A Comprehensive Literature Review on Children's Databases for Machine Learning Applications

SADAM AL-AZANI [1], **SADIQ M. SAIT** [1,2], **(Senior Member, IEEE),**
AND KHALED A. AL-UTAIBI [2,3], **(Member, IEEE)**

[1]SDAIA-KFUPM Joint Research Center for Artificial Intelligence (JRC-AI), King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
[2]Department of Computer Engineering, College of Computing and Mathematics, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
[3]Computer Science and Software Engineering Department, University of Ha'il, Ha'il 53962, Saudi Arabia

Corresponding author: Sadam Al-Azani (sadam.azani@kfupm.edu.sa)

ABSTRACT The COVID-19 pandemic can be attributed as a main factor to accelerate the current digital transformation and to encourage innovation and technological adoption. Consequently, the care provided to our children, one of the significant aspects of life, needs to be adapted with the life's changes. Children are our future and our most precious resources. They need our attention in all life domains including health, education, safety and social interaction. Nowadays, technologies have been incorporated with machine learning and it has been proven that they are more powerful, reliable and profitable. Machine learning methods have been applied by many children-related studies to generate predictive models for different applications. The efficacy of the generated models mainly relies on the constructed databases. This article carries out a comprehensive survey on available children's databases constructed for machine-learning-based solutions with their methodologies, characteristics, challenges, and applications. First, it provides an overview of the available studies and classifies them based on their applications. Next, it defines a set of attributes and evaluates them while also shedding light on their pros and cons. The primary concerns related to collection, development and distribution of children's databases are also discussed. This study can be considered as a guideline for researchers in multidisciplinary fields to construct reliable databases and to develop more advanced techniques.

INDEX TERMS Children's databases, speech and language analysis, affect detection, activity recognition, pain assessment, database construction, data-driven models.

I. INTRODUCTION

Technology has always been resorted to, and adapted to improve the lives of individuals. In recent past, Artificial Intelligence (AI) and Machine Learning (ML) technologies have been extensively used for this purpose. There is a large reliance of data by these technologies, and for this purpose several databases have been constructed and are in use [1]. The focus mainly has been on data from, and on, adults. Some efforts have also been made to build databases for children. Furthermore, the advances in technology including Internet of Things, sensors, social networks, hyper performance computing (HPC), and cloud data storage make it possible to collect data and train machines to produce outcomes, but with

The associate editor coordinating the review of this manuscript and approving it for publication was Zhan Bu.

deep learning and machine learning, they can also generate new knowledge [2].

There is no doubt that developing intelligent systems for children is more challenging than adults due to children's growth, development and aging factors which change over time [3]. And this definitely affects the performance of intelligent recognition systems influenced by these traits including face recognition [4]–[6], speech recognition [7]–[10] and physiological signals [9]. It is know that collection of data from children is a daunting task. Many individuals, devices, etc., need to be deployed to collect the correct data and signals at the right time in order to use it diligently to derive meaning and recommend actions.

The analysis of children's emotions, behaviors, activities, demographics, and pains contributed to the development of children, their protection, care, treatment, etc., to name a few. To enhance the development of technologies for children,

psychologists, therapists and educators need to work with computer scientists. Whenever we as human beings, want to achieve a goal, we need first make observations, collect data (sensing), build a mental model and then reason depending on the goal, and plans. Usually, to make perfect decisions, we do not just relay on one source of information. We, naturally, use our different senses to understand and analyze certain situations. Parents, caregivers and therapists cannot assess the pain of children only by hearing their cry. For better assessment, they also need to see facial expressions, measure physiological signals (temperature, heart rate, etc.,), among others. Likewise, educators need to understand their students interactions not just using the material of examinations but also using their facial expressions (during class) and their activities either individually or in-groups.

With regards to healthcare, intelligent systems based on machine learning have been applied in areas such as children speech and language development for those who have different kinds of speech or hearing diseases [11], [12]. In addition to this, Autism Spectrum Disorder (ASD), as a mental disorder affecting linguistic, communicative, cognitive, skills as well as social skills and abilities [13], is considered in the literature to be diagnosed and treated using ML with the involvement of therapists and psychologists. Early diagnosis of ASD helps a great deal in ameliorating the disease [14], [15] by helping in identifying of factors that may put children at risk for ASD and other developmental disabilities. ML based systems have been applied to protect children either physically or mentally.

All children interact with toys at different ages. Toys evolved with times. In the fifties, they were made of clay and wood, and little motion was added to limbs later (in dolls). Then, came flexibility and motion of limbs or closing of eyes when laid down, followed by embedding of electronics in battery-operated ones that delivered everything from music to motion to entertain children. Remote control followed, and now we have drones and robots. And, the most recent talk has been on embedding of AI in toys to make them reactive to situations for both ensuring the safety of children while learning from their behaviors. Recently, UNICEF World Economic Forum (WEF)[1] sought to work alongside partners to set and lead the global agenda on AI technology related to children. They also aimed at engaging interested parties to build AI-based technology that helps realize and upholds child rights [16]. On the practical side, WEF developed Smart Toy Awards, launched in January 2021, focusing on providing ethical, responsible and innovative AI-powered toys, based on guidelines and criteria developed by the Generation AI Project community at WEF.

A. MOTIVATION AND OBJECTIVES
This study provides the first of its kind review that focuses on intelligent systems related to children based on machine/deep leaning. Reviewing the constructed databases is an effective

[1] https://www.weforum.org/

way to explore the major trends in technologies incorporating ML and the types of pursued problems and fields. Therefore, we review state-of-the-art databases along with their attributes, objectives, challenges and applications. Furthermore, this study aims at providing guidelines for researchers to develop standard and reliable databases and more advanced techniques. Another factor motivating this study is the lack of effort on conducting studies related to children, and scarcity of available databases for them. We advocate the development towards a new generation of databases that can be used in children-related technologies.

B. CONTRIBUTION
This study extensively and specifically presents the current efforts on developing children's databases and the corresponding techniques as well as the associated concerns. We can summarize the main contributions as follows:

1) We provide a structured review of the current methodologies for developing children's databases. Each database is described in terms of its main objective, collection procedure, content, environment, and validation. We categorize the reviewed studies, regarding their objectives into five main groups, namely: (i) speech and language analysis, (ii) affective (emotion and sentiment) analysis, (iii) pain assessment, (iv) activity recognition, and, (v) biometrics analysis.

2) We provide a taxonomy for the children-related technology depicted in Figure 1. We also define evaluation criteria to characterize the children's databases and compare them accordingly.

3) We discuss the main requirements and the primary challenges for constructing children's databases. This includes data resources, environmental settings, annotation, ethical concerns, documentation, verification, and distribution. We also present the limitations of the available databases and identify future research opportunities.

C. MANUSCRIPT ORGANIZATION
The manuscript's structure is depicted in Figure 2. The designed methodology undertaken to conduct this study is presented in Section II. The databases are classified based on their objectives in Sections III to VII. Discussions are presented in Section VIII. Finally, the study is concluded in Section IX.

II. RESEARCH METHODOLOGY
This section details the methodology employed to perform this study. Also discussed are the research questions, the search process for relevant studies, the inclusion/exclusion criteria, the evaluation criteria, and the scope of the study. It also summarizes the relevant available reviews studies.

A. RESEARCH QUESTIONS
- RQ1: What research topics based on AI and ML have been investigated for children?
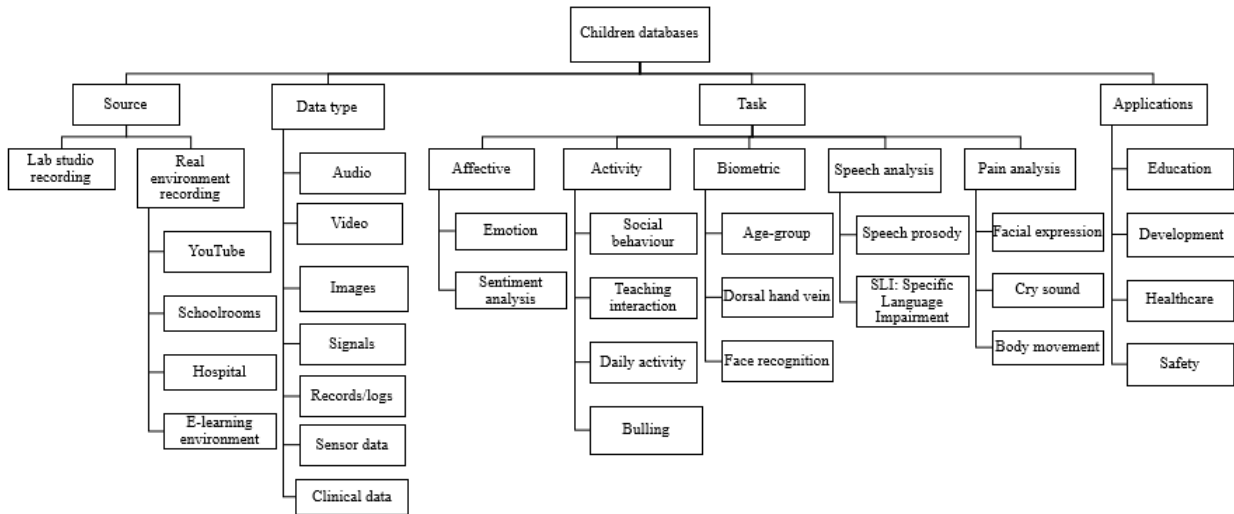
FIGURE 1. Children's databases taxonomy.

FIGURE 2. Manuscript's organization.
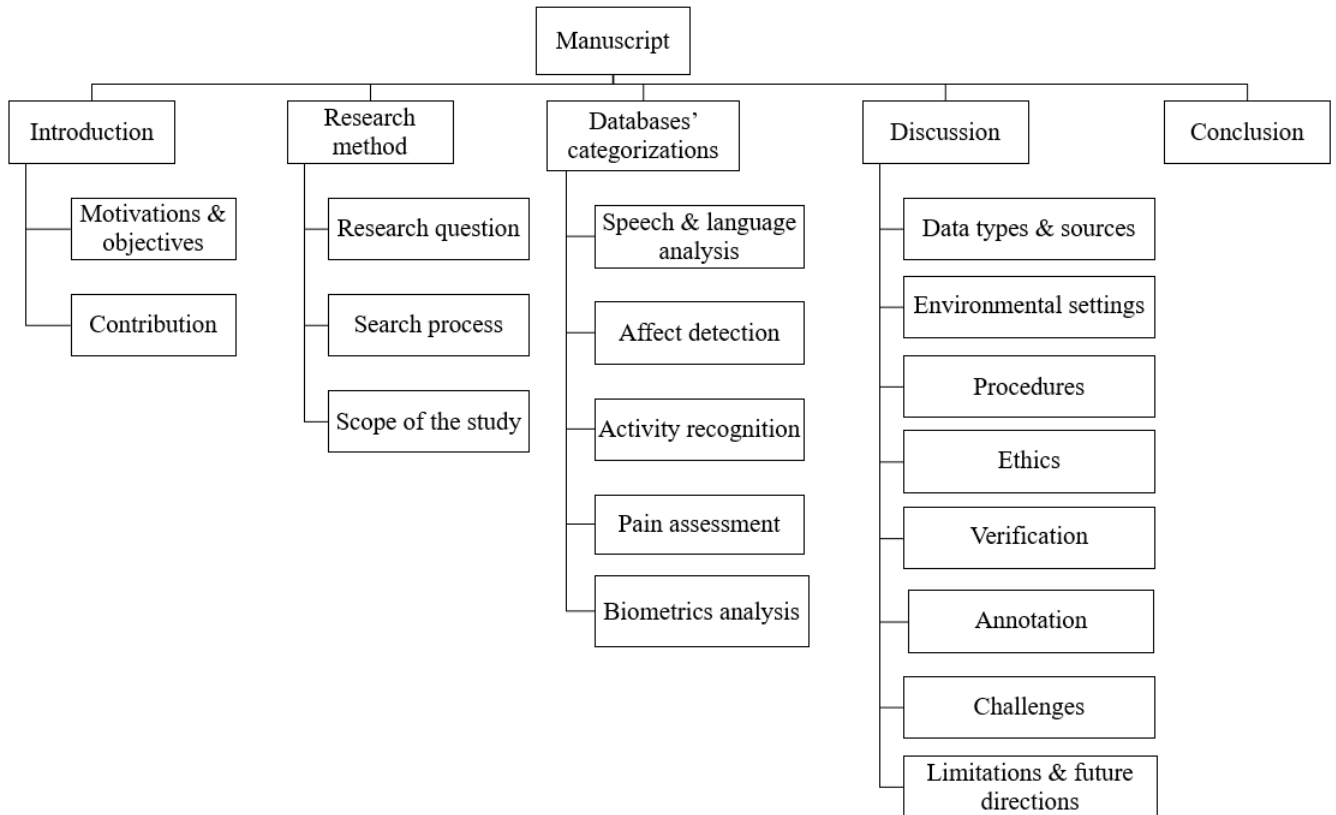
- RQ2: What are the major trends in contemporary ML-based technology for children?
- RQ3: What are the available procedures/protocols for collecting data from children and how can the available

online data for children be utilized and adapted to create the required databases?

- RQ4: What are the characteristic of available databases? In order to answer this research question, we define a

set of attributes to evaluate and compare the reviewed studies, including:

1) Year: this attribute defines when the reviewed study was published.
2) Task: it defines the main objectives of the conducted study and constructed database.
3) Age: it defines the age of the participating children, in days, weeks, months or years. Some of the reviewed literature just provides the children's age-group or their school grades (elementary, primary, etc.,) without identifying the age.
4) Environment: it defines where database was recorded or collected.
5) Modality: it defines the type of data. Modality might be Images (I), Videos (V), Audio (A), physiological signals, etc. This also reflects whether the database is uni-modal or multimodal.
6) Samples: it defines number of samples. Samples might be images, audios, videos, signals, sessions, utterances, recordings, or a combination of these. It also defines the annotation level of video and audio recordings. They might be annotated at video, frame, audio or utterance levels.
7) Children: this attribute defines the number of the participating children and their gender.
8) Labels: it defines the number and the name of classes or groups. This attribute also reflects whether the database is binary or is a multiple classification task.
9) Agreement: it defines whether the researchers got ethical consents for conducting the study or for collecting the data for the databases. If so, it also determines from where.
10) Race or Language: it defines the race or language of the participating children.
11) Availability: it determines whether the database is publicly available or not with providing the database's link.
12) Application/Domain: it determines whether the database for specific or general applications and determines the domain including healthcare/treatment, protection and development.

- RQ5: What are the major challenges in developing data-driven techniques for children?

B. SEARCH PROCESS

The search process was carried out at both general and specific levels. At the general level, we started with querying Google Scholar with the following search terms: ("Children" OR "Child" OR "Infant" OR "Baby") AND ("Database" OR "Dataset" OR "Corpus"). The process of applying the search terms results in a large number of papers. We selected the ML-based studies related to children. They are classified into two types. The first type of papers (**DS-Type I**) construct children's databases. This type is our main target.

The second type (**DS-Type II**) evaluated or used available children's databases constructed in the first type (DS-Type I). We tracked the second type of papers (DS-Type II) by looking at their references to find the main target of papers (DS-Type I). At the specific search level, we found that the retrieved databases are related to "emotion recognition", "activity detection", "pain assessment", "pulling detection", and, "sentiment analysis". Therefore, we used these terms for the search process and repeated the same procedure of the general search level for finding the related papers.

