein, S.; Olsen, J.E.; Eeles, A.L.; Cheong, J.L.Y.; Spittle, A.J.; Ball, G. Automated Identification of Abnormal Infant Movements from Smart Phone Videos. *PLoS Digit. Health* **2024**, *3*, e0000432. [\[CrossRef\]](#)
17. Chu, S.K.; Jayabalan, P.; Kibler, W.B.; Press, J. The Kinetic Chain Revisited: New Concepts on Throwing Mechanics and Injury. *PM&R* **2016**, *8*, S69–S77. [\[CrossRef\]](#)
18. Almansoof, H.S.; Nuhmani, S.; Muaidi, Q. Role of Kinetic Chain in Sports Performance and Injury Risk: A Narrative Review. *J. Med. Life* **2023**, *16*, 1591–1596. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Grogans, S.E.; Bliss-Moreau, E.; Buss, K.A.; Clark, L.A.; Fox, A.S.; Keltner, D.; Cowen, A.S.; Kim, J.J.; Kragel, P.A.; MacLeod, C.; et al. The Nature and Neurobiology of Fear and Anxiety: State of the Science and Opportunities for Accelerating Discovery. *Neurosci. Biobehav. Rev.* **2023**, *151*, 105237. [\[CrossRef\]](#)
20. Matsumoto, D.; Hwang, H.C. Cultural Similarities and Differences in Emblematic Gestures. *J. Nonverbal Behav.* **2013**, *37*, 1–27. [\[CrossRef\]](#)

21. Tomasello, M.; Camaioni, L. A Comparison of the Gestural Communication of Apes and Human Infants. *Hum. Dev.* **1997**, *40*, 7–24. [\[CrossRef\]](#)
22. Sørvoll, M.; Øberg, G.K.; Girolami, G.L. Motor Control and Skill Acquisition in Pediatric Physical Therapy: An Enactive Proposal. *Front. Psychol.* **2023**, *14*, 1226593. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Menti, A.B.; Rosemberg, C.R. The Role of Gestures in the Construction of Meaning in Classroom Context. *Íkala Rev. Leng. Cult.* **2017**, *22*, 455–475. [\[CrossRef\]](#)
24. Néda, Z.; Ravasz, E.; Brechet, Y.; Vicsek, T.; Barabási, A.-L. The Sound of Many Hands Clapping. *Nature* **2000**, *403*, 849–850. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Bates, E.; Camaioni, L.; Volterra, V. The Acquisition of Performatives Prior to Speech. *Merrill-Palmer Q. Behav. Dev.* **1975**, *21*, 205–226.
26. Nacher, V.; Jaen, J. Evaluating the Accuracy of Pre-Kindergarten Children Multi-Touch Interaction. In *Human-Computer Interaction—INTERACT 2015*; Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., Winckler, M., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 9297, pp. 549–556. [\[CrossRef\]](#)
27. Lei, Y.; Shu, D.; Yu, M.; Shi, D.; Li, J.; Chen, Y. Evaluation Method of Motor Coordination Ability in Children Based on Machine Vision. *Tsinghua Sci. Technol.* **2025**, *30*, 633–649. [\[CrossRef\]](#)
28. Konrad, J.D.; Marrus, N.; Lohse, K.R.; Thuet, K.M.; Lang, C.E. Motor Competence Is Related to Acquisition of Error-Based but Not Reinforcement Learning in Children Ages 6 to 12. *Heliyon* **2024**, *10*, e32731. [\[CrossRef\]](#)
29. Ward-Miller, S.; Marino, P.; Casey, D.; Levy, J.; Reeves, M.; Veirun, M.; Fitzpatrick, J.J. Use of puppets to teach resourcefulness skills to women who self-injure: An exploratory study. *Arch. Psychiatr. Nurs.* **2019**, *33*, 94–101. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Adugna, T.D.; Ramu, A.; Haldorai, A. A Review of Pattern Recognition and Machine Learning. *J. Mach. Comput.* **2024**, *4*, 210–220. [\[CrossRef\]](#)
