This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3128787, IEEE Transactions on Affective Computing

Automatic emotion recognition in clinical scenario: a systematic review of methods

Lucia Pepa *†, Luca Spalazzi †, Marianna Capecci *, and Maria Gabriella Ceravolo *
*Dept. of Experimental and Clinical Medicine, Università Politecnica delle Marche, Ancona, Italy
†Dept. of Information Engineering, Università Politecnica delle Marche, Ancona, Italy
{l.pepa | l.spalazzi | m.capecci | m.g.ceravolo}@staff.univpm.it

Abstract—BACKGROUND - Automatic emotion recognition has powerful and interesting opportunities in the clinical field, but several critical aspects are still open, such as heterogeneity of methodologies or technologies tested mainly on healthy people. This systematic review aims to survey automatic emotion recognition systems applied in real clinical contexts (i.e. on a population of people with a pathology).

METHODS - The literature review was conducted on the following scientific databases: IEEE *Xplore*®, ScienceDirect®, Scopus®, PubMed®, ACM®. Inclusion criteria were the presence of an automatic emotion recognition algorithm and the enrollment of at least 2 patients in the experimental protocol. The review process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Moreover, the works were analysed according to a reference model in the form of a class diagram, to highlight the most important clinical and technical aspects and relationships among them.

RESULTS - 52 scientific papers passed the inclusion criteria. Based on our findings, most clinical applications involved neuro-developmental, neurological and psychiatric disorders with the aims of diagnosing, monitoring, or treating emotional symptoms. The study design seems to be mostly related to the aim of the study (it is generally observational for monitoring and diagnosis, interventional for treatment), the most adopted signals are video and audio, and supervised shallow learning emerged as most used approach for emotion recognition algorithm.

DISCUSSION - Tiny samples, absence of a control group and of tests in real-life conditions emerged as important clinical limitations. Under a technical point of view, a great heterogeneity of performance metrics, datasets and algorithms challenges the comparability, robustness, reliability and reproducibility of results. Suggested guidelines are identified and discussed to help scientific community to overcome limitations and provide direction for future works.

Index Terms—Emotion recognition, Clinical applications, Neurological disorders, Psychiatric disorders, Machine Learning, Artificial Intelligence.

1 Introduction

Automatic recognition of human emotions opens up powerful and interesting opportunities in many application fields [1], as the clinical scenario. It can be applied to improve the user experience in assistive robotics for frail people [2] and serious games for motor rehabilitation [3]. Furthermore, it may be useful to objectively study emotional symptoms in neurological or psychiatric disorders [4]–[6]. For these reasons, automatic emotion recognition is rapidly evolving and attracting the attention of researchers [7], nonetheless, several critical aspects are still open. 1) First of all, the recognition of emotions has been proposed in a great variety of clinical scenarios, but often these works are still far from their application in real cases [5], [8]. The presented results, in fact, often lack concrete evidence of their applicability to real cases [5], [6], [8]. For instance, sometimes the proposed approaches have been tested on healthy subjects, making these experiments not very relevant for a clinical application [5], [8]. 2) The followed methodologies are very heterogeneous and experimental standards widely accepted by the scientific community have not yet emerged [6], [8]. Therefore, the presented results are often difficult to reproduce and to compare with each other [7], [9]. 3) Finally, the technologies (sensors and signals) and the algorithms adopted are very heterogeneous [4], [5]. Hence, the current literature shows difficulties in proposing robust and appropriate approaches for automatic emotion recognition in specific clinical scenarios. On the other hand, the need for clinicians to have at their disposal tools for recognizing emotions is becoming increasingly felt [7]. Therefore, the current paper aims at providing a systematic review of automatic emotion recognition approaches applied in real clinical context.

This is not the first review dedicated to the automatic recognition of emotions, however previous reviews are focused exclusively on technical aspects, such as sensors, signals [10], [11] or algorithms [12], and do not take into account any specific field of application, in particular there are no recent reviews focusing on clinical scenarios. On the other hand, surveys dealing with technologies that can help in assessing or treating emotional symptoms [4]–[6] are neither focused on automatic emotion recognition, nor on datasets of pathological subjects. Hence, the literature lacks a comprehensive survey dealing with both clinical and technical aspects of automatic emotion recognition with respect to pathologies.

As a consequence, this systematic review aims to answer the following research questions:

- **RQ1** What are the pathologies and clinical scenarios where automatic emotion recognition can be applied and for which clinical purposes?
- **RQ2** Given a certain pathology and the related clinical

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3128787, IEEE Transactions on Affective Computing

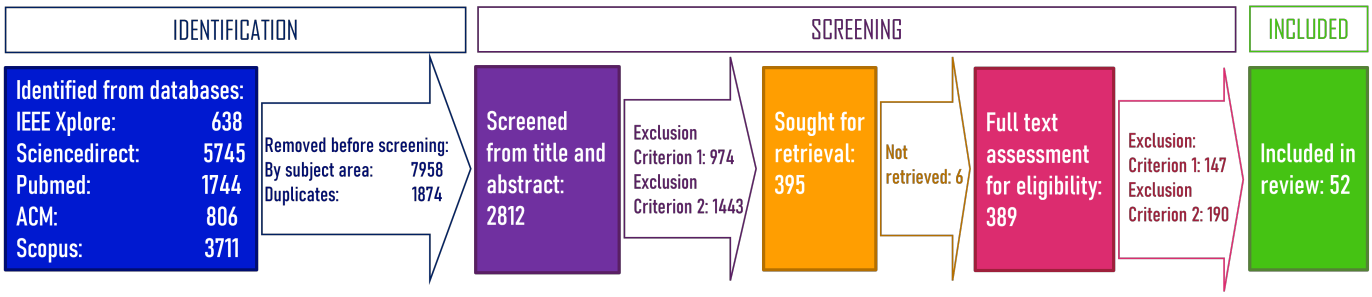


Fig. 1: Overview of the systematic review.

purpose, what are the relevant aspects when designing an experimental protocol to test an emotion recognition system on a pathological population?

RQ3 Given a certain pathology and the related clinical purpose, what is the most appropriate emotion model to adopt?

RQ4 What sensors and signals can be adopted to deal with a given pathology?

RQ5 What kind of algorithms can be used for a given pathology and how can they be evaluated?

RQ6 What are the problems still open and future directions?

Such questions suggested the adoption of the *reference model* depicted in Figure 2 as a class diagram. This model presents the categories through which the various reviewed works were classified. It also highlights the relations between the various categories that emerged during the systematic review. Finally, it represents a map to navigate within this systematic review: i.e. given a category or a relationship, the map shows which graphics and tables should be consulted in order to find the related results. Results obtained from this systematic review show that automatic emotion recognition was mainly applied in neurological and psychiatric disorders with the aim of monitoring, diagnosing or treating emotional symptoms. These clinical aims are correlated with several categories of the reference model, indicating that the final clinical purpose to be addressed by the emotion recognition system is important to guide the design of the system itself as well as the research study to evaluate its effectiveness and reliability. The most important limitations concern study design (tiny samples, absence of a control group and tests under real-life conditions), lack of robustness in training, validating, and testing algorithms.

This review is structured as follows. Section 2 describes the eligibility criteria for papers to be included in the survey and defines the reference model to analyze the selected studies. Section 3 reports the results of this analysis. Section 4 discusses the results in terms of answers to RQs 1-5. Guidelines and future directions are drawn in Section 5, in answer to RQ6. Conclusions are given in Section 6.

2 Materials and Methods

2.1 Literature search methodology

The literature review was conducted on the following scientific databases: IEEE *Xplore*®, ScienceDirect®, Scopus®, PubMed®, ACM®. The review process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA[1]) guidelines [13]. The following query was used:

((“emotion recognition” OR “affective computing”) AND (“healthcare” OR “disease” OR “pathology” OR “disorder” OR “clinical” OR “rehabilitation”)).

Two review authors (LP and MC) independently screened and assessed records for eligibility. We resolved disagreements on study eligibility through consensus, and, when necessary, we met with a third review author not involved in the particular assessment (LS) for discussion. We retrieved full-text articles of potentially relevant reports and linked together multiple reports of the same study. We corresponded with investigators, when appropriate, to clarify study eligibility or to seek further information, such as missing data.

2.2 Inclusion and exclusion criteria

Inclusion criteria to select papers were the following:

- Inclusion criterion 1: the methodology or the aim of the work include an automatic emotion recognition system;
- Inclusion criterion 2: the investigated sample included at least 2 patients.

Conversely, research works that fall into at least one of the following *exclusion criteria* were not included:

- Exclusion criterion 1: there is not an automatic emotion recognition algorithm.
- Exclusion criterion 2: the automatic emotion recognition algorithm is not tested on patients.

Finally, this review only considers articles published in peer-reviewed scientific journals before August 2020 and written in English.

The queries returned a total of 12644 scientific papers from the five databases. 7958 were excluded by automation tools and 1874 out of the remaining were duplicates. Automation tools were used in Scopus and ScienceDirect databases, by restricting the returned papers to “Computer Science” and “Engineering” subject areas. Indeed, a first screening revealed that excluding these areas all the papers returned by search keywords seemed to fall into exclusion

1. http://www.prisma-statement.org/PRISMAStatement/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3128787, IEEE Transactions on Affective Computing

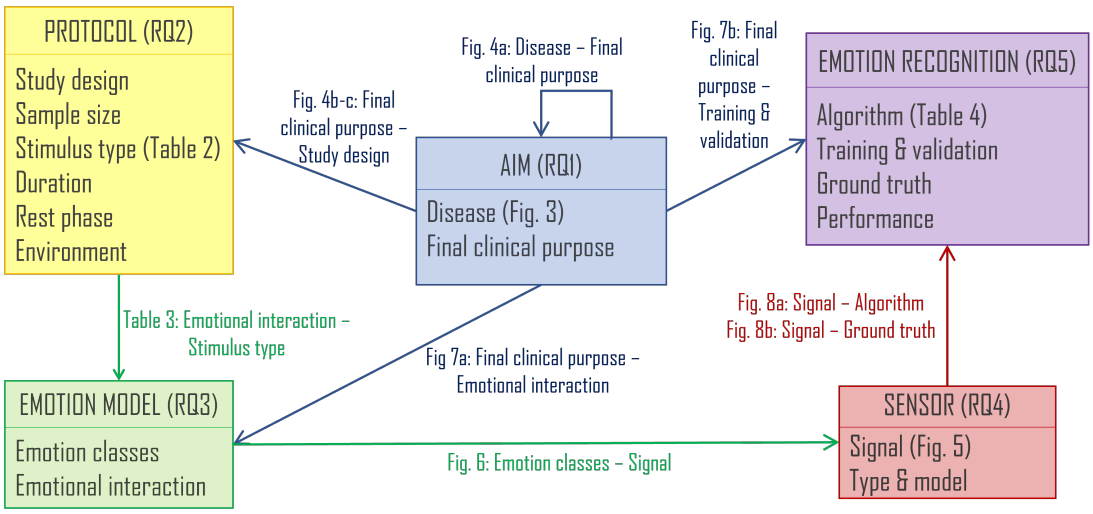


Fig. 2: Reference model for the literature review.

criterion 1: they were clinical study where "emotion recognition" was intended as the human ability to recognize emotions. Hence, a total of 2812 papers were screened from their title and abstract: 2417 of them were excluded, 6 were not retrieved, resulting in 389 manuscripts assessed for eligibility. Finally, 52 works passed the inclusion criteria. Fig. 1 shows the flow diagram of the screening procedure.

LP and MC independently extracted trial data in included studies, and LS arbitrated any conflicts not due to extractor error. We collated multiple publications for the same trial and used the most complete report (i.e. the one with outcomes most relevant to the review or with the most recent outcomes) as the primary reference.

LP and MC performed independently the assessment of risk of bias in the included studies, recording in detail and discussing the study design, but no formal or quantitative assessment of the methodological quality was performed because the studies were not controlled or controlled but not randomized or blinded. In other words, all studies can be classified of 'high risk' according to the Cochrane's 'Risk of bias' tool [14].

2.3 Reference model

Data extraction from the 52 selected works was performed according to the *reference model* reported in Figure 2 in the form of a class diagram. The model classes were designed to allow a structured response to research questions, and to deeply examine the works both from a clinical and a technological perspective. The *clinical perspective* includes the *aim* of the work, the experimental *protocol* adopted and the related *emotion model*. The *technical perspective* includes the *sensors* used, and the *emotion recognition algorithms*. Each concept concerns a RQ and is in turn described by a set of attributes that are considered important for a deep and complete analysis of that particular aspect of the emotion recognition system. The attributes of each class are the variables extracted from papers: *disease* and *final clinical purpose* for the "Aim" of the work; *study design*, *sample size*, *stimulus*, *duration*, *rest phase*, and *environment* for the experimental "Protocol"; *emotion classes* and *emotional interaction* for the "Emotion model"; *signal* and *type & model* for the adopted "Sensor"; *algorithm*, *training & validation*, *ground truth*, and *performance* for the "Emotion recognition" technique. Correlations between concept attributes that were found as a result of this systematic review are expressed by associations in Figure 2. The labels of such associations detail what attributes are correlated and what Figure or Table shows the found correlation. The direction of associations indicates the topic that influences (start) and the topic being influenced (end). When a concept attribute needed a Figure or Table for a more detailed and structured comprehension, this information is reported aside the concept in round brackets, with the related Figure or Table. Table 1 lists all the acronyms adopted in Figures and Tables as well as their meaning, in order to facilitate the reader.