C. SCOPE OF THE STUDY

A large number of references were retrieved by applying the above search process. By applying the inclusion and exclusion criteria and aligning with the scope of this study, we came up with 58 databases constructed for children. They are categorized by their main objectives into five main groups. The distribution of the retrieved studies for each of the years is presented in Figure 3.

The general framework of the children's technology and related main components for developing ML-based systems is depicted in Figure 4. The framework comprises several components. First, the main source is children, in their different stages or age-groups. In this study, we consider the following age-groups, according to "healtychildren.org"[2]:

- Baby/Infant: newborns aged between the age from one day to 12 months.
- Toddler: the age between one to three years.
- Preschool: the age between three to five years.
- Gradeschooler: the age between five to 12 years. It is also called "Childhood".
- Teens/Adolescence: the age between 12 to 18 years.

There are several data modalities that can be collected from children, including: speech, visuals, physiological and records. Each modality has several data types depicted in the figure. There are some standards and ethics need to be considered for constructing the children' databases and the corresponding technology. The figure also shows the main computer research area involved in developing this technology and the corresponding techniques. Feedback, then, can be provided to the different beneficiaries, including, the children themselves and their parents, researchers, therapists, caregivers and educators.

In addition to this, this study takes into account other databases which were not mainly constructed for children but also for different human being age-groups including adults and elders.

In order to differentiate and distinguish our study/survey from other relevant previous ones conducted for children, the salient points can be summarized as follows: Saraswathy *et al.* [17] conducted a study in 2012 to review works conducted on analyzing infant cry signals. In the same context, Zamzmi *et al.* [18] provided, in 2017 a review study

---

[2]https://www.healthychildren.org/English/ages-stages/Pages/default.aspx
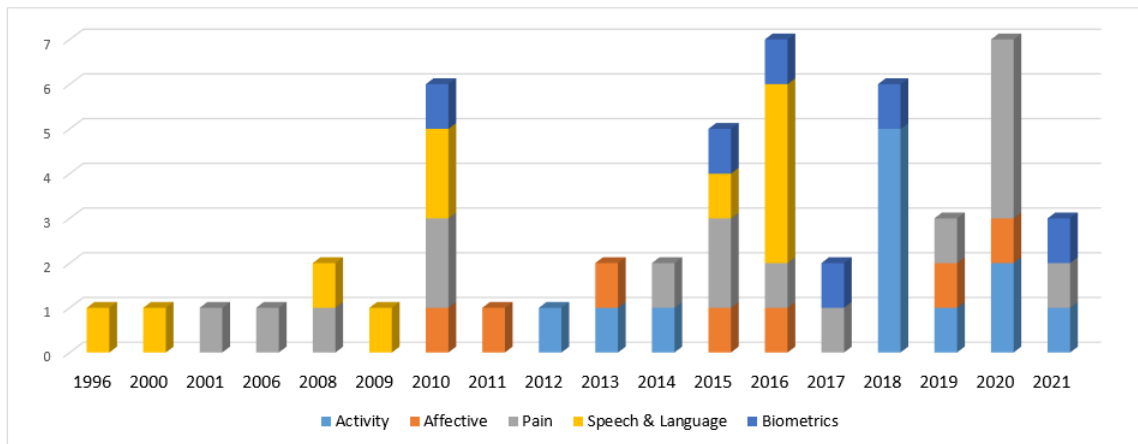
FIGURE 3. Distribution of the retrieved children' databases for different years.

on the automated pain recognition approaches. The review study conducted by Beckman [19], in 2017, mainly focused on two issues related to automatic speech recognition. The first concern was to understand the differences between adult and children speech. The second issue was to identify the annotation schemes and analysis techniques that are able to successfully capture relevant aspects of the speech variability. In the subsequent year, McKechnie *et al.* [20] conducted a systematic literature survey to review automated speech analysis tools developed for analysing and modifying speech of normal children and children with speech sound disorders learning a foreign language. Recently, in 2021, Mihaescu *et al.* [21] presented a review on the most used publicly available educational data mining sources along with their associated tasks, used algorithms, experimental results and main findings. The authors are not aware of any review study conducted for children's databases for their characteristics, methods, challenges, and applications.

III. SPEECH AND LANGUAGE ANALYSIS
Speech signals not only convey the message and the linguistic information, they also contain information about speakers' demographic characterizations, health and emotional stats, among others [22]. This section focuses on data-driven recognition systems which are able to assess speech and vocalizations for predicting the level of children's development and the detection of developmental anomalies. Automatic acoustic analysis has become a mature discipline and now it is possible to accurately recognize children with speech-related disorders [23]–[25]. In case of constructing speech development and assessment databases, samples are recorded from disordered speech children, referred to as ''Cases'', and from healthy children, referred to as ''Controls''. The participants are then asked to pronounce texts. The text (tokens, words, phrases, sentences, numbers, etc.) should be chosen carefully. For example, the vocabulary of [26] database was prepared based on a 57-word set chosen from a well-known

handbook for speech therapy in Spanish whereas, the sentences in [11] database were obtained from SPECO software [27] developed for children requiring speech therapy. The controls are then used to train the ML models which are used to either assess the cases or develop them under the supervisions of researchers or therapists, as depicted in Figure 5. These models are supposed to be used as plug-ins in the a screening tool for assessment or development.

The main characteristics of the developed databases for speech and language analysis are shown in Table 1. We are aware of 11 databases mainly constructed children's speech and language development and analysis. Oller *et al.* [23] reported that it is possible to track children's development using their acoustic features and to recognize children with speech-related disorders and differentiate them from the healthy one. This was based on data-driven models generated using a database of day-long audio recordings collected from children who were typically developing language, had language delay and autism.

Similarly, VanDam *et al.* [25] reported that children who demonstrate hard-of-hearing patterns are similar to the typically-developing children rather than those who are language-delayed or have autism. This was based on a database collected from 273 toddlers and preschoolers who were typically developing and exhibited language delay or autism.

Miller *et al.* [28] collected two datasets of 478 children's speech recordings from telephone channels (TEL) and direct recordings via microphone (MIC) with the joint efforts of Southwestern Bell Technology Resources and Central Institute for the Deaf. Participating children were asked to read words displayed on a computer screen. For children who ages ranged between 5 and 6 years, who were not able to read texts, their speech were elicited by pre-recorded imitation.

In order to help speech handicapped children to learn the correct prosodic pronunciation of sentences, Sztahó *et al.* [11] developed a sentence intonation teaching and training system
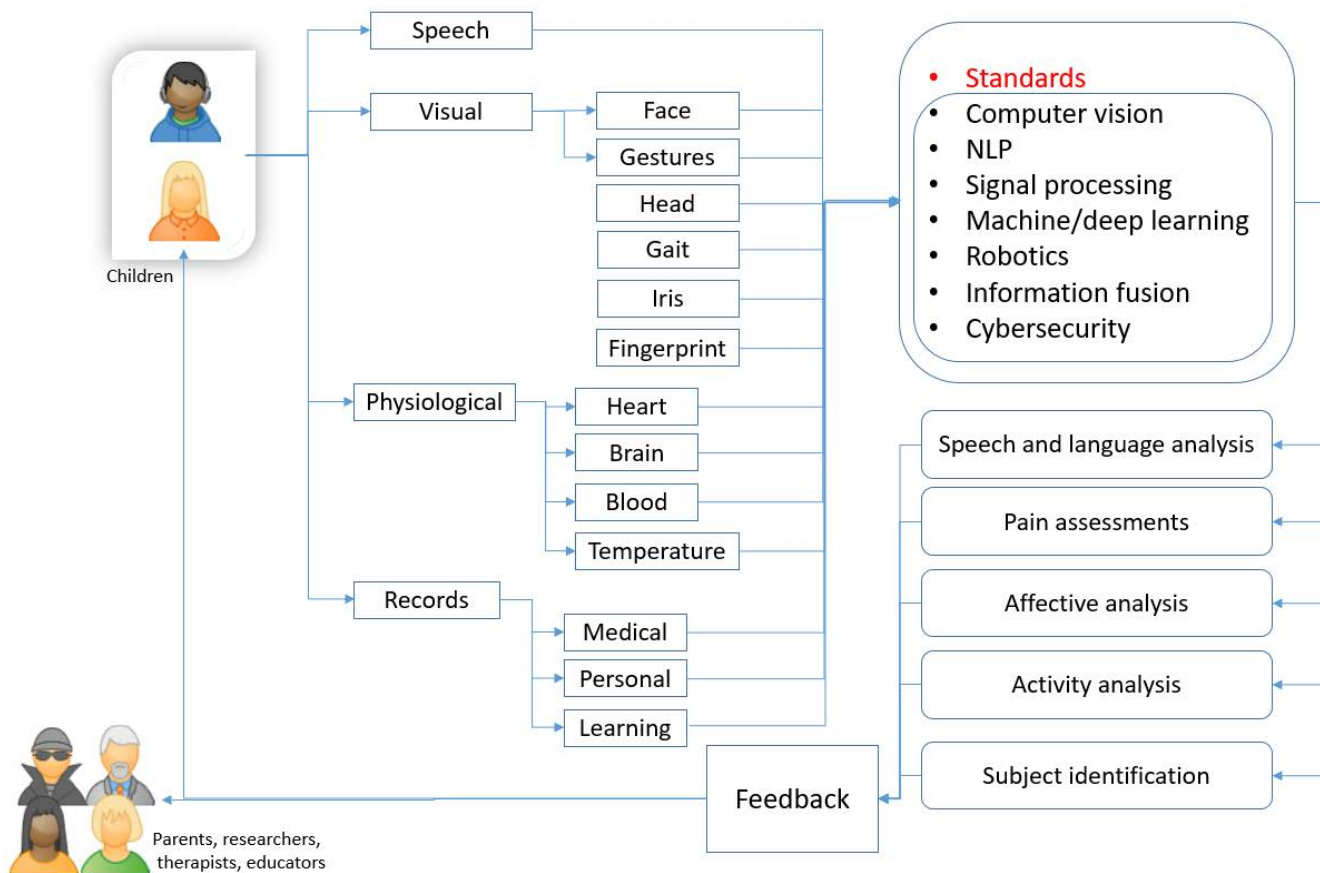
FIGURE 4. Framework of the children-related technology incorporating ML.
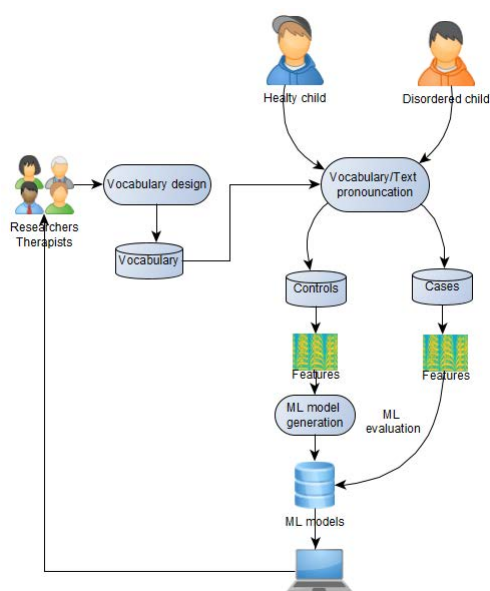

FIGURE 5. A framework for children speech analysis and development.

and accordingly, a dataset from 59 correctly speaking children was recorded. In their other work [12], they recorded a

database from 19 speech impaired children classified manually into five intonation classes according to fundamental frequency curves. Children were sitting in front of a microphone and read sentences prepared from two text materials: sentences and dialogues. The previous database [11] was used to train Hidden Markov models and this database was used to evaluate the generated models. Similarly, LANNA research group in the Faculty of Electrical Engineering at Czech Technical University in Prague compiled four databases [29] mainly for medical research and linguistics from healthy children and children with Specific Language Impairment (SLI). The first, LANNA-H-CH database, was recorded from healthy children's speech (''Controls'') while the second database, H-CH with DEFECT was recorded from children with minor speech errors or from children visiting a speech and language therapist (''controls with defect''). Both LANNA-H-CH and LANNA-H-CH with DEFECT databases were recorded in kindergarten and in the first four grades of elementary school. On the other hand, the third and the fourth databases, LANNA-SLI-CH-I and LANNA-SLI-CH-II, respectively, were recorded from children with SLI ''Cases''. However, they differ in that, LANNA-SLI-CH-I considers the inclusion of SLI while LANNA-SLI-CHI determines the degree of severity of the children's diagnosis.

TABLE 1. Children's databases for speech and language development. ASR: Automatic Speech Recognition, SMR: Sentence Modality Recognition, SPR:Speech Prosody Recognition, SLI:Specific Language Impairment, NM: Not mentioned, ack.: acknowledgment, (b: boys, g: girls).

Ref	Year	Task	Age	Env.	Samples	# Children: gender	Labels	Agreement	Language/ Race	Availability	Application
[28]	1996	ASR	5-18 years	Lab at museum	NM	914	25 words and phrase, 6 telephone numbers, spontaneous answers to eight questions, etc.	Parents or legal guardian	American English	No	Development
OGI[32]	2000	ASR	5-16 years	schoolroom	70683 utterance	1117: 605 b, 512 g	205-words	Ack.	English	No	General
Impaired-DB[26]	2008	SLI	11-21 years	School	228 utterances	14: 7 b, 7 g	57-word vocabulary	Ack.	Spanish	No	Development
[11]	2009	SMR-SPR	8-13 years	School + Lab[a]	~ 8000[b]	59	5-intonions: Descending, Falling, Ascending-falling, Floating, Rising	Schools + ack.	Hungarian	No	Development
[12]	2010	SMR-SPR	10-14 years	Lab	560	19	5-intonions: Descending, Falling, Ascending-falling, Floating, Rising	Schools + ack.	Hungarian	No	Development
[23]	2010	vocal development	10-48 months	House	1486 Recordings, 3.139 Million utterances	232	3: Typical, Delayed, Autistic	Parents	English	No	Development, Healthcare
[25]	2015	Vocal analysis	10-48 months	House	1913 Recordings, 4.307 million utterances	273	4: Typical, Delayed, Hard of hearing, Autistic	Institution +parents	English	No	Healthcare
LANNA-H-CH[29]	2016	SLI	4-10 years	schoolroom	70 R, 4620 Utterances	70: 25 b, 45 g	13-parts	Parents + ack.	Czech	Yes[c]	Healthcare
LANNA-H-CH with defect[29]	2016	SLI	4-10 years	schoolroom	33 Recordings, 2178 Utterances	33: 17 b, 16 g	13-parts	Parents + ack.	Czech	Yes[c]	HealthCare
LANNA-SLI-CH-I[29]	2016	SLI	4-12 years	therapist's office	86 R, 5676 Utterances	46: 33 b, 13 g	13-parts	Parents + ack.	Czech	Yes[c]	Healthcare
LANNA-SLI-CH-II[29]	2016	SLI	6-12 years	therapist's office	133 R, 8819 Utterances	72: 46 b, 26 g	13-parts	Parents + ack.	Czech	Yes[c]	Healthcare

[a] https://www.tmit.bme.hu/beszedkommunikacio-es-intelligens-interakciok-laboratorium?language=en
[b] The average samples per class was 1,600 samples
[c] https://lindat.mff.cuni.cz/repository/xmlui/handle/11372/LRT-1597#

They were recorded at a surgery of speech and language therapists and at the hospital for pathological speech of pediatric patients with SLI. Databases for autism spectrum disorders such as [30], [31] are gathered based on online surveys provided by parents that contain some or all of the following: Demographics, the Social Communication Questionnaire, and the Pediatric Evaluation of Disability Inventory-Computer Adaptive Test Social/Cognitive, Daily Activities, and Responsibility domains.

