information

Review

Computer Vision Tasks for Ambient Intelligence in Children's Health

Danila Germanese [1], Sara Colantonio [1,*], Marco Del Coco [2], Pierluigi Carcagnì [2] and Marco Leo [2]

[1] Institute of Information Science and Technologies (ISTI), National Research Council (CNR), Via G. Moruzzi 1, 56124 Pisa, Italy; danila.germanese@isti.cnr.it
[2] Institute of Applied Sciences and Intelligent Systems (ISASI), National Research Council (CNR), Via Monteroni snc University Campus, 73100 Lecce, Italy; marco.delcoco@cnr.it (M.D.C.); pierluigi.carcagni@cnr.it (P.C.); marco.leo@cnr.it (M.L.)
* Correspondence: sara.colantonio@isti.cnr.it

Abstract: Computer vision is a powerful tool for healthcare applications since it can provide objective diagnosis and assessment of pathologies, not depending on clinicians' skills and experiences. It can also help speed-up population screening, reducing health care costs and improving the quality of service. Several works summarise applications and systems in medical imaging, whereas less work is devoted to surveying approaches for healthcare goals using ambient intelligence, i.e., observing individuals in natural settings. Even more, there is a lack of papers providing a survey of works exhaustively covering computer vision applications for children's health, which is a particularly challenging research area considering that most existing computer vision technologies have been trained and tested only on adults. The aim of this paper is then to survey, for the first time in the literature, the papers covering children's health-related issues by ambient intelligence methods and systems relying on computer vision.

Keywords: computer vision; ambient intelligence; body motion analysis; facial expression recognition; children's healthcare

Citation: Germanese, D.; Colantonio, S.; Del Coco, M.; Carcagnì, P.; Leo, M. Computer Vision Tasks for Ambient Intelligence in Children's Health. *Information* **2023**, *14*, 548. https://doi.org/10.3390/info14100548

Academic Editor: Xin Ning

Received: 31 July 2023
Revised: 15 September 2023
Accepted: 3 October 2023
Published: 6 October 2023

1. Introduction

Computer vision (CV) offers powerful tools to assist healthcare applications, especially when coupled with artificial intelligence and machine learning. CV applications can provide objective evidence of the presence of pathologies or an assessment that is not dependent on clinicians' skills and experiences. They can also help speed-up population screening reducing health care costs and improving the quality of service [1].

According to the related scientific literature [2], there are two distinct levels in which CV can be effectively exploited: physician-level diagnostics (medical imaging) and medical scene perception (ambient intelligence). In medical imaging, the interior body is represented for clinical diagnosis and medical intervention, whereas ambient intelligence covers techniques aimed at recognizing human activity and their physical, motor and mental status while moving and acting in physical spaces.

In its broader sense, ambient intelligence is an umbrella term that encompasses intelligent and ubiquitous sensing, smart computing, and human-centred interfaces combined to deliver environments that are sensitive and responsive to people's presence and activities. In healthcare, ambient intelligence can refer to a continuous, non-invasive awareness of activity and health status of individuals, patients and people in need in a physical space that can assist doctors, nurses and other healthcare workers with clinical tasks such as patient monitoring, automated documentation and protocol compliance monitoring [3,4]. Cameras and visual sensors are key ingredients of ambient intelligence, as they convey precious information about the activity and the behaviour of people in an environment [3]. Visual data can be also processed to unobtrusively measure individuals' vital signs and to

support visual analyses of disease signs and symptoms [5]. CV tools come into play here as key enablers of physicians' and caregivers' tasks based on visual inspection.

Regarding the related scientific literature, whereas several works summarise applications and systems in medical imaging [6], less work is devoted to surveying approaches for ambient intelligence [7,8]. More importantly, most of the current literature focuses on CV for ambient intelligence in adult and older adult care, whereas there is a lack of papers that comprehensively review work on CV for children's health. This is an emerging and cogent topic, which is receiving a growing attention by health organizations and healthcare institutions lately [9,10]. The perspectives that ambient intelligence and innovative health technologies may open in paediatric care are manifold and can strongly benefit from research and technology advancement [10,11]. Particularly, CV coupled with artificial intelligence and machine learning can support several clinical tasks, for disease detection or well-being monitoring. Among them, the clinical tasks most commonly considered comprise detection and assessment of

- Neurocognitive impairment (e.g., based on Prechtl General Movement Assessment—GMA) or early signs of neurocognitive developmental disorders (e.g., Autism Spectrum Disorders—ASD or Attention Deficit Hyperactivity Disorders—ADHD).
- Dysmorphisms (e.g., cleft lip) or physical or motor impairments (e.g., gait and walking disorders) due to genetic disorders or surgery.
- The well-being and health status of newborns (e.g., vital signs and sleep monitoring in the nursery or in the Neonatal Intensive Care Unit—NICU) and children.

To support these clinical tasks, CV tools need to be able to perform low-level tasks such as face detection and head pose estimation, gaze tracking and analysis, motion detection and tracking (e.g., legs and arms), and measurement of physiological signs (e.g., heart rate). These low-level tasks underpin more complex inferences for the detection and assessment of activities, vital signs or disease symptoms.

So far, some survey papers have analysed the state of the art on one specific low-level task (e.g., body motion, gaze tracking, head pose estimation), which might be related to neurological diseases or motor impairments. For instance, methods and systems aimed at an early neurological disorder diagnosis have been recently collected in [12]. Prechtl general movement assessment by CV was summarized in [13,14]. A review of works dealing with gait deviations (also in children) in individuals with intellectual disabilities has been proposed in [15]. Data-driven detection techniques that quantify behavioural differences between autism cases and controls are reported in [16,17].

This paper aims to fill the aforementioned gap by providing, for the first time in the literature, a comprehensive overview of the papers covering children's health-related issues by ambient intelligence methods and systems relying on computer vision. A coarse taxonomy for the paper can be recovered by dividing works according to the part they concentrate on, e.g., the face for extracting gaze direction and facial expressions, or on the whole body, e.g., for gait analysis, posture estimation, and human–object interaction.

The proposed taxonomy is schematized in Figure 1 and the paper is then arranged accordingly as follows: Section 2 discusses works that introduced CV-based systems relying on tasks related to children's head and face such as face analysis and head-pose estimation; whereas Section 3 deals with works involving CV tasks aimed at analysing the human body or part of it. These sections map each CV method to the clinical problem it addresses, providing the reader with the background clinical motivation.

Section 4 then discusses some challenges to be addressed to reach a level of performance that allows the instruments to be effectively used in clinical practice and, finally, Section 5 will conclude the paper.

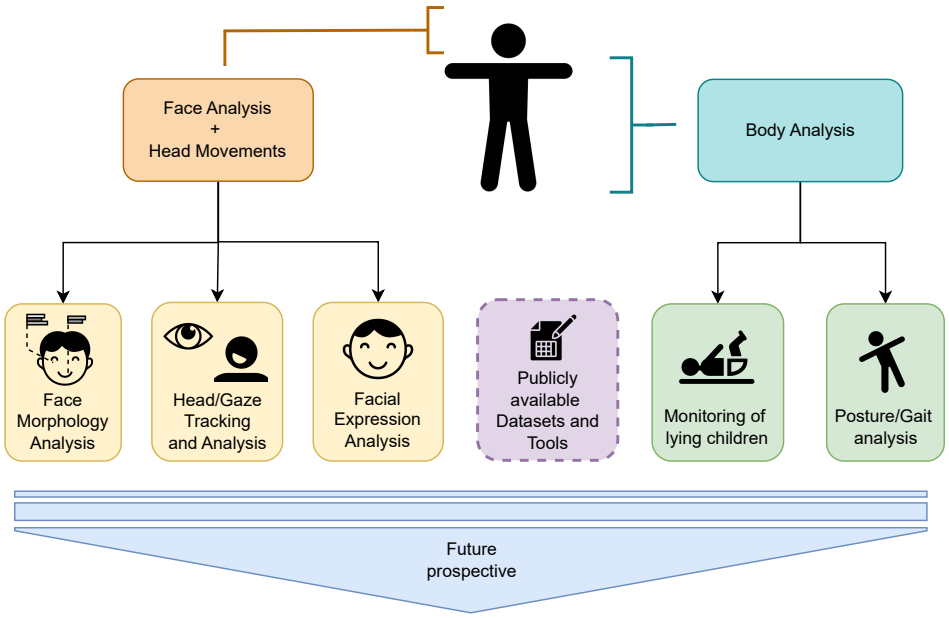

Figure 1. Schema of the proposed taxonomy.

2. Face Analysis and Head Movements

Children's faces contain a variety of valuable information regarding their state of health. Indeed, due to physiological or behavioural responses, certain pathological conditions alter the expression or appearance of children's faces.

Contactless approaches, such as computer vision methods, may detect and analyse the most relevant facial features, thus providing clinicians (or parents, teachers, caregivers, etc.) with unobtrusive and objective information on children's health status.

In the literature, many efforts have been made in this field, documented by a plethora of research papers that have been reported and discussed in this section. The studies range from the analysis of children's face morphology to recognize genetic disorders, to head and gaze tracking as a tool for large-scale screening of neurocognitive problems, to children's facial expression recognition.

For each reported work, the clinical aim, the used methods, the performances and eventually the limits of the study have been pointed out.

A selection of papers was initially made by using the following queries in the research databases:

- **Scopus**
  QUERY "TITLE-ABS-KEY ((newborn OR baby OR children OR toddler OR infant) AND (face OR facial) AND (analysis OR detection OR recognition OR tracking) AND "computer vision" AND PUBYEAR > 2014 AND PUBYEAR < 2024 that returned 158 documents;
- **Web of Science Core Collection**
  ((((ALL=(children)) OR ALL=(infant)) OR ALL=(baby)) OR ALL=(newborn)) AND ((ALL=(face)) OR ALL=(facial)) AND ((ALL=(analysis)) OR ALL=(detection) OR ALL=(recognition)) AND ALL=(computer) AND ALL=(vision)), refined in the YEARS from 2015 to 2023, that returned 197 documents;
- **Scholar**
  allintitle: children OR newborn OR babies OR infants OR face OR facial OR analysis OR recognition OR detection OR tracking OR "computer vision", refined in the YEARS from 2015 to 2023, that returned 158 documents.

Among all the scientific papers retrieved from the above-mentioned databases, a further selection was conducted based mainly on the scientific content (many works were in fact not relevant for the purposes of the proposed survey), the type of publication

(journals were preferred to conferences in case of comparable ideas) and finally on the number of citations (articles prior to 2020 with less than 10 citations were not considered).

The remaining documents were split depending on the task: (i) analysis of the morphology of the face (see Section 2.1); (ii) head pose estimation and/or gaze tracking (see Section 2.2); and (iii) facial expression recognition (see Section 2.3).

In the following subsections, some tables have been used to organize information within them. Section 2.4 concentrates on papers exploiting multimodal data to capture and analyse as many aspects of face-related behaviours. In each table, (i) the proposed computer vision approach used to perform the analysis, (ii) the clinical task, (iii) the obtained performance and (iv) the test data population in terms of cardinality and age of children are reported. Regarding the performance, the scores and the metrics released by the authors are reported. In the case of non-quantitative measurements, the term 'qualitative' has been used in the relative table cell. Then, Section 2.5 reports some datasets that have been made public and available to researchers and data scientists to enable them to train and validate their methods. Finally, in Section 2.6, the most recent and promising methods that attempt to address automatic face analysis challenges are introduced and discussed.

2.1. Face Morphology Analysis

Facial morphology refers to a series of many different complex traits, each influenced by genetic and environmental factors. In Table 1, a summary of the selected work for face morphology analysis is reported.

Table 1. Summary of the selected work for face morphology analysis. 'acc' = accuracy (correct predictions/total number of predictions with respect to the clinical goal occurrences tested); CNN = Convolutional Neural Network; RMSE = root-mean-square error (it measures the differences between values predicted by a model and the values provided by experts); SVR = Support Vector Regression.

Work (Year)	Method	Clinical Task	Metrics	Dataset Population/Age (h = hours, w = weeks, m = months, y = years)
[18,19] (2014, 2018)	Extraction of features related to nasolabial symmetry	Quantification of facial asymmetry in children pre- and post- primary cleft lip repair	Qualitative	50 infants and 50 children (age: 8–10 y)
[20,21] (2016)	Geometrical approach + landmarks identified by computer-based template mesh deformation	Quantification of facial asymmetry in children with unilateral cleft lip nasal deformity	Qualitative + Symmetry Scores	49 infants (age 4–10 m)
[22] (2019)	Face2Gene CLINIC app (based on a CNN)	Recognition of facial dysmorphisms due to genetic disorders	acc = 72.5%	51 children
[23] (2019)	CNN + SVR	Estimation of postnatal gestational age	7.98 days RMSE	130 newborns (gestational age: 28–40 w)

In the literature, an early interesting clinical task has been the **quantification of facial asymmetry** in children with unilateral cleft lip nasal deformity. The authors in [20,21] developed a computer vision-based method using a template mesh deformed to fit a target mesh using a geometric point detector. The clinical task was the quantification of facial asymmetry in children with unilateral cleft lip nasal deformity. To accomplish that, the authors (1) identify the three-dimensional midfacial plane in children with an unrepaired cleft lip, (2) quantify nasolabial symmetry (by assessing, for instance, symmetry scores for cleft severity) and (3) determine the correlation of these measures to clinical expectations. A total of 35 infants (ages 4 to 10 months) with unrepaired unilateral cleft lips and 14 infant

controls were enrolled in this study. Significant differences in symmetry scores were found between cleft types, and before and after surgery.

Also, Mercan and colleagues [18,19] aimed at developing a computer vision-based approach to analyse 3D facial images of 50 infants and 50 children (aged 8–10 years) before and after primary cleft lip repair. They assessed a specific set of features related to unilateral cleft lip nasal deformity: dorsal deviation, columellar deviation, nasal tip asymmetry, and blunting of the alar-cheek junction. They also showed a correlation between this set of measures related to nasolabial symmetry and aesthetic appraisal, demonstrating that computer vision analysis techniques can quantify nasal deformity at different stages.

Another application for automatic facial morphology analysis is the **estimation of the postnatal gestational age**, to assess whether or not infants are premature, which helps clinicians to decide on suitable post-natal treatment. The work of Torres et al. [23] focused on the development of a novel system for postnatal gestational age estimation using small sets of images of a newborn's face, foot and ear. A Convolutional Neural Network with two-stage architecture predicts broad classes of gestational age; then, it fuses the outputs of these discrete classes with the baby's weight to make fine-grained predictions of gestational age using Support Vector Regression.

Recently, most studies aimed at analysing children's face morphology by computer vision methods focus on recognising facial dysmorphisms due to **genetic disorders** instead. This is a complex recognition problem: several genetic disorders can cause facial dysmorphism that can eventually be combined with dysfunctions in other organs [24]. Based on facial features, a geneticist or a paediatrician can reach a possible diagnosis and order appropriate tests for confirmation of the same. Nonetheless, while some of these syndromes can be associated with distinctive facial features, others can be harder to detect at first sight. A computer vision approach, aimed at automatically analysing the face of children with facial dysmorphism, may avoid a delay in diagnosis by supporting geneticists and paediatricians in recognizing the facial gestalt of genetic syndromes. In [22], the authors aim at testing a computer vision approach to identify dysmorphic syndromes in Indian children. Fifty-one children with definite chromosomal abnormalities or microdeletion/duplication syndromes, or single gene disorders, with recognizable facial dysmorphism were enrolled in the study. Their facial photographs (frontal and lateral) were uploaded in the Face2Gene CLINIC app [25], where a deep convolutional neural network compares a patient's gestalt to its database for syndrome suggestion. Of the 51 patients, the software predicted the correct diagnosis in 37 patients (72.5%). The method works quite well to classify facial dysmorphism available during training. The open challenge becomes then to handle "unseen" cases since there is a vast number of genetic disorders causing dysmorphism and providing all of them during model training becomes unrealistic [26].