In particular, it was found that the disease influences the final clinical purpose, i.e. for a certain category of pathologies, selected works share the clinical need addressed (Fig. 4a). The final clinical purpose generally guides and influences the choice of the study design (Fig. 4b-4c), the kind of emotional interaction for participants (Fig. 7a), and the training and validation phase of the emotion recognition algorithm (Fig. 7b). To induce a certain emotional interaction, the stimulus type should be chosen appropriately (Table 3). The detection of a high number of emotion classes is mostly accomplished by means of video signal (Fig. 6). Finally, some correlations were found between the adopted signal and the kind of algorithm (Fig. 8a) or labeling methodology (Fig. 8b). These associations that emerged from the surveyed literature will be deeply analysed in Sections 3 and 4.

Tables 5-6 summarize the results obtained by applying the reference model to the selected papers. In particular, Table 5 reports the results related to the clinical aspects, i.e. *Aim*, *Emotion Model* and *Protocol*. Instead, Table 6 reports the results related to purely technological aspects, i.e. *Sensor* and *Emotion Recognition Algorithm*.

3 Results

In this section, each concept of the reference model, its related attributes and possible values are described in-depth

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3128787, IEEE Transactions on Affective Computing

TABLE 1: Acronyms adopted in text and tables.

Acronym	Definition
ADHD	Attention Deficit/Hyperattention Disorder
AS	Asperger Syndrome
ASD	Autism Spectrum Disored
BD	Bipolar Disorder
BN	Bayesian Network
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electrencephalogram
EMG	Electromyogram
ES	Experimental Subjects
HC	Healthy controls
HR	Heart Rate
ICG	Impedence Cardiogram
KNN	K-Nearest Neighbour
LDA	Linear Discriminant Analysis
LSTM	Long-Short Term Memory
MLP	Multilayer Perceptron
NN	Neural Network
PCG	Phono Cardiogram
PD	Parkinson's Disease
PPG	Photoplethysmogram
NB	Naive Bayes
RF	Random Forest
RSP	Respiration
SLI	Speech Language Impairment
ST	Skin Temperature
SVM	Support Vector Machines
UD	Unipolar Disorder

as well as a summary of the contribution of selected works for each topic. In particular, subsection 3.1 presents results related to RQ1, subsection 3.2 to RQ2, subsection 3.3 to RQ3, subsection 3.4 to RQ4, and subsection 3.5 to RQ5.

3.1 RQ1: Aim

The research aim is described in terms of the investigated **disease** and the clinical need that the proposed system is going to address, thus named **final clinical purpose**.

3.1.1 Disease

The diseases that attracted the attention of automatic emotion recognition methods belong mainly to the neurodevelopmental, neurological or psychiatric medical fields except for 1 work addressing automatic emotion recognition in **metabolic syndrome** [15]. In details, automatic emotion recognition was applied in **Autism Spectrum Disorder** (ASD) [16]–[40], **Parkinson's Disease** (PD) [41]–[45], **apathy** [46], **stroke** [47]–[49], **cognitive impairment** and **dementia** [50], [51], **consciousness disorder** [52], **schizophrenia** [53], **bipolar and unipolar disorder** [54]–[65], **depressive disorders** [63], [64], **obsessive compulsive** [65], **acrophobia** [66]. Fig. 3 details the number of selected works for each pathology.

3.1.2 Final clinical purpose

For the aforementioned diseases, the automatic emotion recognition was experimented or applied to achieve the following final clinical purposes:

- *monitoring*: keeping under observation the patient emotional states [15]–[24], [35]–[37], [41]–[44], [47], [48], [52], [63]–[65];

- *diagnosis* or *differential diagnosis*: the identification of the nature of an emotion problem (i.e. a diagnostic tool) or the distinguishing of a particular disease or condition from others that present similar clinical features based on the recognition of emotional phenomena or of mood and behavior pathology (i.e. a differential diagnostic tool) [23]–[27], [45], [46], [49], [53]–[62];

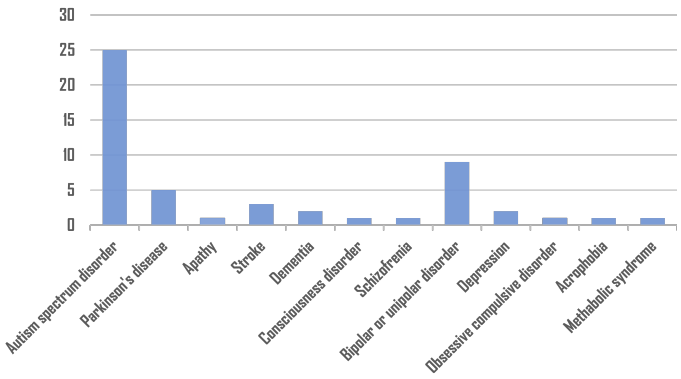


Fig. 3: Number of works with respect to disease.

- *treatment*: the emotion recognition algorithm is used in order to build or study the effect of a system or an architecture able to treat emotion or behaviour disorders. [28]–[34], [38]–[40], [50], [51], [66].

The term "final" highlights that reviewed works were categorized on the basis of the application scenario once system development is finished, even if currently the research is at a prior stage, such as the algorithm validation. Fig. 4a shows the distribution of works with respect to final clinical purpose and disease category.

3.2 RQ2: Experimental protocol

Designing an experimental protocol to arise specific emotions is difficult and involves many variables, given the inner and subjective nature of emotions [67]. These difficulties are exacerbated in a clinical scenario, since patients may present altered emotion recognition and emotional interaction.

3.2.1 Study design

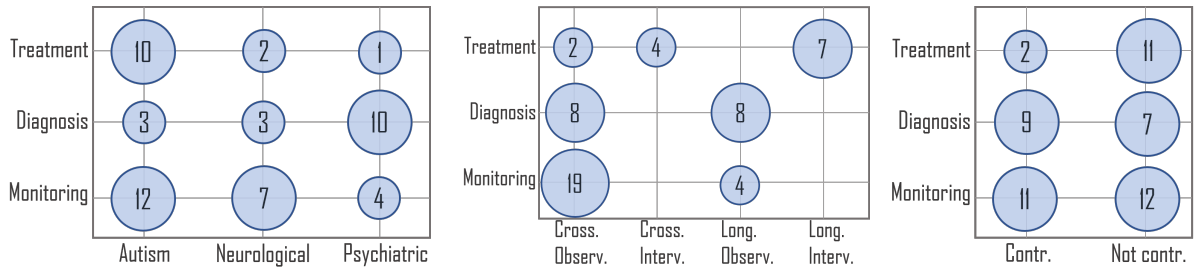
The following study designs emerged from the selected literature:

- *Observational or interventional*. A study is interventional when it is specifically designed to evaluate the impacts of treatment on disease [28]–[31], [33], [38], [39], [50], [51], [66]. Otherwise it is observational [15]–[27], [34]–[37], [40]–[49], [52]–[65].

- *Cross-sectional or longitudinal*. A study is longitudinal when multiple assessments are made prospectively over time to assess disease progression or improvement following therapy [17], [29], [30], [33], [38], [39], [47], [51], [54]–[61], [63], [64], [66]. Otherwise it is cross-sectional [15], [16], [18]–[28], [31], [32], [34]–[37], [40]–[46], [48]–[50], [52], [53], [62], [65].

- *Controlled or not controlled*. A study is controlled if two groups are used for comparison purpose [15], [18], [19], [22], [25]–[27], [33], [37], [41], [42], [44]–[46], [48], [49], [51]–[53], [57], [62], [63]. Otherwise it is not controlled [16], [17], [20],

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3128787, IEEE Transactions on Affective Computing



(a) (b) (c)

Fig. 4: Distribution of works by clinical purpose and disease (a) or clinical purpose and study design (b-c). Study design abbreviations: Cross-sectional (Cross.), Longitudinal (Long.), Observational (Observ.), Interventional (Interv.), Controlled (Contr.), Not controlled (Not contr.)

[21], [23], [24], [28]–[32], [34]–[36], [38]–[40], [43], [47], [50], [54]–[56], [58]–[61], [64]–[66].

Fig. 4b shows the relationship between the study design and the final clinical purpose, as emerged from the conducted review. 33 out of 52 surveyed works showed a cross-sectional design, while 19 a longitudinal one. Among the 33 cross-sectional studies, the final clinical purpose was monitoring in 19 cases [15], [16], [18]–[24], [35]–[37], [41]–[44], [48], [52], [65], diagnosis or differential diagnosis in 8 [25]–[27], [45], [46], [49], [53], [62], and treatment in 6 [28], [31], [32], [34], [40], [50]. All cross-sectional studies were observational, except for 4 interventional studies presenting systems to treat behavior or emotional disorders in subjects with ASD [28], [31], [32] and in frail elderly [50]. Among the 19 longitudinal studies, 4 were applied to monitor evolution of emotional symptoms in ASD [17], stroke [47] or psychiatric disorders [63], [64], 8 to diagnose the presence of dangerous mood states in bipolar disorder [54]–[61], while 7 studies were aimed at evaluating a treatment system [29], [30], [33], [38], [39], [51], [66].

Finally, only 22 out of 52 studies (42%) were controlled with respect to healthy subjects or in other conditions [15], [18], [19], [22], [25]–[27], [33], [37], [41], [42], [44]–[46], [48], [49], [51]–[53], [57], [62], [63] (Fig. 4c).

3.2.2 Sample size

The overall group of selected studies were pilot studies in which no sample size definition was calculated in advance. Indeed, most of the surveyed works (63%) enrolled less than 20 participants [16]–[19], [21], [24], [28]–[31], [35], [36], [38]–[40], [42], [43], [47], [48], [51], [52], [54]–[56], [64]–[66], just 8 works (15%) enrolled more than 50 participants [22], [26], [27], [49], [50], [57], [62], [63].

In emotion recognition applied to clinical scenarios, the importance of the population sample derives from the fact that the experimental subjects, i.e. people with a disease diagnosis, have different psychometric characteristics compared to the healthy population [68], [69]. Methods and algorithms exhibit different characteristics depending on the population and the results may not be applicable or directly transferable from healthy people to diseased people [36], [68]. For this reason, some works aim to propose automated methods for collecting emotional data or sharing a clinical dataset among researchers that can be used to develop and compare algorithms. For example, in the context of the 2018 Audio/Visual Emotion Challenge, Ciftci et al. [57], made available a bipolar disorder dataset with 46 patients and 49 healthy controls, that was in turn used by other research groups [58]–[61]. Kalantarian et al. [36] developed an automatic method to crowd-source facial emotion labeled data and build a benchmark dataset.

3.2.3 Stimulus type

Researchers tested a wide and heterogeneous range of stimuli to induce emotional state. A first differentiation can be made between studies that monitored participants' daily activities [30], [54]–[56] (i.e. emotional stimuli are completely ecological and uncontrolled), and studies that adopted specific stimuli that were chosen or built to induce specific emotions. As regards the latter, the list of all the specific stimuli adopted in the selected papers is reported in Table 2. Furthermore, in order to provide a more comprehensive and structured overview, in this survey all the stimuli are expressed as combinations of basic *"channels"*, namely: audio, still images, moving images, and verbal (here *verbal* means an instruction from or a verbal interaction with other subjects). These channels along with daily activities are used in Table 5 to fill the column *Stimulus channels*. Most of the surveyed works adopted sources not specifically designed for the study of emotions to gather such stimuli. Only some authors used standard databases containing audio and visual stimuli specifically designed for these studies. A remarkable example is presented by the Center for Study of Emotion and Attention at University of Florida[2] that provides various types of stimuli, among them the International Affective Picture System (IAPS) [70] and the International Affective Digital Sounds (IADS) [71]. Another example is represented by the Swiss Center for Affective Sciences at Université de Genève[3] that provides a database of pictures: the Geneva Affective PicturE Database (GAPED) [72].

3.2.4 Duration

Two kinds of information about duration were analysed in the selected papers: the duration of the entire experiment

2. https://csea.phhp.ufl.edu/Media.html
3. https://www.unige.ch/cisa/research/materials-and-online-research/research-material/

TABLE 2: List of stimuli to induce specific emotions.

STIMULUS	CHANNELS	DESCRIPTION	WORKS
Videos	Moving images, or audio and moving images	– Clips from films, Internet, or a combination of IAPS and IADS	[15], [22], [25], [26], [33], [35], [41], [43], [44], [49], [52], [62], [63]
Images	Still images	– Pictured book	[27]
		– Face expressions	[29], [36], [45]
		– Standard databases such as IAPS and GAPED	[21]
Sounds	Audio	– Pre-recorded emotional sentences to reproduce	[27], [42]
		– Songs	[18]
Verbal	Verbal	– Instructions	[45], [53]
		– Interactions between participants	[28]
		– Interactions with therapist or parents	[39], [46], [64]
		– Interactions in learning environments	[38]
Cognitive tasks	A combination of audio, still images, moving images, and verbal	– Sentences to read	[48]
		– Anagrams and Pong task	[16]
		– Stroop test and public speaking	[23]
		– Simulated e-learning environment with various images, mathematical games and tests	[24], [37]
Video games	A combination of audio, moving images, and verbal		[19], [34], [40]
Virtual reality	A combination of audio, moving images, and verbal		[20], [31], [47], [66]
Robot Interaction	A combination of audio, still & moving images, verbal, and, eventually, daily	– NAO robot	[32]
		– Matilda robot	[50]
Hybrid	A combination of several of the above activities	– Interactions with parents, daily activities	[30]
		– Robot interaction, cognitive tasks	[51]
		– Interactions with therapist, cognitive tasks	[17], [57]–[61]
Other	Deep brain stimulation		[65]

and of single stimuli. Some selected papers reported both durations, hence they were written in Table 5 in the form *total(stimulus)*. Others reported only one duration (experiment or stimulus): if they provided enough information for the estimation of both conditions, the value is preceded by $\sim$, otherwise only the specified duration is presented, as *total* or *(stimulus)* if the total or stimulus duration is described, respectively. The durations of the whole experiments are in the range of ***[2-120] min***, while for the single stimuli are in the range of ***[2-390] s***. Median and inter-quartile range of experiment duration are ***22.5 min*** and ***50 min*** respectively, while they are ***61.5*** *s* and ***115 s*** for stimulus duration.