Oregon Graduate Institute (OGI) Kids' Speech corpus [32] contains both read speech (prompted) and spontaneous speech of words and sentences recorded from 1100 children from kindergarten through grade 10 at the Northwest Regional School District near Portland, Oregon. In addition to this, information on each child speaker were recorded including, gender, age, language spoken and physical maladies that could affect children's speech. This makes it applicable for other tasks, including age detection and gender recognition.

IV. AFFECT DETECTION

This section focuses on databases constructed for "affect detection" in children, including emotion recognition and sentiment analysis. Affective computing is an interdisciplinary research field which spans from computer science to psychology, and from social sciences to cognitive science [33], [34]. It concerns developing techniques that are able to recognize, analyze and interpret human emotions, moods and opinions. Automatic affect detection is receiving a significant amount of attention due its wide applications and importance. It plays a significant role in human beings development, therapy, skills' improvement, machine-human interaction, and decision making. The main characteristics of databases developed for children's affect detection are presented in Table 2.

A. BASIC EMOTIONS

The Ekman model(1970) [35]–[37], defined six basic universal emotions: anger, disgust, fear, happiness, sadness, and, surprise, with their aspects. Contempt is another emotion that was considered later by researchers [38]. The Ekman model was then extended to include 12 additional emotions. Accordingly, different data-driven approaches have been developed to detect emotions. We are aware of just six databases which are either developed especially for children or contain children-related subjects.

Radboud Faces Database (RaFD) [39] is made up of facial images with three gaze directions for both adults and children. The dataset comprises 10 Caucasian Dutch children. The participants were asked to express several emotional states: anger, disgust, fear, happiness, sadness, surprise, contempt, plus neutral states. These were captured using five cameras from five different angles. The Child Affective Facial Expression [40] (CAFE) is another image database collected from children's faces. The participating children were engaged by unscripted play to express Ekman's basic emotions. Dartmouth [41] database is another facial images database in which participating children were asked to express eight emotions. National Institute of Mental Health Child Emotional Faces Picture Set (NIMH-ChEFS) [42] is built using 482 children's facial images expressing five affect states with direct and averted gaze conditions. The children were asked to pose the considered emotions: angry, fearful, happy, sad and neutral according to the Ekman and Friesen (1975) model. A theater teacher and a neuroscientist from the NIH research group worked together to record a series of two-hour sessions conducted over two-week period in 2004. Participating children were enrolled in classes at the children's theater. The participating children were selected carefully by the theater's teachers.

B. APPLICATION-BASED EMOTIONS

Databases under this category might be constructed for both basic and complex affective states, yet for specific applications or under certain circumstances.

EmoReact [43] is an audio-visual database comprising 1102 videos collected from YouTube. The focus here was on 17 affective states related to learning and education defined in [34]. The collected videos presenting children's affective states during learning a subject, when: 1) they need to know about the subject, (2) they were asked a question about the subject, (3) they were answering a question, (4) were listening and reacting to a learner, and (5) expressing their opinions individually about the subject.

C. SENTIMENT ANALYSIS

Sentiment analysis is one of the most active research areas in natural language processing and is widely studied in data mining, Web mining, and text mining [44], [45]. It is defined, according to Liu [46], as "the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes". To our knowledge, only one study analyzed the sentiments of children. Paz *et al.* investigated the behavior of third party punishment-type behavior of participating children with developmental disorders: ASD, attention deficit hyperactivity disorder, learning disabilities, intellectual disability and a database was built for this purpose [47]. Each of the participating children watched two videos of 40 seconds each, for football matches. In Video-1, a football-player from the participant's country scores a goal with his hand. In Video-2, a player from another country does the same against the country of the participant. Then, the participating children were asked to express their opinions on the events. It was reported that the ASD group showed negative feelings for both videos. However, other groups showed positive opinion in Video-1 and negative in Video-2. This suggests that children with ASD respect rules regardless of whether those who break them belong or not to their own group, possibly due to lower degrees of empathy.

TABLE 2. Children's databases for affect detection. NM: Not Mentioned, SA: Sentiment analysis.

Ref	Year	Task	Age	Modality	Env.	Samples	Children, Gender	Labels	Agreement	Race	Availability	Application
RaFD[39]	2010	Emotion	8-12	Images	Home	80 I	10: 6b, 4g	8:anger, contempt, disgust, fear, happiness, sadness, surprise, neutral	NM	Caucasian	Yes	General
NIMH-ChEFS[42]	2011	Emotion	10-17	Images	Lab	482 I	22b, 37g	5: anger, fear, happiness, sadness , neutral	Yes	Chinese	Yes [a]	General
Dartmouth [41]	2013	Emotion	6-16	Images	Lab	640 I	80:40b, 40g	8:angry, afraid, content, disgusted, happy, sad, surprised, neutral	Yes	Dartmouth, UK	Upon request [b]	General
CAFE [40]	2015	Emotion	2-8	Images	Lab	1192 I	154: 90, 64b	7: anger, disgust, fear, happiness, sadness, surprise, neutral	Yes	NM	Yes [c]	General
EmoReact[43]	2016	Emotion	4-14	Audio, Video	Youtub	1102 V	63: 32g, 31b	17:anger, disgust, fear, happiness, sadness, surprise, neutral, valence and nine complex emotions including curiosity, uncertainty, excitement and frustration	No	different races	Upon request by email	learning and education
BRED[48]	2019	Emotion	6-12	Video	CRI-child's room	215	30	6:Anger, disgust, fear, happiness, sadness, surprise	University	NM	Yes [d]	
[47]	2020	SA	9-17	Text, clinical records	Clinic	NM	37: 28b, 9g	3: positive, negative, neutral	Yes	Argentine	No	Healthcare

[a] http://devepi.duhs.duke.edu/instruments.html
[b] https://lab.faceblind.org/k_dalrymple/ddcf
[c] https://nyu.databrary.org/
[d] https://zenodo.org/record/3233060

V. ACTIVITY RECOGNITION

Automatic activity recognition of children has been receiving significant attention due to its various applications in areas including child-computer interaction, healthcare, education, safety monitoring, and social behavior analysis. It has become a hot research topic due to the availability of various data acquisition technologies including sensors [49]–[51]. Several databases have been constructed to analyze children's activities. We categorize them as daily-life activities, social behaviors, learning activities and school violence (bullying). The main characteristics of the children's activities databases are depicted in Table 3.

A. DAILY-LIFE ACTIVITIES

Analysis of Children's daily-life activities is available in the literature, and has attracted researchers' attention. Suzuki *et al.* [9] presented a daily-life activity recognition system to support parents and kindergarten staff. The participants were asked to wear a wireless 3-axis acceleration sensor WAA-001 based on Bluetooth. The sensor was attached to the upper arm of participants, who were asked to execute seven activities: (standing, walking, running, sitting, sleeping (or lying), climbing-up, and climbing-down). During the activities the arm's Electromyogram (EMG) signals and voices were recorded.

Hosseini *et al.* [52] designed an Android smartwatch application to record accelerometer and gyroscope signals in real-time and transfer the anonymous data to a web server for activity prediction. Each child was asked to wear the smartwatch and was instructed to perform six different activities (running, walking, standing, sitting, lying down, and stair climbing). Each activity was recorded for a duration of 10 minutes.

Kokkoni *et al.* [53] recorded videos of three participating children in eight sessions for four weeks using five cameras. Two of the participating children were healthy, with ages 10 to 11 months while the third child was diagnosed with Down syndrome and was of age of 24 months. In another study, Pacheco *et al.* [54] collected a database from six infants (7.8 - 23.7 months) using the similar procedure of [53] in a pediatric rehabilitation environment. One of the children was diagnosed with Down Syndrome. In addition to the participating children, robots and adults appeared in recordings. The participating children wore a white suit tagged with "AR" (Activity Recognition) on different parts.

In another attempt, Efthymiou *et al.* [55] recorded videos from 39 subjects, 25 of whom were children (6-9 years old). The participants were instructed to perform 12 daily activities. They played a pantomime game with robotic agent (NAO robot). The participating children and the robot alternately performed a pantomime presented on a computer screen. In order to collect more data from background

movements of participants, they were also asked to perform random movements.

B. SOCIAL BEHAVIORS

The analysis of children's communication and social behaviors has a great deal of importance in children development. It helps in early diagnosis and treatment of developmental disorders [56], [57]. Rehg *et al.* [58] analyzed a set of main behaviors of children aged 18 to 30 months. Observations included smiling, eye contacts and engagement during their interaction with adults. They constructed a Multimodal Dyadic Behavior (MMDB) database for this purpose based on semi-structured play interaction protocol (Rapid-ABC) [59] of five stages, namely: greeting, ball, book, hat, and tickle. In a similar context, Marrus *et al.* [60] developed a system based on both standard questionnaire data and videos for measuring quantitative variation in reciprocal social behavior in toddlers aged 18 to 30 months. The behavior of a subject, during watching videos, was evaluated with that observed in a short video of a young child manifesting a highly competent level of social communication.

C. LEARNING ACTIVITIES

Automatic analysis of students interaction, communication and/or collaboration with either teachers, learning/teaching materials, or, with each other has a crucial role to improve the learning process [61]–[63]. In addition to interaction, communication and collaboration, learning activities also include the use of digital and non-digital tools. The Multimodal Learning Analytics Math (MMLA-MATH) database [64], is a multimodal collection of audio, video, and digital pen-based data recorded from 18 high-school students during solving mathematics problems. The students were grouped into three groups of girls and three groups of boys. Each group of the participating children met for two sessions. The mathematical problems varied in level of difficulty and were prepared carefully by educators. When all students understood how it had been solved and agreed upon the solution, the leader of a group submitted the solution to the computer. The main phases of solving the math tasks during recording the data started by: clarifying the meaning $\Rightarrow$ working on the calculation $\Rightarrow$ proposing the solution $\Rightarrow$ submitting the answer $\Rightarrow$ and ended by: explaining the answer. The database was created to predict the dominant expert student of each group and to predict correctly the solved problems. The first description of this database appeared in 2006 [65] while the released version of the database was in 2014 [64]. The selection of math subject rather than other subjects can be attributed to that mathematics solution thinking and answer explaining are embodied in physical activity [66], [67].

In other case, the interaction of students and teachers were recorded for building MUlti-modal Teaching and Learning Analytics (MUTLA) database [68]. MUTLA is a multimodal data recorded from primary- and middle-school students during solving problems of varying difficulty levels in five subjects: Chemistry, Chinese, English, Math, and Physics from the Squirrel AI Learning System (SAIL). MUTLA includes students' learning logs/records stored in SAIL, brainwave data collected by electroencephalogram (EEG) headset devices, and videos recorded by web cameras during their studying on SAIL.

D. BULLYING DETECTION

School bullying is one of the popular violence around the world affecting children both mentally and physically [69], [70]. Bullied children are at an increased risk [71]; they might escape from schools, physically harm themselves and/or commit crimes [10], [71]. Some applications such as "Stop Bullies", "Campus Safety", "ICE BlackBox", "TipOff", and "Back Off Bully", have been developed for bullying prevention. For those applications, a user at a risk event needs to press a key on the smartphone, and then the smartphone will send different information (phones, videos, texts, the Global Positioning System (GPS)) to certain receiver(s). Such applications are not reliable as they are human-driven [10], [72]. In addition, children at risk cases might not be able to notify responsible persons or prevent themselves from getting hurt.

There is important need to move from man-driven techniques to information-driven techniques to automatically detect children's emotions in violence events and immediately notify officials [73], [74]. In this section, we presented the databases constructed for children bullying detection. Violence simulation environment was designed by psychologists to record video, speech, heart rate variability (HRV), and other physiological signals (body temperatures, postures, breathing rates, saliva samples, etc.,) in [10]. Participating children were students in second and sixth grades at Normaalikoulu Elementary School in Oulu, Finland. They were divided into groups of three students and alternately played as bullies and victims. Five experimental scenarios were conducted to record in the database. Emotional speech task was one of them, in which 12 students were asked to read three sentences and express five emotions.

Ye *et al.* [72] developed a school bullying detection system based on children's emotional speech and motions. Experimental data were recorded for physical violence and verbal bullying. For physical violence recording data, 3D accelerations and 3D gyros signals were collected from participating subjects through fixing a single movement sensor on their waists. The situation was designed to collect both school daily-life and bullying activities. Daily-life activities cover walking, running, jumping, falling down, playing, and standing whereas, bullying activities cover hitting, punishing, and pushing. For verbal recordings, verbal bullying speeches expressing negative emotions and daily-life conversations expressing positive emotions were recorded. They were labelled as "Bullying" and "Non-bullying". In another work, Ye *et al.* [75] presented a method for school violence detection using two sensors. The participants were asked to act three kinds of school violence activities, namely beating,

TABLE 3. Children's databases for activity recognition. IRB: Institutional Review Board, CRI: Child Robot Interaction, Ack.: Acknowledgment, NM: Not mentioned.

Ref	Year	Task	Age	Modality	Env.	Samples	# Children: gender	Labels	Agreement	Race	Availability	Application
Child-group[9]	2012	Daily-life activity	5 and 8 years	EMG, Voice	Kindergarten	12 minutes	3: 1b, 2g	7: Standing, Walking, Running, Sitting, Sleeping (lying), Climb-up and Climb-down	Parent + university + ack.	NM	No	General
MMDB[58]	2013	Activity social behavior	18-30 months	Audio, Video, Physiological Speech, images, digital pen	Child Study Lab	160 sessions	121	3: smiling, eye contact and engagements	Parents and IRB	English	Upon request[a]	Assessment and therapy
MMLA-Math [64], [65]	2014	Activity: learning analytics	15-17 years			192 problem-solving episodes	18: 9b, 9g	9-Groups	IRB approval	English	Upon request[b]	Education
[52]	2018	Daily-life activity	8 -14 years	Sensor data: accelerometer and gyroscope	Free env.	NM	25: 15b, 10g	6: running, walking, standing, sitting, lying down, and stair climbing	IRB Approval	NM	No	Healthcare
Child-group[55]	2018	Daily-life activity	6-10 years	Video	CRI-Child's room		39: 25 children	12: painting a wall, cleaning a window, driving a bus, swimming, dancing, working out, playing the guitar, digging a hole, wiping the floor, ironing a shirt, hammering a nail, reading a book	NM	NM	No	General
MUTLA[68]	2019	Activity: learning analytics	primary and middle schools	Video, EEG, logs	school	2170 segments	156	NA	schools + ack.	Chinese and English	No	Education
[53]	2020	Daily-life activity	10-11 months, 24 months	Video	Lab, GEAR-Env. (CRI)	8 sessions	3	4: crawling, sitting, standing, walking	Parents + university	NM	No[c]	Development
[54]	2021	Daily-life activity	7.8-23.7 months	Video	Lab	1603 instances	6	4: crawling, walking, sitting, and standing	NM	NM	No	Development
Speech dataset[10]	2018	Bullying detection	2nd and 6th grades	Audio	School	382 clips	12: 6b, 6g	5: happiness, sadness, fear, anger, and neutral	Parents + University	Finnish	No	Safety, Protection
[72]	2018	Bullying detection and daily-life activities	2nd and 6th grades	motion signal, audio	School	1160	NM	9 activities + 2 bullying: walking, running, jumping, falling down, playing, standing, hitting, pushing, and pushing down + (Bullying, Non-bullying)	Ack.	Finnish	No	Safety, Protection
[75], [76]	2020	Bullying detection and daily-life activities	2nd and 6th grades	motion signal, video	School	1160	8	9 activities + 2 bullying: [72]	Ack	Finnish	No	Safety, Protection

[a]http://cbs.ic.gatech.edu/mmdb/dataset.php
[b]http://mla.ucsd.edu/data/
[c]Based on the University of Delaware's Institutional Review Board approval, the database cannot be shared

pushing up, and pushing down, and six daily activities similar to those discussed by Ye et al. [72].