2.2. Head and Gaze Tracking and Analysis

In Table 2, a summary of the selected work for head pose estimation and/or gaze tracking is reported.

Problems in **neurocognitive development**, ASD in particular, are associated with disorders in the processing of social information, difficulties in social interaction, and atypical attention and gaze patterns. Atypical eye gaze is an early-emerging symptom of ASD and holds promise for autism screening. Traditionally, gaze tests rely on manual assessments of children's visual fixations to pictorial stimuli, but are very time-consuming and difficult to standardize.

Many studies aimed at developing faster and low-cost solutions to reproduce the two principal visual-based ASD clinical diagnostic tests: (i) the analysis of gaze fixation patterns, which represent the region of an individual's visual focus and (ii) the analysis of visual scanning methods, which corresponds to the way in which individuals scan their surrounding environment [27–29]. For instance, the framework and computational tool proposed in [27] for measuring attention was tested on a population of 104 children (age 16–31 months), 22 of them diagnosed with ASD. The computer vision algorithm

detailed in [30] was used to automatically track children's gaze and head position from a recorded video. The latter was registered using an iPad front-facing camera while they watched a movie displaying dynamic, social and non-social stimuli on the device screen. The authors detected and tracked 51 facial landmarks, thus allowing for the detection of head, mouth, and eye position to assess the direction of attention. They estimated the head positions relative to the camera by computing the optimal rotation parameters between the detected landmarks and a 3D canonical face model. The study showed that children in the ASD group paid less attention to the video stimulus and to the social as compared to the non-social stimuli, and often fixated their attention on one side of the screen.

Table 2. A partial summary of the selected work for head pose estimation and/or gaze tracking. 'acc' = accuracy (correct predictions/total number of predictions with respect to the clinical goal occurrences tested); CNN = Convolutional Neural Network; R-CNN = Region proposal CNN [31]; ResNet-101 = Residual Network with 101 layers [32].

Work (Year)	Method	Clinical Task	Metrics	Dataset Population/Age (h = hours, w = weeks, m = months, y = years)
[33] (2019)	OpenFace	Early detection of ASD signs	Qualitative	6 children
[27] (2021)	Computation of 51 facial landmark + computation of rotation parameters between the landmarks and a 3D canonical face model	ASD diagnosis	Qualitative	104 toddlers (age: 16–31 m)
[34] (2021)	Faster R-CNN algorithm to fine-tune a pre-trained ResNet-101	Monitoring of paediatric patients in critical settings	acc = 84%	59 paediatric patients

Also, robot-assisted tools have been of interest in intervention for children with ASD, showing impressive results both in the diagnosis and therapeutic intervention when compared to classical methods. The study reported in [33] aimed at early detecting ASD signs in naturalistic behavioural observation through child–robot interaction. The proposed system is composed of a responsive robotic platform, a flexible and scalable vision sensor network, and an automated face analysis algorithm based on machine learning models. The latter is developed using state-of-art trained neural models, available by Dlib3 [35] and OpenFace [36] and involves face detection, recognition, segmentation and tracking, landmarks detection and tracking, head pose, eye gaze and visual focus of attention estimation. The authors also present a proof-of-concept test, with the participation of three typically developing children and three children at risk of suffering from ASD.

Gaze detection and tracking may also be useful to **monitor paediatric patients**, especially in critical settings (e.g., Intensive Care Unit, ICU). The authors in [34] used the Faster R-CNN algorithm to fine-tune a pre-trained ResNet-101 model [37] to automatically detect and track eye regions for paediatric ICU patients monitoring. The last two layers of the CNN were fine-tuned during training with 59 images and annotations for the eye and mouth regions. The mouth landmark was included to improve model performance: it was found in earlier testing that the mouth and eyes were often confused by object detectors because of their similar shape and intensity profile on the face. By explicitly training the model to detect both landmarks, the mouth serves as negative training data for the eye localisation task. With a localization rate of 84%, this study demonstrated the potential of convolutional neural networks for eye localization and tracking in a paediatric ICU setting.

2.3. Facial Expressions Analysis

In Table 3, a summary of the selected work for facial expressions analysis is reported.

Table 3. Summary of the selected work for face expressions analysis. AAM = Active Appearance Models; 'acc' = accuracy (correct predictions/total number of predictions with respect to the clinical goal occurrences tested); AU = Action Units; AUC = Area Under the Curve; CERT = Computer Expression Recognition Toolbox; CLNF = Conditional Local Neural Field; CNN = Convolutional Neural Network; HOG = Histogram Of Gradients; ICC = Intra class Correlation Coefficient; LBP = Local Binary Patterns; LSTM = Long Short-Term Memory network; LTP = Local Ternary Patterns; PCA-LMNN = Principal Components Analysis with Large Margin Nearest Neighbor; RB = radon Barcodes; ResNet = Residual Network [32]; ResNet-152 = Residual Network with 152 layers [32]; SVM = Support Vector Machines; VGG-16 = VGG Net with 16 layers [38].

Work (Year)	Method	Clinical Task	Metrics	Dataset Population/Age (h = hours, w = weeks, m = months, y = years)
[39] (2013)	AAM + HOG features; comparison PCA-LMNN vs. Laplacian Eigenmap and SVM vs. K-Nearest Neighbour	Assessment of the dynamics of face-to-face interactions with the mother	ICC	12 infants (age: 1 m–1 y)
[40] (2016)	CERT	Lie detection	Qualitative	Children (age: 6–11 y)
[41] (2017)	HOG computation + Landmark Detection by CLNF + Facial AU intensities computation	Diagnosis and evaluation of ASD children	Entropy score + Similarity metrics	children (age: 5–6 y)
[42] (2019)	Geometric and appearance features/facial landmark-based template matching + SVM	Pain assessment	AUC = 0.87/0.97	22 infants (age: 1 m–1 y)
[43] (2019)	Neonatal CNN	Pain assessment	acc = 97%	84 neonates (age: 18 h–41 w)
[44] (2019)	CNN + ResNet	Robot assisted therapy	acc = 72%	children (age: 6–12 y)
[45] (2019)	Mean Supervised Deep Boltzmann Machine	Emotion detection and recognition	acc = 75%	154 children (age: 2–8 y)
[46] (2020)	Texture and geometric descriptors: LBP, LTP and RB + SVM	Pain assessment	acc = 95%	26 neonates (age: 18–36 h)
[47,48] (2020, 2021)	Deep spatiotemporal geometric facial features + Recurrent Neural Network	ASD meltdown crisis management	acc = 85.8%	children (age: 4–11 y)
[49] (2021)	Facial landmark detection and tracking	Early detection of ASD symptoms	AUC from 0.62 to 0.73	104 toddlers (age: 1–2 y)
[50] (2021)	PainCheck Infant	Pain assessment	Correlation with standard scores: r = 0.82–0.88; $p < 0.0001$	40 infants (age: 2–9 m)
[51] (2021)	YOLO face detector + VGG-16 for facial features extraction + LSTM	Pain assessment	acc = 79%	58 neonates (age: 27–41 w)
[52] (2021)	ResNet-152	Emotion detection and recognition	Balanced acc = 79.1%	154 children (age: 2–8 y)
[53] (2022)	VGG-16 network	Emotion detection and recognition	AUC = 0.82	123 children (age: 1 m–5 y)
[54] (2022)	Strain-based, geometric-based, texture-based and gradient-based facial features	Pain assessment	acc = 95.56%	31 neonates
[55] (2022)	Progressive lightweight shallow learning (ShallowNet)	Emotion detection and recognition	Acc = 99.06%	12 children (age: 6–12 y)

Over the past decade, research on automatic analysis of children's facial expressions has made great strides. One of the challenges in this area has been the recognition of regular facial expressions. Another has been the attention to micro-expressions or compound ones, i.e., the combination of several facial expressions, which can be critical for the success of an automated system. The computational analysis of facial expressions can overcome the limitations of human perception and provide fast and objective results in a wide range of clinical tasks.

For instance, commonly used screening tools for **autism spectrum disorder (ASD)** generally rely on subjective caregiver questionnaires. While behavioural observation carried out by specialists is more accurate, it is also expensive, time-consuming and requires considerable expertise. Many efforts have been made in the field of CV to overcome such limitations and automatically recognise ASD children's facial expressions to

- *Handle meltdown crisis.* Studies such as [47,48] consider the safety of autistic children during a meltdown crisis. Meltdown signals are not associated with a specific facial expression, but with a mixture of abnormal facial expressions related to complex emotions. Through the evaluation of a set of spatio-temporal geometric facial features of micro-expressions, the authors demonstrate that the proposed system can automatically distinguish a compound emotion of autistic children during a meltdown crisis from the normal state and timely notify caregivers.
- *Support specialists in diagnosing and evaluating ASD children.* In [41], the authors propose a CV module consisting of four main components aimed at face detection, facial landmark detection, multi-face tracking and facial action unit extraction. The authors highlight how the proposed system could provide a noninvasive framework to apply to pre-school children in order to understand the underlying mechanisms of the difficulties in the use, sharing and response to emotions typical of ASD.
- *Computationally analyse how children with ASD produce facial expressions with respect to their typically developing peers.* In [56–58], the authors propose a framework aimed at computationally assessing how ASD and typically developing children produce facial expressions. Such a framework, which works on a sequence of images captured by a webcam under unconstrained conditions, locates and tracks multiple landmarks to monitor facial muscle movements involved in the production of facial expressions (thus performing a type of virtual electromyography). The output from these virtual sensors is then fused to model the individual's ability to produce facial expressions. The results correlate with psychologists' ratings, demonstrating how the proposed framework can effectively quantify the emotional competence of children with ASD to produce facial expressions.
- *Early detect symptoms of autism.* Despite advances in the literature, it is still difficult to identify early markers that can effectively detect the manifestation of symptoms of ASD. Carpenter and colleagues [49] collected videos of 104 young children (22 with ASD) watching short movies on a tablet. They then used a CV approach to automatically detect and track specific facial landmarks in the recorded videos to estimate the children's facial expressions (positive, neutral, all others) and differentiate between children with and without ASD. In these cases, children with ASD were more likely to show 'neutral' facial expressions, while children without ASD were more likely to show 'all other' facial expressions (raised eyebrows, open mouth, engaged, etc.).

Another fundamental goal in healthcare involves **detecting and monitoring pain and discomfort** in children.

Children are particularly vulnerable to the effects of pain and discomfort, which can lead to abnormal brain development, yielding long-term adverse neurodevelopmental outcomes. Nowadays, the evaluation of pain in patients depends mainly on the continuous monitoring of the medical staff when the patient is unable to verbally express his/her experience of pain, as is the case of babies. Therefore, the need to provide alternative methods for its evaluation and detection.

For instance, PainCheck Infant [50] is a mobile point-of-care application that uses automated facial evaluation and analysis to assess procedural pain in infants. Based on an artificial intelligence algorithm, it enables the detection of six facial action units (AUs) that indicate the presence of pain: AU4 (forehead lowering), AU9 (nose wrinkling), AU15 (lip corner pressing), AU20 (horizontal mouth stretching), AU25 (lip parting) and AU43 (eye closure). These facial actions, as classified by the Baby Facial Action Coding System [59], represent specific muscle movements (contraction or relaxation). The authors reported the good psychometric properties of PainCheck Infant after collecting video recordings from 40 infants (aged 2–9 months).

The authors in [60] also proposed an infant monitoring system to detect a broader spectrum of facial expressions consisting of discomfort, unhappiness, joy and neutral. They also aimed at detecting some states, including sleep, pacifier and open mouth. The proposed system was based on combining expression detection using Fast R-CNN with compensated detection using a Hidden Markov Model. The experimental results showed an average precision for discomfort detection up to 90%.

The studies reported in [42,46] focus on texture and geometric descriptors to analyse infants' faces and detect expressions of discomfort. In particular, Martinez et al. [46] used three different texture descriptors for pain detection: Local Binary Patterns, Local Ternary Patterns and Radon Barcodes. A Support Vector Machine (SVM) based model was implemented for their classification. The proposed features gave a promising classification accuracy of around 95% for the infant COPE image database [61,62]. In [42], a two-phase classification workflow was developed: phase 1, subject-independent, derived geometric and appearance features; phase 2, subject-dependent, incorporated template matching based on facial landmarks. Finally, to detect comfort or discomfort facial expressions, an SVM classifier was applied to the video frames. Videos of 22 infants were used to evaluate the proposed method. Experiments showed AUC of 0.87 for the subject-independent phase and 0.97 for the subject-dependent phase.

However, there is a view among some researchers that pain is a multimodal emotion, often expressed through several different modalities. For this reason, in [51] Salekin and colleagues show that there is a need for a multimodal assessment of pain, particularly in the case of post-operative pain (acute and prolonged pain). They integrated visual and vocal signals using a multimodal spatio-temporal approach. For neonatal face analysis, the proposed algorithm first detects the face region in each video frame using a pre-trained YOLO-based [63] face detector. Then, a VGG-16 [38] network extracts visual features from the face. Finally, they used LSTM [64] with deep features to learn the temporal pattern and dynamics typical of postoperative discomfort. Experimental results on a real-world dataset (known as USF-MNPAD-I—University of South Florida Multimodal Neonatal Pain Assessment Dataset, consisting of 58 neonates with a gestational age that ranges from 27 to 41 weeks [65]) show that the proposed multimodal spatio-temporal approach achieves the highest AUC (0.87) and accuracy (79%), averaging 6.67% and 6.33% higher than unimodal approaches. The work of Zamzmi et al. [54] also presented a comprehensive multimodal pain assessment system that fuses facial expressions, crying sounds, body movement and vital signs. In terms of face analysis, the proposed system acquires video of infants being monitored in the neonatal intensive care unit and implements four feature extraction methods, namely strain-based, geometric-based, texture-based, and gradient-based, to extract relevant features from the newborns' faces. The system achieved an accuracy of 95.56%.

The area of children's **social interactions** is also considered clinically relevant, since the ability to produce and decode facial expressions in both childhood and adolescence promotes social competence, whereas deficits characterise several forms of psychopathology. However, even in this area, the study of facial expressions has been hampered by the labour-intensive and time-consuming nature of human coding. Therefore, some efforts have been made to automatically analyse children's facial expressions in order to study their social interactions.

For example, primary social interactions, namely the family context, are the focus of the studies reported in [39,66]. In the latter, the intensity of twelve infants' facial expressions is detected and measured in order to model the dynamics of face-to-face interactions with their mothers. Certified Facial Action Coding System (FACS) coders manually coded facial AUs related to the positive and negative affect from the video. Then, relevant facial features were tracked using Active Appearance Models (AAM) and registered to a canonical view before extracting Histogram of Oriented Gradients (HOG) features. Finally, using these features, the authors compared two dimensionality reduction approaches (Principal Components Analysis with Large Margin Nearest Neighbour and Laplacian Eigenmap) and two classifiers, SVM and K-Nearest Neighbour.

In [40,67], the pro-social and antisocial behaviour of children is studied. In particular, lie detection is carried out. Zanette et al. [40] first collected video recordings of a group of children (6–11 years old). Non-verbal behaviour was analysed using the Computer Expression Recognition Toolbox (CERT), which uses FACS to automatically code children's facial expressions while lying. The results showed the reliability of CERT in detecting differences in children's facial expressions when telling antisocial versus prosocial lies.