3.2.5 Rest phase

Picard et al. [67] firstly introduced a rest phase aimed at reaching a reference condition, as near as possible to a "relaxed" or "rest" state. However, the conducted review highlights the lack of a standard approach with respect to this topic. A major reason may be the scientific difficulty to define and ultimately induce an actual condition of "rest" in a person. Indeed, a great variety of methods were adopted, such as reading age-appropriated material [16], watching video clips [21], [23], [63], "relaxing" images (geometrical shapes [24] neutral [41], [49], nature [44], white/black screen [22]) with eventually a background music [44], [49], or asking to remain calm without any stimulus [15], [20], [31], [34], [51]. Actually, there is no difference between this phase and the stimulus phase: participants are stimulated to induce a relaxed/neutral state, which is itself an emotional state. The difference seems to be the aim to define an emotional baseline/reference, that should facilitate the discrimination of target emotions (i.e. emotions the study wants to classify) [16], [20], [31], [34], [44], [51], [63], to allow participants to relax and calm down [22], [24], [41], [49], or both [21], [23]. As a matter of fact, other studies include a neutral state among the emotions to be recognized. Table 5 reports the duration of the rest phase when present. Median and interquartile range are 3(1.25) min.

3.2.6 Environment

The experimental environment in emotion recognition is another important factor that requires a trade off. On the one side, a natural and uncontrolled environment facilitates the flow of emotions in participants and furthermore it is the nearest condition to a real world application. On the other side, the blindness to sensors placement, data acquisition and participants reactions, movements or activities makes hard to gather useful data, especially in initial stages of research. In order to analyse this topic, in the current review, Table 5 categorizes the environment as:

- *controlled*: a typical laboratory setting is adopted and researchers have full supervision on the experimental protocol [15]–[18], [20]–[28], [31]–[34], [36]–[49], [51]–[53], [57]–[66];
- *semicontrolled*: the environment is enough familiar to participants, who are also let free to behave, move and do what they want. Researchers observe what happens and can intervene [19], [35];
- *uncontrolled*: experiments are carried out in a real world setting (e.g. at participants' home) and researchers can neither supervise nor observe what happens [29], [30], [50], [54]–[56].

3.3 RQ3: Emotion model

3.3.1 Emotional interaction

Given the great heterogeneity of stimuli discussed in Section 3.2.3, the emotional interaction that participants may

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3128787, IEEE Transactions on Affective Computing

TABLE 3: List of emotional interactions and related stimuli.

EMOTIONAL INTERACTION	STIMULUS	WORKS
Imitation		
Intonation imitation	Sounds — Pre-recorded emotional sentences	[27], [42]
Facial mimic imitation	Images — Face expressions	[29], [45]
Passive	Sounds, Images, Videos	[15], [18], [21], [22], [25], [26], [28], [30], [35], [41], [43], [44], [49], [52], [62], [63], [65]
Active	Verbal, Cognitive tasks, Video games, Virtual reality, Robot interaction, Daily activities	[16], [17], [19], [20], [23], [24], [28], [30]–[34], [36]–[40], [46]–[48], [50], [51], [53]–[62], [64], [66]

undergo with the emotion recognition system or with the environment (e.g. people, games, activities of daily life) may be of different kinds. As summarized in Table 3, a first group of emotional interactions that were found in surveyed works is the imitation: participants are presented with audio of recorded sentences or face images and they are asked to imitate the intonation or the facial mimic respectively. These interactions are named *intonation imitation* and *facial mimic imitation*. A second kind of emotional interaction, named *passive*, is related to stimuli that did not require an interaction of the participant, such as watching video clips or listening to music. The last one, named *active*, identifies an active emotional interaction: the participant is not simply subjected to a stimulus, but he/she has an active role of interaction with the stimulus itself. As reported in Table 3, all the other stimulus types fall in this category, such as verbal interactions, cognitive tasks, playing games or virtual reality sessions, robot interaction, and daily activities. Fig. 7a reports the distribution of works with respect to emotional interaction and final clinical purpose, highlighting a prevalence of active interaction when the aim is to treat or diagnose emotional disorders.

3.3.2 Emotion Classes

In this review, many works selected the emotional states to recognize starting from already established clinical models of emotions, such as the categorical [73] or dimensional [74] models. In Fig. 6, the right column indicates the type of emotions categorization from which the works got inspired (the number of works is indicated between round brackets), while the central column indicates the actual number of recognized emotional states. The links with their numeric labels indicate how many works recognized a certain number of emotional states, starting from a certain model of emotion.

The categorical model of emotions by Ekman [73] (*happiness*, *surprise*, *sadness*, *fear*, *disgust*, and *anger*) inspired 23 surveyed works for the selection of emotion classes to recognize. Some works adopted exactly the 6 emotional states proposed by Ekman [41], [49], [62], but more frequently researchers focused on a sub-group of them [15], [18], [22], [24], [25], [33], [35], [42], [43], [45], [48], [53], [64], [66], or on a super-group with the addition of *neutral* or *contempt* emotional states [29], [30], [36]–[40].

Other works were inspired by the dimensional model of Russell [74] thus focusing on *valence* or *arousal* classification [17], [21], [23], [27], [32], [44], [46], [50], [52], [65].

Finally, application-specific or domain-specific emotional categorization were also adopted by researchers, they were chosen because considered significant for the addressed disease [16], [26], [28], [51], [63], e.g. psychiatric mood states [54]–[61], or for the final clinical purpose [19], [20], [31], [34], [47]. Table 5 lists all the classes taken into account in selected works.

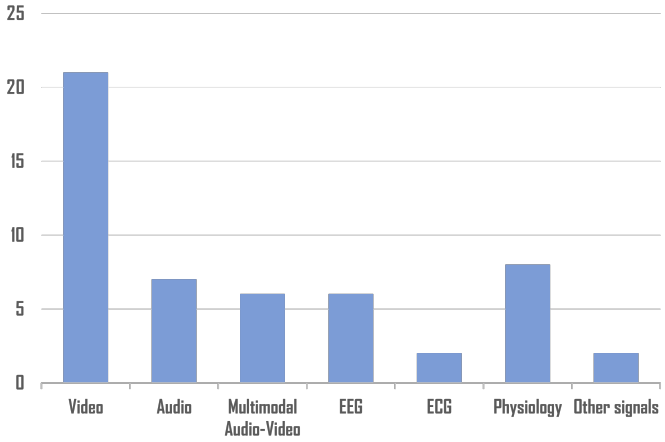


Fig. 5: Number of works with respect to signal.

3.4 RQ4: Sensors

The class Sensor in Fig. 2 aims at modeling both the physical variable being measured (signal), and the type and model of the adopted sensors.

3.4.1 Signal

A variety of sensing modalities were proposed among selected papers, as detailed in Fig. 5. they can be roughly distinguished in audio-video and physiological signals. *Video* based emotion recognition requires a camera to acquire video frames with facial images, then facial expressions are extracted by tracking specific facial landmarks through image processing techniques, and finally emotions are estimated on the base of the detected expressions, generally following the Facial Action Coding System [75]. *Audio* based emotion recognition leverages on signals acquired by a microphone and processed in order to extract spoken words and/or acoustic features about pitch and intonation. Recent advances in camera and microphone sensor quality, audio and image processing techniques, have made audio-video emotion recognition very popular and highly investigated: video signal was used in 21 works [19], [24]–[26], [28]–[30], [33], [35]–[40], [45], [46], [50], [53], [61], [63], [65], audio was adopted alone by 7 works [17], [27], [42], [48], [54], [60], [62], and in combination with smartphone position and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3128787, IEEE Transactions on Affective Computing

acceleration in 1 work [56] (grouped as "audio", in Fig. 5). Multimodal approaches were also adopted: 4 works fused information from both audio and video [57]–[59], [64], in 1 case also in combination with physiological signals [32], (grouped as "audio-video", in Fig. 5).

In spite of this great success, there are some drawbacks in using audio and video signals in emotion recognition. First of all, the voice intonations and the facial expressions of a person are usually mediated by the social contexts in which the person is, as a consequence these expressions might hide the real inner emotions felt. Real-time video processing may be computationally expensive and, furthermore, continuously sensing the video signal is difficult in real life, given the need to be in front of a camera. Two works adopted the video signal in uncontrolled environments, but the video was sensed just during the daily session of emotional training [29], [30].

In order to overcome such limitations of audio-video based emotion recognition, *physiological signals* were taken into account as possible reliable and sincere vehicles of emotions. The signals used in selected works include electrocardiogram (ECG), electroencephalogram (EEG), respiration (RS), skin conductance (SC), heart rate (HR), interbeat interval time series (RR), Blood Volume Pressure (BVP), skin temperature (ST). All the works adopting one or more physiological signals are grouped as "physiology" in Fig. 5. EEG [18], [20], [41], [43], [49], [52] and ECG [23], [55] signals are the only modalities used alone to perform emotion recognition. More commonly a multimodal paradigm is adopted in scientific works, where more physiological signals contribute to automatic emotion recognition [15], [16], [21], [31], [34], [44], [51], [66]. A general common approach for processing physiological signals involves preprocessing (e.g. application of filters in the time or frequency domain), feature extraction, and eventually feature selection. To deepen these processing aspects generally requires a dedicated review study [76]–[78], hence they are not discussed in this review.

Finally, 2 works adopted other types of signals, i.e. joystick pressure [47], or thermal imaging [22]. Tables 6 lists signals sensed in the selected works.

The relation between Sensor and Emotion classes expressed in Fig. 2 was deeply examined through Fig. 6. The column on the left shows the number of works (in curly brackets) adopting a certain signal, grouped as in Fig. 5. The signal boxes are linked to the central column, representing the number of recognized emotional states. These links with their numeric labels indicate how many works recognized a certain number of emotional states adopting a certain signal modality. 14 out of 21 studies adopting video signal classified more than 5 different emotional states, mainly belonging to Ekman or ad-hoc emotion categorization. 6 out of 8 works detecting 1 or 2 emotional states adopted physiological signals.

3.4.2 Type and model

Different kind of sensors were adopted in surveyed works. The video-based approach needs a camera, that can be found on the market with high heterogeneity of models, prices, technical characteristics. In this review, some researchers leveraged on low cost and ease of use, which is the

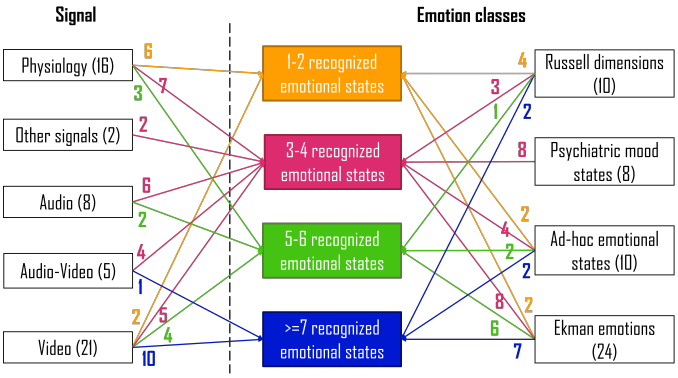
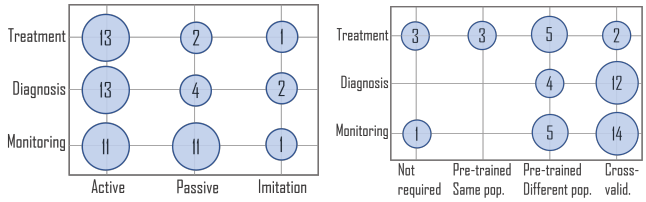


Fig. 6: Relation between Signal and Emotion classes. The right column indicates the number of works (in round brackets) adopting a certain emotions model. The central column indicates the actual number of recognized emotional states. The links from the right to the center, with their numeric labels, indicate how many works recognized a certain number of emotional states, starting from a certain emotions model. The left column shows the number of works (in round brackets) using a certain signal. The links from left to center, with their numeric labels, indicate how many works recognized a certain number of emotional states adopting a certain signal modality.



(a) (b)

Fig. 7: Distribution of works with respect to emotional interaction and final clinical purpose (a), and to training & validation and final clinical purpose (b).

case of webcam [19], [24], [29], [32], [37], smartphone/tablet camera [25], [26], [36], [40], [46], or other commercial cameras [17], [38], [50], [53], [57]–[61], [63]–[65]. On the contrary, other researchers opted for high quality signals, integrated API and functions to perform advanced processing, like Microsoft Kinect [33], [45], [51], Google Glass [30], [39], Intel RealSense SR300 programmable camera [35], Samsung ultramobile PC with a Logitech USB camera [28].

To measure audio signal, surveyed works adopted the microphone embedded in electronic devices [54], [56]–[58], [62], [64] (e.g. pc, smartphones, webcam, cameras, etc.), or a dedicated microphone [27], [42], [48]. Sensing physiological signals requires a trade-off between sensors obtrusiveness and signal quality, which is an important drawback of this methodology. Another typical aspect of physiological signals is that they can be used alone or in combination (in order to increase classification accuracy), thus complicating the above mentioned trade-off. Some commercial sensors adopted among surveyed works were Emotiv EPOC [20],

TABLE 4: List of algorithms used in emotion recognition.