VI. PAIN ASSESSMENT

Manual monitoring of children to assess pain is not the most efficient method because pain occurs at irregular intervals, and there is a lot of reliance on the observer's subjectivity. It is also time-consuming and costs to get accurate data are high [18], [77], [78]. Several machine learning techniques and approaches have been proposed and can be integrated with traditional methods to assess and manage pain in children. They are categorized, in this section, based on facial pain expression, crying sounds, and body movements. The main characteristics of the available databases constructed for pain assessment are shown in Table 4.

A. FACIAL PAIN EXPRESSION

Classification of Pain Expression (COPE) [79] is a database of images created for actual face expressions of pain classification. Newborns' facial expressions were photographed with four stimuli: transport from one crib to another (rest/cry), air stimuli (air puff on the nose), friction (from cotton and alcohol rubbed on the heel), and pain (the puncture of a heel lance).

In order to evaluate the use of pain management strategies as well as infant's pain and distress from facial expressions, Harrison et al. [80] collected 142 videos from YouTube presenting intramuscular injections in infants. They were collected using search terms "baby injection" and "baby vaccine". Moreover, Sun et al. [81] constructed a database of 22 facial expression videos recorded from infants at the Maxima Medical Center in Veldhoven, The Netherlands. Three infants out of the participating ones were born premature. Li et al. [82] presented an infant monitoring system based on videos for the analysis of infants' pain. They constructed three databases namely: Train-Data (Image), Data-Clinic (Video's frames) and Data-YouTube (Video's frames). The Train-Data is composed of 16,165 images collected from the Internet and labeled according to their defined expressions (discomfort, unhappy, joy, and neutral) and states (sleep, pacifier, and open mouth). The Data-Clinic database was recorded at a hospital clinic, from 11 infants with challenging situations, including object occlusions and large head poses. Discomfort expression was captured when infants experiencing pain from a heel prick, placing an intravenous line, or a vaccination. On the other hand, the other expressions and states' videos were captured when infants stayed at the hospital seeking for medical care. On the other hand, the Data-YouTube dataset is composed of 67 videos collected from YouTube, reflecting the defined expressions and states in their study.

B. CRYING SOUND

Cry is babies' first and main oral communication to express their emotions, feelings and needs. Cry is the way to notify children's carer or their parents that they are in a risky situation, or that something is wrong. Infants might cry several times with a different time period. Crying conveys a lot of information including emotions, health situation, diseases (abnormalities) [83]. According to Cleveland Clinic'S children Hospital,[3] newborns cry when they are: hungry, tired, too cold or too hot, over-stimulated, or sick. They also might cry when they either need their diaper changed or need to be comforted. Different approaches have been presented to analyze children cries [84]–[86]. Different databases were constructed for classifying crying of children.

Baeck and Souza [87] recorded two type of cries in their database, pain and manipulation. During recording, caregivers were asked not to speak or calm the infant. To make sure that crying was not due to hunger, the infants were breast-fed for at least 20 minutes prior to recording. Recording manipulation cries started at the confirmation of the beginning of the cry. On the other hand, the recording of pain cries started around five seconds before the puncture. The duration of recording for both types of cries was 60 seconds.

Abdulaziz et al. [88] recorded a database of pain and non-pain cries from infants aged less than 12 months. Pain cries resulted from the pain stimulus carried out during routine immunization. Recordings resulting from anger or hunger were considered as non-pain cries. A database named "Donate a cry" [89] available on GitHub website was constructed from children's crying recorded by mobile phone (android and IoS applications). The recordings were performed and uploaded by volunteers. They also determined the reasons for the cries. Chang et al. [90] recorded three types of cries from healthy (normal) infants of ages 1 to 10 days. These are: hunger-induced, sleep-induced, and pain-induced. Healthy infants, here, means children who did not experience complications during birth and had normal birth weights, and gestational ages.

Dunstan Baby Language is an infant needs language discovered by Priscilla Dunstan et al. in 2006. Dunstan observed some sounds produced from her son before crying. So, a research was conducted to validate and generalize the produced sounds from infants worldwide. Accordingly, they defined five universal words used by the infants and the corresponding needs, namely NEH (Hungry), HEH (Discomfort), EAIRH (Lower Wind/Gas), EH (Burp Me), and OWH (Tired). Dunstan Baby Language is a video containing several examples of the five types of infants' cries. These examples were recorded in studio conditions. Several datasets were extracted this language such as [91].

C. BODY MOVEMENTS

Body movements are another way of communication to children use to indicate that they are in pain. Zamzami et al. [92] collected videos (referred here as "Zamzami-pain-2015") under two different pain conditions: acute and chronic. The former pain recordings were carried out from

---

[3]https://my.clevelandclinic.org/health/articles/9706-pregnancy-newborn-behavior

nine participating subjects during heel lancing procedures. The infant with chronic pain was monitored during the post-operative recovery for around two hours in the presence of nurses who score the pain using Neonatal Pain, Agitation, Sedation Scale (NPASS) scoring tool at different intervals. Zamzami-Pain-2015 database [92] was extended in another work of Zamzami et al. [93] in which a multimodal database of facial expression, body movement, and vital signs data was recorded from 18 infants during acute painful procedures (referred here as "Zamzami-multi-2016"). The database was recorded in Neonatal Intensive Care Unit (NICU) of Tampa General Hospital. The painful procedure was recorded during seven day period before and after the actual painful procedure (heel lancing). Nurses were trained prior to painful recording to document Neonatal Infant Pain Scale (NIPS) pain scores for each time period to provide the ground truth labels as 0, 1, or 2.

VII. BIOMETRICS ANALYSIS

The automated recognition of individuals based on their biological and behavioral traits including fingerprint, face, iris, palmprint, retina, hand geometry, voice, signature and gait is referred as biometric recognition [97]. Each of these traits has different applications. For examples, face recognition is applied for facial emotion recognition, face expression, gender detection, age detection, etc. In this section, we explore the children's databases developed for biometric recognition which are classified into subject identification, tracing missing children, and age-group identification. Table 5 presents the characteristics of those databases.

A. SUBJECT IDENTIFICATION

A set of training samples of one ore more biometric traits are used to train a system and generate models using a set of features. The task is to identify a child's identity from a new given sample.

There are several databases built for children, adults and elders using facial images such as VADANA [98], and MORPH [99]. There are also databases constructed specifically for children identification. Kumar et al. [100] collected dorsal hand vein images from 60 children. For each child, 20 images were acquired, 10 for the right hand and 10 for the left. In order to analyze the changes in dorsal vein patterns during the age, the data was collected in two periods: 960 images were collected in September 2017 and 240 images were collected in July 2019 for the same children. It was reported that there was no difference in the dorsal hand vein patterns of children during the considered period. It was recommended to consider longer period from five to eight years to analyze the changes in dorsal vein patterns.

Mothers have the capabilities to distinguish their newborn babies' cries from crying of others within the first days of the babies life. In this context, children can be recognized from their crying. Messaoud and Tadj [101] recorded spontaneous cries mainly in a hunger context from 13 healthy children,

who did not have family support, and reported 71.4% accuracy rate for identifying those children from their cries.

B. MISSING CHILDREN IDENTIFICATION

Biometric traits, particularly face, fingerprint and/or iris, were also applied for tracing missing children. Given a face image of a recovered child at age age_{probe}, the task is to look for a gallery of missing children with known identities and

TABLE 4. Children's databases for pain assessments. Ack.: Acknowledgment, NM: Not Mentioned, b:boys, g: girls.

Ref	Year	Task	Age	Modality	Env.	Samples	# Children: gender	Labels	Agreement	Race	Availability
[87]	2001	Cry	15-30 days	Audio	Hospital	27 audio	27	2:pain and manipulation	Ack. Hospital	Portugal	No
COPE[79]	2006	pain	18 h- 3days	Images	Hospital	200 images	26: 13g, 13b	5: pain, non-pain, cry, air puff, friction	Parents + Health System Inc	Caucasian	Yes
[88]	2010	Cry	<12months	Audio	Hospital + Home	180 audios, 881 utterances	NM	2: pain, non-pain	NM	Libya	No
[80]	2014	Pain	<12 months	Video	YouTube	142 videos	NM	5 FLACC scores	Hospital	Different	No
[89]	2015	Cry	1 day- >2years	Audio	House	1128 recordings		9:hungry, needs burping, belly pain, discomfort, tired, lonely, cold/hot, scared, Unknown	Yes	NM	Yes[a]
[92]	2015	Pain	32-41 gestations weeks	Video		10V and NIPS scores ~ 116	10: 5b, 5g	2: pain, non-pain	Parents + University + Ack.	Caucasian	No
[93]	2016	Pain	32-41 gestation weeks	V1,V2,Audio	Hospital	490 A	18	3-scores	Parents + University + Ack.	Caucasian	No
[90]	2017	Cry	1-10 days	Audio	Hospital	22 videos, 16,378 frames	37: 15b, 22g	3: Hungry, sleepy, pain	Parents + Hospital	Taiwan	No
[81]	2019	pain	2 days- 13 months	Video	Hospital	22	22	2: comfort, discomfort	Parents	NM	No
Data-Train[82]	2021	Pain	majority <2 years	Images	Internet	16165 Image		4-expressions + 3-states	NM	Asia, Europe and Africa	No
Data-Clinic[82]	2021	Pain	infants	Video	Hospital	11 Video, 18,323 frame		4-expressions + 3-states	Parents + clinic	NM	No
Data-YouTube[82]	2021	Pain	Infants	Video	YouTube	67 Video, 13393 frame	67	4-expressions + 3-states	NM	NM	No
Baby Chillanto[94]	2008	Cry	<9months	Audio	Hospital	2267	NM	5: asphyxia,deaf, hunger,normal, pain	NM	Mexican, Cuban	Upon request
Baby2020[95]	2021	Cry	<7months	Audio	Home, Hospital	5540	100	3: hungry, Sleepy, Wakeup Hunge, pain, wet/discomfort	Ack. Hospital		No
IICC SNCU[96]	2020	Cry		Audio	Hospital	519					No

[a] https://github.com/gveres/donateacry-corpus

age $age_{gallery}$ at which they were either lost or stolen in an attempt to unite the recovered child with his/her family [102]. Tracing missing children using facial images is not as accurate as using fingerprints and iris [103]. However, it is more popular because parents kept photographs of their kids' faces while they might not keep photographs for their irises or fingerprints.

Databases constructed and employed for tracing missing children have special aspects. For example, there should be a number of images/samples for each participating child/subject during a specified period of age. To analyze the capability of face recognition technology to trace lost children, several databases have been created [6], [104], [105].

In-the-Wild Child Celebrity (ITWCC) database [6] comprises a collection of facial images collected from 304 subjects aged from five months to 32 years. Each subject has at least two images less than 16 years. That means it analyzes the children growth and development with age range from five months to 16 years. ITWCC also provides meta and demographic information for each individual, including: a unique photo identifier, subject name, age, race, gender, data of the photo, etc. Best-Rowden *et al.* [104] constructed Newborns, Infants, and Toddlers Longitudinal (NITL) facial images database to investigate the feasibility of face recognition of children as they age. The database was collected during the period of March 2015 to March 2016 in Saran Ashram Hospital, Dayalbagh India. It was reported that facial recognition techniques are still immature for young children aged less than three years while they were able to recognize faces of children older than three years.

Children Multimodal Biometric Database (CMBD) [103] of face, fingerprint and iris was collected from students of kindergarten classes of two schools in India and few home-schooled young children. For iris images, five samples for both left and right irises were simultaneously recorded using Cross-Match iris scanner. For fingerprint images, five samples were collected for left hand, right hand, and two thumbs respectively for each child. Additionally, ten facial samples were captured for each child. All children were normal and did not wear contact lenses during recording.

Children Longitudinal Face (CLF) dataset [105] consists of 3,682 face images of 919 participating children aged between two and 18 years. Each child had at least four face images acquired over a time span of up to six years.

C. AGE-GROUP IDENTIFICATION

Age/age-group can be identified using biometric traits. This task can be considered as a regression task in which the system can detect the age of child in years. It can be, also, considered as a classification problem to detect the age-group of a subject as child, adult, senior adult, elder [106] or as infant, baby, toddler, preschool, etc. For example, Safavi *et al.* [107] utilized OGI kids database [32], which mainly developed for speech recognition, to develop a technique for identifying children age-groups from their speech.

VIII. DISCUSSION

It can be noticed that, the trend of technologies for incorporating ML for children can be divided into two timelines, before and after 2010. Till 2010, the focus was mainly on speech analysis and pain assessments. Five out of six databases were mainly based on speech modality while the remaining one is on pain assessments using facial images. On the other hand, since 2010, there are notable amount of works and several databases developed for various objectives using different modalities.

Figure 6 depicts the distributions of the main tasks of the reviewed studies per children's development stages. Pain assessment was considered mainly for infants and then for toddlers, only. In other words, there is no study that evaluated pains for life stages such as preschool, grade-school or teens. It may also be noted that none of the reviewed studies focused on the preschool stage. Furthermore, bullying detection was only considered for grade-school stage.

A. DATA TYPES AND SOURCES

As shown in Tables 1 to 5, different data types/modalities were gathered from children including images, speech, videos, sensor data, physiological signals as well as from personal and medical records. For each modality, there were different methods to collect data. For instance, facial images can be collected from Internet or captured directly from participating subjects using digital cameras. Likewise, physiological signals, including heart rate, brain's electroencephalogram, blood and temperature signals, etc., can be generated using different sensors/devices including electrocardiography (ECG), photoplethysmography (PPG), ballistocardiography (BCG) [108].

User-generated content, including videos, images and speech, uploaded by users to social media networks, are considered as reliable sources for making decisions and also for changing current or for establishing new policies. Researchers have relied on Youtube videos to build their databases and test/analyze their hypotheses for different applications such as health [80] and affective states [43], [109]. The user-generated databases are known as natural databases. Videos, images or speech data can be also recorded/captured in a specific designed studio environment in which the participants are asked to act or enact some predefined scenarios. Such databases are known as either acted or induced databases.