Regarding expression recognition aiming at **emotion detection**, most of the works in this area have used deep neural networks for automatic classification of children's facial expressions, such as [53], where a VGG-16 network [38] was used. Here, the authors trained the network on adult videos and refined the network using two publicly available databases of toddler videos that differ in context, head pose, lighting, video resolution, and toddler age: FF-NSF-MIAMI [68,69] and CLOCK [70] databases. The resulting AU detection system, which the authors call Infant AFAR (Automated Facial Action Recognition), is available to the research community for further testing and applications.

In [55], the authors present an advanced lightweight shallow learning approach to emotion classification by using the skip connection for the recognition of facial behaviour in children. In contrast to previous deep neural networks, they limit the alternative path for the gradient in the early part of the network by a gradual increase with the depth of the network. They show that the progressive ShallowNet is not only able to explore more feature space, but also solves the overfitting problem for smaller data, using the LIRIS-CSE [71] database to train the network.

Nagpal et al. [45] incorporated supervision into the traditionally unsupervised Deep Boltzmann machine [72] and proposed an average supervised deep Boltzmann machine for classifying an input face image into one of the seven basic emotions [73]. The proposed approach was evaluated on two child face datasets: Radboud Faces [74] and CAFÉ [75].

However, emotion recognition classifiers traditionally predict discrete emotions. Nevertheless, a method for dealing with compound and ambiguous labels is often required to classify emotion expressions. In [52], Washington and colleagues explored the feasibility of using crowdsourcing to obtain reliable soft-target labels and evaluate an emotion detection classifier trained with such labels. Reporting an emotion probability distribution, which takes into account the subjectivity of human interpretation, may be more useful than an absolute label for many applications of affective computing. For the experiments, they used the Child Affective Facial Expression (CAFE) data set [75] and a ResNet-152 neural network [37] as a classifier.

In healthcare, social robotics is experiencing a rapid increase in applications. Some of these applications include **robot-assisted therapy** for children [76]. Empathy, or the ability to correctly interpret the manifestations of human affective states, is a critical capability of social robots. The study reported in [44] proposes a method based on deep neural networks that fuses information from the skeleton of the body posture with facial expressions for the automatic recognition of emotions. The network is composed of two different branches, one focusing on facial expressions and the other focusing on body posture. The two branches are then combined at a later stage to form the branch for the recognition of the whole body expression. The authors evaluated their method on a sophisticated child–robot interaction database (aged 6 to 12 years) of previously collected emotional expressions.

2.4. Multimodal Analysis

Several papers combine different types of data (e.g., gaze tracking and facial morphology data, or head pose estimation and expression classification, etc.) to capture and analyse as many aspects of the condition under study as possible. Some of the most relevant ones are resumed in Table 4

Table 4. Summary of the selected work for multimodal face analysis. 'acc' = accuracy (correct predictions/total number of predictions with respect to the clinical goal occurrences tested); ADHD = Attention deficit hyperactivity disorder; ICC = Intra Class Correlation coefficient.

Work (Year)	Method	Clinical Task	Metrics	Datase Population/Age (h = hours, w = weeks, m = months, y = years)
[30] (2015)	Facal Expressions + Head Pose by IntraFace Software 2015	Detection of early indicators of ASD	ICC	20 toddlers (age: 16–30 m)
[77] (2020)	facial Expression An. + Gaze Tracking by Classical computer vision methods	Detection of early indicators of ASD	acc = 97.12%	10 children (age: 6–11 y)
[78] (2021)	facial Expression An. + Gaze Tracking + 3D Body Pose by Classical computer vision methods	ADHD diagnosis	acc = 80.25%	children (age: 6–12 y)

For instance, several studies have focused on analysing facial features for detecting early symptoms of ASD and on the automatic diagnosis of attention deficit hyperactivity disorder (ADHD) based on children's attention patterns and facial expressions [78].

For example, to detect early indicators of ASD, the authors in [30] analysed both facial expressions and head postures of twenty 16- to 30-month-old children with and without autism. They extracted 49 facial landmarks using the IntraFace software [79]; with regard to the analysis of facial expressions, three classes of emotions were taken into account: Neutral, Positive (Happy) and Negative (Anger, Disgust and Sad). However, the facial expression classifier was trained on the standard Cohn–Kanade dataset [80], which contains video sequences from a total of 123 subjects between the ages of 18 and 50.

Xu et al. [77] and Nag and colleagues [81] also attempted to find notable indicators for early detection of ASD in both facial expressions and gaze patterns. The system proposed in [77] provides participants with three modes of virtual interaction—videos, images and virtual interactive games. Computer vision-based methods are used to automatically detect the subject's emotion and attentional characteristics in the three interaction modes. The system is intended to aid in the early detection of autism. The system's accuracy has been verified through experiments on the publicly available dataset and data collected from 10 children with ASD.

2.5. Publicly Available Datasets

Large amounts of adult facial image datasets were available for research purposes, but very few equivalent datasets for children can be found in the literature. The most relevant datasets reporting infant, toddler, and children faces are reported here and listed in Table 5:

- *COPE Database* [61,62]: This database contains 204 photographs of 26 newborns (between 18–36 h old) who were photographed while experiencing the pain of a heel lance and a variety of stressors, including being moved from one cot to another (a stressor that produces crying that is not in response to pain), a puff of air on the nose (a stressor that produces eye squinting), and friction on the outer lateral surface of the heel (a stressor that produces facial expressions of distress similar to those of pain). In addition to these four facial displays, the database contains images of the newborns in a neutral resting state. All subjects were born in a large Midwestern hospital in the

United States. All newborns involved in the study were Caucasian, evenly divided between the sexes (13 boys and 12 girls), and in good health.

- *CAFE Database* [75]: The CAFE set is a collection of 1192 photographs of 2- to 8-year-old children posing with the six basic emotions defined by Ekman [82]: sadness, happiness, surprise, anger, disgust and fear. It also includes a seventh neutral expression. Such a set is also racially and ethnically diverse, with 27 African American, 16 Asian, 77 Caucasian/European American, 23 Latino, and 11 South Asian children. Photographs include enough face variability to allow independent researchers to determine and study the natural variation in human facial expressions. The children were asked to pose with their mouths open and closed for each expression except surprise. Surprised faces were open-mouthed only. Open-mouthed, disgusted faces usually included a tongue protrusion.
- *CLOCK Database* [70]: This database was generated by a multi-site longitudinal project known as CLOCK (Craniofacial microsomia: Longitudinal Outcomes in Children pre-Kindergarten), which examined the neurodevelopmental and phenotypic outcomes of children with craniofacial microsomia (CFM) and demographically matched controls [83]. Two age-appropriate emotion induction tasks were used to elicit positive and negative facial expressions. In the positive emotion task, an experimenter blew bubbles at the infant. In the negative emotion task, an experimenter presented the infant with a toy car, allowed the infant to play, then removed the car and covered it with a clear plastic container. Each video was approximately 2 min long (745 K and 634 K recorded frames). The video resolution was 1920 × 1080. FACS coders manually annotated for nine action units: AU1 (inner brow raised), AU2 (outer brow raised), AU3 (inner brow pulled together), AU4 (lowered eyebrow), AU6 (raised cheek), AU9 (nose), AU10 (nose wrinkle), AU9 (nasal wrinkling), AU12 (corner of lips pulled back), AU20 (lip stretching) and (lip stretching) and AU28 (lip sucking).
- *LIRIS-CSE Database* [71]: It features video clips and dynamic images consisting of 26,000 frames depicting 12 children from diverse ethnic backgrounds. This database showcases children's natural, unforced facial expressions across various scenarios, featuring six universal or prototypical emotional expressions: happiness, sadness, surprise, anger, disgust, and fear as defined by Ekman [73]. The recordings were made in unconstrained environments, enabling free head and hand movements while sitting freely. In contrast to other public databases, the authors assert that they were capable of gathering children's natural expressions as they happened due to the unconstrained environment. The database has been validated by 22 human raters.
- *GestATional Database* [23]: It comprises 130 neonates recruited between October 2015 and October 2017. Clinical staff at Nottingham University NHS Trust Hospital, Nottingham, UK carried out recruitment and sorted the neonates into five groups based on their prematurity status. The data gathered included: (i) images of the neonates' faces, feet, and ears; (ii) case report forms with important information such as the baby's gestational age, days of life at the time of the visit, current weight, Ballard Score, the mother's medical history, and information related to the delivery. It is important that technical term abbreviations are explained when they are first used, and that a logical flow of information is maintained with causal connections between statements.
- *FF-NFS-MIAMI Database* [68,69]: It is a database documenting spontaneous behaviour in 43 four-month-old infants. Infants' interactions with their mothers were recorded during a Face-to-Face/Still-Face (FF/SF) protocol [84]. The FF/SF protocol elicits both positive and negative effects. It assesses infant responses to parent unresponsiveness, an age-appropriate stressor. AUs were manually annotated from the video by certified FACS coders for four action units: AU4 (brow lowering), AU6 (cheek raising), AU12 (lip corner pulling) and AU20 (lip stretching). The combination of AU6 and AU12 is associated with a positive effect; AU4 and AU20 are associated with a negative effect. The video resolution is 1288 × 964. There are 116,000 manually annotated frames in 129 videos of 43 infants.

- *USF-MNPAD-I Database* [65]: The University of South Florida Multimodal Neonatal Pain Assessment (USF-MNPAD-I) Dataset was collected from 58 neonates (27–41 weeks gestational age) while they were hospitalised in the NICU, undergoing procedural and postoperative procedures. It comprises video footage (face, head, and body), audio (crying sounds), vital signs (heart rate, blood pressure, oxygen saturation), and cortical activity. Additionally, it includes continuous pain scores, following the NIPS (Neonatal Infant Pain Scale) scale [85], for each pain indicator and medical notes for all neonates. This dataset was obtained as a component of a continuous project centred on creating avant garde automated approaches for tracking and evaluating neonatal pain and distress.

Table 5. Datasets reporting children's faces.

Dataset	Reference	Number of Subjects	Type of Data	Age of Subjects	Year	Publicly Available
COPE	[61,62]	26	Images	Neonates: (age: 18–36 h)	2005	Yes
CAFE	[75]	154	Images	Children (age: 2–8 years)	2014	Yes
CLOCK	[70]	80	Video	Children (age: 4–5 years)	2017	No
LIRIS-CSE	[71]	12	Video	Children (age: 6–12 years)	2019	Yes
GestATional	[23]	130	Images	Neonates (gestational age: 28–40 weeks)	2019	No
FF-NFS-MIAMI	[68,69]	43	Video	Infants	2020	No
USF-MNPAD-I	[65]	58	Video, audio, physiological, contextual, information	Neonates (age: 27–41 weeks)	2021	Yes

2.6. New Computer Vision Perspectives for More Accurate Face Analyses

Face detection, head pose estimation and facial expression recognition are challenging tasks, whose success can be hindered by varying conditions such as facial occlusion, lighting, unusual expressions, distance from the cameras, skin type, complex real-world background, low data resolution and noise. These challenges, which are well known when dealing with adult face analyses, might even be exacerbated in the case of children and newborns.

Among the most recent and promising methods that attempt to address such challenges, those based on deep learning models are gathering more and more momentum, as they guarantee remarkable results in terms of accuracy and robustness. To mention a few, DeepFace [86] and FaceNet [87], on which OpenFace [36] is based, have been some of the pioneering solutions that have demonstrated state-of-the-art performance and paved the way for further breakthroughs in the field.

DeepHeadPose [88] (code at https://github.com/natanielruiz/deep-head-pose) (All the links in this section were accessed on 15 September 2023), dense head pose estimation [89] (code available at https://github.com/1996scarlet/Dense-Head-Pose-Estimation), SPIGA (Shape Preserving Facial Landmarks with Graph Attention Networks, [90]) (code available at https://github.com/andresprados/SPIGA.img2pose [91] (code available at: https://github.com/vitoralbiero/img2pose), 6DRepNet [92] (code available at: https://github.com/thohemp/6DRepNet) are some of the tools, mainly based on deep learning models, that have been implemented specifically for head pose estimation and tracking [93,94] and that showed very promising results.

However, these methods have been trained, implemented and tested mainly on adult faces, so further development is needed to test their generalisability to newborn and infant faces.

Regarding face detection, we note that ArcFace [95] (code available at: https://github.com/1996scarlet/ArcFace-Multiplex-Recognition), RetinaFace [96] (code available at: https://github.com/1996scarlet/ArcFace-Multiplex-Recognition) and FaceYolov5 [97] (code available at: https://github.com/deepcam-cn/yolov5-face/tree/master) performed exceptionally in detecting adult faces. Nevertheless, as several studies ([70,98,99]) reported, face recognition methods designed for adults fail when applied to the neonatal population due to the unique craniofacial structure of neonates' faces as well as the large variations in pose and expression as compared to adults. Therefore, further research in this domain should concentrate on designing algorithms trained specifically on datasets collected from the neonatal population, as pointed out by Zamzmi et al. in [43], where a novel Neonatal Convolutional Neural Network for assessing neonatal pain from facial expression is described.

Regarding expression recognition aiming at **emotion detection**, most of the research has focused on adult face images so far [100–103], with no dedicated research on automating expression classification for children. As infants' faces have different proportions, less texture, fewer wrinkles and furrows, and unique facial actions with respect to adults, automated detection of facial action units in infants is challenging. More thorough experiments are needed to assess the applicability and robustness of the cited methods when tested on newborn and child data. Furthermore, emotion recognition classifiers typically forecast isolated emotions. A strategy for addressing complex, compound emotions may involve integrating multiple modalities and other types of sensors (e.g., thermal cameras), thus including temporal, auditory, and visual data, to enhance the precision and robustness of the models, as demonstrated in [104].

3. Body Analysis

Introducing automatic methods to analyse the movements of babies and children (behavioural coding) is becoming increasingly needed. On the one side, when reported by parents or general practitioners, it relieves the workload on specialized health professionals, reducing costs and time to obtain a diagnosis. On the other side, it enables the possibility of continuous screening of a larger population, making early diagnosis of eventual diseases even before symptoms become evident to non-expert observers possible [12]. These automatic methods leverage human pose estimation algorithms. Deep Learning architectures have obtained significant results for human pose estimation in the last few years, but they have been trained on images picturing adults. The estimation of the pose of children (infants, toddlers, children) is sparsely studied despite it can be extremely useful in different application domains [105]. In this section, the works dealing with the estimation of the body posture of babies and children are reported and discussed. In particular, at first, existing benchmarks are introduced and subsequently, the most relevant works in the literature introducing algorithmic pipelines exploited for the healthcare of young subjects are discussed. For each work, the clinical aim and eventually how they addressed the additional bias of dealing with children have been pointed out.

A coarse selection of related papers was initially carried out by using the following queries in the research databases:

- **Scopus**
  QUERY "TITLE-ABS-KEY ((children OR infants OR babies) AND (body OR limbs OR head) AND (motion OR movements) AND computer AND vision) AND PUBYEAR > 2014 AND PUBYEAR < 2024 that returned 105 documents;
- **Web of Science Core Collection**
  ((((ALL=(children)) OR ALL=(infants)) OR ALL=(babies)) AND ALL=(computer) AND ALL=(vision) AND ((ALL=(motion)) OR ALL=(movements)) AND ((ALL=(body)) OR

(ALL=(limbs)) OR (ALL=(head))), refined in the YEARS from 2015 to 2023, that returned 60 documents;
and

- **Scholar**
  allintitle: children OR babies OR infants OR motion OR movements "computer vision", refined in the YEARS from 2015 to 2023, that returned 132 documents.

Among all the documents retrieved from the databases, a fine selection was therefore conducted based mainly on the scientific content (some documents were in fact not relevant for the purposes of the proposed survey), the type of publication (journals were preferred to conferences in case of comparable ideas) and finally also on the number of citations (articles prior to 2020 with less than 10 citations were not considered).