31. Ahmadi, M.N.; Trost, S.G. Device-Based Measurement of Physical Activity in Pre-Schoolers: Comparison of Machine Learning and Cut Point Methods. *PLoS ONE* **2022**, *17*, e0266970. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.-E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 172–186. [\[CrossRef\]](#)
33. Anzulewicz, A.; Sobota, K.; Delafield-Butt, J.T. Toward the Autism Motor Signature: Gesture Patterns during Smart Tablet Gameplay Identify Children with Autism. *Sci. Rep.* **2016**, *6*, 31107. [\[CrossRef\]](#)
34. Javed, H.; Lee, W.; Park, C.H. Toward an Automated Measure of Social Engagement for Children with Autism Spectrum Disorder—A Personalized Computational Modeling Approach. *Front. Robot. AI* **2020**, *7*, 43. [\[CrossRef\]](#)
35. Groos, D.; Adde, L.; Støen, R.; Ramampiaro, H.; Ihlen, E.A.F. Towards Human-Level Performance on Automatic Pose Estimation of Infant Spontaneous Movements. *Comput. Med. Imaging Graph.* **2022**, *95*, 102012. [\[CrossRef\]](#)
36. Liu, Z.; Li, J.; Zhang, Y.; Wu, D.; Huo, Y.; Yang, J.; Zhang, M.; Dong, C.; Jiang, L.; Sun, R.; et al. Auxiliary Diagnosis of Children with Attention-Deficit/Hyperactivity Disorder Using Eye-Tracking and Digital Biomarkers: Case-Control Study. *JMIR MHealth UHealth* **2024**, *12*, e58927. [\[CrossRef\]](#)
37. Hashi, A.O.; Hashim, S.Z.M.; Asamah, A.B. A Systematic Review of Hand Gesture Recognition: An Update from 2018 to 2024. *IEEE Access* **2024**, *12*, 143599–143626. [\[CrossRef\]](#)
38. Yassen, M.; Jusoh, S. A Systematic Review on Hand Gesture Recognition Techniques, Challenges and Applications. *PeerJ Comput. Sci.* **2019**, *5*, e218. [\[CrossRef\]](#)
39. Abed Ibrahim, L.; Fekete, I. What Machine Learning Can Tell Us About the Role of Language Dominance in the Diagnostic Accuracy of German LITMUS Non-Word and Sentence Repetition Tasks. *Front. Psychol.* **2019**, *9*, 2757. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Lanna, L.C.; Oro, M.G. Touch Gesture Performed by Children under 3 Years Old When Drawing and Coloring on a Tablet. *Int. J. Hum.-Comput. Stud.* **2019**, *124*, 1–12. [\[CrossRef\]](#)
41. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *Syst. Rev.* **2021**, *10*, 89. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Adde, L.; Brown, A.; Van Den Broeck, C.; DeCoen, K.; Eriksen, B.H.; Fjørtoft, T.; Groos, D.; Ihlen, E.A.F.; Osland, S.; Pascal, A.; et al. In-Motion-App for Remote General Movement Assessment: A Multi-Site Observational Study. *BMJ Open* **2021**, *11*, e042147. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Al-azzawi, S.S.; Khaksar, S.; Hadi, E.K.; Agrawal, H.; Murray, I. HeadUp: A Low-Cost Solution for Tracking Head Movement of Children with Cerebral Palsy Using IMU. *Sensors* **2021**, *21*, 8148. [\[CrossRef\]](#)
44. Andarage, I.S.N.; Fernando, D.; Lokuarachchi, B.A.; Athuluwage, M.G.; Wijewickrama, P. Early Childhood Action Monitoring and Analytics System (ECAMS). In *Proceedings of the 2023 IEEE 28th Pacific Rim International Symposium on Dependable Computing (PRDC)*, Singapore, 18–21 December 2023; pp. 346–351.