Sensing data was also considered in the construction of databases for children. One of the main concerns for recording sensing data is the number of sensors. In general, the more number of sensors are, the more reliable the database is, and more accurate the generated models will be [110]–[112]. Another concern related to sensing data is the positions of sensors while recording data. Waist, ankle, wrist, arm, chest and leg are possible parts of a child's body to attach sensors. In case of using a single sensor, waist is recommended as the best place for motion sensors. This is because the waist is
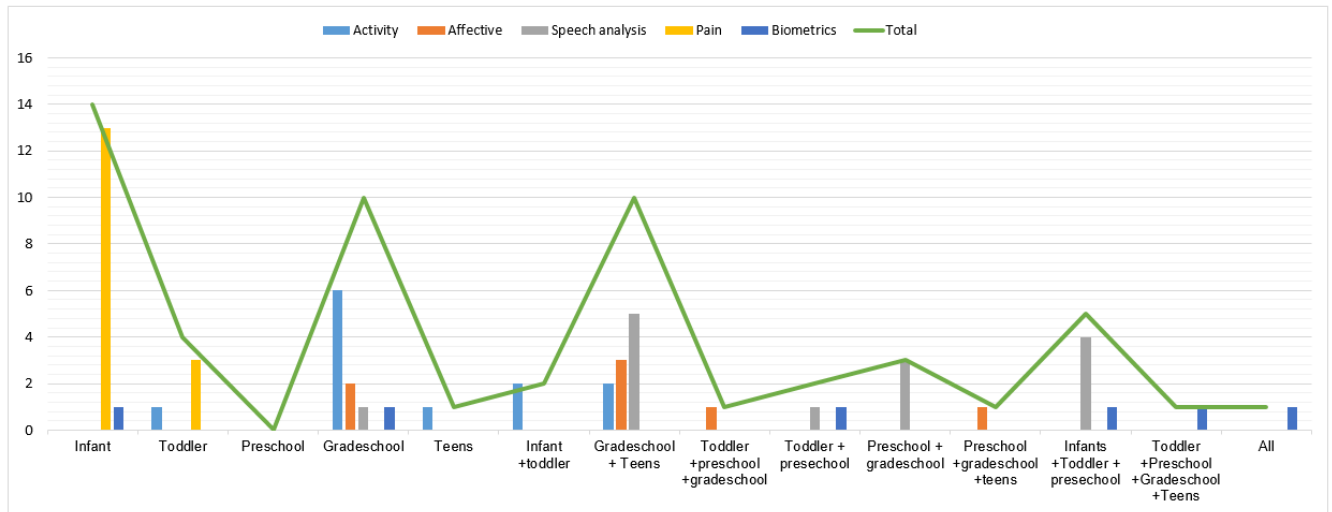
FIGURE 6. Distributions of the reviewed databases' objectives per children's development stages.

close to the center of mass of a whole human body, and the torso occupies the most mass of a human body [72], [113].

Brain signals have been collected for emotion analysis in [114], [115]. However, collecting brain signals from children is not recommended as they utilize EEG which poses a difficulty when attaching it to the child's heads for a long time [9]. Ye *et al.* [75] reported that the motions of wrists and arms are very random and are difficult for recognition of daily-life activities. This might not be valid for children due to the short sizes of their arms and the upper arm's EMG signal convey part of brain's EEG signals [9]. A study was conducted by Chowdhury *et al.* [116] to compare activity recognition systems for sensor data positioned in leg, waist and chest individually and in combinations. For single sensors the highest accuracy rate was obtained using waist sensor data which was followed by leg sensor data. However, combining leg and waist senor data achieved the highest results and even higher than the fusion of waist, leg and chest sensor data.

B. ENVIRONMENTAL SETTINGS

As shown in Tables 1 to 5, databases were recorded in different environments including hospital (clinics, therapists offices), schools (kindergarten, primary, middle, secondary), universities, homes, and in research labs. It is possible to record data in multiple environments [29], [88]. For example, in [88] the ''pain cry'' instances were recorded at a local pediatric clinic, in Darnah, Libya. However, non-pain cries were recorded in quiet rooms at infants' homes. Similarly, some of children's speech instances of LANNA [29] database were recorded in schoolroom while other instances were recorded in a speech and language therapist's consulting room.

Specifically, the information in the MMLA-MATH Database [64], discussed in Section V, was recorded in a room equipped with five video cameras, four audio sources, and three digital pens. The mathematical problems were

displayed to students' groups via tabletop computer screen. MMDB Database [58], discussed also in Section V, was recorded in Child Study Lab (CSL) at Georgia Tech. CSL is a child-friendly 300-square feet laboratory space which was equipped with the following sensing capabilities:

- Visual sensing (Cameras): Two frontal view Basler cameras, one overhead view Kinect (RGB-D) camera, eight side view and three overhead view AXIS cameras.
- Vocal sensing (Microphones): One omnidirectional and a cardioid microphone, ceiling mounted, two wireless lapel microphones, worn by both the child and the adult.
- Physiological sensing: Four Affectiva Q-sensors for electrodermal activity and accelerometry, worn on right and left wrists of both the adult and the child.

For Dartmouth [41] image database, three cameras were positioned in front of a child with a distance of 130 meters in front and with three different angles (0, 30 and 60 degrees). Participating subjects were seated in front of a black felt backdrop and wore black hats to cover their hair and ears.

Additionally, Baeck *et al.* [87] recorded their database from infants in the 'Follow Up' sector of the Fernandes Figueira Institute, a public pediatric hospital in Rio de Janeiro. The participating infants were placed in a canvas bed on their back. Recordings were done in the presence of infants' caregivers.

Infants Cry Database [90] was recorded in the department of Obstetrics and Gynecology at National Taiwan University Hospital, Yunlin Branch, Taiwan. Cries were recorded with infants in a supine position using a SONY HDR-PJ10 HD digital video recorder with a built-in microphone. The microphone was held about 40 centimeters away from the infant's mouth. Each recorded file is of duration between 10 and 60 seconds. Efthymiou *et al.* [55] designed a child's room with three Kinect V2 sensors for their recordings. One sensor

TABLE 5. Children's databases for biometrics. Ack.: Acknowledgment, NM: Not Mentioned, b:boys, g: girls.

Ref	Year	Task	Age	Modality	Env.	Samples	Children: Gender	Labels	Agreement	Race	Availability
[101]	2010	Subject identification	<6 months	Cry speech	Hospital	1615 utterances	13: 6b, 7g	13 children	NM	NM	No
ITWCC [6]	2015	Longitudinal Face	5m-32 years (5m-16 years)	Images		1705 I	304	304	NM	Different	Upon request
NTTL [104]	2016	Longitudinal Face	0-4 years	Images	Hospital	3144	314	314	Parents	India	No
CMBD [103]	2017	Longitudinal	2-4 years	Face, fingerprint, iris	schoolroom + home		NM		Ethics board approval + school administrations +Ack.	India	No
CLF [105]	2018	Longitudinal Face	2-18 years	Images		3682	919: 315g 604b	919	Ack. school	NM	No
[100]	2021	Subject identification	5-12 years	Dorsal Hand Vein Image		1200	60	60	Ack.	NM	No

environment, called Grounded Early Adaptive Rehabilitation (GEAR) to provide motor interventions. It was composed of physical (hardware) and cyber (software) components. The physical component consisted of playground environment, an open-area body weight support (BWS) device, two mobile robots, and camera network connected to server for processing data collection. The cyber component contained interaction system that read child's motion and video data, identified the child's action and behavioral models for the child-robot interaction and recommended the most appropriate robot action in support of given motor training goals for the child. In order to keep the child engaged and interacting with the robots, a human operator remotely controlled the robots. The datasets of TEL and MIC [28], descried in Section III, were recorded in a laboratory facility designed at a local science and technology museum. The constructed laboratory was provided with three standard telephone lines. In order to collect TEL recordings, an IAC booth was supplied with telephone handsets. Another booth was lined with anechoic foam wedges for collecting MIC recordings.

C. DATABASE COLLECTION'S PROCEDURES

Several methodologies have been specifically designed to collect data/information for children's databases. Table 6 presents these methodologies and categorizes the reviewed studies, accordingly. From the discussion thus far, it is evident that not all data/information for databases was recorded in specific studio labs. Researchers also utilized available user-generated data uploaded to Internet to build their databases by collecting either videos from YouTube [43], [80]–[82] or images from Internet [82].

Transfer learning was also adopted to generate data-driven models for children [55], [117], [118]. Given a source dataset D^s of X^s attributes and Y^s labels, and a target dataset D^t of X^t attributes without labels, with the assumption that the distributions of source and target domains are not matching, transfer learning aims at making both distribution matching through learning domain invariant features [119]. Hosseini et al. [117] employed deep domain adaptation to transfer an activity recognition models, trained on a popular adult dataset for activity recognition, UCI-HAR [120]. The target dataset was collected from 20 children aged from 8 to 14 years, in their prior work [121].

Transfer learning was also employed by Efthymiou et al. [55] such that pre-trained model on the Sports1M database [122] was fine-tuned to classify the considered activities in their study. Also, Khan [118] utilized the available database COCO [123] of object recognition to monitor children during the following situations: (a) detection of face covered, (b) detection of blanket removed, (c) detection of frequent moving, (d) awake/sleep detection.

D. ETHICAL PERSPECTIVES

Ethical guidelines/standards need to be followed during databases' construction. Development procedures also need to be reviewed by responsible committees in order to protect

was located at the ceiling facing down to the interaction area and the remaining two sensors were located at each side of the room. Kokkoni et al. [53] designed a smart pediatric learning

TABLE 6. Available methods for children's databases' constructions.

Procedure	Tasks	References
Asking children to read texts and recording their speech	Speech and language analysis	[11], [12], [23], [25], [28]
Speech recorded via telephone channels (TEL)	Speech and language analysis	[28]
Whole-day, naturalistic recordings of children's vocalizations and auditory environments	Speech and language analysis	[23], [25]
Asking children to wear sensors (smartwatch, arm band) and perform different activities as instructed	Activity analysis	[9], [52]
Collecting records/logs from the learner records store of e-learning platforms	Activity analysis	[68]
Recording videos during learning or solving problems	Activity analysis, Emotion recognition	[43] [64], [68]
Asking children to pose various facial emotion expression	Emotion	[39] [41]
Semi-structured play interactions between children and adults	Activity	[58]
School violence simulation event designed by psychologists	Bullying detection	[10], [72], [75]
Presenting videos to children with Neurodevelopmental Disorders then asking them to express their opinions on the events	SA, Autism analysis	[47]
Human Robot Interaction	Activity analysis, emotion	[48], [53]–[55]
Recording either videos, audios, or both during taking injection, immunizations, vaccination, etc.	Pain assessment	[79]–[82], [87], [88], [90], [94]–[96]
Dunstan Baby Language	Pain assessment	[91]
Online survey	Speech and language analysis	[30], [31]
User-generated data	Emotion recognition, Pain assessment	[43], [80]–[82], [82]
Transfer learning	Activity analysis	[55], [117], [118]

children's rights and to ensure that the collected data will not affect them now or in the future. To a great extent, researchers are aware of the importance of obtaining consents from parents or legal guardians, before collecting material for databases, as shown in Tables 1 to 5. Additionally, approvals are obtained from organizations or institutions including, schools and hospitals in which data is collected. Some studies, also, obtained approvals from specific committees in universities prior to conducting the studies or collecting the data. For example, Dalrymple *et al.* [41] took into account the ethical guidelines of the Protection of Human Subjects at Dartmouth College when collecting data for the Dartmouth database. In order to construct the database for analyzing school violence, Han *et al.* [10] first obtained approval from the students' parents and the experimental settings were reviewed and permitted by the Ethics Committee of University of Oulu as well. In addition to written informed consent of infants' parents, the database collection procedure of [90] database was approved by the Ethical Review Committee of National Taiwan University Hospital. Similarly, besides the informed consent of infants' parents, the procedures of recording Zamzami-pain-2015 [92] and Zamzami-multi-2016 [93] databases were in compliance with protocols and ethical directives for research involving human subjects at the University of South Florida. Grill *et al.* [29] got approval to collect information for LANNA database,

and to conduct their research from the Ethics Committee of Motol University Hospital in Prague, Czech Republic. This is in addition to written consent from participating children's parents. Different databases, including MMDB [58], MMLA-Math [52], [64], [65] were recorded under the institutional review board (IRB) approvals. IRB is an organization that reviews and approves (or disapproves) any research study involving human subjects.[4] NIMH-ChEFS [42], Dartmouth [41], MMDB [58] and CAFE [40] databases got their approvals from parents and legal guardians to make the participating children's pictures publicly available for research purposes. Therefore, getting IRB approval alone is not sufficient to conduct research requiring children's participation. Parents' approval is also required.

Other studies acknowledged the participating children [32], their parents, nurses, other peoples [41], schools [11], [12], hospitals and/or universities [41] for helping collect the data/information for their databases, or assistance in conducting the studies. Although, Harrison *et al.* [80] collected the videos for pain assessment from Youtube, an approval from Children's Hospital of Eastern Ontario (CHEO) Research Ethics Board in Ottawa, Canada was obtained for conducting the study.

---

[4]https://dimewiki.worldbank.org/IRB_Approval

E. VERIFICATION

A databases should be verified to give researchers in academia, the clinical and industrial domains confidence for the use of the databases and the required trust on the reported findings. This phase ensures that the tasks of developing a database were carried out in a systematic and valid manner. Validation tasks differ upon the procedure and the application of the database. The following list defines a set of activities need to be considered during databases' constructions and Figure 7 depicts the common verification activities.

- Participants Selection: It is essential to clearly define the inclusion and exclusion criteria for selecting participants. Researchers in turn need to verify whether the candidate participants are in accordance with the inclusion criteria or not. Although these criteria were not clearly documented in most of the reviewed papers, we summarize, in this part, some of the considered criteria. Paz *et al.* [47] excluded two children as they did not complete all the required neuro-psychological tests during recording. Oller *et al.* [23] evaluated the children with autism and language delay by diagnosing clinicians. The children who passed the evaluation were selected for participation. Furthermore, one of the criteria of participants' selection in [124] database was the equal distribution of children in terms of age, sex, and regional language.

- Children's Performance: This task verifies the performance of the participating children with considered standards or proposed instructions. It involves obtaining available standards and evaluating them to if it is possible to use them. Some studies utilized the available standards whereas, others defined their own instructions to be followed by participating children. The designed instructions needed to be verified by domain's experts. Different approaches have been followed to verify the children's performance. For example, the participating subjects for RaDF database [39] were trained by two certified Facial Action Coding System (FACS) [125] specialists to express their emotions. Similarly, two clinical experts were involved to verify the collected videos/frames of [81] database as "comfort" or "discomfort". A team of neuro-scientists worked together with teachers at the children's theater group to review the Ekman and Friesen details and reviewed the procedures with each participating child for NIMH-ChEFS database [42]. Each speech utterance in OGI database [32] was independently verified by two individuals and was rated as "good", "questionable" or "bad". An utterance was considered as "good" when the word was clearly intelligible with no significant background noise or extraneous speech otherwise it was considered as either "questionable" or "bad".

- Recording procedures and protocols: It is crucial to design systematic and well-defined data collection procedure and settings. Therefore, evaluators need to be involved to verify the recording procedure and the

environmental settings. The recording procedure also need to be reviewed by ethical committees.

F. ANNOTATION

Given a set of labels $Y = \{y_1, y_2, \ldots, y_n\}$, n is the number of classes, $n = 2$ for binary classification problems and $n > 2$ for multi-class problems, and given a set of samples $X = \{x_1, x_2, \ldots, x_m\}$, where m is the number of database's samples, the task is to assign the appropriate label y_i for each sample x_j in a valid manner. Different scenarios have been proposed to annotate databases. Annotation task can be conducted manually by domain experts and/or by trained individuals according to standards/instructions. Several pain assessment standards have been proposed to evaluate and document pain in children including the Children's Hospital of Eastern Ontario Pain Scale (CHEOPS) [126], Face, Legs, Activity, Cry, Consolability (FLACC) tool [127], Neonatal Infant Pain Scale (NIPS), Neonatal Pain, Agitation, Sedation Scale (N-PASS) [128], to name a few. These standards/tools were utilized to annotate different databases. Nurses were trained to provide the ground truth labels before recording COPE [79], Zamzami-pain-2015 [92] and Zamzami-multi-2016 [93] databases, according to NIPS, NPASS, or defined instructions. Children's pains in [80] database were assessed by crying duration and pain scores using the FLACC tool. A total of 20 faculty and staffs working in the Center for Developmental Epidemiology (CDE) at Duke University Medical Center participated in rating NIMH-ChEFS database [42]. Li *et al.* [82] defined a set of infants' expressions and states with their corresponding Facial Action Units. Accordingly, pediatricians and nurses were trained to annotate the Data-Clinic and Data-YouTube videos databases at frame level.