This led to the following content organization: at first, in Section 3.1, documents describing datasets and common tools exploited for pose estimation are reported. Then, the remaining documents have been split depending on how the infants were acquired, i.e., lying in a bed/crib or standing/walking and two different subsections are used to describe them accordingly. Similarly to Section 2, in each subsection some tables have been used to organize information within them. Finally, in Section 3.4, new research directions for more accurate infants' pose estimation are reported and discussed.

3.1. Common Datasets and Tools for Human Pose Estimation

Healthcare would enjoy powerful and reliable algorithms fine-tuned on specific goals (i.e., pathology classification or evaluation of its stage); nevertheless interdisciplinary specialists could be advantaged in having tools oriented to more generic processing and providing semi-raw data ready for further analysis. This includes many tools that allow the research community to set up pipelines aiming at the final goal of analysing infant movements. Such an approach enables a wider range of scientists to analyse child movement patterns and, at the same time, represents a starting point for the image-processing research community. Among these tools, the most common library for human pose detection is OpenPose. It is a real-time multi-person human pose detection library [106] that maps 25 points on the body including shoulders, elbows, wrists, hips (+mid-hip), knees, ankles, heels, big toes, little toes, eyes, ears, and nose. It was trained on adults but, as reported in the following section, it has been largely used also on infants with or without a specific domain adaptation learning phase. It is available at https://github.com/CMU-Perceptual-Computing-Lab/openpose (accessed on 15 September 2023). OpenPose has then been integrated into AutoViDev [107], a system specifically created for automated video action recognition. It provides a highly modular implementation of 188 primitives, on which users can flexibly create pipelines. It also supports automated tuners and an easy-to-use GUI to help researchers/practitioners develop prototypes. AutoVideo is released under MIT license at https://github.com/datamllab/autovideo (accessed on 13 February 2023).

Another efficient deep architecture for markerless pose estimation and semantic features detection is DeepLabCut (DLC) [108]. Open source Python code for selecting training frames, checking human annotator labels, generating training data in the required format, and evaluating the performance on test frames is available at https://github.com/DeepLabCut/DeepLabCut (accessed on 13 February 2023).

Some useful annotation tools are introduced to build image databases for computer vision research. The early one introduced is LabelMe [109] that works online at http://labelme2.csail.mit.edu/Release3.0/index.php (accessed on 13 April 2023). Another is Kinovea, designed for sports analysis, open-source and freely available at www.kinovea.org (accessed on 29 Match 2023). Finally, the tool in [110] is more oriented to pose estimation, it is interactive and it relies on a heuristic weakly supervised human pose (HW-HuP) solution to estimate 3D human poses in contexts where no ground truth 3D pose data are accessible, even for fine-tuning.

Unfortunately, all the above-mentioned instruments suffer from a bias in terms of target patients. They are designed and/or trained for adults, reducing their reliability when

applied to infants. This allows specifically child-oriented tools to shine in this landscape. A tool specifically designed for the semi-automatic annotation of baby joints, namely Movelab, has been recently introduced in [111] instead. It consists of a GUI that allows users to browse videos and to choose an algorithm for baby pose detection among MediaPipe Pose [112] and two ResNet architectures fine-tuned on a proprietary dataset of 600 videos of children lying on a bed.

AVIM is another tool, developed using the OpenCV image processing library and specifically designed for an objective analysis of infants from 10 days to the 24th week of age. It acquires and records images and signals from a webcam and a microphone and allows users to perform audio and video editing [113]. It is similar to MOVIDEA, a software developed using MATLAB [114]. Both tools rely on manual annotation of interesting points of the body and provide cinematic measurements.

All these discussions raise a problem that concerns artificial intelligence and which becomes more serious in highly specialised environments: lack of data. Retrieving data, providing accurate annotations and complying with all regulations could be extremely tedious and time-consuming. In addition, the specific care required by the category of children and the need for long-term monitoring make datasets from this sector extremely rare. Some of the most relevant ones are resumed in Table 6. Some of them have been introduced to help in pose estimation and markerless joint detection and tracking. Under this umbrella, we can cite the Moving INfants In RGB-D (MINI-RGBD) [115] dataset that was generated by mapping real infant movements to the Skinned Multi-Infant Linear body model (SMIL) with realistic shapes and textures and generating RGB and depth images with precise ground truth 2D and 3D joint positions. The dataset is available for research purposes at http://s.fhg.de/mini-rgbd (accessed on 23 August 2023).

Another relevant contribution to this topic has been recently provided in [116], where hybrid synthetic and real infant pose (SyRIP) were collected and made publicly available. It came with a multi-stage invariant representation learning strategy that could transfer the knowledge from the adjacent domains of adult poses and synthetic infant images into a fine-tuned domain-adapted infant pose (FiDIP) estimation model. The code is available at https://github.com/ostadabbas/Infant-Pose-Estimation (accessed on 23 August 2023).

Other relevant datasets for infant body parsing and pose estimation from videos are the BHT dataset [117], the AIMS dataset [118] and the Youtube-infant dataset [119]. BHT consists of 20 movement videos of infants aged from 0–6 months. YouTube-infant has 90 infant movement videos collected from YouTube. Both datasets contain annotations for five classes: background, head, arm, torso and leg. Pose annotation was made by the LabelMe online annotation tool and the dataset comes with BINS scores describing neurological risks associated with each infant [120]. The AIMS dataset contains 750 real and 4000 synthetic infant images with Alberta Infant Motor Scale (AIMS) pose labels [121]. Code and data referenced in [119] are provided at https://github.com/cchamber/Infant_movement_assessment/ (accessed on 23 August 2023).

In [122,123], the BabyPose dataset, consisting of 16 depth videos of 16 preterm infants recorded during the actual clinical practice in a neonatal intensive care unit (NICU), has been introduced. Each video lasts 100 s (at 10 fps). Each frame was annotated with the limb-joint locations. Twelve joints were annotated, i.e., left and right shoulder, elbow, wrist, hip, knee and ankle. The database is freely accessible at https://zenodo.org/record/3891404 (accessed on 13 February 2023).

Concerning autism-related behaviours, the Self-Stimulatory Behaviour Dataset (SSBD) [124] collected stimming behaviour videos of children available on public domain websites and video portals, such as Youtube, Vimeo, Dailymotion, etc. The dataset contains 75 videos grouped into three categories each containing 25 videos. The mean duration of a video is 90 s. The resolution of the videos varies, but is greater than 320 × 240 pixels. Videos are related to Armflapping, Headbanging, and Spinning repetitive behaviours. The dataset can be found at https://github.com/antran89/clipping_ssbd_videos (accessed on 23 August 2023) with annotated data.

Table 6. Datasets reporting children's poses. BINS stands for Bayley Infant Neurodevelopmental Screener [120], AIMS stands for Alberta Infant Motor Scale [121].

Dataset	Reference	Number of Subjects	Type of Data	Age of Subjects	Year	Publicly Available
SSBD	[124]	75	RGB videos (attributes of the behaviour)	Neonates: (age: 0–7 m)	2013	Yes
MINI-RGBD	[115]	12	Synthetic videos (RGB, Depth, 2D-3D joint positions)	Neonates: (age: 0–7 m)	2018	Yes
BHT	[117]	20	RGB images (body parts segmentation)	Neonates: (age: 0–6 m)	2019	No
babyPose	[122,123]	16	depth Videos (limb-joint locations)	Neonates: (Gestation Period: 24–37 w)	2019	Yes
Youtube-infant	[119]	104	Videos: (BINS score)	Neonates: (age: 6–9 w)	2020	Yes
SyRip	[116]	140	Synthetic and Real Images: (fully 2D body joints)	Neonates: (age: 0–12 m)	2021	Yes
AIMS	[118]	NA	Synthetic and Real Images: (AIMS pose label)	Neonates: (age: 0–6 m)	2022	No

3.2. Monitoring of Lying Children

Early diagnosis plays a key role in most healthcare scopes, including neurological disorder. It is clear that a diagnosis in the first weeks of a child's life is crucial, especially in preterm infants, to recognise signs of possible lesions in the developing brain and to plan timely and appropriate rehabilitation interventions. Unfortunately, this can only be achieved by monitoring the child lying down, with two main constraints: on the one hand, it is mandatory to use completely non-invasive methods, and on the other hand, considering the specific movement dynamics under investigation, ad hoc datasets are required, underlining what was discussed in the previous section. Beyond this critical analysis, further monitoring can be carried out to track vital signs or movements of discomfort that represent manifestations of the child's distress. All these kind of analysis are generally performed using a camera mounted at the top of the crib at a neonatal intensive care unit (NICU) as shown in the typical setup for data acquisition and processing is reported in Figure 2.

Figure 2. A typical experimental setup for children monitoring in a NICU. Image has been taken from [125].

Most relevant work is resumed in Table 7.

Table 7. A summary of selected work concerning the analysis of the body of infants lying in a crib. SVM = Support Vector Machine, AUC = Area Under Curve, FVGAN = factorized video generative adversarial network, GMA = Prechtl General Movement Assessment [126], acc = accuracy (correct predictions / total number of predictions with respect to the clinical goal occurrences tested), RF = Random Forest, RMSE = root-mean-square error (it measures the differences between values predicted by a model and the values provided by experts, b.p.m. = beats per minute, r.p.m. = respirations per minute.

Work (Year)	Method	Clinical Task	Metrics	Dataset Population/Age (w = weeks, y = years)
[127] (2019)	Optical flow + SVM	Discomfort Moments	acc = 0.86 AUC of 0.94	11 (34 w)
[128] (2019)	Skin detection + Motion magnification	Vital Sign	Limit of agreement = $[-8.3, +17.4]$ b.p.m., $[-22; +23.6]$ r.p.m.	10 (23 w–40 w)
[129] (2019)	Motion Detection	Vital Sign	acc = 87%	1/unknown
[119] (2020)	OpenPose + kinematic features + Naïve Bayesian Class	Neuromotor Risk	RMSE = 0.92	19 (10 w)
[130] (2021)	OpenPose	Reaching Trajectories	95% confidence in hands tracking	12 (48 w)
[131] (2022)	basic tracking primitives	GMA	qualitative	8 (3 m–5 m)
[132] (2022)	DeepLabCut + kinematic features + RF	Neuromotor Risk	acc = 0.62	142 (40 w)
[133] (2023)	FVGAN +SiamParseNet	GMA	96.46 ± 2.11	161 (49 w–60 w)

The easiest approaches rely on optical flow's motion information to estimate pixel motion vectors between frames. One of the former applications related to infants' health care was recognising comfort or discomfort. In [127], the authors calculated the motion acceleration rate and 18 time- and frequency-domain features characterizing motion patterns and provided them with a support vector machine (SVM) classifier. The method was evaluated using 183 video segments for 11 infants from 17 heel prick events. The experimental results show an AUC of 0.94 for discomfort detection and an average accuracy of 0.86 when combining all proposed features, which is promising for clinical use.

A more effective computer-aided pipeline to characterize and classify infants' motion from 2D video recordings has been proposed in [132]. The authors used data from 142 preterm infants, acquired from a viewpoint perpendicular to the plane where the infants lay, at 40 weeks of gestational age. The final goal was detecting anomalous motion patterns. The ground truth was built starting from brain MRI evidence at birth and neurological examinations 30 months after the video recording. DeepLabCut was exploited, but it was fine-tuned to detect a small set of meaningful landmark points (nose, hands and feet) on the infants' bodies. The authors discussed these choices. They wrote that classical full-body pose estimation algorithms, if not fine-tuned on infants' poses, have proven to not always be appropriate for infants since they are trained and implemented for detecting adults' poses. Since fine-tuning requires a significant amount of data, the authors focused only on some key points that provide meaningful information regarding infants' motion, guaranteeing this way a higher per-point accuracy and a higher control on the interpretability of the results. Starting from the trajectories of the detected landmark points, quantitative parameters describing infants' motion patterns were extracted and

classified between normal or abnormal motion patterns by means of different shallow and deep classifiers. Despite the accurate setup, the mean overall accuracy was not over 60%. The problems were the unbalanced dataset and the sparsity of landmarks tracked over time. Obtaining a dense body motion analysis of babies is particularly challenging indeed, as the body part dimensions between infants and adults vary significantly. Similarly, in [119], a framework that predicts the neuromotor risk level of 19 infants (more or less than 10 weeks of age) was proposed. The training was conducted using 420 YouTube video segments. OpenPose was used to extract pose information. Due to differences between adults and infants in their appearance and pose, pose tracking using OpenPose was initially limited in performance. Therefore, the authors specialized OpenPose for infants by creating a dataset of infant images with labels of joint positions. Root-mean-square error on joint positioning decreased from 0.05 by standard OpenPose to 0.02 after the specialization of the algorithm on infants. The adapted pose estimator allowed authors to extract movement trajectories from videos of infants moving. Finally, the authors combined many features into one estimate for assessing neuromotor risk, demonstrating a correlation between the score and the risk associated with each infant by clinicians.

Recently, in [133], a method to assess the general movement assessment (GMA) of infant movement videos has been proposed. It uses a semi-supervised model, termed SiamParseNet (SPN), which consists of two branches, one for intra-frame body parts segmentation and another for inter-frame label propagation. Another important contribution is the adoption of two training strategies to alternatively employ different training modes to achieve optimal performance. Factorized video GAN was exploited to augment training. Similarly, in [131], the automated analysis of general movements was achieved using low-cost instrumentation in the home. Videos from a single commercial RGB-D sensor were processed using DeepLabCut to estimate the 2D trajectories of selected points and then to reconstruct 3D trajectories by aligning data recorded with the depth sensor. Eight infants were recorded in the home at 3, 4, and 5 months of age.

The potential ability of computer vision to accurately characterize infant reaching motion is the topic of the paper in [130,134]. Analysing reaching motion (fast movement towards a given target, usually a toy) may contribute to the early diagnosis and assessment of infants at risk for upper extremity motor impairments. In [130] the analysed videos obtained were from 12 infants (5 with developmental disorders) of about 12 months of age or less. The total number of reaching actions analysed was 65. The x and y coordinates of hand key points were obtained from OpenPose and compared with those manually annotated (frame-by-frame), resulting in 95% confidence intervals. The authors concluded that OpenPose may be used for markerless automatic tracking of infant reaching motion from recorded videos, but did not provide evidence of the ability to automatically classify disorders.

In [134], a lightweight network was tested on videos of infants (up to 12 months of age) performing reaching/grabbing actions collected from an online video-sharing platform and semiautomatically annotated by exploiting the toll kinovea. A total of 193 reaches performed by 21 distinct subjects were processed with a precision of 0.57–0.66 and recall of 0.72–0.49 for reaching and no-reaching action, respectively.

Persistent asymmetrical body behaviour in early life provides a prominent prodromal risk marker of neurodevelopmental conditions like autism spectrum disorder and congenital muscular torticollis. The authors in [135] proposed a computer vision method for assessing bilateral infant postural symmetry from images, based on 3D human pose estimation, adapted to the challenging setting of infant bodies. In particular, the HW-HuP interactive annotation tool was modified to correct 3D poses predicted on infants in the SyRIP dataset. A Bayesian estimator of the ground truth derived from a probabilistic graphical model of fallible human raters was proposed.