45. De Froy, A.; Rollins, P.R. The cross-racial/ethnic gesture production of young autistic children and their parents. *Dev. Lang. Impair.* **2023**, *8*, 1–17. [\[CrossRef\]](#)

46. Goldstein, T.R.; Stutesman, M.; Thompson, B. Moving with Puppets: Preschool Children's Gesture with Puppets during Pretense. *Cogn. Dev.* **2022**, *63*, 101198. [[CrossRef](#)]
47. Kracht, C.L.; Hendrick, C.; Lowe, A.; Roman, H.; Staiano, A.E.; Katzmarzyk, P.T.; Beyl, R.; Redman, L.M. Evaluation of indoor activities for moderate-to-vigorous physical activity in preschoolers. *J. Sports Sci.* **2024**, *42*, 1776–1784. [[CrossRef](#)]
48. Latrèche, K.; Kojovic, N.; Pittet, I.; Natraj, S.; Franchini, M.; Smith, I.M.; Schaer, M. Gesture Imitation Performance and Visual Exploration in Young Children with Autism Spectrum Disorder. *J. Autism Dev. Disord.* **2026**, *56*, 695–708. [[CrossRef](#)]
49. Liu, F.; Qiu, K.; Wang, H.; Dong, Y.; Yu, D. Decreased Wrist Rotation Imitation Abilities in Children with Autism Spectrum Disorder. *Front. Psychiatry* **2024**, *15*, 1349879. [[CrossRef](#)]
50. Maitre, N.L.; Kjeldsen, C.P.; Duncan, A.F.; Guzzetta, A.; Jeanvoine, A. Automated detection of abnormal general movements from pressure and positional information in hospitalized infants. *Pediatr. Res.* **2025**, *97*, 598–607. [[CrossRef](#)] [[PubMed](#)]
51. Malbog, A.D.V.; Irinco, A.J.M.; Morales, R.J.L.; Salas, L.T.; Maaño, R.A.; Belleza, R.C.; Panganiban, J.F.B. Implementation of Visual Tracking Model for Assessing Attentive Behavior in Children with Autism Spectrum Disorder Through Eye-Movement. In Proceedings of the 2024 IEEE International Conference on E-health Networking, Application & Services (HealthCom), Nara, Japan, 18–20 November 2024; pp. 1–4.
52. Na, L.; Wang, Y.; Li, X. Child facial recognition technology in early assist and intervention in hospital outpatient clinics. In Proceedings of the 2024 IEEE 7th Eurasian Conference on Educational Innovation (ECEI), Bangkok, Thailand, 26–28 January 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 105–108. [[CrossRef](#)]
53. Palmer, K.K.; Cox, E.R.; Scott-Andrews, K.Q.; Robinson, L.E. Structured observations of child behaviors during a mastery-motivational climate motor skill intervention: An exploratory study. *Int. J. Environ. Res. Public Health* **2022**, *19*, 15484. [[CrossRef](#)] [[PubMed](#)]
54. Pal, M.; Rubini, P. Gesture Recognition for Autistic Children Using Person Pose Estimation and Supervised Learning. In Proceedings of the 2021 IEEE 3rd PhD Colloquium on Ethically Driven Innovation and Technology for Society, Bangalore, India, 13 November 2021; pp. 1–2.
55. Shic, F.; Dommer, K.J.; Benton, J.; Li, B.; Snider, J.C.; Nyström, P.; Falck-Ytter, T. Remote, Tablet-Based Assessment of Gaze Following: A Nationwide Infant Twin Study. *Front. Psychol.* **2023**, *14*, 1223267. [[CrossRef](#)] [[PubMed](#)]
56. Spin-Neto, R.; Hauge Matzen, L.; Hermann, L.; de Carvalho E Silva Fuglsig, J.M.; Wenzel, A. Head motion and perception of discomfort by young children during simulated CBCT examinations. *Dentomaxillofacial Radiol.* **2021**, *50*, 20200445. [[CrossRef](#)]
57. Stevens, W.R.; Borchard, J.M.; Sleeper, P.; Dempsey, D.; Jeans, K.A.; Jo, C.-H.; Tulchin-Francis, K. Inclusive Community Playgrounds Benefit Typically Developing Children: An Objective Analysis of Physical Activity. *Front. Sports Act. Living* **2023**, *4*, 1100574. [[CrossRef](#)]