On the other hand, rather than domain experts, annotators (crowd-workers) from crowd sourcing platforms such as Amazon's Mechanical Turk (AMT) [129] have been recruited for annotating several "gold standard" databases. This helps in acceleratin the process of developing databases, mitigating the issue of scarcity of individuals with domain experiences, and reducing costs. However, there are growing concerns related to the reliability and validity of the annotation task [1], [130], [131].

Children's parents have also participated in the annotation task. For example, in "Donate a cry" database [89], parents were asked to assess their infants' cries and assign the appropriate cry's reason. However, it is not sufficient to rely on the annotation provided by the participants and there is a need to be evaluated/validated by experts, particularly for subjective tasks such as pain assessments and affective states.

Annotation task can be performed at different levels including uttered phoneme [26], frame level [53] or video level [43]. For children's activities in [53] database, the annotators watched the five synchronized video streams and assigned the appropriate action label as long as the action was clearly visible from at least one camera at any particular time frame. Pacheco *et al.* [54] also provided coordinates of a bounding
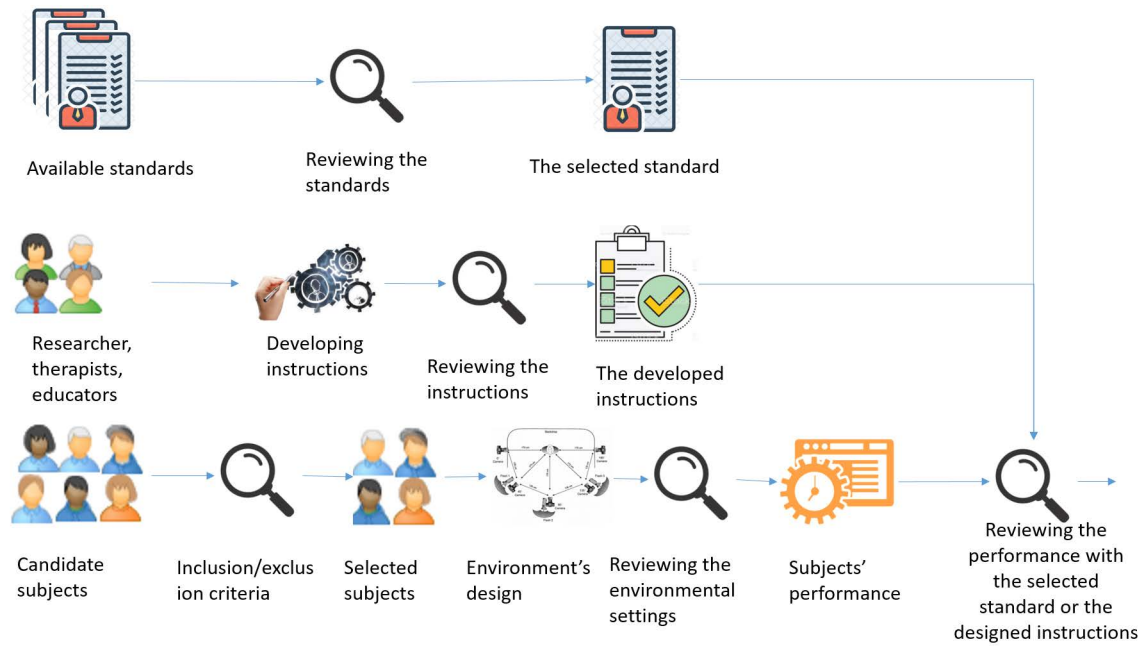
FIGURE 7. Activities for children's database verification.

box containing the infant's body. The bounding box was assigned either when changing action or moving by more than 50% out of the frame of the previous bounding box using Kinovea tool.[5]

G. CHALLENGES
In general, for several reasons, developing ML-based techniques is more challenging for children than for adults. First of all, children's growing process is far more complex than adults due to changes in the shape and size of children's facial components, voice, motion and activities [6]. In addition to this, collecting data from children is difficult due to their rapid movements and children do not listen well to the experimenters commands/instructions [52]. They may not perform as instructed.

Another issue is related to community awareness and cultures. Some humans still do not accept or believe that technology can contribute to developing children skills, in their treatments, [9] or in protecting them. Therefore, obtaining parents' consents to gather data from their children is not a straightforward task, especially for children with special needs [29].

Due to this difficulty of selecting participating children with special needs, researchers recruited the same subjects of a certain database to participate for developing another database, for example [23] and [25] databases. It was reported that SLI affects around six percent of the pediatric population [132], [133]. Demographic bias presents another issue. Tomblin *et al.* [134] reported that boys are more affected by SLI than girls. Similarly, ASD is more than four times more common among boys than girls [135].

[5]https://www.kinovea.org/

Naturally, diseases diagnostic databases suffer from skewed data distribution [136]. Thus, such databases heavily affect the performance of predictive models [137]. Several techniques have been proposed to mitigate this issue at different levels including, at the data level and/or algorithmic level [138]. The data-level-based techniques include oversampling, under-sampling, and hybrid of oversampling and under-sampling techniques. On the other hand, the algorithmic-level-based techniques include cost-sensitive learning, one-class learning, and ensemble learning [139]. Data augmentation techniques also contribute to alleviate this issue and the scarcity of the training samples.

Additionally, constructing children's databases might take a long period of time depending on the task and the application. For example, Oller *et al.* [23] started collecting their database in 2006 and completed in 2009. Similarly, recording children's speech of LANNA [29] database took seven years, from 2003 to 2010, and Kumar *et al.* [100] collected the dataset in 22 months.

Basak *et al.* [103] reported some unconventional challenges for capturing images for children's faces, irises and fingerprints. The time required for capturing iris images is more than the allowed time by sensor's in-built threshold. So, a constant adult needs to be available to motivate children to look into the camera and not blink or move their eyes for a couple of seconds. Another issue is that sensors cannot clearly capture children's fingerprints, particularly for ages less than three years and for those who have excessive dry skin. To overcome such issues a small amount of moisturizer with the help of adult supervision had to be applied. Several facial images were ignored due to issues relating to pose, illumination, and blurriness variations.

It is recommended to prepare studio environment for recording databases in popular area visited by large number of people. For example, the datasets of [28] were collected at the St. Louis Science Center, a local science and technology museum which was one of the most attractive places in the nation at that time.

Synchronization is another challenge related to multimodal databases [33], [140]. Such databases entail different types of data including physiological signals, audios and videos. Synchronization is the alignment task of those signals with each others. For generating more efficient systems, these signals/streams need to be aligned precisely and reliably. This issue gets more complicated in case of heterogeneous multimodal databases such as synchronizing different sampled signals of accelerometer and multi-track audio with multi-video streams [141].

H. LIMITATIONS AND FUTURE DIRECTIONS

Here, we present the main limitations of the available databases and future research opportunities towards reliable, reproducible and reusable databases.

- Most of the reviewed databases are not publicly available. Researchers constructed in-house database of small number of samples to evaluate their approaches. The developed models and the reported findings using such databases maybe questionable.
- Most of the reviewed studies were built upon one modality. With the advancement of deep learning techniques and their capability for automated representation learning from nearly all types of data [142], [143], there is an urgent need to construct much larger annotated databases from different modalities for researchers to develop more advanced techniques and compare their findings with others.
- The process of database's construction from data collection to database deployment or distribution needs to be documented in detail. This in turn, will make the databases reliable and reproducible. Here, we define the most important and common information that needs to be provided when constructing new databases.

  - Database Statistics: The database needs to be detailed in terms of number of children, number of samples, number of classes, number of samples per class, and the number of samples per gender. The number of samples per subject during a specified period of age is also important for some applications such as tracing missing children.
  - Children's Demographics: Details of the participating children need to be provided including their genders, races, languages and ages.
  - Environmental and Recording Settings: the environment in which the database was recorded needs to be described. This also includes types of recording devices, number of devices, etc.

  - Annotation Procedure: As mentioned above, annotation is one of the most important tasks for constructing databases. Therefore, this task needs to be described in detail and verified by providing the number of the involved annotators, the educational-level of the annotators, the agreement measures to validate the inter-annotator agreement, such as Kappa score [144], Finn's metric [145], SScore [146] and Krippendorff's alpha statistic [147].
  - Ethics: In addition to determining the availability of the ethical consents for conducting the study or collecting the database and from whom, it also includes the distributing of the dataset within research community. Another aspect that needs to be taken into account is the possibility of presenting the data, especially images of the participating children, in the written research media such as in presentations or in scientific papers.

- It is not sufficient just to build databases for binary classification tasks. For example, there is an important need to go beyond recognizing children with speech-related disorders from the healthy children and constructing databases for diagnosing the disorders' types. It is also a future research opportunity to do in-depth analysis to diagnose the autism levels, effects and factors. This is also true for pain assessment. It is not enough to detect whether an infant is crying or not. There is a need to deeply analyze the pain's effects and the children's needs. Infants' pain can be analyzed in early medical diagnoses. Likewise, it becomes essential to move from basic emotional states towards more complex affective states. For some applications such as children bullying detection, it is recommended to construct databases using physiological and emotional signals.
- It is an essential to analyze the development of children during different ages in terms of their activities, learning, pain assessment, therapies, and thinking manner, among others. Developing the corresponding technology requires constructing databases for each of those applications for children at their different life stages.
- Exploring the effects of sensors' numbers and to which of child's body parts they are to be attached is another aspect of future work that needs to be considered.

IX. CONCLUSION

This study provides a comprehensive review of state-of-the-art databases developed for use in children-related technology incorporating machine learning. It gives interested researchers from multidisciplinary fields an overview of the databases' methodologies, characteristics, objectives, applications and challenges. We started with defining the main tasks considered in the literature and categorized the reviewed studies based on them. We also evaluated the reviewed studies

according to our defined evaluation attributes. Furthermore, the paper discusses the main steps for constructing children's databases for ML-based systems. This survey showed that there are obvious advantages for applying ML for children care, development and protection. It was noticed that different children life stages were considered, yet they were not covered for all the reviewed tasks. There are significant opportunities for future work to fill these research gaps. It was also found that in order to develop more advanced techniques and seed further research, it is necessary to construct large databases that can be used for generating deep-learning-based solutions and benchmark results. There have been attempts to develop techniques based on multi-view, multi-sensors, multimodal and/or multi-labels but with insignificant efforts. Our survey also presented the evaluation rules to verify and validate the process of developing children's databases.

This study showed that it necessary to get informed consent from parents on behalf of their children prior to collecting their data or conducting any research related to children. Approval also might be obtained from other related organizations if necessary. In order to create reliable and reproducible databases, the process of database construction should be clearly documented.

ACKNOWLEDGMENT

REFERENCES

[1] A. Paullada, I. D. Raji, E. M. Bender, E. Denton, and A. Hanna, "Data and its (dis) contents: A survey of dataset development and use in machine learning research," *Patterns*, vol. 2, no. 11, Nov. 2021, Art. no. 100336.

[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, Feb. 2015.

[3] P. Grother and P. Ngan, "FRVT: Performance of face identification algorithms," NIST Interagency Rep., Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep., 2014, vol. 8009.

[4] A. Lanitis and C. J. Taylor, "Robust face recognition using automatic age normalization," in *Proc. IEEE 10th Medit. Electrotechnical Conf. Inf. Technol. Electrotechnol. Medit. Countries.*, vol. 2, May 2000, pp. 478–481.

[5] K. Ricanek and E. Boone, "The effect of normal adult aging on standard pca face recognition accuracy rates," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, vol. 4, Jul./Aug. 2005, pp. 2018–2023.

[6] K. Ricanek, S. Bhardwaj, and M. Sodomsky, "A review of face recognition against longitudinal child faces," in *Proc. BIOSIG*, A. Brömme, C. Busch, C. Rathgeb, and A. Uhl, Eds. Bonn, Germany: Gesellschaft für Informatik e.V., 2015, pp. 15–26.

[7] J. G. Wilpon and C. N. Jacobsen, "A study of speech recognition for children and the elderly," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. Conf.*, vol. 1, May 1996, pp. 349–352.

[8] D. Elenius and M. Blomberg, "Comparing speech recognition for adults and children," in *Proc. FONETIK* 2004, pp. 156–159.

[9] S. Suzuki, Y. Mitsukura, H. Igarashi, H. Kobayashi, and F. Harashima, "Activity recognition for children using self-organizing map," in *Proc. 21st IEEE Int. Symp. Robot Hum. Interact. Commun.*, Sep. 2012, pp. 653–658.

[10] T. Han, J. Zhang, Z. Zhang, G. Sun, L. Ye, H. Ferdinando, T. Seppänen, X. Yu, and S. Yang, "Emotion recognition and school violence detection from children speech," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, pp. 1–10, Dec. 2018.

[11] D. Sztahó, K. Nagy, and K. Vicsi, "Automatic sentence modality recognition in children's speech, and its usage potential in the speech therapy," in *Cross-Modal Anal. Speech, Gestures, Gaze Facial Expressions*. Berlin, Germany: Springer, 2009, pp. 266–275.

[12] D. Sztaho, K. Nagy, and K. Vicsi, "Subjective tests and automatic sentence modality recognition with recordings of speech impaired children," in *Development of Multimodal Interfaces: Active Listening and Synchrony*. Berlin, Germany: Springer, 2010, pp. 397–405.

[13] F. Thabtah, "Autism spectrum disorder screening: Machine learning adaptation and DSM-5 fulfillment," in *Proc. 1st Int. Conf. Med. Health Informat.*, May 2017, pp. 1–6.

[14] F. N. Buyukoflaz and A. Ozturk, "Early autism diagnosis of children with machine learning algorithms," in *Proc. 26th Signal Process. Commun. Appl. Conf. (SIU)*, May 2018, pp. 1–4.

[15] T. Akter, M. Shahriare Satu, M. I. Khan, M. H. Ali, S. Uddin, P. Lio, J. M. W. Quinn, and M. A. Moni, "Machine learning-based models for early stage detection of autism spectrum disorders," *IEEE Access*, vol. 7, pp. 166509–166527, 2019.

[16] (2021). UNICEF and WORLD ECONOMIC FORUM. *Children and AI Where are the Opportunities and Risks?*. [Online]. Available: https://www.unicef.org/innovation/sites/unicef.org.innovation/files/201%8-11/Children%20and%20AI_Short%20Verson%20%283%29.pdf

[17] J. Saraswathy, M. Hariharan, S. Yaacob, and W. Khairunizam, "Automatic classification of infant cry: A review," in *Proc. IEEE Int. Conf. Biomed. Eng. (ICoBE)*, Feb. 2012, pp. 543–548.

[18] G. Zamzmi, R. Kasturi, D. Goldgof, R. Zhi, T. Ashmeade, and Y. Sun, "A review of automated pain assessment in infants: Features, classification tasks, and databases," *IEEE Rev. Biomed. Eng.*, vol. 11, pp. 77–96, 2018.