ALGORITHM FAMILY	ALGORITHM	WORKS
Supervised Learning	Shallow Learning	
	— Decision Trees (DT)	[44]
	— Support Vector Machines (SVM)	[16], [24], [25], [27], [33], [45], [47], [52], [54], [55]
	— k-Nearest Neighbour (KNN)	[20], [22]
	— Naive Bayes (NB), Logistic Regression (LR)	[30]
	— Random Forest (RF)	[34]
	— Bayesian Networks (BN)	[28], [53]
	— Neural Network (NN)	[15], [59], [63]
	— Mixture classification	[43]
	— Comparison	[18], [21], [41], [42], [48], [49], [56], [57], [66]
	Deep Learning	[17], [29], [32], [46], [60], [64], [65]
	Ensemble Learning	[19], [26], [58], [61], [62]
Knowledge-based	— Threshold	[51]
	— Kalman filter	[23]
	— Fuzzy logic system	[31], [50]
Commercial	— Affectiva	[37], [39], [40]
	— RealSense SDK	[35]
	— Multiclassifier: Azure Faces API, Sightsound and Amazon Rekognition	[36]
	— Microsoft Oxford API	[38]

[41], [43], [49], [51], Neurosky EEG [18], and Acticap Express Bundle for EEG signal [66], whereas the Shimmer Research ECG sensor [23] or a sensorized t-shirt by Smartex were adopted for ECG signal [55]. Measurement units embedding multiple sensors were chosen when using a combination of physiological signals, such as Biopac MP150 [16], [31], [34], [51], Procomp Infinity [21], Shimmer Multi-Sensory [66], Microsoft Band 2 [44], Empatica E4 [32], or a custom prototype [15].

3.5 RQ5: Emotion Recognition

3.5.1 Algorithm

Automatic emotion recognition can be accomplished using several techniques. The most diffused approach is the supervised shallow (features-based) learning: it requires feature extraction after data windowing or segmentation, then each set of features calculated in each window/segment must be labeled with the correct class and sent in input to the algorithm. The algorithms applied by surveyed works in this approach are: Decision Trees (DT) [44], Support Vector Machines (SVM) [16], [24], [25], [27], [33], [45], [47], [52], [54], [55], k-Nearest Neighbour (KNN) [20], [22], Naive Bayes (NB), Logistic Regression (LR) [30], Random Forest (RF) [34], Bayesian Networks (BN) [28], [53], Neural Network (NN) [15], [63], Extreme Learning Machines [59], Mixture classification [43]. They have different characteristics, hence a comparison among some of them was often found among selected works [18], [21], [41], [42], [48], [49], [56], [57], [66]. Different types of NNs may be used in deep learning approaches, where the network is composed by one or more hidden layers that are trained taking in input labeled raw data [17], [29], [46], [60], [64], [65], or with the addition of some basic features extracted from commercial tools [32]. Recently, multi-classifiers (a special case of ensemble learning) [79] are gaining importance, where the overall classification algorithm is composed of multiple subsystems that may or may not be feature-based [19], [26], [58], [61], [62]. Finally, unsupervised learning is another opportunity: algorithms are trained with not-labeled input data, thus learning by themselves underlying patterns and clusters for classification with no or minimal human supervision. However, no work has been found adopting this approach.

In some cases, a knowledge based approach, instead of a learning based approach, was adopted to build the classifiers, such as a threshold [51], a Kalman filter [23] or a Fuzzy logic system [31], [50].

Finally, commercial off-the-shelf algorithms were also used as whole or part of the emotion recognition system, such as RealSense SDK [35], Microsoft Oxford emotion API [38], Affectiva [37], [39], [40], or a combination of multiple classifiers (Azure Faces API, Sightsound and Amazon Rekognition) [36].

An overview of the surveyed algorithms is proposed in Table 4. Fig. 8a shows the number of works adopting a certain type of algorithm for each signal modality (grouped as in Fig. 5). While the graph confirms that shallow learning algorithms are mostly diffused, it highlights a correlation between video and audio signals and deep learning algorithms. Finally, video is the only signal for which commercial algorithms are available.

3.5.2 Training and validation

A training, validation, and possibly test phase must be accomplished when a learning based emotion recognition algorithm is adopted. According to [80], the distinction between test and validation datasets is crucial. Indeed, in this context, learning consists in discovering the best approximation of the true classifier, so that it will perform well, even on new examples beyond the training dataset. This means to choose it by means of a *validation process*. After that, when an approximation of the classifier has been chosen, it must be evaluated by means of a *test process* with a test dataset that is distinct from both the training dataset and validation dataset, in order to avoid the so called *peeking* phenomenon [80]. Therefore, the presence of a test phase indicates that the classifier tuning is finished and is a more advanced developmental stage. Surveyed works may be distinguished based upon the phases included in the study:

- case 1: training, validation, and test;

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2021.3128787, IEEE Transactions on Affective Computing

- case 2: training and validation;
- case 3: test;

For works belonging to "case 1" and "case 2", the "Training" column of Table 6 reports the adopted validation and eventually test methods. Concerning "case 1", the percentage of data used for training, validation and test are reported, since all the works in this category used a unique dataset for all the phases [15], [19], [32], [57]–[61]. Concerning "case 2", selected works adopted k-fold [16], [17], [20]–[22], [24], [41], [42], [44], [47], [49], [52], [56], [62], [66] or leave-one-out [20], [27], [53]–[55], [63] cross validation.

Works belonging to "case 3" are indicated by the "pre-trained" keyword [25], [26], [28]–[30], [33], [34], [45], [46], [48], [52], [64], since the training and validation phases were done in previous works. Many works labeled as "pre-trained" use a commercial classification API [35]–[40]. All these works used the algorithm as it is, just Washington et al. [30] performed a further tuning before applying it to the target population (indicated by adding the word "tuning" in Table 6). Training phase is critical for all learning problems and strongly depends on training dataset, but this is especially true in clinical scenarios, where pathological populations exhibit peculiar behaviours and characteristics not only with respect to healthy controls, but also among them (due to different stages or symptoms of the disease). The present review analyzed this aspect by considering the homogeneity of training dataset with respect to the target population. When the algorithm is "pre-trained", Table 6 reports "same type" if the population from which the training dataset was extracted coincides with the target population [30], [34], [39], [52], and "different type" otherwise [25], [26], [28], [29], [33], [35]–[38], [40], [45], [46], [48], [64]. Instead, "same type" is implicit for works belonging to "case 1 and 2", since training data belong to target population. Some few works also performed training and validation in two different population in order to investigate the differences between them [42], [44].

3.5.3 Ground truth methodology

The problem of assigning a label with the correct class to each observation (i.e. sample of data or set of features in input to the algorithm) is closely related to the training of supervised algorithms. Data labeling is challenging in emotion recognition, since there is no universally accepted and claimed gold standard for the evaluation of one's emotional state. According to the papers included in this review, emotions are labeled, as also reported in Table 6, in the following ways:

- "self": asking enrolled subjects to self-assess their emotions [15], [19], [41], [43], [44], [49], [51], [63], [66];
- "expert": requiring the evaluation from external expert raters: a combination of therapists, doctors, teachers, pathologists or even parents (especially for children with ASD) [16], [17], [19], [20], [22], [24]–[26], [28], [32]–[36], [39], [46]–[48], [53]–[61], [64], [65];
- "stimulus": based upon a predefined classification of the type of stimulus (e.g., a sad song, a happy video, etc.) and regardless of the real emotion felt by the participant [21], [23], [27], [30], [42], [45], [52];

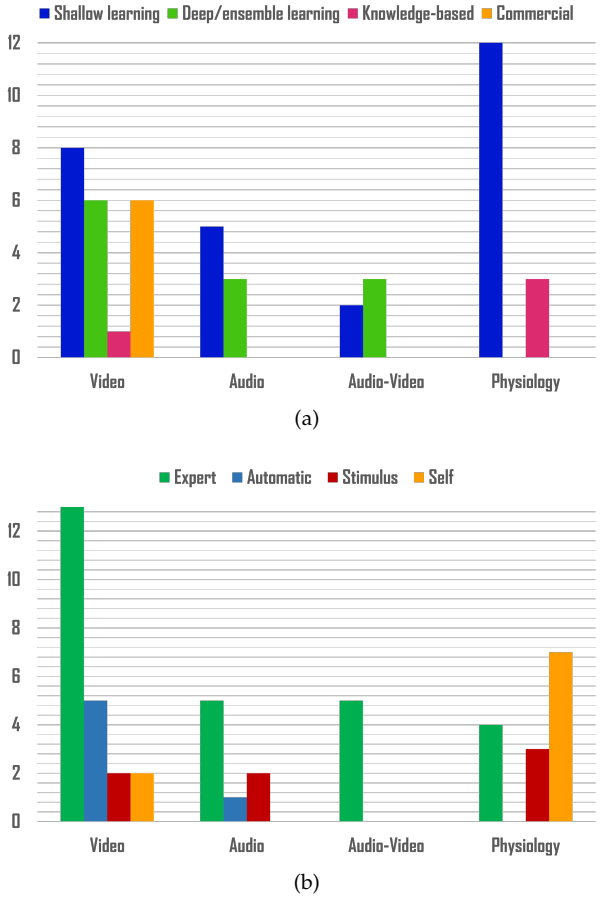


Fig. 8: Number of works grouped by signal and algorithm (a), and by signal and ground truth methodology (b).

- "automatic": automatically, using the emotion recognition algorithm itself previously trained [29], [37], [38], [40], [50], [62] or based on experience-driven rules [50].

The distribution of labeling methods with respect to adopted signal is deepened in Fig. 8b, showing a prevalence of expert and automatic labeling for video signal.

3.5.4 Performance

The absence of standard approaches and datasets makes hard to fairly compare the performance of the surveyed emotion recognition algorithms. However, it is worth considering the performance in order to investigate the evaluation methodologies and to eventually individuate possible trends in scores values. Table 6 reports the name of the adopted metric and its score. With respect to the training methods described in subsection 3.5.2, the reported scores are representative of the validation phase for works that performed just a validation, and of the test phase for works that are "pretrained" or that split the dataset in train, validation, and test folds. When more scores are obtained in a work, i.e. when the emotion model considers more than one emotion, the average value is reported. Most part of reviewed works always included *accuracy* among performance metrics [16]–[19], [21]–[23], [25], [28]–[30], [33], [34], [36], [41]–[45], [47], [49], [52], [53], [55], [56], [61]–[64], [66], however other metrics were also found, such as *F1-score* [20], [81], *recognition rate* [24], a *correlation* measure [26], [32],

unweighted average recall [27], [48], [57]–[60], *recall* [35], *area under the curve* [54], [65]. The works that aim to evaluate the effectiveness of a system in improving some clinical characteristics do not report the performance of the emotion recognition algorithm. These cases are indicated with *Not available*.

4 Discussion

This section discusses the results summarizing answers to RQ 1-5.

4.1 RQ1: Aim

The surveyed literature revealed a recent increased attention to emotion recognition in clinical scenario, given that several surveyed works were published after 2015. Most selected papers are aimed at measuring feasibility and validating automatic emotion recognition in neurological and psychiatric disorders in order to provide a reliable system for treating, diagnosing, and monitoring emotional features. Many works (58%) concern the ASD and PD. This interest is motivated by some important clinical needs: ASD [82] and PD [83] are two of the most frequent neuropsychiatric diseases of childhood and elderly, respectively, and both are characterized by emotional behavior disorders and emotions recognition disturbances, which mainly contribute to disability in daily life and restriction of participation.

ASD is a neuro-developmental disorder affecting social communication and interaction. Evidence supports that intensive behavioral intervention can help ASD patients to manage their deficits and effectively interact with other people [84]. Some critical issues of behavioral intervention that may be overcome by an emotionally intelligent training tool applicable at patients' home are the transfer of obtained improvements to real life situations and the bottleneck of therapists' availability. Indeed, in ASD scenario, the "treatment" purpose stands out with respect to all other pathologies since automatic emotion recognition is experimented in systems for emotional training, aimed at reducing emotional disability and improve emotion recognition skills. Most of them are video games [28], [29], [31], [33], [39], [40], but also augmented reality [30] and robot interaction [32] were investigated. Furthermore, some virtual reality platforms were proposed with the specific aim to improve driving skills [20], [34], essential for plain independence. Early diagnosis is crucial to increase the chance of positive outcome for these people, hence, some studies presented diagnostic tools based on automatic emotion recognition [25]–[27]. Finally, some researchers investigated the development of emotion recognition algorithms for monitoring purpose [16]–[19], [21]–[24], [35], [85].

PD is a neurodegenerative disorder that comprises a lot of motor and non-motor symptoms affecting the emotional sphere. Attention to non-motor symptoms is fairly recent and research studies share the aim of monitoring emotional states. Clinical studies have found a strong association between health-related quality of life and "psychosocial functioning" [86]. Psychosocial functioning is in turn related to both mood and ability to recognize and express emotions that are impaired in PD [87]. Indeed, a reduced facial expressivity, called hypomimia, as well as a difficulty to recognize others expressions, called alexithymia, are two symptoms needing in depth investigation. Automatic and quantitative methods of emotion detection [41]–[45] are aiding researchers to deepen these aspects.

29% of papers studied the reliability of emotion recognition algorithm for diagnosis or monitoring emotions in psychiatric disorders: mainly bipolar disorders [54]–[62], but also other depressive disorders [63], [64], schizophrenia [53], obsessive compulsive [65], and acrophobia [66]. Diagnosis" and "differential diagnosis" are the most frequent final clinical purposes found in works on psychiatric diseases, especially for bipolar or unipolar disorders. Indeed, high rates of misdiagnosis, delayed diagnosis, and lack of recognition of treatment effect often lead patients with psychiatric disorders to have a chronic course with high disability, unemployment rates, and mortality [88]. Therefore, an objective, rapid and reliable recognition of emotional state and interepisodic dysfunctional domains may be useful in preventing acute phase of the disease and dramatic consequences.