A less debated area of research is devoted for measuring vital signs (especially in the neonatal intensive care unit) in a contactless fashion by exploiting RGB or RGBD data. These solutions are aimed at avoiding trauma and pain observed in traditional sensors-

based monitoring when removing the strong adhesive bond between the electrode and epidermis of pre-term infants. A preliminary study of a proposed non-contact system based on photoplethysmography (PPGi) and motion magnification is reported in [128]. The proposed non-contact system framework involved skin colour and motion magnification, region of interest (ROI) selection, spectral analysis and peak detection. Non-contact heart rate (HR) and respiratory rate (RR) in 10 infants were monitored and compared with ECG data. The authors concluded that the non-contact technique requires further investigations to improve the accuracy necessary for use with neonates. One of the main factors of failure was spotted in the reduced ROI to be analysed with respect to experiments involving adults. A similar approach was also proposed in [129], but just a baby was used to create a dataset and two different motion detection methods (based on frame differences and background subtraction, respectively) were applied individually and integrated to achieve better accuracy. For the same aim in [136], authors used depth information captured by two RGB-D cameras in order to reconstruct a 3D surface of a patient's torso with high spatial coverage. The volume was computed based on an octree subdivision technique of the 3D space. Finally, respiratory parameters were calculated from the estimated volume-time curve, but experiments were carried out only on a baby mannequin with an artificial test lung for infants. The lung was branched to a mechanical ventilator. Recently, in [125] remotely monitored both HR and RR of neonates in the NICU using colour and motion-based methods. The most interesting contribution of the paper is the use of YOLO V3 weights to achieve a baby detection model, detecting this way ROI automatically.

3.3. Posture/Gait Analysis

Problems related to standing infants are linked to motor deficits or temporary or chronic illnesses, problems involving the way a child walks, stands or sits that require precise quantitative assessment to evaluate both the severity of the pathology and the effectiveness of clinical treatments. From this perspective, the spectrum of dynamics involved is much broader and includes monitoring how they walk or sit and how they perform specific actions.

In this area, we can find works presenting tools for the diagnosis of motor impairments, for assessing temporal or chronic diseases, and for evaluating the efficacy of drugs or the outcomes of therapeutic sessions. Most relevant work is resumed in Table 8. It is worth noting that by examining these works, once again, it becomes clear how the demand for specific datasets can be pivotal for the development and assessment of specific algorithms.

Table 8. A Summary of selected work for posture/gait analysis. AUC = Area Under Curve, acc = accuracy (correct predictions/total number of predictions with respect to the clinical goal occurrences tested, ASD = Autism Spectrum Disorders.

Work (Year)	Method	Clinical Task	Metrics	Dataset Population/Age (in years)
[137] (2020)	OpenPose + Motion Param + CNN	Gait Analysis to predict surgery	AUC = 0.71	1026 (5–11)
[138] (2020)	AutoViDev + arms and legs time series distance	assessing coordination	qualitative	24 (1)
[139] (2022)	Optical flow + RGB + 3D CNN	ASD/Healthy	acc = 86.04%	60 (3–6)
[140] (2022)	OpenPose + motion parameters	Evaluating dystrophy	qualitative	11 (13)

The work in [140] focuses on Duchenne muscular dystrophy (DMD) and is aimed at developing a digital platform to enable innovative outcome measures. Eleven participants were involved (the median age was 13). Six participants were ambulant and five non-ambulant. Each participant was acquired AT HOME while performing tasks decided by medical experts. Video analysis was then performed using OpenPose software and different

parameters, such as trajectory, smoothness and symmetry of movement, and voluntary or compensatory movements were extracted. Data from the videos of DMD participants were compared to data from the healthy control on four tasks: walking, Hands-to-head while standing, Hands-to-head while sitting and Sit-to-stand then hands-to-head while standing. Front and side views were used.

In [139], videos of children with ASD in an uncontrolled environment were analysed by a multi-modality fusion network (RGB and optical flow) based on 3D CNNs. The final goal was recognizing autistic behaviours in videos. The method is based on I3D architecture pre-trained on a large-scale action recognition dataset and fine-tuned on a small dataset of stereotypic actions. The child was detected by Yolov5 [141] and tracked by DeepSORT algorithms[142]. Optical flow extraction was performed by the RAFT algorithm [143]. Extensive experiments on different deep learning frameworks were performed to propose a baseline. The best-gathered accuracy was 86.04% using a fusion of RGB and flow streams.

The authors in [137] analysed clinical **gait analysis** videos from young patients (average patient age was 11 years). For each video, they used OpenPose to extract time series of anatomical landmarks. Next, these time series were processed to create features for supervised machine learning models (CNN, RF, and RR) to predict gait parameters and clinical decisions. The approach relying on CNN for classification outperformed the others with an AUC of 0.71 in correctly predicting surgery decisions.

An interesting research was conducted in [138], in which authors observed the coordination patterns in 11-month-old pre-walking infants with a range of cruising (moving sideways in an upright posture while holding onto support) and crawling experiences. Computer vision tasks were delegated to the AutoViDev system. Subsequently, authors identified infants' coordination patterns demonstrating how infants learn to assemble solutions in real-time as they encounter new problems. This evolutionary model could be used to assess motor or neurological impairments.

3.4. New Research Directions for More Accurate Infants Pose Estimation

In this section, up-to-date computer vision strategies for human pose estimation are reported and discussed with reference to their possible application on infants and viable research directions. First, it is important to observe that all the listed strategies have been trained and tested on adults, and then an assessment of their performance on children is the first pathway to be suggested to the research community hoping that their efficiency in terms of outcomes (and sometimes also having a reduced computational workload) will be kept also on datasets involving children. Recently, transformer-based solutions have shown great success in 3D human pose estimation. Under this premise, a breakthrough work is the one introducing PoseFormer [144], a pure transformer-based approach for 3D pose estimation from 2D videos. The spatial transformer module encodes the local relationships between the 2D joints and the temporal transformer module captures global dependencies across the arbitrary frames regardless of the distance. Extensive experiments show that the PoseFormer model achieved state-of-the-art performance on popular 3D pose datasets. Code is available at https://github.com/zczcwh/PoseFormer (accessed on 4 September 2023). Another important related achievement that deserves a mention is PoseAug [145], a novel auto-augmentation framework that learns to augment the available training poses towards greater diversity and thus enhances the generalization power of the trained 2D-to-3D pose estimator. It has been conceived to address the existing problem of inferior generalization performance to new datasets of existing 3D human pose estimation methods. In other words, it augments the diversity of 2D–3D pose pairs in the training data. The code is available at https://github.com/jfzhang95/PoseAug (accessed on 4 September 2023). Both methods can speed up the clinical assessment and diagnosis of children due to their capability to localize joints with higher precision independently from specific acquisition setups and camera views. In [146], a Spatio-Temporal Criss-cross attention (STC) block has been introduced to improve joint correlation computation for comparing trajectories into the 3D space, including spatial and temporal analysis.

The system works very well on complicated pose articulation (as those of children are, especially while they lie in a bed). These systems are highly complex: to overcome this, a tokenization mechanism can allow us to operate on temporally sparse input poses but still generate dense 3D pose sequences as proposed in [147]. The code and models can be accessed at https://github.com/goldbricklemon/uplift-upsample-3dhpe (accessed on 4 September 2023). This could particularly help with children where occlusions often appear and reduce the availability of 2D data. Viable alternatives to transformers, also beyond CNN, have been also recently proposed. For example, in [148], capsule networks (CapsNets) have been introduced for 3D human pose estimation, ensuring viewpoint-equivariance and drastically reducing both the dataset size and the network complexity, while retaining high output accuracy. Its peculiarities make this approach very suitable for modelling children's poses with few shots, even using domestic setups.

4. Discussion

Protecting and safeguarding children's health is a key priority that benefits society as a whole. The World Health Organization and the United Nations Children's Fund have specialised in improving children's health, including care before and after birth [149]. There is evidence from longitudinal studies showing that the benefits of healthy childhood development extend to older ages [150].

Advances in the use of health technologies have the potential to bring further benefits to neonatal and paediatric healthcare, as it is recognised by the same health organisations [10]. Ambient intelligence and CV, as a means of unobtrusive, contactless, remote monitoring of children's physical, motor and mental health status and activities in both healthcare and private settings, can assist in a range of clinical tasks and thereby contribute to a better understanding of child physiology and pathophysiology [151].

The scientific and technological communities have become more aware of this potential in recent years, as is evidenced by the rising trend in the number of publications we have retrieved in the past five years (i.e., 54 out of the 65 papers retrieved have been published in the last 5 years).

Research in the field is spread across several countries. In Italy, there is a vibrant and lively community of scientists and scholars working to advance the scientific frontiers. Initially, research has mostly focused on monitoring and improving the interaction with children with ASD [41,56,105,152,153]. Most recently, attention has moved to the prediction of neurological development disorders [12,132], especially in relation to general movements [111,114,123] and for preterms [122,132].

Although great strides have been made in the past few years, several open issues and challenges still need to be addressed, also in relation to ethical and legal concerns, to reach a significant level of performance and to allow the instruments to be effectively used in clinical practice, as we will discuss in the following subsections.

4.1. Gaps and Open Challenges

Among existing challenges, the lack of task-specific public datasets, missing in several areas, represents one of the most critical issues. On the one side, this lack wastes the energy of several research groups to build datasets from scratch and, besides, it makes difficult a fair algorithms comparison. Of course, collecting datasets of children is even more challenging due to several reasons:

- Privacy and ethical concerns: Collecting data from children requires strict adherence to privacy laws and regulations, such as the General Data Protection Regulation (GDPR) and the Children's Online Privacy Protection Act (COPPA) in the United States. These laws require obtaining explicit consent from parents or guardians and ensuring the anonymity and security of children's personal information. Meeting these requirements can be complex and time-consuming.
- Parental consent: Obtaining parental consent for data collection can be difficult, especially if it involves sensitive information or requires active participation from

children. Parents may be concerned about the potential risks of data misuse or the potential impact on their child's privacy. Building trust and addressing these concerns is crucial, and it often involves clear communication and transparency about data handling practices.

- Limited accessibility: Children may have limited access to technology or may not be able to provide consistent or reliable data due to various factors like socioeconomic disparities, geographical location, or cultural norms. This can result in biased or incomplete datasets, which can negatively impact the performance and fairness of AI models.
- Dynamic and diverse nature of children's behaviour: Children's behaviour, cognition, and language skills undergo rapid development and change over time. Creating a dataset that adequately captures this dynamic nature requires extensive longitudinal studies, which can be resource-intensive and time-consuming.
- Ethical considerations in data collection: Collecting data from vulnerable populations, such as children, requires special care to ensure their well-being and protection. Researchers must consider the potential emotional or psychological impact on children and ensure that the data collection process is designed ethically and with sensitivity.
- Limited sample size: Children constitute a smaller population subset compared to adults, making it challenging to gather a sufficiently large and diverse dataset. Limited data can lead to overfitting, where the AI model performs well on the training data but fails to generalize to new examples.
- Consent withdrawal and data management: Children's participation in data collection should be voluntary, and they or their parents should have the right to withdraw consent at any time. Managing and removing data associated with withdrawn consent can be challenging, especially if it has already been incorporated into AI training models.

The use of large, shared datasets is a long way but there are other gaps and limitations in this topic that should be addressed. From the analysis of the literature, it has emerged that a large number of metrics are used to assess introduced machine learning and computer vision methods. How to select the most suited metric for each specific task is a big challenge and that choice should be shared and therefore universally accepted among research groups. In fact, even if data would have been available, experimental baselines must share common reference metrics. This is a big challenge, especially in the case of face analysis. Indeed, by observing the tables in Sections 2 and 3, it is possible to see a large number of used metrics, depending on the specifically addressed task. Some metrics look at the broad clinical problem (normal/atypical), whereas other ones concentrate on the finer visual task (e.g., landmark positioning) demanding a supervisor (human or automatic) to make a diagnosis. This way methods become not easily comparable, and it is not trivial to understand which one might help in clinical practice. On the other side, there are still many qualitative evaluations that do not help the clinician follow up on clinical practice since subjectiveness is not pushed away but even strengthened since it masters the process automatization. Another limitation that slows the development of effective machine learning approaches involving children is the need for long-term follow-up: many medical conditions require long-term follow-up to accurately observe and evaluate the clinical evolution. This extended time frame is necessary to assess the effectiveness of treatments, the progression of diseases, or the occurrence of relevant events. Waiting for this follow-up period adds to the time required for verification. This can also affect the statistical significance and sample size. Efforts in this research direction could also allow researchers to deploy foundation models, which are at the edge of machine learning research right now and are pushing ahead the so-called generalist medical AI [154].

Addressing these challenges requires interdisciplinary collaboration between researchers, ethicists, and legal experts to ensure that the collection and use of children's data for AI model training aligns with ethical and legal guidelines while prioritising the privacy and well-being of children. In the following section, we overview some of the most common ethical and legal concerns and suggest possible solutions where available.

4.2. Ethico-Legal Considerations

The debate on ethical and legal issues in neonatology and paediatrics is broad and long-standing and has been addressed in a large body of literature in the field, both from a general perspective [155] and in specific scenarios [156,157]. The ethical mandates to which clinical practice should adhere include respect for parental autonomy, the primacy of the best interests of the child, doing no harm, and the right to be informed and to give consent.

As far as ambient intelligence and CV are concerned, an ethical approach to technology development and a thorough understanding of all the relevant ethical, legal and social issues raised by monitoring technologies should be a top priority for researchers and innovators. This is the only way to ensure immediate acceptance and long-term use. The debate in this respect is more mature when ambient intelligence and assistive technologies are targeted at older adults and their caregivers [158]. In neonatology and paediatrics, a child-focused approach is certainly the way forward to ensure a safe, effective, ethical and equitable future for these technologies.

Most of the papers published to date have addressed ethics and bioethics for any technological aid in clinical practice [159,160], while some recent publications have addressed the ethical and legal implications of the use of artificial intelligence in child education [161,162], child entertainment [163] and child care [164,165].

Overall, we can identify some key ethical and legal issues that researchers in CV and ambient intelligence in childcare should be aware of. These are privacy, extensive validation, transparency and accountability, and are discussed below along with some recommendations to address them.

Privacy: the privacy and confidentiality of children and their parents are treated with high standards, as already introduced in the previous section. This hinders the rapid development of technology to some extent but ensures that children's dignity and respect are properly taken into account. It is worth noting that when ambient intelligence comes into play, privacy becomes an issue not only for patients and parents but also for clinicians and caregivers. Addressing this issue at the technical level requires the adoption of privacy-preservation approaches such as those based on privacy-preserving visual sensors (e.g., depth or thermal sensors) or those based on ad hoc techniques able to ensure context-aware visual privacy and retain all the information contained in RGB cameras [166]. This may help reduce the feeling of intrusion in parents and caregivers.

Extensive validation: scientists are aware of the inherent limitations of data-inductive techniques, such as those CV methods that use machine learning approaches. The accuracy of these methods is closely related to the type and quality of data used to train and develop them. For this reason, it is very important to perform extensive technical and clinical validation of such methods to verify their ability to generalise and handle unknown conditions. Standardised external validation and multi-centre studies should be carefully planned, together with standardised evaluation metrics, to demonstrate the reliability of the methods developed, particularly in terms of generalisability, safety and clinical value.

Transparency: the use of technology should be made clear and transparent, thus avoiding any grey areas and uncertainties in their adoption. This entails accounting for the relevant details about the data used, the actors involved, the choices and processes enacted during development along with the main scope and limitations of the CV and ambient intelligence tools. In addition, meaningful motivations behind their outputs should be provided, especially when they are used to support diagnostic and prognostic processes. Only this way, end-users and beneficiaries, mainly children, caregivers, clinicians, nurses and parents can really be aware and empowered by the CV- and AI-powered technologies and gather trust in them [167–169]. The final goal is actually to contribute to collaborative decision-making, by augmenting caregivers and recipients with powerful information-processing tools.

Accountability: healthcare professionals are responsible for justifying their actions and decisions to patients and their families, and are liable for any potential positive or negative impact on the patient's health. The use of decision support technologies, such as those

based on CV and ambient intelligence, should be clearly modelled in the legal framework of medical liability to avoid any grey area when clinicians decide to use the results of a tool or follow a suggestion received. This is still a very controversial issue. On a technical level, CV applications can implement traceability tools that document their entire development lifecycle, making it easier to deal with cases where something goes wrong.