[19] M. E. Beckman, A. R. Plummer, B. Munson, and P. F. Reidy, "Methods for eliciting, annotating, and analyzing databases for child speech development," *Comput. Speech Lang.*, vol. 45, pp. 278–299, Sep. 2017.

[20] J. McKechnie, B. Ahmed, R. Gutierrez-Osuna, P. Monroe, P. McCabe, and K. J. Ballard, "Automated speech analysis tools for children's speech production: A systematic literature review," *Int. J. Speech-Lang. Pathol.*, vol. 20, no. 6, pp. 583–598, Oct. 2018.

[21] M. C. Mihaescu and P. S. Popescu, "Review on publicly available datasets for educational data mining," *WIREs Data Mining Knowl. Discovery*, vol. 11, no. 3, p. e1403, May 2021.

[22] M. Benzeghiba, R. De Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, C. Ris, R. Rose, V. Tyagi, and C. Wellekens, "Automatic speech recognition and speech variability: A review," *Speech Commun.*, vol. 49, nos. 10–11, pp. 763–786, Oct. 2007.

[23] D. K. Oller, P. Niyogi, S. Gray, J. A. Richards, J. Gilkerson, D. Xu, U. Yapanel, and S. F. Warren, "Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 30, pp. 13354–13359, 2010.

[24] M. VanDam, S. E. Ambrose, and M. P. Moeller, "Quantity of parental language in the home environments of hard-of-hearing 2-year-olds," *J. Deaf Stud. Deaf Educ.*, vol. 17, no. 4, pp. 402–420, Oct. 2012.

[25] M. VanDam, D. K. Oller, S. E. Ambrose, S. Gray, J. A. Richards, D. Xu, J. Gilkerson, N. H. Silbert, and M. P. Moeller, "Automated vocal analysis of children with hearing loss and their typical and atypical peers," *Ear Hearing*, vol. 36, no. 4, pp. e146–e152, 2015.

[26] O. Saz, W. Rodríguez, E. Lleida, and C. Vaquero, "A novel corpus of children's disordered speech," in *Proc. 1st Workshop Child, Comput. Interact.*, 2008, pp. 1–6.

[27] G. Szaszák and K. Vicsi, "Speech recognition supported by prosodic information for fixed stress languages," in *Proc. Int. Conf. Text, Speech Dialogue*. Berlin, Germany: Springer, 2007, pp. 262–269.

[28] J. D. Miller, S. Lee, R. M. Uchanski, A. F. Heidbreder, B. B. Richman, and J. Tadlock, "Creation of two children's speech databases," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. Conf. Proc.*, vol. 2, May 1996, pp. 849–852.

[29] P. Grill and J. Tučková, "Speech databases of typical children and children with SLI," *PLoS ONE*, vol. 11, no. 3, Mar. 2016, Art. no. e0150365.

[30] W. J. Coster, J. M. Kramer, F. Tian, M. Dooley, K. Liljenquist, Y.-C. Kao, and P. Ni, "Evaluating the appropriateness of a new computer-administered measure of adaptive function for children and youth with autism spectrum disorders," *Autism*, vol. 20, no. 1, pp. 14–25, Jan. 2016.

[31] M. Costa, A. Costa, V. Julián, and P. Novais, "A task recommendation system for children and youth with autism spectrum disorder," in *Proc. Int. Symp. Ambient Intell.* Cham, Switzerland: Springer, 2017, pp. 87–94.

[32] K. Shobaki, J.-P. Hosom, and R. A. Cole, "The OGI kids' speech corpus and recognizers," in *Proc. 6th Int. Conf. Spoken Lang. Process.*, 2000, pp. 564–567.

[33] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Inf. Fusion*, vol. 37, pp. 98–125, Sep. 2017.

[34] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affective Comput.*, vol. 1, no. 1, pp. 18–37, Jan. 2010.

[35] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *J. Personality Social Psychol.*, vol. 17, no. 2, p. 124, 1971.

[36] P. Ekman, Wallace V. Friesen, M. O'sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti, and K. Scherer, "Universals and cultural differences in the judgments of facial expressions of emotion," *J. Personality Social Psychol.*, vol. 53, no. 4, p. 712, 1987.

[37] E. L. Rosenberg and P. Ekman, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford, U.K.: Oxford Univ. Press, 2020.

[38] N. J. Shoumy, L.-M. Ang, K. P. Seng, D. M. M. Rahaman, and T. Zia, "Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals," *J. Netw. Comput. Appl.*, vol. 149, Jan. 2020, Art. no. 102447.

[39] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, "Presentation and validation of the radboud faces database," *Cognit. Emotion*, vol. 24, no. 8, pp. 1377–1388, 2010.

[40] V. LoBue and C. Thrasher, "The child affective facial expression (CAFE) set: Validity and reliability from untrained adults," *Frontiers Psychol.*, vol. 5, p. 1532, Jan. 2015.

[41] K. A. Dalrymple, J. Gomez, and B. Duchaine, "The Dartmouth database of children's faces: Acquisition and validation of a new face stimulus set," *PLoS ONE*, vol. 8, no. 11, Nov. 2013, Art. no. e79131.

[42] H. L. Egger, D. S. Pine, E. Nelson, E. Leibenluft, M. Ernst, K. E. Towbin, and A. Angold, "The NIMH child emotional faces picture set (NIMH-ChEFS): A new set of children's facial emotion stimuli," *Int. J. Methods Psychiatric Res.*, vol. 20, no. 3, pp. 145–156, Sep. 2011.

[43] B. Nojavanasghari, T. Baltrušaitis, C. E. Hughes, and L.-P. Morency, "EmoReact: A multimodal approach and dataset for recognizing emotional responses in children," in *Proc. 18th ACM Int. Conf. Multimodal Interact.*, Oct. 2016, pp. 137–144.

[44] M. Farhadloo and E. Rolland, "Fundamentals of sentiment analysis and its applications," in *Sentiment Analysis and Ontology Engineering*. Cham, Switzerland: Springer, 2016, pp. 1–24.

[45] S. Al-Azani and E.-S.-M. El-Alfy, "Early and late fusion of emojis and text to enhance opinion mining," *IEEE Access*, vol. 9, pp. 121031–121045, 2021.

[46] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.

[47] E. V. Paz, M. Martino, M. Hyland, M. Corletto, C. Puga, M. Peralta, N. Deltetto, T. Kuhlmann, D. Cavalié, M. Leist, B. Duarte, and I. Lascombes, "Sentiment analysis in children with neurodevelopmental disorders in an ingroup/outgroup setting," *J. Autism Develop. Disorders*, vol. 50, no. 1, pp. 162–170, Jan. 2020.

[48] P. P. Filntisis, N. Efthymiou, P. Koutras, G. Potamianos, and P. Maragos, "Fusing body posture with facial expressions for joint recognition of affect in child–robot interaction," *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 4011–4018, Oct. 2019.

[49] S. Boughorbel, J. Breebaart, F. Bruekers, I. Flinsenberg, and W. Ten Kate, "Child-activity recognition from multi-sensor data," in *Proc. 7th Int. Conf. Methods Techn. Behav. Res.*, 2010, pp. 1–3.

[50] Y. Nam and J. W. Park, "Child activity recognition based on cooperative fusion model of a triaxial accelerometer and a barometric pressure sensor," *IEEE J. Biomed. health Informat.*, vol. 17, no. 2, pp. 420–426, Jan. 2013.

[51] A. García-Dominguez, C. E. Galván-Tejada, L. A. Zanella-Calzada, H. Gamboa-Rosales, J. I. Galván-Tejada, J. M. Celaya-Padilla, H. Luna-García, and R. Magallanes-Quintanar, "Feature selection using genetic algorithms for the generation of a recognition and classification of children activities model using environmental sound," *Mobile Inf. Syst.*, vol. 2020, pp. 1–12, Feb. 2020.

[52] A. Hosseini, S. Fazeli, E. V. Vliet, L. Valencia, R. Habre, M. Sarrafzadeh, and A. Bui, "Children activity recognition: Challenges and strategies," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 4331–4334.

[53] E. Kokkoni, E. Mavroudi, A. Zehfroosh, J. C. Galloway, R. Vidal, J. Heinz, and H. G. Tanner, "GEARing smart environments for pediatric motor rehabilitation," *J. Neuroeng. Rehabil.*, vol. 17, no. 1, pp. 1–15, Dec. 2020.

[54] C. Pacheco, E. Mavroudi, E. Kokkoni, H. G. Tanner, and R. Vidal, "A detection-based approach to multiview action classification in infants," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 6112–6119.

[55] N. Efthymiou, P. Koutras, P. P. Filntisis, G. Potamianos, and P. Maragos, "Multi-view fusion for action recognition in child-robot interaction," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 455–459.

[56] J. N. Constantino and C. P. Gruber, "Social responsiveness scale: SRS-2," Western Psychol. Services Torrance, CA, USA, Tech. Rep., 2012.

[57] R. L. Spitzer, K. K. Md, and J. B. W. Williams, "Diagnostic and statistical manual of mental disorders," *Amer. Psychiatric Assoc*, vol. 21, 2013.

[58] J. Rehg, G. Abowd, A. Rozga, M. Romero, M. Clements, S. Sclaroff, I. Essa, O. Ousley, Y. Li, C. Kim, and H. Rao, "Decoding children's social behavior," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, Jun. 2013, pp. 3414–3421.

[59] O. Ousley, R. Arriaga, G. Abowd, and M. Morrier, "Rapid assessment of social-communicative abilities in infants at risk for autism," Center Behav. Imag., Georgia Tech., Atlanta, GA, USA, Tech. Rep. CBI-100, 2012.

[60] N. Marrus, S. Kennon-McGill, B. Harris, Y. Zhang, A. L. Glowinski, and J. N. Constantino, "Use of a video scoring anchor for rapid serial assessment of social communication in toddlers," *J. Visualized Exp.*, vol. 133, p. e57041, Mar. 2018.

[61] J. Schneider, D. Börner, P. van Rosmalen, and M. Specht, "Augmenting the senses: A review on sensor-based learning support," *Sensors*, vol. 15, no. 2, pp. 4097–4133, Feb. 2015.

[62] B. Schneider and P. Blikstein, "Unraveling Students' interaction around a tangible interface using multimodal learning analytics," *J. Educ. Data Mining*, vol. 7, no. 3, pp. 89–116, 2015.

[63] H. J. Pijeira-Díaz, H. Drachsler, S. Järvelä, and P. A. Kirschner, "Investigating collaborative learning success with physiological coupling indices based on electrodermal activity," in *Proc. 6th Int. Conf. Learn. Anal. Knowl.*, 2016, pp. 64–73.

[64] S. Oviatt, A. Cohen, N. Weibel, K. Hang, and K. Thompson, "Multimodal learning analytics data resources: Description of math data corpus and coded documents," in *Proc. 3rd Int. Data-Driven Grand Challenge Workshop Multimodal Learn. Analytics*, vol. 414, 2014.

[65] A. M. Arthur, R. Lunsford, M. Wesson, and S. Oviatt, "Prototyping novel collaborative multimodal systems: Simulation, data collection and analysis tools for the next decade," in *Proc. 8th Int. Conf. Multimodal Interfaces (ICMI)*, 2006, pp. 209–216.

[66] A. Sriramulu, J. Lin, and S. Oviatt, "Dynamic adaptive gesturing predicts domain expertise in mathematics," in *Proc. Int. Conf. Multimodal Interact.*, 2019, pp. 105–113.

[67] R. Abdu, G. van Helden, R. Alberto, and A. Bakker, "Multimodal dialogue in small-group mathematics learning," *Learn., Culture Social Interact.*, vol. 29, Jun. 2021, Art. no. 100491.

[68] F. Xu, L. Wu, K. Thai, C. Hsu, W. Wang, and R. Tong, "MUTLA: A large-scale dataset for multimodal teaching and learning analytics," 2019, *arXiv:1910.06078*.

[69] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "ECG pattern analysis for emotion detection," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 102–115, Jan./Mar. 2012.

[70] L. Ye, H. Ferdinando, T. Seppanen, T. Huuki, and E. Alasaarela, "An instance-based physical violence detection algorithm for school bullying prevention," in *Proc. Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Aug. 2015, pp. 1384–1388.

[71] L. R. Barhight, J. A. Hubbard, and C. T. Hyde, "Children's physiological and emotional reactions to witnessing bullying predict bystander intervention," *Child Develop.*, vol. 84, no. 1, pp. 375–390, Jan. 2013.

[72] L. Ye, P. Wang, L. Wang, H. Ferdinando, T. Seppänen, and E. Alasaarela, "A combined motion-audio school bullying detection algorithm," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 32, no. 12, Dec. 2018, Art. no. 1850046.

[73] J. Wang, R. J. Iannotti, and T. R. Nansel, "School bullying among adolescents in the united states: Physical, verbal, relational, and cyber," *J. Adolescent Health*, vol. 45, no. 4, pp. 368–375, Oct. 2009.

[74] T. Vaillancourt, R. Faris, and F. Mishna, "Cyberbullying in children and youth: Implications for health and clinical practice," *Can. J. Psychiatry*, vol. 62, no. 6, pp. 368–373, Jun. 2017.

[75] L. Ye, J. Shi, H. Ferdinando, T. Seppänen, and E. Alasaarela, "A multi-sensor school violence detecting method based on improved relief-F and D-S algorithms," *Mobile Netw. Appl.*, vol. 25, no. 5, pp. 1655–1662, Oct. 2020.

[76] L. Ye, L. Wang, H. Ferdinando, T. Seppänen, and E. Alasaarela, "A video-based DT–SVM school violence detecting algorithm," *Sensors*, vol. 20, no. 7, p. 2018, Apr. 2020.

[77] S. Brown and F. Timmins, "An exploration of nurses' knowledge of, and attitudes towards, pain recognition and management in neonates," *J. Neonatal Nursing*, vol. 11, no. 2, pp. 65–71, Sep. 2005.

[78] R. P. Riddell and N. Racine, "Assessing pain in infancy: The caregiver context," *Pain Res. Manage.*, vol. 14, no. 1, pp. 27–32, 2009.

[79] S. Brahnam, C.-F. Chuang, F. Y. Shih, and M. R. Slack, "Machine recognition and representation of neonatal facial displays of acute pain," *Artif. Intell. Med.*, vol. 36, no. 3, pp. 211–222, Mar. 2006.

[80] D. Harrison, M. Sampson, J. Reszel, K. Abdulla, N. Barrowman, J. Cumber, A. Fuller, C. Li, S. Nicholls, and C. M. Pound, "Too many crying babies: A systematic review of pain management practices during immunizations on Youtube," *BMC Pediatrics*, vol. 14, no. 1, pp. 1–8, Dec. 2014.

[81] Y. Sun, C. Shan, T. Tan, X. Long, A. Pourtaherian, S. Zinger, and P. H. N. de With, "Video-based discomfort detection for infants," *Mach. Vis. Appl.*, vol. 30, no. 5, pp. 933–944, Jul. 2019.

[82] C. Li, A. Pourtaherian, L. van Onzenoort, W. E. T. A. Ten, and P. H. N. de With, "Infant facial expression analysis: Towards a real-time video monitoring system using R-CNN and HMM," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 5, pp. 1429–1440, May 2021.

[83] H. A. Patil, "'Cry baby': Using spectrographic analysis to assess neonatal health status from an infant's cry," in *Advances in Speech Recognition*. Boston, MA, USA: Springer, 2010, pp. 323–348.