Automatic emotion recognition was studied in frail subjects to diagnose apathy [46], a pervasive neuropsychiatric symptom of the majority of neurocognitive, neurodegenerative, and psychiatric disorders or to build an effective robotic companion that may help aging population with mild cognitive impairment and dementia [50], [51] to improve social, cognitive, and motor functions. Ageing population is at the center of the looming healthcare crisis, increasing healthcare costs, and expected serious shortage of healthcare workers. Companion robots may help for supporting aged care facilities to meet this challenge and improve the quality of care of older people mental and physical health outcomes, as well as to support healthcare workers in personalizing care.

In post-stroke patients, monitoring is useful to evaluate the effects of brain damage in emotion perception [48], [49], or to build motivating treatment tools [47].

Huang et al. [52] applied automatic emotion recognition to assess residual consciousness in patients with disorders of consciousness. They investigated a theme widely discussed by neurologists, the "interoception and emotion" [89]. There is a wide range of disturbances of consciousness ranging from a deep coma, a vegetative state, to a state of minimal consciousness or emergency state of minimal consciousness in which the partial motor response to pain or verbal stimuli suggests an activation of the emotional system [90] that is used to promote awakening [91]. However, it cannot be conclusive due to its pilot design.

Automatic emotion recognition was used also in metabolic syndrome [15]. Recently, the prevalence of metabolic syndrome (diabetes mellitus) in particular, has rapidly increased and become a major health issue. Environmental and physiological effects such as stress, behavioral and metabolic disturbances, infections, and nutritional deficiencies have recognized as contributing factors to develop metabolic diseases. Patlar et al. [15] presents a multivariate methodology for the modeling of stress on metabolic syndrome patients developing a supporting system to cope with patients' anxiety and stress. Indeed, stress was associated with major cardiovascular events [92], hence an early

TABLE 5: Works classification according to clinical perspective. A double line separate works dealing with neurodevelopmental disorders, and works dealing with neurological and psychiatric disorders. Study design abbreviations: cross-sectional (cross.), longitudinal (long.), observational (observ.), interventional (interv.), controlled (contr.), not controlled (not contr.)

Paper	Disease	Study design	Final clinical purpose	Population	Emotions Classes	Stimulus channels	Duration [Total (single stimulus)]	Rest	Environment	Emotional interaction
Liu et al. 2008 [16]	ASD	observ. cross. not contr.	monitoring	6 ES	engagement, anxiety, liking	still & moving images	1 h (3-4 min)	3 min	controlled	active
Gong et al 2017 [17]	ASD	observ. long. not contr.	monitoring	4 ES	very negative, negative, neutral, positive	verbal, audio, still & moving images	1 h	none	controlled	active
Bairavi et al. 2018 [18]	ASD, ADHD	observ. cross. contr.	monitoring	18 ES	happy, sad, neutral	audio	$\sim$ 5 min (30 s)	none	controlled	passive
Dawood et al. 2018 [19]	ASD	observ. cross. contr.	monitoring	6 ES, 5 HC	anxiety, boredom, uncertainty, confidence, engaged	audio, moving images	2h	none	semicontrolled	active
Fan et al. 2018 [20]	ASD	observ. cross. not contr.	monitoring	20 ES	engagement, enjoyment, boredom, frustration	audio, moving images	1 h	3 min	controlled	active
Sarabadani et al. 2018 [21]	ASD	observ. cross. not contr.	monitoring	15 ES	arousal, valence	still images	$\sim$ 80 min (2 min)	5 min	controlled	passive
Rusli et al. 2020 [22]	ASD	observ. cross. contr.	monitoring	23 ES, 34 ASD	happy, fear, sad	audio, moving images	$\sim$ 20 min (1-5 min)	2 min	controlled	passive
Kushki et al. 2015 [23]	ASD	observ. cross. not contr.	monitoring	24 ES	low/high arousal	still & moving images, audio, verbal	35 min (3-5 min)	5 min	controlled	active
Chu et al. 2018 [24]	ASD	observ. cross. not contr.	monitoring	15 ES	calm, happy, anxious, angry	audio, still & moving images	$\sim$ 1 h (3-10 min)	3 min	controlled	active
Hashaemi et al. 2018 [25]	ASD	observ. cross. contr.	diagnosis	15 ES, 18 HC	happy	moving images, verbal	$\sim$ 5 min (30-60s)	none	controlled	passive
Li et al. 2019 [26]	ASD	observ. cross. contr.	diagnosis	49 ES, 39 HC	expressions, valence, arousal	audio, moving images	$\sim$ 10 min	none	controlled	passive
Ringeval et al. 2016 [27]	SLI, ASD	observ. cross. contr.	diagnosis, differential diagnosis	35 ES, 70 HC	positive, neutral, and negative valence	audio, still images	$\sim$ 6 min	none	controlled	intonation imitation, active
Madsen et al. 2008 [28]	ASD	interv. cross. not contr.	treatment	3 ES	agreeing, concentrating, disagreeing, interested, thinking, confused	verbal	not said	none	controlled	passive & active
Tsangouri et al. 2016 [29]	ASD	interv. long, not contr.	treatment	9 ES	6 Ekman emotions, neutral	still images	20 min	none	uncontrolled	facial mimic imitation, active
Washington et al. 2017 [30]	ASD	interv. long. not contr.	treatment	14 ES	Ekman emotions, neutral, contempt	daily activities, verbal	20 min	none	uncontrolled	passive & active
Kuriakose et al. 2017 [31]	ASD	interv. cross. not contr.	treatment	9 ES	anxiety	audio, moving images	2 h	3 min	controlled	active
Rudovic et al. 2018 [32]	ASD	interv. cross. not contr.	treatment	35 ES	valence, arousal, engagement	verbal, still images, interaction with robot	25 min	none	controlled	active
White et al. 2018 [33]	ASD	interv. long. contr.	treatment	20 ES, 20 HC	happiness, fear, anger, neutral	audio, still & moving images	60-90 min	none	controlled	active
Bian et al. 2019 [34]	ASD	observ. cross. not contr.	treatment	20 ES	engagement	audio, moving images	90 min	3 min	controlled	active
Tang et al. 2017 [35]	ASD	observ. cross. not contr.	monitoring	6 ES	happiness	audio, moving images	$\sim$ 15 min (2 min)	none	semicontrolled	passive

Continue from Table 5

Paper	Disease	Study design	Final clinical purpose	Population	Classes	Stimulus channels	Duration	Rest phase	Environment	Emotional interaction
Kalantarian et al. 2019 [36]	ASD	observ. cross. not contr.	monitoring	8 ES	6 Ekman emotions, neutral	still images	(90s)	none	controlled	active
Banire et al. 2020 [37]	ASD	observ. cross. contr.	monitoring	20 ES, 26 HC	6 Ekman emotions, contempt, neutral	still images	5 min (2 s)	none	controlled	active
Bharantharaj et al. 2016 [38]	ASD	interv. long. not contr.	treatment	9 ES	6 Ekman emotions, contempt, neutral	verbal	15 min	none	controlled	active
Vahabzadeh et al. 2018 [39]	ASD	interv. long. not contr.	treatment	8 ES	Ekman emotions, engagement, contempt	verbal	not available	none	controlled	active
Garcia et al. 2019 [40]	ASD	observ. cross. not contr.	treatment	3 ES	6 Ekman emotions, contempt, neutral	audio, still & moving images	$\sim$ 10 min (3-5 min)	none	controlled	active
Yuvaraj et al. 2014 [41]	PD	observ. cross. contr.	monitoring	20 ES, 20 HC	6 Ekman emotions	still & moving images, audio	1 h (36-45 s)	10 s	controlled	passive
Zhao et al. 2014 [42]	PD	observ. cross. contr.	monitoring	5 ES, 7 HC	anger, happiness, sadness, fear, neutral	audio	25 min (15-30 s)	none	controlled	intonation imitation
Khrishna et al. 2019 [43]	PD	observ. cross. not contr.	monitoring	16 ES	happy, sad, angry, calm	audio, moving images	$\sim$ 5 min(30 s)	none	controlled	passive
Pepa et al. 2019 [44]	PD	observ. cross. contr.	monitoring	10 ES, 18 HC	arousal, valence	audio, moving images	1 h (30-95 s)	3-5 min	controlled	passive
Bandini et al. 2017 [45]	PD	observ. cross. contr.	diagnosis	17 ES, 17 HC	happiness, sadness, anger, disgust, neutral	still images, verbal	$\sim$ 2 min (3-12 s)	none	controlled	facial mimic imitation, active
Happy et al. 2019 [46]	Apathy	observ. cross. contr.	diagnosis	18 ES, 27 HC	positive, negative, neutral	verbal	(1 min)	none	controlled	active
Rivas et al. 2015 [47]	stroke	observ. long. not contr.	monitoring	2 ES	tiredness, tension, pain, satisfaction	audio, moving images	45 min	none	controlled	active
Sidorova et al. 2019 [48]	stroke	observ. cross. contr.	monitoring	2 ES, 2 HC	anger, happiness, neutral, sadness	still images	not said	none	controlled	active
Bong et al. 2017 [49]	stroke	observ. cross. contr.	diagnosis	19 LES, 19 RES, 19 HC	6 Ekman emotions	audio, still & moving images	1-2 h (46-60 s)	3 min	controlled	passive
Khosla et al. 2013 [50]	Frail older people	interv. cross. not contr.	treatment	70 ES	valence (6 levels)	audio, verbal, still & moving images, daily activities	not available	none	uncontrolled	active
Fan et al. 2017 [51]	Mild cognitive impairment, dementia	interv. long. contr.	treatment	4 ES, 7 HC	engagement	Still & moving images	60 min	3 min	controlled	active
Huang et al. 2019 [52]	consciousness disorder	observ. cross. contr.	monitoring	10 HC, 8 ES	positive, negative	audio, moving images	HC: $\sim$ 1h (30 s), ES: $\sim$ 35 min (30 s)	none	controlled	passive
Wang et al. 2008 [53]	Schizophrenia	observ. cross. contr.	diagnosis	12 ES, 12 HC	happiness, sadness, anger, fear	verbal	not said	none	controlled	active
Karam et al. 2014 [54]	BD	observ. long. not contr.	diagnosis	6 ES	depression, hypomania, euthymia	daily activities	not applicable	none	uncontrolled	active
Valenza et al. 2014 [55]	BD	observ. long. not contr.	diagnosis	8 ES	depression, hypomania, mixed-state, euthymia	daily activities	not applicable	none	uncontrolled	active
Grunerbl et al. 2015 [56]	BD	observ. long. not contr.	diagnosis	10 ES	depression, euthymia, mania	daily activities	not applicable	none	uncontrolled	active

Continue from Table 5

Paper	Disease	Study design	Final clinical purpose	Population	Emotions Classes	Stimulus channels	Duration [Total (single stimulus)]	Rest	Environment	Emotional interaction
Ciftci et al. 2018 [57]	BD	observ. long. contr.	diagnosis	46 ES, 49 HC	remission, hypomania, mania	still images, verbal	∼ 5-15 min	none	controlled	active
Ebrahim et al. 2018 [58]	BD	observ. long. not contr.	diagnosis	46 ES	mania, hypomania, remission	still images, verbal	∼ 5-15 min	none	controlled	active
Syed et al. 2018 [59]	BD	observ. long. not contr.	diagnosis	46 ES	mania, hypomania, remission	still images, verbal	∼ 5-15 min	none	controlled	active
Ren et al. 2019 [60]	BD	observ. long. not contr.	diagnosis	46 ES	mania, hypomania, remission	still images, verbal	∼ 5-15 min	none	controlled	active
Abaei et al. 2020 [61]	BD	observ. long. not contr.	diagnosis	46 ES	mania, hypomania, remission	still images, verbal	∼ 5-15 min	none	controlled	active
Huang et al. 2018 [62]	UD, BD	observ. cross. contr.	differential diagnosis	15 BD, 15 UD, 15 HC	6 Ekman emotions	audio, moving images, verbal	∼ 40 min	none	controlled	passive & active
Gavrilescu et al. 2019 [63]	Depression	observ. long. contr.	monitoring	56 ES, 72 HC	anxiety	audio, moving images	∼ 20 min (1 min)	5 min	controlled	passive
Jiang et al. 2020 [64]	Mild Depressive Disorder	observ. long. not contr.	monitoring	12 ES	anger, happy, sad, neutral	verbal	15 min	none	controlled	active
Cohn et al. 2018 [65]	Obsessive-compulsive disorder	observ. cross. not contr.	monitoring	2 ES	valence	Deep brain stimulation	∼ (2 min)	none	controlled	passive
Balan et al. 2020 [66]	Acrophobia	interv. long. not contr.	treatment	8 ES	fear	moving images	not said	none	controlled	active
Patlar et al. 2020 [15]	Methabolic syndrome	observ. cross. contr.	monitoring	15 ES, 15 HC	calmness, happiness, fear, disgust, anger	still & moving images, audio	∼ 1h	not said	controlled	passive

stage of comprehensive intervention that uses also an alert for stress may reduce the risk of complication in subjects with metabolic disorders.