5. Conclusions

This paper surveyed, for the first time in the literature, the works covering children's health-related issues by ambient intelligence methods and systems relying on computer vision. A taxonomy has been introduced by dividing works according to the part they concentrate on, e.g., the face for extracting gaze direction and facial expressions, or the whole body, for gait analysis, posture estimation, and human–object interaction. For each research area, publicly available datasets and new computer vision perspectives have been discussed with particular attention on some challenges that still need to be addressed to reach a level of performance that allows the instruments to be effectively used in clinical practice.

In the coming years, we expect to see a significant increase in work in this area, both from the ethico-legal community and from the scientific and technological community. In particular, with regard to scientific and technological advances, future developments are expected to take place in several directions:

- The collection and availability of larger datasets, also covering longer periods of children monitoring;
- The improvement of current solutions thanks to more precise and advanced methods, also based on foundational vision models;
- The integration of different types of visual sensors, such as thermal cameras that might provide relevant information for instance about the development of the thermoregulatory system of newborns;
- The integrated processing of multimodal data, such as audio signals (e.g., to monitor children's crying), IoT data (e.g., from smart mattresses) and videos, thereby allowing, for example, a comprehensive monitoring of the health and well-being status of newborns in nurseries or in NICUs;
- The optimization of computing and sensing facilities to enable technology diffusion in resource-limited and most needy countries.

Overall, considering the new perspectives that CV and machine learning tools can open, we deem it relevant to stress that researchers and innovators should strive to comply with several mandates at technical, socio-ethical and organizational levels. Solutions should strictly comply with existing and emerging regulations, such as that the Artificial Intelligence Act (COM/2021/206 final—available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206, accessed on 15 September 2023). Only this way, they can aspire to have real-life adoption and, thus, have an actual impact. Currently, innovation endeavours in this field are still in their early stages, but we are sure they can benefit from the more mature discussion going on in the field of ambient intelligence and Active and Assisted Living, towards a really beneficial application for children, parents, caregivers and society at a large [158].

Author Contributions: Conceptualization, S.C. and M.L.; methodology, all authors; formal analysis, D.G., M.D.C., M.L. and P.C.; investigation, all authors; writing—original draft preparation, D.G. and M.L.; writing—review and editing, all authors; supervision, S.C. and M.L.; funding acquisition, S.C. and M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by Future Artificial Intelligence Research—FAIR CUP B53C220036 30006 grant number PE0000013 and by the Cost Action 19121 GoodBrother—"Network on Privacy-Aware Audio- and Video-Based Applications for Active and Assisted Living".

Data Availability Statement: No new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Leo, M.; Farinella, G.M. *Computer Vision for Assistive Healthcare*; Academic Press: Cambridge, MA, USA, 2018.
2. Esteva, A.; Chou, K.; Yeung, S.; Naik, N.; Madani, A.; Mottaghi, A.; Liu, Y.; Topol, E.; Dean, J.; Socher, R. Deep learning-enabled medical computer vision. *NPJ Digit. Med.* **2021**, *4*, 5. [CrossRef]
3. Aleksic, S.; Atanasov, M.; Agius, J.C.; Camilleri, K.; Cartolovni, A.; Climent-Peerez, P.; Colantonio, S.; Cristina, S.; Despotovic, V.; Ekenel, H.K.; et al. State of the Art of Audio- and Video-Based Solutions for AAL. *arXiv* **2022**, arXiv:2207.01487.
4. Haque, A.; Milstein, A.; Li, F.-F. Illuminating the dark spaces of healthcare with ambient intelligence. *Nature* **2020**, *585*, 202. [CrossRef] [PubMed]
5. Andreu, Y.; Chiarugi, F.; Colantonio, S.; Giannakakis, G.; Giorgi, D.; Henriquez, P.; Kazantzaki, E.; Manousos, D.; Marias, K.; Matuszewski, B.J.; et al. Wize Mirror—A smart, multisensory cardio-metabolic risk monitoring system. *Comput. Vis. Image Underst.* **2016**, *148*, 3–22. [CrossRef]
6. Chaddad, A.; Peng, J.; Xu, J.; Bouridane, A. Survey of Explainable AI Techniques in Healthcare. *Sensors* **2023**, *23*, 634. [CrossRef] [PubMed]
7. Dunne, R.; Morris, T.; Harper, S. A survey of ambient intelligence. *ACM Comput. Surv.* **2021**, *54*, 1–27. [CrossRef]
8. Leo, M.; Carcagnì, P.; Mazzeo, P.L.; Spagnolo, P.; Cazzato, D.; Distante, C. Analysis of facial information for healthcare applications: A survey on computer vision-based approaches. *Information* **2020**, *11*, 128. [CrossRef]
9. Dimitri, P. Child health technology: Shaping the future of paediatrics and child health and improving NHS productivity. *Arch. Dis. Child.* **2019**, *104*, 184–188. [CrossRef]
10. Sacks, L.; Kunkoski, E.; Noone, M. Digital Health Technologies in Pediatric Trials. *Ther. Innov. Regul. Sci.* **2022**, *56*, 929–933. [CrossRef]
11. Senechal, E.; Jeanne, E.; Tao, L.; Kearney, R.; Shalish, W.; Sant'Anna, G. Wireless monitoring devices in hospitalized children: A scoping review. *Eur. J. Pediatr.* **2023**, *182*, 1991–2003. [CrossRef]
12. Leo, M.; Bernava, G.M.; Carcagnì, P.; Distante, C. Video-Based Automatic Baby Motion Analysis for Early Neurological Disorder Diagnosis: State of the Art and Future Directions. *Sensors* **2022**, *22*, 866. [CrossRef]
13. Silva, N.; Zhang, D.; Kulvicius, T.; Gail, A.; Barreiros, C.; Lindstaedt, S.; Kraft, M.; Bölte, S.; Poustka, L.; Nielsen-Saines, K.; et al. The future of General Movement Assessment: The role of computer vision and machine learning—A scoping review. *Res. Dev. Disabil.* **2021**, *110*, 103854. [CrossRef] [PubMed]
14. Marcroft, C.; Khan, A.; Embleton, N.D.; Trenell, M.; Plötz, T. Movement recognition technology as a method of assessing spontaneous general movements in high risk infants. *Front. Neurol.* **2015**, *5*, 284. [CrossRef] [PubMed]
15. Hallemans, A.; Van de Walle, P.; Wyers, L.; Verheyen, K.; Schoonjans, A.; Desloovere, K.; Ceulemans, B. Clinical usefulness and challenges of instrumented motion analysis in patients with intellectual disabilities. *Gait Posture* **2019**, *71*, 105–115. [CrossRef]
16. Washington, P.; Park, N.; Srivastava, P.; Voss, C.; Kline, A.; Varma, M.; Tariq, Q.; Kalantarian, H.; Schwartz, J.; Patnaik, R.; et al. Data-driven diagnostics and the potential of mobile artificial intelligence for digital therapeutic phenotyping in computational psychiatry. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **2020**, *5*, 759–769. [CrossRef] [PubMed]
17. de Belen, R.A.J.; Bednarz, T.; Sowmya, A.; Del Favero, D. Computer vision in autism spectrum disorder research: A systematic review of published studies from 2009 to 2019. *Transl. Psychiatry* **2020**, *10*, 333. [CrossRef]
18. Mercan, E.; Morrison, C.; Stuhaug, E.; Shapiro, L.; Tse, R. Novel computer vision analysis of nasal shape in children with unilateral cleft lip. *J. Cranio-Maxillo Surg. Off. Publ. Eur. Assoc. Cranio-Maxillo Surg.* **2018**, *46*, 35–43. [CrossRef]
19. Wu, J.; Tse, R.; Shapiro, L. Automated face extraction and normalization of 3d mesh data. In Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 750–753.
20. Wu, J.; Heike, C.; Birgfeld, C.; Evans, K.; Maga, M.; Morrison, C.; Saltzman, B.; Shapiro, L.; Tse, R. Measuring Symmetry in Children With Unrepaired Cleft Lip: Defining a Standard for the Three-Dimensional Midfacial Reference Plane. *Cleft Palate-Craniofacial J. Off. Publ. Am. Cleft Palate-Craniofacial Assoc.* **2016**, *53*, 695–704. [CrossRef]
21. Wu, J.; Liang, S.; Shapiro, L.; Tse, R. Measuring Symmetry in Children With Cleft Lip. Part 2: Quantification of Nasolabial Symmetry Before and After Cleft Lip Repair. *Cleft Palate-Craniofacial J. Off. Publ. Am. Cleft Palate-Craniofacial Assoc.* **2016**, *53*, 705–713. [CrossRef]
22. Narayanan, D.L.; Ranganath, P.; Aggarwal, S.; Dalal, A.; Phadke, S.; Mandal, K. Computer-aided Facial Analysis in Diagnosing Dysmorphic Syndromes in Indian Children. *Indian Pediatr.* **2019**, *56*, 1017–1019. [CrossRef]
23. Torres Torres, M.; Valstar, M.; Henry, C.; Ward, C.; Sharkey, D. Postnatal gestational age estimation of newborns using Small Sample Deep Learning. *Image Vis. Comput.* **2019**, *83–84*, 87–99. [CrossRef] [PubMed]
24. Winter, R. What's in a face? *Nat. Genet.* **1996**, *12*, 124–129. [CrossRef] [PubMed]
25. Gurovich, Y.; Hanani, Y.; Bar, O.; Fleischer, N.; Gelbman, D.; Basel-Salmon, L.; Krawitz, P.; Kamphausen, S.; Zenker, M.; Bird, L.; et al. DeepGestalt-identifying rare genetic syndromes using deep learning. *arXiv* **2018**, arXiv:1801.07637.
26. Hustinx, A.; Hellmann, F.; Sümer, Ö.; Javanmardi, B.; André, E.; Krawitz, P.; Hsieh, T.C. Improving Deep Facial Phenotyping for Ultra-rare Disorder Verification Using Model Ensembles. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 5018–5028.
27. Bovery, M.; Dawson, G.; Hashemi, J.; Sapiro, G. A Scalable Off-the-Shelf Framework for Measuring Patterns of Attention in Young Children and Its Application in Autism Spectrum Disorder. *IEEE Trans. Affect. Comput.* **2021**, *12*, 722–731. [CrossRef] [PubMed]

28. Chang, Z.; Di Martino, M.; Aiello, R.; Baker, J.; Carpenter, K.; Compton, S.; Davis, N.; Eichner, B.; Espinosa, S.; Flowers, J.; et al. Computational Methods to Measure Patterns of Gaze in Toddlers With Autism Spectrum Disorder. *JAMA Pediatr.* **2021**, *175*, 827–836. [CrossRef]
29. Varma, M.; Washington, P.; Chrisman, B.; Kline, A.; Leblanc, E.; Paskov, K.; Stockham, N.; Jung, J.Y.; Sun, M.W.; Wall, D. Identification of social engagement indicators associated with autism spectrum disorder using a game-based mobile application. *J. Med. Internet Res.* **2021**, *24*, e31830. [CrossRef]
30. Hashemi, J.; Campbell, K.; Carpenter, K.; Harris, A.; Qiu, Q.; Tepper, M.; Espinosa, S.; Schaich Borg, J.; Marsan, S.; Calderbank, R.; et al. A scalable app for measuring autism risk behaviors in young children: A technical validity and feasibility study. In Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare, London, UK, 14–16 October 2015; pp. 23–27.
31. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
33. Ramirez-Duque, A.; Frizera-Neto, A.; Bastos, T.F. Robot-Assisted Autism Spectrum Disorder Diagnostic Based on Artificial Reasoning. *J. Intell. Robot Syst.* **2019**, *96*, 267–281. [CrossRef]
34. Prinsen, V.; Jouvet, P.; Al Omar, S.; Masson, G.; Bridier, A.; Noumeir, R. Automatic eye localization for hospitalized infants and children using convolutional neural networks. *Int. J. Med. Inform.* **2021**, *146*, 104344. [CrossRef]
35. King, D.E. Dlib-ml: A machine learning toolkit. *J. Mach. Learn. Res.* **2009**, *10*, 1755–1758.
36. Baltrusaitis, T.; Robinson, P.; Morency, L.P. OpenFace: An open source facial behavior analysis toolkit. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016; pp. 1–10.
37. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, Montreal, QC, Canada, 7–12 December 2015; pp. 91–99.
38. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
39. Zaker, N.; Mahoor, M.H.; Mattson, W.I.; Messinger, D.S.; Cohn, J.F. A comparison of alternative classifiers for detecting occurrence and intensity in spontaneous facial expression of infants with their mothers. In Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 22–26 April 2013; pp. 1–6.
40. Zanette, S.; Gao, X.; Brunet, M.; Bartlett, M.S.; Lee, K. Automated decoding of facial expressions reveals marked differences in children when telling antisocial versus prosocial lies. *J. Exp. Child Psychol.* **2016**, *150*, 165–179. [CrossRef]
41. Del Coco, M.; Leo, M.; Carcagnì, P.; Spagnolo, P.; Mazzeo, P.L.; Bernava, M.; Marino, F.; Pioggia, G.; Distante, C. A Computer Vision Based Approach for Understanding Emotional Involvements in Children with Autism Spectrum Disorders. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 1401–1407.
42. Sun, Y.; Shan, C.; Tan, T.; Long, X.; Pourtaherian, A.; Zinger, S.; de With, P. Video-based discomfort detection for infants. *Mach. Vis. Appl.* **2019**, *30*, 933–944. [CrossRef]
43. Zamzmi, G.; Paul, R.; Salekin, M.S.; Goldgof, D.; Kasturi, R.; Ho, T.; Sun, Y. Convolutional Neural Networks for Neonatal Pain Assessment. *IEEE Trans. Biom. Behav. Identity Sci.* **2019**, *1*, 192–200. [CrossRef]
44. Filntisis, P.P.; Efthymiou, N.; Koutras, P.; Potamianos, G.; Maragos, P. Fusing Body Posture With Facial Expressions for Joint Recognition of Affect in Child–Robot Interaction. *IEEE Robot. Autom. Lett.* **2019**, *4*, 4011–4018. [CrossRef]
45. Nagpal, S.; Singh, M.; Vatsa, M.; Singh, R.; Noore, A. Expression Classification in Children Using Mean Supervised Deep Boltzmann Machine. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–17 June 2019; pp. 236–245.
46. Martinez, A.; Pujol, F.; Mora, H. Application of Texture Descriptors to Facial Emotion Recognition in Infants. *Appl. Sci.* **2020**, *10*, 1115. [CrossRef]
47. Jarraya, S.K.; Masmoudi, M.; Hammami, M. A comparative study of Autistic Children Emotion recognition based on Spatio-Temporal and Deep analysis of facial expressions features during a Meltdown Crisis. *Multimed. Tools Appl.* **2021**, *80*, 83–125. [CrossRef]
48. Jarraya, S.K.; Masmoudi, M.; Hammami, M. Compound Emotion Recognition of Autistic Children During Meltdown Crisis Based on Deep Spatio-Temporal Analysis of Facial Geometric Features. *IEEE Access* **2020**, *8*, 69311–69326. [CrossRef]
49. Carpenter, K.; Hahemi, J.; Campbell, K.; Lippmann, S.; Baker, J.; Egger, H.; Espinosa, S.; Vermeer, S.; Sapiro, G.; Dawson, G. Digital Behavioral Phenotyping Detects Atypical Pattern of Facial Expression in Toddlers with Autism. *Autism Res. Off. J. Int. Soc. Autism Res.* **2021**, *14*, 488–499. [CrossRef]
50. Hoti, K.; Chivers, P.T.; Hughes, J.D. Assessing procedural pain in infants: A feasibility study evaluating a point-of-care mobile solution based on automated facial analysis. *Lancet Digit. Health* **2021**, *3*, 623–634. [CrossRef]
51. Salekin, S.; Zamzmi, G.; Goldgof, D.; Kasturi, R.; Ho, T.; Sun, Y. Multimodal spatio-temporal deep learning approach for neonatal postoperative pain assessment. *Comput. Biol. Med.* **2021**, *129*, 104150. [CrossRef]
52. Washington, P.; Kalantarian, H.; Kent, J.; Husic, A.; Kline, A.; Leblanc, E.; Hou, C.; Mutlu, C.; Dunlap, K.; Penev, Y.; et al. Training affective computer vision models by crowdsourcing soft-target labels. *Cogn. Comput.* **2021**, *13*, 1363–1373. [CrossRef]