[84] K. Ashwini, P. M. D. R. Vincent, K. Srinivasan, and C.-Y. Chang, "Deep learning assisted neonatal cry classification via support vector machine models," *Frontiers Public Health*, vol. 9, pp. 1–13, Jun. 2021.

[85] D. Widhyanti and D. Juniati, "Classification of baby cry sound using Higuchi's fractal dimension with K-nearest neighbor and support vector machine," *J. Phys., Conf.*, vol. 1747, no. 1, Feb. 2021, Art. no. 012014.

[86] P. Bhasha, T. P. Kumar, K. K. Baseer, and V. Jyothsna, "An IoT-based BLYNK server application for infant monitoring alert system to detect crying and wetness of a baby," in *Proc. Int. Conf. Intell. Smart Comput. Data Anal. (ISCDA)*, vol. 1312. Singapore: Springer, 2021, p. 55.

[87] H. E. Baeck and M. N. Souza, "Study of acoustic features of newborn cries that correlate with the context," in *Proc. Conf. 23rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Oct. 2001, pp. 2174–2177.

[88] Y. Abdulaziz and S. M. S. Ahmad, "Infant cry recognition system: A comparison of system performance based on mel frequency and linear prediction cepstral coefficients," in *Proc. Int. Conf. Inf. Retr. Knowl. Manage. (CAMP)*, Mar. 2010, pp. 260–263.

[89] *Donateacry-Corpus*. Accessed: Sep. 4, 2021. [Online]. Available: https://github.com/gveres/donateacry-corpus

[90] C.-Y. Chang, C.-W. Chang, S. Kathiravan, C. Lin, and S.-T. Chen, "DAG-SVM based infant cry classification system using sequential forward floating feature selection," *Multidimensional Syst. Signal Process.*, vol. 28, no. 3, pp. 961–976, Jul. 2017.

[91] T. N. Maghfira, T. Basaruddin, and A. Krisnadhi, "Infant cry classification using CNN–RNN," *J. Phys., Conf.*, vol. 1528, no. 1, Apr. 2020, Art. no. 012019.

[92] G. Zamzami, G. Ruiz, D. Goldgof, R. Kasturi, Y. Sun, and T. Ashmeade, "Pain assessment in infants: Towards spotting pain expression based on infants' facial strain," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, May 2015, pp. 1–5.

[93] G. Zamzmi, C.-Y. Pai, D. Goldgof, R. Kasturi, T. Ashmeade, and Y. Sun, "An approach for automated multimodal analysis of infants' pain," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 4148–4153.

[94] O. F. Reyes-Galaviz, S. D. Cano-Ortiz, and C. A. Reyes-García, "Evolutionary-neural system to classify infant cry units for pathologies identification in recently born babies," in *Proc. 7th Mex. Int. Conf. Artif. Intell.*, Oct. 2008, pp. 330–335.

[95] J. Chunyan, M. Chen, L. Bin, and Y. Pan, "Infant cry classification with graph convolutional networks," in *Proc. IEEE 6th Int. Conf. Comput. Commun. Syst. (ICCCS)*, Jan. 2021, pp. 322–327.

[96] A. Sharma and D. Malhotra, "Speech recognition based IICC—Intelligent infant cry classifier," in *Proc. 3rd Int. Conf. Smart Syst. Inventive Technol. (ICSSIT)*, Aug. 2020, pp. 992–998.

[97] A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," *Pattern Recognit. Lett.*, vol. 79, pp. 80–105, Aug. 2016.

[98] G. Somanath, M. Rohith, and C. Kambhamettu, "VADANA: A dense dataset for facial image analysis," in *Proc. IEEE Int. Conf. Comput. Vis. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 2175–2182.

[99] K. Ricanek and T. Tesafaye, "MORPH: A longitudinal image database of normal adult age-progression," in *Proc. 7th Int. Conf. Automat. Face Gesture Recognit. (FGR)*, Apr. 2006, pp. 341–345.

[100] R. Kumar, R. C. Singh, and S. Kant, "Dorsal hand vein recognition using very deep learning," in *Macromolecular Symposia*, vol. 397, no. 1. Hoboken, NJ, USA: Wiley, Jun. 2021, Art. no. 2000244.

[101] A. Messaoud and C. Tadj, "A cry-based babies identification system," in *Proc. Int. Conf. Image Signal Process.* Berlin, Germany: Springer, 2010, pp. 192–199.

[102] D. Deb, D. Aggarwal, and A. K. Jain, "Identifying missing children: Face age-progression via deep feature aging," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 10540–10547.

[103] P. Basak, P. De, M. Agarwal, A. Malhotra, M. Vatsa, and R. Singh, "Multimodal biometric recognition for toddlers and pre-school children," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 627–633.

[104] L. Best-Rowden, Y. Hoole, and A. Jain, "Automatic face recognition of newborns, infants, and toddlers: A longitudinal evaluation," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2016, pp. 1–8.

[105] D. Deb, N. Nain, and A. K. Jain, "Longitudinal study of child face recognition," in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 225–232.

[106] S. A. Azani and E.-S. M., "Multimodal age-group recognition for opinion video logs using ensemble of neural networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 4, pp. 1–8, 2019.

[107] S. Safavi, M. Russell, and P. Jančovič, "Identification of age-group from children's speech by computers and humans," in *Proc. 15th Annu. Conf. Int. Speech Commun. Assoc.*, 2014.

[108] D. Hai, C. Chen, R. Yi, S. Gou, B. Yu Su, C. Jiao, and M. Skubic, "Heartbeat detection and rate estimation from ballistocardiograms using the gated recurrent unit network," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 451–454.

[109] S. Al-Azani and E.-S.-M. El-Alfy, "Enhanced video analytics for sentiment analysis based on fusing textual, auditory and visual information," *IEEE Access*, vol. 8, pp. 136843–136857, 2020.

[110] D. Coskun, O. D. Incel, and A. Ozgovde, "Phone position/placement detection using accelerometer: Impact on activity recognition," in *Proc. IEEE 10th Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process. (ISSNIP)*, Apr. 2015, pp. 1–6.

[111] A. Mannini, S. S. Intille, M. Rosenberger, A. M. Sabatini, and W. Haskell, "Activity recognition using a single accelerometer placed at the wrist or ankle," *Med. Sci. Sports Exercise*, vol. 45, no. 11, p. 2193, 2013.

[112] H. Qi, X. Wang, S. S. Iyengar, and K. Chakrabarty, "Multisensor data fusion in distributed sensor networks using mobile agents," in *Proc. 5th Int. Conf. Inf. Fusion*, 2001, pp. 1–16.

[113] C.-C. Yang and Y.-L. Hsu, "A review of accelerometry-based wearable motion detectors for physical activity monitoring," *Sensors*, vol. 10, no. 8, pp. 7772–7788, 2010.

[114] Y. Tomita, H. Fukai, and Y. Mitsukura, "Factor analysis for EEG response to ride quality," *Information*, vol. 14, no. 2, pp. 593–600, 2011.

[115] S.-I. Ito, Y. Mitsukura, and M. Fukumi, "Study on model for predicting the intra-individual difference in left prefrontal pole electroencephalogram variability and its evaluation," *Electron. Commun. Jpn.*, vol. 94, no. 5, pp. 9–16, May 2011.

[116] A. K. Chowdhury, D. Tjondronegoro, V. Chandran, and S. G. Trost, "Physical activity recognition using posterior-adapted class-based fusion of multiaccelerometer data," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 3, pp. 678–685, May 2017.

[117] A. Hosseini, D. Zamanzadeh, L. Valencia, R. Habre, A. A. T. Bui, and M. Sarrafzadeh, "Domain adaptation in children activity recognition," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 1725–1728.

[118] T. Khan, "An intelligent baby monitor with automatic sleeping posture detection and notification," *AI*, vol. 2, no. 2, pp. 290–306, Jun. 2021.

[119] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[120] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proc. Esann*, vol. 3, 2013, p. 3.

[121] A. Hosseini, C. Buonocore, S. Hashemzadeh, H. Hojaiji, H. Kalantarian, C. Sideris, A. Bui, C. King, and M. Sarrafzadeh, "Feasibility of a secure wireless sensing smartwatch application for the self-management of pediatric asthma," *Sensors*, vol. 17, no. 8, p. 1780, Aug. 2017.

[122] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1725–1732.

[123] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2014, pp. 740–755.

[124] P. Rosenbeck, B. Baungaard, C. Jacobsen, and D.-J. Barry, "The design and efficient recording of a 3000 speaker scandinavian telephone speech database: Rafael. 0," in *Proc. 3rd Int. Conf. Spoken Lang. Process.*, 1994.

[125] P. Ekman, W. V. Friesen, and J. C. Hager, *Facial Action Coding System: Facial Action Coding System: Manual: CD-ROM.* Research Nexus, 2002.

[126] P. A. McGrath, "Multidimensional pain assessment in children," *Adv. Pain Res. Therapy*, vol. 18, pp. 387–393, Jan. 1985.

[127] T. Voepel-Lewis, J. R. Shayevitz, and S. Malviya, "The FLACC: A behavioral scale for scoring postoperative pain in young children," *Pediatr Nurs*, vol. 23, no. 3, pp. 293–297, 1997.

[128] P. Hummel, M. Puchalski, S. Creech, and M. G. Weiss, "N-PASS: Neonatal pain, agitation and sedation scale–reliability and validity," in *Proc. Pediatric Academic Societies Annu. Meeting*, 2003.

[129] M. Buhrmester, T. Kwang, and S. D. Gosling, "Amazon's mechanical turk: A new source of inexpensive, yet high-quality data?" in *Methodological Issues and Strategies in Clinical Research.* Washington, DC, USA: American Psychological Association, 2016, pp. 133–139.

[130] S. Sen, M. E. Giesel, R. Gold, B. Hillmann, M. Lesicko, S. Naden, J. Russell, Z. Wang, and B. Hecht, "Turkers, scholars 'arafat' and 'peace' cultural communities and algorithmic gold standards," in *Proc. 18th ACM Conf. Comput. Supported Cooperat. Work Social Comput.*, 2015, pp. 826–838.

[131] M. Geva, Y. Goldberg, and J. Berant, "Are we modeling the task or the annotator? An investigation of annotator bias in natural language understanding datasets," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, Hong Kong, Nov. 2019, pp. 1–6.

[132] L. B. Leonard, *Children With Specific Language Impairment.* Cambridge, MA, USA: MIT Press, 2014.

[133] D. V. Bishop, *Uncommon Understanding (Classic Edition): Development and Disorders of Language Comprehension in Children.* London, U.K.: Psychology Press, 2014.

[134] J. B. Tomblin, N. L. Records, P. Buckwalter, X. Zhang, E. Smith, and M. O'Brien, "Prevalence of specific language impairment in kindergarten children," *J. Speech, Lang., Hearing Res.*, vol. 40, no. 6, pp. 1245–1260, Dec. 1997.

[135] M. J. Maenner, K. A. Shaw, and J. Baio, "Prevalence of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2016," *MMWR Surveill. Summaries*, vol. 69, no. 4, p. 1, 2020.

[136] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, pp. 1–54, Dec. 2019.

[137] E.-S.-M. El-Alfy and S. Al-Azani, "Empirical study on imbalanced learning of Arabic sentiment polarity with neural word embedding," *J. Intell. Fuzzy Syst.*, vol. 38, no. 5, pp. 6211–6222, May 2020.

[138] S. Al-Azani and E.-S.-M. El-Alfy, "Imbalanced sentiment polarity detection using emoji-based features and bagging ensemble," in *Proc. 1st Int. Conf. Comput. Appl. Inf. Secur. (ICCAIS)*, Apr. 2018, pp. 1–5.

[139] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Classification with class imbalance problem: A review," *Int. J. Adv. Soft Comput. its Appl.*, vol. 7, no. 3, pp. 1–30, 2015.

[140] D. Bannach, O. Amft, and P. Lukowicz, "Automatic event-based synchronization of multimodal data streams from wearable and ambient sensors," in *Proc. Eur. Conf. Smart Sens. Context.* Berlin, Germany: Springer, 2009, pp. 135–148.

[141] J. Six and M. Leman, "Synchronizing multimodal recordings using audio-to-audio alignment," *J. Multimodal User Interfaces*, vol. 9, no. 3, pp. 223–229, Sep. 2015.

[142] M. A. Alsheikh, A. Selim, D. Niyato, L. Doyle, S. Lin, and H.-P. Tan, "Deep activity recognition models with triaxial accelerometers," in *Proc. Workshops 30th AAAI Conf. Artif. Intell.*, 2016, pp. 1–6.

[143] S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. Abdelzaher, "DeepSense: A unified deep learning framework for time-series mobile sensing data processing," in *Proc. 26th Int. Conf. World Wide Web*, Apr. 2017, pp. 351–360.

[144] M. L. McHugh, "Interrater reliability: The kappa statistic," *Biochem. Med.*, vol. 22, no. 3, pp. 276–282, 2012.

[145] J. T. Colorado and J. Eberle, "Student demographics and success in online learning environments," Emporia State Res. Stud., Tech. Rep., 2010, pp. 4–10, vol. 46.

[146] K. Bousmalis, L.-P. Morency, and M. Pantic, "Modeling hidden dynamics of multimodal cues for spontaneous agreement and disagreement recognition," in *Proc. Face Gesture*, Mar. 2011, pp. 746–752.

[147] K. Krippendorff, "Estimating the reliability, systematic error and random error of interval data," *Educ. Psychol. Meas.*, vol. 30, no. 1, pp. 61–70, 1970.

SADAM AL-AZANI received the B.Sc. degree (Hons.) in computer science from Thamar University, Dhamar, Yemen, in 2004, and the M.Sc. and Ph.D. degrees in computer science from the King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, in 2014 and 2019, respectively. He was a full-time Teaching Assistant with Thamar University, for a period of five years. He is currently a Research Scientist III with the SDAIA-KFUPM Joint Research Center for Artificial Intelligence (JRC-AI), KFUPM. His research interests include artificial intelligence, natural language processing, video analytics, machine/deep learning, social network analysis, and pattern recognition. He has served as a technical program committee member and a reviewer for a number of international conferences and journals.

SADIQ M. SAIT (Senior Member, IEEE) was born in Bengaluru. He received the bachelor's degree in electronics engineering from Bengaluru University, in 1981, and the master's and Ph.D. degrees in electrical engineering from the King Fahd University of Petroleum and Minerals (KFUPM), in 1983 and 1987, respectively. He is currently a Professor of computer engineering and the Director of the Center for Communications and IT Research, Research Institute, KFUPM. He has authored over 300 research articles, contributed chapters to technical books, and lectured in over 25 countries. He is also the principle author of two books. He received the Best Electronic Engineer Award from the Indian Institute of Electrical Engineers, Bengaluru, in 1981.

KHALED A. AL-UTAIBI (Member, IEEE) was born in Riyadh, Saudi Arabia, in 1973. He received the B.S., M.S., and Ph.D. degrees in computer engineering from the King Fahd University of Petroleum and Engineering, Dhahran, Saudi Arabia, in 1997, 2002, and 2019, respectively. He worked as a Lecturer with the Department of Computer Engineering, King Fahd University of Petroleum and Minerals, from 2002 to 2004. In 2004, he joined the Department of Computer Engineering, University of Ha'il, Ha'il, Saudi Arabia, as a Lecturer, where he became an Assistant Professor, in 2020. His current research interests include image cryptography, optimization algorithms, machine learning, and artificial intelligence.

● ● ●