4.2 RQ2: Experimental protocol

Concerning the study design, most are cross-sectional in line with studies purpose, i.e. validating a tool for monitoring or diagnosis. Among the trials aimed at studying systems for the treatment of emotional and behavioral disorders the interventional longitudinal designs prevail. From a clinical point of view important limitations concern the low number of enrolled subjects (27 studies, 63%, enrolled less than 20 subjects) and the absence of a control sample (31 studies, 58%, were not controlled). In particular, many studies with a "treatment" final clinical purpose enrolled a not controlled population (Fig. 4c). Moreover, many selected works provide a poor or absent description of clinical characteristics of the enrolled population, making hard to interpret and exploit obtained results. Works concerning bipolar disorders are an exception: 66.7% of the studies enrolled more than 45 subjects [57]–[62] , 88.9% are longitudinal studies [54]–[61], suggesting a reliable study design. Experimental protocols conducted in uncontrolled/real environment are very rare. These results may indicate that many works concern just preliminary or pilot studies. Furthermore, few works met the inclusion criteria, compared to the number of works selected by keywords and to the literature on emotion recognition. Altogether, these considerations may reflect difficulties in coordinating multidisciplinary team and in applying the same algorithms in healthy subjects as well as in subjects with diseases. Getting ethics clearances for studies with people with a disorder may be an obstacle too.

Emotional stimuli are very heterogeneous concerning the type and the duration, as detailed in Sections 3.2.3 and 3.2.4 and in Table 2. The analysis of times and duration can be discussed in terms of the duration of the entire experiment, and the duration of single stimuli. A long experimental protocol may help controlling multiple variables and ensuring greater accuracy, however it conveys the risk to get participants bored and tired, especially in presence of an underlying disease. The optimal duration of single stimuli should be discussed in light of a crucial issue: how long human beings take to experience a certain emotional state after being exposed to a stimulus? The answer varies depending on stimulus type and duration and on several latent factors as the individual's personality, starting mood, internal feeling, confidence with the stimulus type, etc.

The use of a rest phase is a means to reduce the variability of findings, referring individual responses to stimuli to their own rest state. However, there is not agreement or clear scientific evidence of the usefulness of such approach.

4.3 RQ3: Emotion model

Selected studies consider different classes of emotions. Subset or superset of Ekman emotions are widely used, but several works define their own specific emotion targets. An important difference among works is the complexity of the adopted models and classes. Regarding the kind of emotional interaction, the conducted analysis revealed that stimulus can be chosen in order to convey a certain type

TABLE 6: Work classification according to technological perspective. A double line separate works dealing with neurodevelopmental disorders and works dealing with neurological and psychiatric disorders.

Paper	Sensor	Signal	Algorithm	Training	Ground truth	Performance
Liu et al. 2008 [16]	Biopac MP150	ECG, PPG, ICG, PCG, EDA, ST, EMG	SVM	10-fold cross validation	expert	Accuracy 82.7%
Gong et al. 2017 [17]	camera	audio	Deep Convolutional NN	10-fold cross validation	expert	Accuracy 61.0%
Bairavi et al. 2018 [18]	Neurosky EEG	EEG	KNN, NB, SVM, RF	not available	not available	Accuracy 82.0%
Dawood et al. 2018 [19]	webcam	video	Convolutional NN + LSTM	80% training, 10% test, 10% validation	self, expert	Accuracy 91.0%
Fan et al. 2018 [20]	Emotiv EPOC	EEG	KNN	leave-one-out + 10-fold cross validation	expert	F1score 87.8%
Sarabadani et al. 2018 [21]	Procomp infinity	ECG, EDA, RSP, ST	KNN, LDA, 3 types of SVM	4 fold cross validation	stimulus	Accuracy 81.5%
Rusli et al. 2020 [22]	FLIR T420 thermal camera	Thermal imaging	KNN	5 fold cross validation	expert	Accuracy 88.0%
Kushki et al. 2015 [23]	Shimmer Research wearable sensor	ECG	Kalman filter	Not applicable	stimulus	Accuracy 95.0%
Chu et al. 2018 [24]	webcam	video	SVM	10-fold cross validation	experts	Recognition rate: 98%
Hashaemi et al. 2018 [25]	iPad camera	video	SVM	pre-trained / different type	expert	Accuracy 90%
Li et al. 2019 [26]	iPad camera	video	Convolutional NN	pre-trained / different type	expert	Correlation coefficient 0.65
Ringeval et al. 2016 [27]	microphone	audio	SVM	leave-one-speaker-out	stimulus	Unweighted average recall 41%
Madsen et al. 2008 [28]	Logitech camera	video	Dynamic BN	pre-trained / different type	expert	Accuracy 77%
Tsangouri et al. 2016 [29]	webcam	video	Convolutional NN	pre-trained / different type	automatic	Accuracy 74%
Washington et al. 2017 [30]	Google Glass	video	logistic regression	pre-trained / same type / tuning	stimulus	Accuracy 75%
Kuriakose et al. 2017 [31]	Biopac MP150	PPG, EDA, ST	Fuzzy logic	Not applicable	Not available	Not available
Rudovic et al. 2018 [32]	webcam, microphone, E4 wrist sensor	video, audio, HR, EDA, ST	Feed-forward NN	40% train, 20%valid, 40% test	expert	Intra-class correlation 57%
White et al. 2018 [33]	Microsoft Kinect	video	SVM	pre-trained / different type	expert	Accuracy 60%
Bian et al. 2019 [34]	Biopac MP150	PPG, EDA, RSP	RF	pre-trained / same type	expert	Accuracy 85%
Tang et al. 2017 [35]	Intel RealSense camera	video	RealSense SDK	pre-trained / different type	expert	Recall 69%
Kalantarian et al. 2019 [36]	smartphone camera	video	multi-classifier: Azure Faces API, Sightsound and Amazon Rekognition	pre-trained / different type	expert	Accuracy 55%
Banire et al. 2020 [37]	webcam	video	Affectiva	pre-trained / different type	automatic	Not available
Bharantharaj et al. 2016 [38]	Ai-ball camera	video	Microsoft Oxford emotion API	pre-trained / different type	automatic	Not available
Vahabzadeh et al. 2018 [39]	Google Smartglasses	video	Affectiva	pre-trained / same type	expert	Not available
Garcia et al. 2019 [40]	smartphone camera	video	Affectiva	pre-trained / different type	automatic	Not available
Yuvaraj et al. 2014 [41]	Emotiv EPOC	EEG	KNN, SVM	10-fold cross validation	self	Accuracy 83%
Zhao et al. 2014 [42]	Isomax EarSet E60P5L microphone	audio	NB, RF, SVM	10-fold cross validation	stimulus	Accuracy 60%
Khrishna et al. 2019 [43]	Emotiv EPOC	EEG	Mixture Classification	Not available	self	Accuracy 89%
Pepa et al. 2019 [44]	Microsoft Band 2	HR, EDA, ST	Decision Tree	5-fold cross validation	self	Accuracy 86%
Bandini et al. 2017 [45]	Microsoft Kinect	video	SVM	pre-trained / different type	stimulus	Accuracy 88%
Happy et al. 2019 [46]	tablet camera	video	CNN	pre-trained / different type	expert	Not available
Rivas et al. 2015 [47]	Gesture Therapy gripper	gripping pressure	SVM	10 fold CV	expert	Accuracy 72%
Sidorova et al. 2019 [48]	Audiotechnica AT 2035 microphone	audio	C4.5, RF, SVM, MLP	pre-trained / different type	expert	Unweighted average recall 61%

Continue from Table 6

Paper	Sensor	Signal	Algorithm	Training	Ground truth	Performance
Bong et al. 2017 [49]	Emotiv EPOC	EEG	KNN, Probabilistic NN	10-fold cross validation	self	Accuracy 66%
Khosla et al. 2013 [50]	camera	video	Fuzzy logic	Not applicable	automatic	Not available
Fan et al. 2017 [51]	Kinect, Emotiv EPOC, Biopac MP150	eye gaze, EEG, EDA	threshold	Not applicable	self	Not available
Huang et al. 2019 [52]	32 channel EEG cap & SynAmps2	EEG	SVM	online: pre-trained / same type, offline: 5 fold cross validation	stimulus	Accuracy 59%
Karam et al. 2014 [54]	smartphone microphone	audio	SVM	leave-one-out	expert	Area under the curve 0.81
Valenza et al. 2014 [55]	Sensorized T-shirt	ECG	SVM and Markov chain	leave-one-acquisition-out	expert	Accuracy 90%
Grunerbl et al. 2015 [56]	smartphone (microphone, GPS, accelerometer)	audio, position, acceleration	NB, KNN, j48, conjunctive rule learner	3 fold cross validation	expert	Accuracy 76%
Ciftci et al. 2018 [57]	camera	audio, video	Partial least square regression, extreme learning machine	50% train, 25%valid, 25% test	expert	Unweighted average recall 47%
Ebrahim et al. 2018 [58]	camera	audio, video	LSTM	50% train, 25%valid, 25% test	expert	Unweighted average recall 44% test
Syed et al. 2018 [59]	camera	audio, video	Weighted extreme learning machines	50% train, 25%valid, 25% test	expert	Unweighted average recall 57.4% test
Ren et al. 2019 [60]	camera	audio	Feed-forward NN, gated recurrent NN, convolutional NN	50% train, 25%valid, 25% test	expert	Unweighted average recall 57.4% test
Abaei et al. 2020 [61]	camera	video	LSTM	50% train, 25%valid, 25% test	expert	Accuracy 57%
Huang et al. 2018 [62]	webcam	audio	Hierarchical spectral clustering + LSTM	5-fold cross validation	automatic	Accuracy 88%
Gavrilescu et al. 2019 [63]	camera	video	Feed forward NN	leave one out cross validation	self	Accuracy 74%
Jiang et al. 2020 [64]	camera: Canon Vixia HF R600	audio, video	Convolutional NN	pre-trained / different type	expert	Accuracy 78%
Wang et al. 2008 [53]	camera	video	BN	leave one out cross validation	expert	Accuracy 60%
Cohn et al. 2018 [65]	camera	video	Convolutional NN	Not available	expert	Area under the curve 0.87
Balan et al. 2020 [66]	Acticap Xpress Bundle, Shimmer3 GSR+ Unit	EEG, EDA, HR	kNN, SVM, RF, LDA, Deep NN	10-fold cross validation	self	Accuracy 79%
Patlar et al. 2020 [15]	CVDiMo	ECG, EDA, ST, blood pressure, blood glucose, SpO2	Feed forward NN	70% training, 15% validation, 15% test	self	Not available

of interaction (Table 3) and that when the clinical aim is different from "monitoring", the emotional interaction is generally active (Fig. 7a).

4.4 RQ4: Sensors

The choice of sensors is critical in the healthcare field and it mainly influences long term adoption of the proposed systems in the clinical practise or in people daily life, since sensors type and model should guarantee some important requirements to patients and caregivers, such as privacy, ease of use, low cost, etc. Furthermore, their usage is related to several aspects, that may be subjective, e.g. people preferences and attitude towards technology, or objective, such as the withdrawn of some sensing devices from the market (e.g. Google glasses, Microsoft Band 2).

For what concerns measured signals, Section 3.4.1 highlighted a relation between signal and emotion model, showing that video is used to classify a wide variety and plurality of emotions, followed by audio and EEG that can classify 3-4 emotions. The video modality seems to be at a more advanced stage in automatic emotion recognition in clinical scenarios, also for other reasons: it is used by the majority of works, it is the only one adopted by commercial off-the-shelf systems, and it is generally labeled by experts or even automatic algorithms. A great heterogeneity of approaches emerged among works adopting physiological signals, that are characterized by: comparison among different algorithms, presence or absence of a rest phase, and different labeling methods (with a prevalence of self-assessment). Finally, they mostly accomplish binary classification tasks on few emotion classes (see Fig. 6).

4.5 RQ5: Emotion recognition

Section 3.5 reports a correlation between signals and algorithms from which it emerges that shallow learning algorithms are by far the preferred, regardless of the signal. This is probably because in this application domain it is difficult to have large datasets available for learning. In fact, deep learning works better than shallow learning only when very large learning datasets are used, otherwise shallow learning works better. Furthermore, it should be noticed that, when used, deep learning is only used for audio-video signals (Fig. 8a). A relation is also present between final clinical purpose and training: treatment studies are pre-trained, indicating the need for an already available and possibly robust emotion recognition algorithm. The phases of training, validation, and test are conducted with great heterogeneity, concerning both the adopted methods and composition of sample population to be used in each phase. Indeed, several works trained the model on healthy participants and then tested or directly used it on patients. Great heterogeneity and variability were found with respect to performance metrics and scores. There are not any evident relations or trends between performance and other topics. This is a possible marker of lack of robustness and reproducibility of obtained results. However, obtained performances are generally quite high, highlighting the potential of these techniques and encouraging further research.

5 Future directions and recommendations

Moving from this systematic review, in response to RQ6, some future directions can be identified and some recommendations can be drawn up for those who want to try their hand at these types of studies and technologies.

The limitations that were found in response to RQ5 suggest that recommendations for a trustworthy and ethical AI may be important to follow [93]–[96]. 1) First of all, the system should be ethical, i.e. adhering to ethical principles and values (e.g., respect for human dignity and freedom, equality, non-discrimination). These principles have several consequences, just to mention two of them: system outcomes should be explainable, and avoid unfair bias. 2) Furthermore, the system should be compliant with all applicable laws and regulations (e.g., the UN Human Rights treaties, the European General Data Protecion Regulation, national laws, FDA guiding principles[4]) and guarantee privacy, safety, and security [96]. 3) Moreover, each system should be comparable with its other competitors. 4) Finally, the system should be reliable (i.e. working properly with a range of inputs and in a range of situations), robust and reproducible (i.e. exhibit the same behaviour when repeated under the same conditions), as it can cause unintentional harm (e.g., a wrong diagnosis or treatment).