58. Liu, L.; Li, S.; Tian, L.; Yao, X.; Ling, Y.; Chen, J.; Wang, G.; Yang, Y. The impact of cues on joint attention in children with autism spectrum disorder: An eye-tracking study in virtual games. *Behav. Sci.* **2024**, *14*, 871. [[CrossRef](#)]
59. Weng, Z.; Bravo-Sánchez, L.; Wang, Z.; Howard, C.; Xenochristou, M.; Meister, N.; Kanazawa, A.; Milstein, A.; Bergelson, E.; Humphreys, K.L.; et al. Artificial intelligence-powered 3D analysis of video-based caregiver-child interactions. *Sci. Adv.* **2025**, *11*, eadp4422. [[CrossRef](#)]
60. Yae, Y.; Yuge, K.; Maeda, T.; Ichinose, F.; Matsuo, M.; Kobayashi, O.; Okanari, K.; Baba, Y.; Yonee, C.; Maruyama, S.; et al. Exploratory evaluation of an eye-tracking system in patients with advanced spinal muscular atrophy type I receiving nusinersen. *Front. Neurol.* **2022**, *13*, 918255. [[CrossRef](#)]
61. Yagi, S.; Nakata, Y.; Nakamura, Y.; Ishiguro, H. Perception of Emotional Expression of Mobile Humanoid Robot Using Gait-Induced Upper Body Motion. *IEEE Access* **2021**, *9*, 124793–124804. [[CrossRef](#)]
62. Yuan, A.; Sabatos-DeVito, M.; Bey, A.L.; Major, S.; Carpenter, K.L.H.; Franz, L.; Howard, J.; Vermeer, S.; Simmons, R.; Troy, J.; et al. Automated movement tracking of young autistic children during free play is correlated with clinical features associated with autism. *Autism* **2023**, *27*, 2530–2541. [[CrossRef](#)]
63. Zhang, B.; Gao, X.; Zhou, J.; Jia, X. The Associations between Evacuation Movements and Children's Physiological Demands Analyzed via Wearable-Based Sensors. *Sensors* **2022**, *22*, 8094. [[CrossRef](#)]
64. Zhou, S.; Hou, L.; Wang, N.; Liu, F.; Wei, N.; Chi, X.; Yu, D.; Zhang, X.; Tong, M. Using Machine Learning to Explore the Characteristics of Eye Movement Patterns and Relationship with Cognition Ability of Chinese Children Aged 1–6 Years. *Front. Hum. Neurosci.* **2023**, *17*, 1220178. [[CrossRef](#)]
65. Wang, H.; Zhang, X.; Xia, Y.; Wu, X. A differential privacy-preserving deep learning caching framework for heterogeneous communication network systems. *Int. J. Intell. Syst.* **2022**, *37*, 11142–11166. [[CrossRef](#)]
66. Wu, X.; Zou, B.; Lu, S.; Lu, C.; Wang, L.; Zhang, Y.; Wang, H. Dynamic security computing framework with zero trust based on privacy domain prevention and control theory. *IEEE J. Sel. Areas Commun.* **2025**, *43*, 2266–2278. [[CrossRef](#)]
67. Khosravi, H.; Shum, S.B.; Chen, G.; Conati, C.; Tsai, Y.-S.; Kay, J.; Knight, S.; Martinez-Maldonado, R.; Sadiq, S.; Gašević, D. Explainable Artificial Intelligence in Education. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100074. [[CrossRef](#)]

68. Contreras-Higuera, W.; Crescenzi-Lanna, L.; Ventura, C. High-Fidelity Temporal Modeling of Facial Expression: A Wavelet and LSTM Approach on Action Units Sequences. In *Proceedings of the 3rd International Workshop on Multimodal and Responsible Affective Computing, Dublin, Ireland, 27 October 2025*; ACM: New York, NY, USA, 2025; pp. 137–145.
69. Contreras-Higuera, W.; Crescenzi-Lanna, L. The Role of Time in Facial Dynamics and Challenges in Automatic Emotion Recognition (2019–2024). *Hum. Behav. Emerg. Technol.* **2025**, *2025*, 7777949. [[CrossRef](#)]
70. Schwab, J.F.; Lew-Williams, C. Language Learning, Socioeconomic Status, and Child-directed Speech. *WIREs Cogn. Sci.* **2016**, *7*, 264–275. [[CrossRef](#)]
71. Marín, V.I.; Tur, G. Ethical Issues in the Use of Technologies in Education Settings: A Scoping Review. *Educ. Knowl. Soc. EKS* **2024**, *25*, e31301. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.