53. Ertugrul, I.O.; Ahn, Y.A.; Bilalpur, M.; Messinger, D.S.; Speltz, M.L.; Cohn, J.F. Infant AFAR: Automated facial action recognition in infants. *Behav. Res. Methods* **2022**, *55*, 1024–1035. [CrossRef]
54. Zamzmi, G.; Pai, C.Y.; Goldgof, D.; Kasturi, R.; Ashmeade, T.; Sun, Y. A Comprehensive and Context-Sensitive Neonatal Pain Assessment Using Computer Vision. *IEEE Trans. Affect. Comput.* **2022**, *13*, 28–45. [CrossRef]
55. Qayyum, A.; Razzak, I.; Moustafa, N.; Mazher, M. Progressive ShallowNet for large scale dynamic and spontaneous facial behaviour analysis in children. *Image Vis. Comput.* **2022**, *119*, 104375. [CrossRef]
56. Leo, M.; Carcagnì, P.; Distante, C.; Mazzeo, P.L.; Spagnolo, P.; Levante, A.; Petrocchi, S.; Lecciso, F. Computational Analysis of Deep Visual Data for Quantifying Facial Expression Production. *Appl. Sci.* **2019**, *9*, 4542. [CrossRef]
57. Leo, M.; Carcagnì, P.; Distante, C.; Spagnolo, P.; Mazzeo, P.L.; Rosato, A.C.; Petrocchi, S.; Pellegrino, C.; Levante, A.; De Lumè, F.; et al. Computational Assessment of Facial Expression Production in ASD Children. *Sensors* **2018**, *18*, 3993. [CrossRef]
58. Leo, M.; Carcagnì, P.; Coco, M.D.; Spagnolo, P.; Mazzeo, P.L.; Celeste, G.; Distante, C.; Lecciso, F.; Levante, A.; Rosato, A.C.; et al. Towards the automatic assessment of abilities to produce facial expressions: The case study of children with ASD. In Proceedings of the 20th Italian National Conference on Photonic Technologies (Fotonica 2018), Lecce, Italy, 23–25 May 2018; pp. 1–4.
59. Oster, H. *Baby Facs: Facial Action Coding System for Infants and Young Children*; Unpublished Monograph and Coding Manual; New York University: New York, NY, USA, 2006.
60. Li, C.; Pourtaherian, A.; van Onzenoort, L.; Ten, W.E.T.A.; de With, P.H.N. Infant Facial Expression Analysis: Towards a Real-Time Video Monitoring System Using R-CNN and HMM. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1429–1440. [CrossRef] [PubMed]
61. Brahnam, S.; Chuang, C.F.; Shih, F.Y.; Slack, M.R. SVM classification of neonatal facial images of pain. In Proceedings of the International Workshop on Fuzzy Logic and Applications, Crema, Italy, 15–17 September 2005; Springer: Berlin/Heidelberg, Germany, 2005; pp. 121–128.
62. Brahnam, S.; Chuang, C.F.; Sexton, R.; Shih, F. Machine assessment of neonatal facial expressions of acute pain. *Decis. Support Syst.* **2007**, *43*, 1242–1254. [CrossRef]
63. Redmon, J.; Farhadi, A. Yolov3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
64. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
65. Salekin, M.S.; Zamzmi, G.; Hausmann, J.; Goldgof, D.; Kasturi, R.; Kneusel, M.; Ashmeade, T.; Ho, T.; Sun, Y. Multimodal neonatal procedural and postoperative pain assessment dataset. *Data Brief* **2021**, *35*, 106796. [CrossRef] [PubMed]
66. Haines, N.; Bell, Z.; Crowell, S.; Hahn, H.; Kamara, D.; McDonough-Caplan, H.; Shader, T.; Beauchaine, T. Using automated computer vision and machine learning to code facial expressions of affect and arousal: Implications for emotion dysregulation research. *Dev. Psychopathol.* **2019**, *31*, 871–886. [CrossRef]
67. Bruer, K.C.; Zanette, S.; Ding, X.P.; Lyon, T.D.; Lee, K. Identifying Liars Through Automatic Decoding of Children's Facial Expressions. *Child Dev.* **2020**, *91*, e995–e1011. [CrossRef] [PubMed]
68. Chen, M.; Chow, S.M.; Hammal, Z.; Messinger, D.S.; Cohn, J.F. A person-and time-varying vector autoregressive model to capture interactive infant-mother head movement dynamics. *Multivar. Behav. Res.* **2020**, *56*, 739–767. [CrossRef]
69. Hammal, Z.; Cohn, J.F.; Messinger, D.S. Head movement dynamics during play and perturbed mother-infant interaction. *IEEE Trans. Affect. Comput.* **2015**, *6*, 361–370. [CrossRef] [PubMed]
70. Hammal, Z.; Chu, W.S.; Cohn, J.F.; Heike, C.; Speltz, M.L. Automatic action unit detection in infants using convolutional neural network. In Proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, TX, USA, 23–26 October 2017; pp. 216–221.
71. Khan, R.A.; Crenn, A.; Meyer, A.; Bouakaz, S. A novel database of children's spontaneous facial expressions (LIRIS-CSE). *Image Vis. Comput.* **2019**, *83*, 61–69. [CrossRef]
72. Salakhutdinov, R.; Hinton, G. Deep boltzmann machines. In Proceedings of the 12th International Conference on Artificial Intelligence and Statistics, Clearwater Beach, FL, USA, 16–18 April 2009; pp. 448–455.
73. Ekman, P. Facial Expressions of Emotion: New Findings, New Questions. *Psychol. Sci.* **1992**, *3*, 34–38. [CrossRef]
74. Langner, O.; Dotsch, R.; Bijlstra, G.; Wigboldus, D.H.; Hawk, S.T.; Van Knippenberg, A. Presentation and validation of the Radboud Faces Database. *Cogn. Emot.* **2010**, *24*, 1377–1388. [CrossRef]
75. LoBue, V.; Trasher, C. The Child Affective Facial Expression (CAFE) Set: Validity and Reliability from Untrained Adults. *Front. Psychol.* **2014**, *5*, 1532. [CrossRef]
76. Belpaeme, T.; Baxter, P.; De Greeff, J.; Kennedy, J.; Read, R.; Looije, R.; Neerincx, M.; Baroni, I.; Zelati, M.C. Child-robot interaction: Perspectives and challenges. In Proceedings of the Social Robotics: 5th International Conference—ICSR 2013, Bristol, UK, 27–29 October 2013; Springer: Cham, Switzerland, 2013; pp. 452–459.
77. Xu, K.; Ji, B.; Wang, Z.; Liu, J.; Liu, H. An Auxiliary Screening System for Autism Spectrum Disorder Based on Emotion and Attention Analysis. In Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, 11–14 October 2020; pp. 2299–2304.
78. Zhang, Y.; Kong, M.; Zhao, T.; Hong, W.; Xie, D.; Wang, C.; Yang, R.; Li, R.; Zhu, Q. Auxiliary diagnostic system for ADHD in children based on AI technology. *Front. Inf. Technol. Electron. Eng.* **2021**, *22*, 400–414. [CrossRef]
79. Xiong, X.; De la Torre, F. Supervised Descent Method and Its Applications to Face Alignment. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 532–539.
80. Kanade, T.; Cohn, J.; Tian, Y. Comprehensive database for facial expression analysis. In Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 28–30 March 2000; pp. 46–53.

81. Nag, A.; Haber, N.; Voss, C.; Tamura, S.; Daniels, J.; Ma, J.; Chiang, B.; Ramachandran, S.; Schwartz, J.N.; Winograd, T.; et al. Toward Continuous Social Phenotyping: Analyzing Gaze Patterns in an Emotion Recognition Task for Children With Autism Through Wearable Smart Glasses. *J. Med. Internet Res.* **2020**, *22*, e13810. [CrossRef] [PubMed]
82. Ekman, P.; Friesen, W. *Pictures of Facial Affect*; Consulting Psychologists Press: Palo Alto, CA, USA, 1976.
83. Luquetti, D.V.; Speltz, M.L.; Wallace, E.R.; Siebold, B.; Collett, B.R.; Drake, A.F.; Johns, A.L.; Kapp-Simon, K.A.; Kinter, S.L.; Leroux, B.G.; et al. Methods and challenges in a cohort study of infants and toddlers with craniofacial microsomia: The CLOCK study. *Cleft Palate-Craniofacial J.* **2019**, *56*, 877–889. [CrossRef] [PubMed]
84. Adamson, L.B.; Frick, J.E. The Still Face: A History of a Shared Experimental Paradigm. *Infancy* **2003**, *4*, 451–473. [CrossRef]
85. Hudson-Barr, D.; Capper-Michel, B.; Lambert, S.; Palermo, T.M.; Morbeto, K.; Lombardo, S. Validation of the pain assessment in neonates (PAIN) scale with the neonatal infant pain scale (NIPS). *Neonatal Netw.* **2002**, *21*, 15–21. [CrossRef]
86. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1701–1708. [CrossRef]
87. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A Unified Embedding for Face Recognition and Clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
88. Ruiz, N.; Chong, E.; Rehg, J.M. Fine-Grained Head Pose Estimation Without Keypoints. In Proceedings of the The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, UT, USA, 18–23 June 2018.
89. Guo, J.; Zhu, X.; Yang, Y.; Yang, F.; Lei, Z.; Li, S.Z. Towards Fast, Accurate and Stable 3D Dense Face Alignment. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020.
90. Prados-Torreblanca, A.; Buenaposada, J.M.; Baumela, L. Shape Preserving Facial Landmarks with Graph Attention Networks. In Proceedings of the 33rd British Machine Vision Conference 2022—BMVC 2022, London, UK, 21–24 November 2022; BMVA Press: Surrey, UK, 2022.
91. Albiero, V.; Chen, X.; Yin, X.; Pang, G.; Hassner, T. img2pose: Face Alignment and Detection via 6DoF, Face Pose Estimation. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021.
92. Hempel, T.; Abdelrahman, A.A.; Al-Hamadi, A. 6d Rotation Representation For Unconstrained Head Pose Estimation. In Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022; pp. 2496–2500. [CrossRef]
93. Abate, A.F.; Bisogni, C.; Castiglione, A.; Nappi, M. Head Pose Estimation: An Extensive Survey on Recent Techniques and Applications. *Pattern Recogn.* **2022**, *127*, 108591. [CrossRef]
94. Asperti, A.; Filippini, D. Deep Learning for Head Pose Estimation: A Survey. *SN Comput. Sci.* **2023**, *4*, 349. [CrossRef]
95. Deng, J.; Guo, J.; Niannan, X.; Zafeiriou, S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
96. Deng, J.; Guo, J.; Zhou, Y.; Yu, J.; Kotsia, I.; Zafeiriou, S. RetinaFace: Single-stage Dense Face Localisation in the Wild. *arXiv* **2019**, arXiv:1905.00641.
97. Qi, D.; Tan, W.; Yao, Q.; Liu, J. YOLO5Face: Why Reinventing a Face Detector. *arXiv* **2021**, arXiv:2105.12931.
98. Bharadwaj, S.; Bhatt, H.S.; Vatsa, M.; Singh, R. Domain Specific Learning for Newborn Face Recognition. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 1630–1641. [CrossRef]
99. Wen, D.; Fang, C.; Ding, X.; Zhang, T. Development of Recognition Engine for Baby Faces. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 3408–3411.
100. Wang, Z.; Zeng, F.; Liu, S.; Zeng, B. OAENet: Oriented attention ensemble for accurate facial expression recognition. *Pattern Recognit.* **2021**, *112*, 107694. [CrossRef]
101. Zhuang, X.; Liu, F.; Hou, J.; Hao, J.; Cai, X. Transformer-Based Interactive Multi-Modal Attention Network for Video Sentiment Detection. *Neural Process. Lett.* **2022**, *54*, 1943–1960. [CrossRef]
102. Yang, H.; Zhu, K.; Huang, D.; Li, H.; Wang, Y.; Chen, L. Intensity enhancement via GAN for multimodal face expression recognition. *Neurocomputing* **2021**, *454*, 124–134. [CrossRef]
103. Zhang, T.; Tang, K. An Efficacious Method for Facial Expression Recognition: GAN Erased Facial Feature Network (GE2FN). In Proceedings of the 2021 13th International Conference on Machine Learning and Computing (ICMLC 2021), Shenzhen China, 26 February–1 March 2021; pp. 417–422. [CrossRef]
104. Schoneveld, L.; Othmani, A.; Abdelkawy, H. Leveraging recent advances in deep learning for audio-Visual emotion recognition. *Pattern Recognit. Lett.* **2021**, *146*, 1–7. [CrossRef]
105. Sciortino, G.; Farinella, G.M.; Battiato, S.; Leo, M.; Distante, C. On the estimation of children's poses. In Proceedings of the International Conference on Image Analysis and Processing, Catania, Italy, 11–15 September 2017; pp. 410–421.
106. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 172–186. [CrossRef] [PubMed]
107. Zha, D.; Bhat, Z.P.; Chen, Y.W.; Wang, Y.; Ding, S.; Chen, J.; Lai, K.H.; Bhat, M.Q.; Jain, A.K.; Reyes, A.C.; et al. Autovideo: An automated video action recognition system. *arXiv* **2021**, arXiv:2108.04212.
108. Mathis, A.; Mamidanna, P.; Cury, K.M.; Abe, T.; Murthy, V.N.; Mathis, M.W.; Bethge, M. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **2018**, *21*, 1281–1289. [CrossRef]