To reach this target, algorithms and datasets should be made publicly available to allow easier and fair comparison among different solutions. Furthermore, some good practices about the composition of the sample of population to be used in the training, validation and testing phase may help, as also recommended from FDA. 1) Firstly, when selecting the classifier, the training dataset should be distinct from the validation one, and both should be distinct from the test dataset. 2) Secondly, each dataset should consist of a statistically appropriate number of samples, taken from a statistically appropriate number of people. Indeed, studies with a small population give unreliable results. 3) Moreover, the three datasets should be homogeneous with each other in the composition, to avoid bad performance due to having trained the classifier with a different type of population with respect to the one employed for testing or use. 4) Performance of pre-trained emotion recognition systems should be monitored, expecially if further tuning or re-training are needed, in order to prevent risks of overfitting, unintended bias, or degradation. 5) A control sample should be enrolled, especially when the final clinical purpose is diagnosis, differential diagnosis or treatment. 6) Finally, there should be shared best practice guidelines for establishing the ground truth of emotional data.

Also from a clinical point of view, the problems and limitations highlighted in response to RQ2 and RQ3 deserve more attention in the future. 1) There should be shared guidelines on experimental protocols. Clinicians may suggest which are suitable sample size in relation to a certain class disease and final clinical purpose, when there is the need to introduce a control sample, and how to adequately describe the enrolled population. 2) Shared approaches to the rest phase. 3) Some recommendations may be useful about the most suitable emotion model with respect to the application context (e.g. which classes to recognize for a certain clinical purpose and a certain category of patients).

6 Conclusion

The work proposed a systematic literature review of scientific studies investigating automatic emotion recognition in a clinical population composed of at least a sample of people with a disease diagnosis. The addressed RQs allowed to deepen the most important clinical and technical aspects and their relationships. This review allowed to individuate some open issues and guidelines for future works. Research efforts should pursue some important improvements in this field: the choice of an appropriate study design, a more rigorous and robust approach to train, validate and test algorithms, code and datasets sharing, in order to improve reliability, reproducibility and robustness of claimed results, which emerged as a great obstacle for adoption of these techniques in the clinical practice in respect to ethical issues.

The conducted review presents some limitations. A first one concerns the screening process: some works that could satisfy inclusion criteria were possibly skipped by search terms and automation tools. Another limitation concerns the data extraction process, variables that could be relevant to analyse may have been left out. For example, some variables related to the population may need investigation, such as age, gender, socio-cultural and socio-economic background and hereditary factors.

4. https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles

References

[1] R. W. Picard, "Affective computing," MIT Press, Cambridge, Tech. Rep., 1995.

[2] C. Ke, V. W. qun Lou, K. C. kian Tan, M. Y. Wai, and L. L. Chan, "Changes in technology acceptance among older people with dementia: the role of social robot engagement," International Journal of Medical Informatics, vol. 141, p. 104241, 2020.

[3] I. Ayed, A. Ghazel, A. J. i Capó, G. Moyà-Alcover, J. Varona, and P. Martínez-Bueso, "Vision-based serious games and virtual reality systems for motor rehabilitation: A review geared toward a research methodology," International Journal of Medical Informatics, vol. 131, p. 103909, 2019.

[4] R. Yuvaraj, M. Murugappan, and K. Sundaraj, "Methods and approaches on emotions recognition in neurodegenerative disorders: A review," in 2012 IEEE Symposium on Industrial Electronics and Applications, 2012, pp. 287–292.

[5] K. Grabowski, A. Rynkiewicz, A. Lassalle, S. Baron-Cohen, B. Schuller, N. Cummins, A. Baird, J. Podgórska-Bednarz, A. Pieniążek, and I. Łucka, "Emotional expression in psychiatric conditions: New technology for clinicians," Psychiatry and Clinical Neurosciences, vol. 73, no. 2, pp. 50–62, 2019.

[6] C. Grossard, O. Grynspan, S. Serret, A.-L. Jouen, K. Bailly, and D. Cohen, "Serious games to teach social interactions and emotions to individuals with autism spectrum disorders (asd)," Computers & Education, vol. 113, pp. 195 – 211, 2017.

[7] R. Arya, J. Singh, and A. Kumar, "A survey of multidisciplinary domains contributing to affective computing," Computer Science Review, vol. 40, p. 100399, 2021.

[8] K. Woodward, E. Kanjo, D. Brown, T. M. McGinnity, B. Inkster, D. MacIntyre, and T. Tsanas, "Beyond mobile apps: a survey of technologies for mental well-being," IEEE Transactions on Affective Computing, pp. 1–1, 2020.

[9] R. V. Aranha, C. G. Corrêa, and F. L. S. Nunes, "Adapting software with affective computing: a systematic review," IEEE Transactions on Affective Computing, pp. 1–1, 2019.

[10] M. Egger, M. Ley, and S. Hanke, "Emotion recognition from physiological signal analysis: A review," Electronic Notes in Theoretical Computer Science, vol. 343, pp. 35 – 55, 2019.

[11] N. J. Shoumy, L.-M. Ang, K. P. Seng, D. Rahaman, and T. Zia, "Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals," Journal of Network and Computer Applications, vol. 149, p. 102447, 2020.

[12] P. J. Bota, C. Wang, A. L. N. Fred, and H. Plácido Da Silva, "A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals," IEEE Access, vol. 7, pp. 140 990–141 020, 2019.

[13] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and D. Moher, "The prisma 2020 statement: an updated guideline for reporting systematic reviews," BMJ, vol. 372, 2021.

[14] J. P. Higgins and D. G. Altman, Assessing Risk of Bias in Included Studies. John Wiley & Sons, Ltd, 2008, ch. 8, pp. 187–241.

[15] F. P. Akbulut], B. Ikitimur, and A. Akan, "Wearable sensor-based evaluation of psychosocial stress in patients with metabolic syndrome," Artificial Intelligence in Medicine, vol. 104, p. 101824, 2020.

[16] C. Liu, K. Conn, N. Sarkar, and W. Stone, "Physiology-based affect recognition for computer-assisted intervention of children with autism spectrum disorder," International Journal of Human-Computer Studies, vol. 66, no. 9, pp. 662 – 677, 2008.

[17] Y. Gong and C. Poellabauer, "Continuous assessment of children's emotional states using acoustic analysis," in 2017 IEEE International Conference on Healthcare Informatics (ICHI), Aug 2017, pp. 171–178.

[18] K. Bairavi and K. B. K. Sundhara, "Eeg based emotion recognition system for special children," in Proceedings of the 2018 International Conference on Communication Engineering and Technology. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–4.

[19] A. Dawood, S. Turner, and P. Perepa, "Affective computational model to extract natural affective states of students with asperger syndrome (as) in computer-based learning environment," IEEE Access, vol. 6, pp. 67 026–67 034, 2018.

[20] J. Fan, J. W. Wade, A. P. Key, Z. E. Warren, and N. Sarkar, "Eeg-based affect and workload recognition in a virtual driving environment for asd intervention," IEEE Transactions on Biomedical Engineering, vol. 65, no. 1, pp. 43–51, Jan 2018.

[21] S. Sarabadani, L. C. Schudlo, A. Samadani, and A. Kushki, "Physiological detection of affective states in children with autism spectrum disorder," IEEE Transactions on Affective Computing, pp. 1–1, 2018.

[22] N. Rusli, S. N. Sidek, H. M. Yusof, N. I. Ishak, M. Khalid, and A. A. A. Dzulkarnain, "Implementation of wavelet analysis on thermal images for affective states recognition of children with autism spectrum disorder," IEEE Access, vol. 8, pp. 120 818–120 834, 2020.

[23] A. Kushki, A. Khan, J. Brian, and E. Anagnostou, "A kalman filtering framework for physiological detection of anxiety-related arousal in children with autism spectrum disorder," IEEE Transactions on Biomedical Engineering, vol. 62, no. 3, pp. 990–1000, March 2015.

[24] H.-C. Chu, W. W.-J. Tsai, M.-J. Liao, and Y.-M. Chen, "Facial emotion recognition with transition detection for students with high-functioning autism in adaptive e-learning," Soft Comput., vol. 22, no. 9, p. 2973–2999, May 2018.

[25] J. Hashemi, G. Dawson, K. L. H. Carpenter, K. Campbell, Q. Qiu, S. Espinosa, S. Marsan, J. P. Baker, H. L. Egger, and G. Sapiro, "Computer vision analysis for quantification of autism risk behaviors," IEEE Transactions on Affective Computing, pp. 1–1, 2018.

[26] B. Li, S. Mehta, D. Aneja, C. Foster, P. Ventola, F. Shic, and L. Shapiro, "A facial affect analysis system for autism spectrum disorder," in 2019 IEEE International Conference on Image Processing (ICIP), 2019, pp. 4549–4553.

[27] F. Ringeval, E. Marchi, C. Grossard, J. Xavier, M. Chetouani, D. Cohen, and B. Schuller, "Automatic analysis of typical and atypical encoding of spontaneous emotion in the voice of children," in Interspeech 2016, 2016, pp. 1210–1214.

[28] M. Madsen, R. el Kaliouby, M. Goodwin, and R. Picard, "Technology for just-in-time in-situ learning of facial affect for persons diagnosed with an autism spectrum disorder," in Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility. New York, NY, USA: Association for Computing Machinery, 2008, p. 19–26.

[29] C. Tsangouri, W. Li, Z. Zhu, F. Abtahi, and T. Ro, "An interactive facial-expression training platform for individuals with autism spectrum disorder," in 2016 IEEE MIT Undergraduate Research Technology Conference (URTC), 2016, pp. 1–3.

[30] P. Washington, C. Voss, A. Kline, N. Haber, J. Daniels, A. Fazel, T. De, C. Feinstein, T. Winograd, and D. Wall, "Superpowerglass: A wearable aid for the at-home therapy of children with autism," Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 1, no. 3, Sept. 2017.

[31] S. Kuriakose and U. Lahiri, "Design of a physiology-sensitive vr-based social communication platform for children with autism," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 8, pp. 1180–1191, 2017.

[32] O. Rudovic, J. Lee, M. Dai, B. Schuller, and R. W. Picard, "Personalized machine learning for robot perception of affect and engagement in autism therapy," Science Robotics, vol. 3, no. 19, 2018.

[33] S. W. White, L. Abbott, A. T. Wieckowski, N. N. Capriola-Hall, S. Aly, and A. Youssef, "Feasibility of automated training for facial emotion expression and recognition in autism," Behavior Therapy, vol. 49, no. 6, pp. 881 – 888, 2018, integration of Technological Advances in Cognitive-Behavior Therapy.

[34] D. Bian, J. Wade, A. Swanson, A. Weitlauf, Z. Warren, and N. Sarkar, "Design of a physiology-based adaptive virtual reality driving platform for individuals with asd," ACM Trans. Access. Comput., vol. 12, no. 1, Feb. 2019.

[35] T. Y. Tang, P. Winoto, and G. Chen, "Emotion recognition via face tracking with realsense(tm) 3d camera for children with autism," in Proceedings of the 2017 Conference on Interaction Design and Children. New York, NY, USA: Association for Computing Machinery, 2017, p. 533–539.

[36] H. Kalantarian, K. Jedoui, P. Washington, Q. Tariq, K. Dunlap, J. Schwartz, and D. P. Wall, "Labeling images with facial emotion and the potential for pediatric healthcare," Artificial Intelligence in Medicine, vol. 98, pp. 77 – 86, 2019.

[37] B. Banire, D. Al Thani, and M. Qaraqe, "Validation of emotions as a measure of selective attention in children with autism spectrum disorder," in Proceedings of the 2020 9th International Conference on Educational and Information Technology, ser. ICEIT 2020. New York, NY, USA: Association for Computing Machinery, 2020, p. 205–210.

[38] J. Bharatharaj, L. Huang, A. M. Al-Jumaily, C. Krageloh, and M. R. Elara, "Experimental evaluation of parrot-inspired robot and adapted model-rival method for teaching children with autism," in 2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV), 2016, pp. 1–6.

[39] A. Vahabzadeh, N. U. Keshav, J. P. Salisbury, and N. T. Sahin, "Improvement of attention-deficit/hyperactivity disorder symptoms in school-aged children, adolescents, and young adults with autism via a digital smartglasses-based socioemotional coaching aid: Short-term, uncontrolled pilot study," JMIR mental health, vol. 5, no. 2, p. e25, 2018.

[40] J. M. Garcia-Garcia, M. d. M. Cabañero, V. M. R. Penichet, and M. D. Lozano, "Emotea: Teaching children with autism spectrum disorder to identify and express emotions," in Proceedings of the XX International Conference on Human Computer Interaction. New York, NY, USA: Association for Computing Machinery, 2019.

[41] R. Yuvaraj, M. Murugappan, N. M. Ibrahim, K. Sundaraj, M. I. Omar, K. Mohamad, and R. Palaniappan, "Detection of emotions in parkinson's disease using higher order spectral features from brain's electrical activity," Biomedical Signal Processing and Control, vol. 14, pp. 108 – 116, 2014.

[42] S. Zhao, F. Rudzicz, L. G. Carvalho, C. Marquez-Chin, and S. Livingstone, "Automatic detection of expressed emotion in parkinson's disease," in 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2014, pp. 4813–4817.

[43] N. Murali Krishna, K. Sekaran, A. V. Naga Vamsi, G. S. Pradeep Ghantasala, P. Chandana, S. Kadry, T. Blažauskas, and R. Damaševičius, "An efficient mixture model approach in brain-machine interface systems for extracting the psychological status of mentally impaired persons using eeg signals," IEEE Access, vol. 7, pp. 77 905–77 914, 2019.