109. Torralba, A.; Russell, B.C.; Yuen, J. Labelme: Online image annotation and applications. *Proc. IEEE* **2010**, *98*, 1467–1484. [CrossRef]
110. Liu, S.; Huang, X.; Fu, N.; Ostadabbas, S. Heuristic weakly supervised 3d human pose estimation in novel contexts without any 3d pose ground truth. *arXiv* **2021**, arXiv:2105.10996.
111. Bernava, G.M.; Leo, M.; Carcagnì, P.; Distante, C. An Advanced Tool for Semi-automatic Annotation for Early Screening of Neurodevelopmental Disorders. In Proceedings of the Image Analysis and Processing—ICIAP 2022 Workshops: ICIAP International Workshops, Lecce, Italy, 23–27 May 2022; Revised Selected Papers, Part II; Springer: Cham, Swizterland, 2022; pp. 154–164.
112. Bazarevsky, V.; Grishchenko, I.; Raveendran, K.; Zhu, T.; Zhang, F.; Grundmann, M. Blazepose: On-device real-time body pose tracking. *arXiv* **2020**, arXiv:2006.10204.
113. Orlandi, S.; Guzzetta, A.; Bandini, A.; Belmonti, V.; Barbagallo, S.D.; Tealdi, G.; Mazzotti, S.; Scattoni, M.L.; Manfredi, C. AVIM—A contactless system for infant data acquisition and analysis: Software architecture and first results. *Biomed. Signal Process. Control* **2015**, *20*, 85–99. [CrossRef]
114. Baccinelli, W.; Bulgheroni, M.; Simonetti, V.; Fulceri, F.; Caruso, A.; Gila, L.; Scattoni, M.L. Movidea: A software package for automatic video analysis of movements in infants at risk for neurodevelopmental disorders. *Brain Sci.* **2020**, *10*, 203. [CrossRef]
115. Hesse, N.; Bodensteiner, C.; Arens, M.; Hofmann, U.G.; Weinberger, R.; Sebastian Schroeder, A. Computer vision for medical infant motion analysis: State of the art and rgb-d data set. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018; pp. 32–49.
116. Huang, X.; Fu, N.; Liu, S.; Ostadabbas, S. Invariant representation learning for infant pose estimation with small data. In Proceedings of the 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), Jodhpur, India, 15–18 December 2021; pp. 1–8.
117. Zhang, Q.; Xue, Y.; Huang, X. Online training for body part segmentation in infant movement videos. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 489–492.
118. Yang, C.Y.; Jiang, Z.; Gu, S.Y.; Hwang, J.N.; Yoo, J.H. Unsupervised Domain Adaptation Learning for Hierarchical Infant Pose Recognition with Synthetic Data. In Proceedings of the 2022 IEEE International Conference on Multimedia and Expo (ICME), Taipei, Taiwan, 18–22 July 2022; pp. 1–6.
119. Chambers, C.; Seethapathi, N.; Saluja, R.; Loeb, H.; Pierce, S.R.; Bogen, D.K.; Prosser, L.; Johnson, M.J.; Kording, K.P. Computer vision to automatically assess infant neuromotor risk. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 2431–2442. [CrossRef]
120. Aylward, G. *Bayley Infant Neurodevelopmental Screener*; Pearson: San Antonio, TX, USA, 1995.
121. Piper, M.C.; Pinnell, L.E.; Darrah, J.; Maguire, T.; Byrne, P.J. Construction and validation of the Alberta Infant Motor Scale (AIMS). *Can. J. Public Health Rev. Can. De Sante Publique* **1992**, *83*, S46–50.
122. Moccia, S.; Migliorelli, L.; Carnielli, V.; Frontoni, E. Preterm infants' pose estimation with spatio-temporal features. *IEEE Trans. Biomed. Eng.* **2019**, *67*, 2370–2380. [CrossRef]
123. Migliorelli, L.; Moccia, S.; Pietrini, R.; Carnielli, V.P.; Frontoni, E. The babyPose dataset. *Data Brief* **2020**, *33*, 106329. [CrossRef]
124. Rajagopalan, S.; Dhall, A.; Goecke, R. Self-stimulatory behaviours in the wild for autism diagnosis. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Sydney, Australia, 2–8 December 2013; pp. 755–761.
125. Khanam, F.T.Z.; Perera, A.G.; Al-Naji, A.; Gibson, K.; Chahl, J. Non-contact automatic vital signs monitoring of infants in a neonatal intensive care unit based on neural networks. *J. Imaging* **2021**, *7*, 122. [CrossRef]
126. Einspieler, C.; Prechtl, H.F. Prechtl's assessment of general movements: A diagnostic tool for the functional assessment of the young nervous system. *Ment. Retard. Dev. Disabil. Res. Rev.* **2005**, *11*, 61–67. [CrossRef] [PubMed]
127. Sun, Y.; Kommers, D.; Wang, W.; Joshi, R.; Shan, C.; Tan, T.; Aarts, R.M.; van Pul, C.; Andriessen, P.; de With, P.H. Automatic and continuous discomfort detection for premature infants in a NICU using video-based motion analysis. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 5995–5999.
128. Gibson, K.; Al-Naji, A.; Fleet, J.; Steen, M.; Esterman, A.; Chahl, J.; Huynh, J.; Morris, S. Non-contact heart and respiratory rate monitoring of preterm infants based on a computer vision system: A method comparison study. *Pediatr. Res.* **2019**, *86*, 738–741. [CrossRef] [PubMed]
129. Hussain, T.; Muhammad, K.; Khan, S.; Ullah, A.; Lee, M.Y.; Baik, S.W. Intelligent baby behavior monitoring using embedded vision in IoT for smart healthcare centers. *J. Artif. Intell. Syst.* **2019**, *1*, 110–124. [CrossRef]
130. Sahin, I.; Modi, A.; Kokkoni, E. Evaluation of OpenPose for Quantifying Infant Reaching Motion. *Arch. Phys. Med. Rehabil.* **2021**, *102*, e86. [CrossRef]
131. Balta, D.; Kuo, H.; Wang, J.; Porco, I.G.; Morozova, O.; Schladen, M.M.; Cereatti, A.; Lum, P.S.; Della Croce, U. Characterization of Infants' General Movements Using a Commercial RGB-Depth Sensor and a Deep Neural Network Tracking Processing Tool: An Exploratory Study. *Sensors* **2022**, *22*, 7426. [CrossRef] [PubMed]
132. Moro, M.; Pastore, V.P.; Tacchino, C.; Durand, P.; Blanchi, I.; Moretti, P.; Odone, F.; Casadio, M. A markerless pipeline to analyze spontaneous movements of preterm infants. *Comput. Methods Programs Biomed.* **2022**, *226*, 107119. [CrossRef]
133. Ni, H.; Xue, Y.; Ma, L.; Zhang, Q.; Li, X.; Huang, S.X. Semi-supervised body parsing and pose estimation for enhancing infant general movement assessment. *Med. Image Anal.* **2023**, *83*, 102654. [CrossRef] [PubMed]

134. Dechemi, A.; Bhakri, V.; Sahin, I.; Modi, A.; Mestas, J.; Peiris, P.; Barrundia, D.E.; Kokkoni, E.; Karydis, K. Babynet: A lightweight network for infant reaching action recognition in unconstrained environments to support future pediatric rehabilitation applications. In Proceedings of the 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), Vancouver, BC, Canada, 8–12 August 2021; pp. 461–467.
135. Huang, X.; Wan, M.; Luan, L.; Tunik, B.; Ostadabbas, S. Computer Vision to the Rescue: Infant Postural Symmetry Estimation from Incongruent Annotations. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–7 January 2023; pp. 1909–1917.
136. Rehouma, H.; Noumeir, R.; Jouvet, P.; Bouachir, W.; Essouri, S. A computer vision method for respiratory monitoring in intensive care environment using RGB-D cameras. In Proceedings of the 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), Montreal, QC, Canada, 28 November–1 December 2017; pp. 1–6.
137. Kidziński, Ł.; Yang, B.; Hicks, J.L.; Rajagopal, A.; Delp, S.L.; Schwartz, M.H. Deep neural networks enable quantitative movement analysis using single-camera videos. *Nat. Commun.* **2020**, *11*, 4054. [CrossRef] [PubMed]
138. Ossmy, O.; Adolph, K.E. Real-time assembly of coordination patterns in human infants. *Curr. Biol.* **2020**, *30*, 4553–4562. [CrossRef] [PubMed]
139. Ali, A.; Negin, F.F.; Bremond, F.F.; Thümmler, S. Video-based behavior understanding of children for objective diagnosis of autism. In Proceedings of the VISAPP 2022—17th International Conference on Computer Vision Theory and Applications, Online, 6–8 February 2022.
140. Ferrer-Mallol, E.; Matthews, C.; Stoodley, M.; Gaeta, A.; George, E.; Reuben, E.; Johnson, A.; Davies, E.H. Patient-led development of digital endpoints and the use of computer vision analysis in assessment of motor function in rare diseases. *Front. Pharmacol.* **2022**, *13*, 916714. [CrossRef]
141. Jocher, G.; Stoken, A.; Borovec, J.; Changyu, L.; Hogan, A.; Diaconu, L.; Ingham, F.; Poznanski, J.; Fang, J.; Yu, L.; et al. ultralytics/yolov5: V3. 1-bug fixes and performance improvements. *Zenodo* **2020**, *1*. [CrossRef]
142. Wojke, N.; Bewley, A.; Paulus, D. Simple online and realtime tracking with a deep association metric. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 3645–3649.
143. Teed, Z.; Deng, J. Raft: Recurrent all-pairs field transforms for optical flow. In Proceedings of the ECCV 2020: 16th European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Proceedings, Part II 16; Springer: Cham, Switzerland, 2020; pp. 402–419.
144. Zheng, C.; Zhu, S.; Mendieta, M.; Yang, T.; Chen, C.; Ding, Z. 3d human pose estimation with spatial and temporal transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 11656–11665.
145. Zhang, J.; Gong, K.; Wang, X.; Feng, J. Learning to Augment Poses for 3D Human Pose Estimation in Images and Videos. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 10012–10026. [CrossRef]
146. Tang, Z.; Qiu, Z.; Hao, Y.; Hong, R.; Yao, T. 3D Human Pose Estimation With Spatio-Temporal Criss-Cross Attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 18–22 June 2023; pp. 4790–4799.
147. Einfalt, M.; Ludwig, K.; Lienhart, R. Uplift and Upsample: Efficient 3D Human Pose Estimation with Uplifting Transformers. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–7 January 2023; pp. 2903–2913.
148. Garau, N.; Conci, N. CapsulePose: A variational CapsNet for real-time end-to-end 3D human pose estimation. *Neurocomputing* **2023**, *523*, 81–91. [CrossRef]
149. Clark, H.; Coll-Seck, A.M.; Banerjee, A.; Peterson, S.; Ameratunga, S.; Balabanova, D.; Bhan, M.K.; Bhutta, Z.A.; Borrazzo, J.; Doherty, T.; et al. A future for the world's children? A WHO-UNICEF-Lancet Commission. *Lancet* **2020**, *395*, 605–658. [CrossRef]
150. Hanson, M.A.; Cooper, C.; Aihie Sayer, A.; Eendebak, R.J.; Clough, G.F.; Beard, J.R. Developmental aspects of a life course approach to healthy ageing. *J. Physiol.* **2016**, *594*, 2147–2160. [CrossRef]
151. van Heerden, A.; Leppanen, J.; Rotheram-Borus, M.J.; Worthman, C.M.; Kohrt, B.A.; Skeen, S.; Giese, S.; Hughes, R.; Bohmer, L.; Tomlinson, M. Emerging Opportunities Provided by Technology to Advance Research in Child Health Globally. *Glob. Pediatr. Health* **2020**, *7*, 1–9. [CrossRef]
152. Magrini, M.; Salvetti, O.; Carboni, A.; Curzio, O. An Interactive Multimedia System for Treating Autism Spectrum Disorder. In Proceedings of the ECCV 2016 Workshops: European Conference on Computer Vision, Amsterdam, The Netherlands, 8–10 and 15–16 October 2016; Hua, G., Jégou, H., Eds.; Springer: Cham, Switzerland, 2016; pp. 331–342.
153. Magrini, M.; Curzio, O.; Carboni, A.; Moroni, D.; Salvetti, O.; Melani, A. Augmented Interaction Systems for Supporting Autistic Children. Evolution of a Multichannel Expressive Tool: The SEMI Project Feasibility Study. *Appl. Sci.* **2019**, *9*, 3081. [CrossRef]
154. Moor, M.; Banerjee, O.; Abad, Z.S.H.; Krumholz, H.M.; Leskovec, J.; Topol, E.J.; Rajpurkar, P. Foundation models for generalist medical artificial intelligence. *Nature* **2023**, *616*, 259–265. [CrossRef]
155. Lantos, J.D.; Meadow, W.L. *Neonatal Bioethics: The Moral Challenges Of Medical Innovation*; The Johns Hopkins University Press: Baltimore, MD, USA, 2008.
156. Liu, J.; Chen, X.X.; Wang, X.L. Ethical issues in neonatal intensive care units. *J. Matern.-Fetal Neonatal Med.* **2016**, *29*, 2322–2326. [CrossRef]
157. Botkin, J.R. Ethical issues in pediatric genetic testing and screening. *Curr. Opin. Pediatr.* **2016**, *28*, 700–704. [CrossRef]

158. Ake-Kob, A.; Blazeviciene, A.; Colonna, L.; Cartolovni, A.; Dantas, C.; Fedosov, A.; Florez-Revuelta, F.; Fosch-Villaronga, E.; He, Z.; Klimczuk, A.; et al. State of the Art on Ethical, Legal, and Social Issues Linked to Audio- and Video-Based AAL Solutions. 2022. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4075151 (accessed on 30 July 2023) [CrossRef]
159. Walsh, V.; Oddie, S.; McGuire, W. Ethical Issues in Perinatal Clinical Research. *Neonatology* **2019**, *116*, 52–57. [CrossRef] [PubMed]
160. Alexander, D.; Quirke, M.; Doyle, C.; Hill, K.; Masterson, K.; Brenner, M. Technology solutionism in paediatric intensive care: Clinicians' perspectives of bioethical considerations. *BMC Med. Ethics* **2023**, *24*, 55.
161. Adams, C.; Pente, P.; Lemermeyer, G.; Rockwell, G. Artificial Intelligence Ethics Guidelines for K-12 Education: A Review of the Global Landscape. In Proceedings of the International Conference on Artificial Intelligence in Education, Utrecht, The Netherlands, 14–18 June 2021; Roll, I., McNamara, D., Sosnovsky, S., Luckin, R., Dimitrova, V., Eds.; Springer: Cham, Switzerland, 2021; pp. 24–28.
162. Adams, C.; Pente, P.; Lemermeyer, G.; Rockwell, G. Ethical principles for artificial intelligence in K-12 education. *Comput. Educ. Artif. Intell.* **2023**, *4*, 100131. [CrossRef]
163. McStay, A.; Rosner, G. Emotional artificial intelligence in children's toys and devices: Ethics, governance and practical remedies. *Big Data Soc.* **2021**, *8*, 1–16. [CrossRef]
164. Boch, S.; Sezgin, E.; Lin Linwood, S. Ethical artificial intelligence in paediatrics. *Lancet Child Adolesc. Health* **2022**, *6*, 833–835. [CrossRef] [PubMed]
165. Thai, K.; Tsiandoulas, K.H.; Stephenson, E.A.; Menna-Dack, D.; Zlotnik Shaul, R.; Anderson, J.A.; Shinewald, A.R.; Ampofo, A.; McCradden, M.D. Perspectives of Youths on the Ethical Use of Artificial Intelligence in Health Care Research and Clinical Care. *JAMA Netw. Open* **2023**, *6*, e2310659. [CrossRef] [PubMed]
166. Ravi, S.; Climent-Pérez, P.; Florez-Revuelta, F. A review on visual privacy preservation techniques for active and assisted living. *Multimed. Tools Appl.* **2023**. [CrossRef]
167. Jovanovic, M.; Mitrov, G.; Zdravevski, E.; Lameski, P.; Colantonio, S.; Kampel, M.; Tellioglu, H.; Florez-Revuelta, F. Ambient Assisted Living: Scoping Review of Artificial Intelligence Models, Domains, Technology, and Concerns. *J. Med. Internet Res.* **2022**, *24*, e36553. [CrossRef]
168. Colantonio, S.; Jovanovic, M.; Zdravevski, E.; Lameski, P.; Tellioglu, H.; Kampel, M.; Florez-Revuelta, F. Are Active and Assisted Living applications addressing the main acceptance concerns of their beneficiaries? Preliminary insights from a scoping review. In Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments, Corfu Island, Greece, 29 June–1 July 2022; pp. 414–421.
169. Lekadir, K.; Osuala, R.; Gallin, C.; Lazrak, N.; Kushibar, K.; Tsakou, G.; Aussó, S.; Alberich, L.C.; Marias, K.; Tsiknakis, M.; et al. FUTURE-AI: Guiding Principles and Consensus Recommendations for Trustworthy Artificial Intelligence in Medical Imaging. *arXiv* **2021**, arXiv:2109.09658.