[44] L. Pepa, M. Capecci, and M. G. Ceravolo, "Smartwatch based emotion recognition in parkinson's disease," in 2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT), 2019, pp. 23–24.

[45] A. Bandini, S. Orlandi, H. J. Escalante, F. Giovannelli, M. Cincotta, C. A. Reyes-Garcia, P. Vanni, G. Zaccara, and C. Manfredi, "Analysis of facial expressions in parkinson's disease through video-based automatic methods," Journal of Neuroscience Methods, vol. 281, pp. 7 – 20, 2017.

[46] S. L. Happy, A. Dantcheva, A. Das, R. Zeghari, P. Robert, and F. Bremond, "Characterizing the state of apathy with facial expression and motion analysis," in 2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019), 2019, pp. 1–8.

[47] J. J. Rivas, F. Orihuela-Espina, L. E. Sucar, L. Palafox, J. Hernández-Franco, and N. Bianchi-Berthouze, "Detecting affective states in virtual rehabilitation," in 2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2015, pp. 287–292.

[48] J. Sidorova, S. Karlsson, O. Rosander, M. Berthier, and I. Moreno-Torres, "Towards disorder-independent automatic assessment of emotional competence in neurological patients with a classical emotion recognition system: application in foreign accent syndrome," IEEE Transactions on Affective Computing, pp. 1–1, 2019.

[49] S. Bong, K. Wan, M. Murugappan, N. Mohamed Ibrahim, Y. Rajamanickam, and K. Mohamad, "Implementation of wavelet packet transform and non linear analysis for emotion classification in stroke patient using brain signals," Biomedical Signal Processing and Control, vol. 36, pp. 102–112, 7 2017.

[50] R. Khosla and M.-T. Chu, "Embodying care in matilda: An affective communication robot for emotional wellbeing of older people in australian residential care facilities," ACM Trans. Manage. Inf. Syst., vol. 4, no. 4, dec 2013. [Online]. Available: https://doi.org/10.1145/2544104

[51] J. Fan, D. Bian, Z. Zheng, L. Beuscher, P. A. Newhouse, L. C. Mion, and N. Sarkar, "A robotic coach architecture for elder care (rocare) based on multi-user engagement models," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 8, pp. 1153–1163, 2017.

[52] H. Huang, Q. Xie, J. Pan, Y. He, Z. Wen, R. Yu, and Y. Li, "An eeg-based brain computer interface for emotion recognition and its application in patients with disorder of consciousness," IEEE Transactions on Affective Computing, pp. 1–1, 2019.

[53] Peng Wang, C. Kohler, E. Martin, N. Stolar, and R. Verma, "Learning-based analysis of emotional impairments in schizophrenia," in 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008, pp. 1–8.

[54] Z. N. Karam, E. M. Provost, S. Singh, J. Montgomery, C. Archer, G. Harrington, and M. G. Mcinnis, "Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech," in 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014, pp. 4858–4862.

[55] G. Valenza, M. Nardelli, A. Lanatà, C. Gentili, G. Bertschy, R. Paradiso, and E. P. Scilingo, "Wearable monitoring for mood recognition in bipolar disorder based on history-dependent long-term heart rate variability analysis," IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 5, pp. 1625–1635, Sep. 2014.

[56] A. Grünerbl, A. Muaremi, V. Osmani, G. Bahle, S. Öhler, G. Tröster, O. Mayora, C. Haring, and P. Lukowicz, "Smartphone-based recognition of states and state changes in bipolar disorder patients," IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 1, pp. 140–148, 2015.

[57] E. Çiftçi, H. Kaya, H. Güleç, and A. A. Salah, "The turkish audio-visual bipolar disorder corpus," in 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia), 2018, pp. 1–6.

[58] M. Ebrahim, M. Al-Ayyoub, and M. Alsmirat, "Determine bipolar disorder level from patient interviews using bi-lstm and feature fusion," in 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS), 2018, pp. 182–189.

[59] Z. S. Syed, K. Sidorov, and D. Marshall, "Automated screening for bipolar disorder from audio/visual modalities," in Proceedings of the 2018 on Audio/Visual Emotion Challenge and Workshop, ser. AVEC'18. New York, NY, USA: Association for Computing Machinery, 2018, p. 39–45.

[60] Z. Ren, J. Han, N. Cummins, Q. Kong, M. D. Plumbley, and B. W. Schuller, "Multi-instance learning for bipolar disorder diagnosis using weakly labelled speech data," in Proceedings of the 9th International Conference on Digital Public Health, ser. DPH2019. New York, NY, USA: Association for Computing Machinery, 2019, p. 79–83.

[61] N. Abaei and H. A. Osman, "A hybrid model for bipolar disorder classification from visual information," in ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 4107–4111.

[62] K. Huang, C. Wu, M. Su, and Y. Kuo, "Detecting unipolar and bipolar depressive disorders from elicited speech responses using latent affective structure model," IEEE Transactions on Affective Computing, pp. 1–1, 2018.

[63] M. Gavrilescu and N. Vizireanu, "Predicting depression, anxiety, and stress levels from videos using the facial action coding system," Sensors (Basel), vol. 19, no. 17, p. 3693, 2019.

[64] Z. Jiang, S. Harati, A. Crowell, H. Mayberg, S. Nemati, and G. Clifford, "Classifying major depressive disorder and response to deep brain stimulation over time by analyzing facial expressions," IEEE Transactions on Biomedical Engineering, pp. 1–1, 2020.

[65] J. F. Cohn, L. A. Jeni, I. Onal Ertugrul, D. Malone, M. S. Okun, D. Borton, and W. K. Goodman, "Automated affect detection in deep brain stimulation for obsessive-compulsive disorder: A pilot study," in Proceedings of the 20th ACM International Conference on Multimodal Interaction, ser. ICMI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 40–44.

[66] O. Balan, G. Moise, A. Moldoveanu, M. Leordeanu, and F. Moldoveanu, "An investigation of various machine and deep learning techniques applied in automatic fear level detection and acrophobia virtual therapy," Sensors (Basel), vol. 20(2), p. 496, 2020.

[67] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 10, pp. 1175–1191, Oct 2001.

[68] B. Taati, S. Zhao, A. B. Ashraf, A. Asgarian, M. E. Browne, K. M. Prkachin, A. Mihailidis, and T. Hadjistavropoulos, "Algorithmic bias in clinical populations—evaluating and improving facial analysis technology in older adults with dementia," IEEE Access, vol. 7, pp. 25 527–25 534, 2019.

[69] A. Dawood, S. Turner, and P. Perepa, "Natural-spontaneous affective-cognitive dataset for adult students with and without asperger syndrome," IEEE Access, vol. 7, pp. 77 990–77 999, 2019.

[70] P. Lang, M. Bradley, and B. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual," University of Florida, Gainesville, FL., Tech. Rep. A-8, 2008.

[71] M. Bradley and P. Lang, "The international affective digitized sounds (2nd edition; IADS-2): Affective ratings of sounds and instruction manual," University of Florida, Gainesville, FL., Tech. Rep. B-3, 2008.

[72] E. S. Dan-Glauser and K. R. Scherer, "The geneva affective picture database (gaped): a new 730-picture database focusing on valence and normative significance," Behavior Research Methods, vol. 43, no. 2, pp. 468–477, 2011.

[73] P. Ekman, E. Sorenson, and W. Friesen, "Pan-cultural elements in facial displays of emotion," Science, vol. 164, no. 3875, pp. 86–88, 1969.

[74] J. A. Russell, "A circumplex model of affect," Journal of Personality and Social Psychology, vol. 39, no. 6, pp. 1161–1178, 1980.

[75] P. Ekman and W. V. Friesen, Facial action coding system: a technique for the measurement of facial movement. Palo Alto: Consulting Psychologists Press, 1978.

[76] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from eeg," IEEE Transactions on Affective Computing, vol. 5, no. 3, pp. 327–339, 2014.

[77] J. Shukla, M. Barreda-Angeles, J. Oliver, G. C. Nandi, and D. Puig, "Feature extraction and selection for emotion recognition from electrodermal activity," IEEE Transactions on Affective Computing, pp. 1–1, 2019.

[78] D. Nikolova, P. Petkova, A. Manolova, and P. Georgieva, "Ecg-based emotion recognition: Overview of methods and applications," in ANNA '18; Advances in Neural Networks and Applications 2018, 2018, pp. 1–5.

[79] M. Mohandes, M. Deriche, and S. O. Aliyu, "Classifiers combination techniques: A comprehensive review," IEEE Access, vol. 6, pp. 19 626–19 639, 2018.

[80] S. J. Russell and P. Norvig, Artificial intelligence: A modern approach. Englewood Cliffs, N.J: Prentice Hall, 1995.

[81] A. Joshi, S. Ghosh, S. Gunnery, L. Tickle-Degnen, S. Sclaroff, and M. Betke, "Context-sensitive prediction of facial expressivity using multimodal hierarchical bayesian neural networks," in 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), May 2018, pp. 278–285.

[82] T. Sappok, M. Heinrich, and J. Böhm, "The impact of emotional development in people with autism spectrum disorder and intellectual developmental disability," Journal of Intellectual Disability Research, vol. 64, no. 12, pp. 946–955, 2020.

[83] S. F. Lerman, G. Bronner, O. S. Cohen, S. Elincx-Benizri, H. Strauss, G. Yahalom, and S. Hassin-Baer, "Catastrophizing mediates the relationship between non-motor symptoms and quality of life in parkinson's disease," Disability and Health Journal, vol. 12, no. 4, pp. 673 – 678, 2019.

[84] C. Conner, S. White, K. Beck, J. Golt, I. Smith, and C. Mazefsky, "Improving emotion regulation ability in autism: The emotional awareness and skills enhancement (ease) program," Autism, vol. 23, no. 5, pp. 1273–1287, 2019.

[85] M. Jazouli, A. Majda, and A. Zarghili, "A $p recognizer for automatic facial emotion recognition using kinect sensor," in 2017 Intelligent Systems and Computer Vision (ISCV), April 2017, pp. 1–5.

[86] J. M. van Uem, J. Marinus, C. Canning, R. van Lummel, R. Dodel, I. Liepelt-Scarfone, D. Berg, M. E. Morris, and W. Maetzler, "Health-related quality of life in patients with parkinson's disease—a systematic review based on the icf model," Neuroscience & Biobehavioral Reviews, vol. 61, pp. 26 – 34, 2016.

[87] G. Mattavelli, E. Barvas, C. Longo, F. Zappini, D. Ottaviani, M. C. Malaguti, M. Pellegrini, and C. Papagno, "Facial expressions recognition and discrimination in parkinson's disease," Journal of Neuropsychology, 2020.

[88] M. Leboyer and D. Kupfer, "Bipolar disorder: new perspectives in health care and prevention," J Clin Psychiatry, vol. 71, no. 12, pp. 1689–1695, 2010.

[89] H. Critchley and S. Garfinkel, "Interoception and emotion," Curr Opin Psychol, vol. 17, pp. 7–14, 2017.

[90] H. Brace, "The feeling of what happens: body and emotions in the making of consciousness," Psychiatric Services, vol. 51, no. 12, p. 1579–1579, 2000.

[91] S. Leonardi, A. Cacciola, R. De Luca, B. Aragona, V. Andronaco, D. Milardi, P. Bramanti, and R. Calabro, "The role of music therapy in rehabilitation: improving aphasia and beyond," Int J Neurosci, vol. 128, no. 1, pp. 90–99, 2018.

[92] L. Morera, G. Marchiori, L. Medrano, and M. Defago, "Stress, dietary patterns and cardiovascular disease: A mini-review," Front Neurosci, vol. 12, no. 13, p. 1226, 2019.

[93] Committee on Technology National Science and Technology Council and Penny Hill Press, Preparing for the Future of Artificial Intelligence. North Charleston, SC, USA: CreateSpace Independent Publishing Platform, 2016. [Online]. Available: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

[94] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. Version 2," Tech. Rep., Dec. 2017. [Online]. Available: http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html.

[95] High-Level Expert Group on AI, "Ethics guidelines for trustworthy ai," European Commission, Brussels, Report, Apr. 2019. [Online]. Available: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[96] R. Schwartz, L. Down, A. Jonas, and E. Tabassi, "A proposal for identifying and managing bias within artificial intelligence," National Institute of Standards and Technology, U.S., Report, Jun. 2021. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270-draft.pdf

Lucia Pepa received in 2012 the Master degree in Electronic Engineering, and in 2016 the Ph.D. degree in E-learning – Technology Enhanced Learning from the Universita' Politecnica delle Marche (UNIVPM), Italy. She is currently postdoc researcher at UNIVPM, her primary research interests involve affective computing and movement analysis through consumer electronics devices.

Luca Spalazzi (PhD'94, MEng'89) is associate professor at the Università Politecnica delle Marche, Italy. He got the Scientific National Italian Habilitation as Full Professor (2020). He worked as a consultant at the IRST-FBK, Trento, Italy (1991-1993, 1997). His research areas include formal methods and artificial intelligence applied to software engineering, cybersecurity and privacy, and e-health. He is currently involved in the Cariverona Project "Remote rehAbilitation for ParkInson's Disease at any stage".

Marianna Capecci (MD, PhD): Associate Professor of Physical and Rehabilitation Medicine at the Department of Experimental and Clinical Medicine - University "Politecnica" of Marche. Main research interests: rehabilitation, Parkinson's disease and movement disorders, telemonitoring of motor and non-motor disorders, tele-rehabilitation, gait and posture analisys. Publications:https://orcid.org/0000-0002-1472-606X

Maria Gabriella Ceravolo , MD, PhD is a Neurologist and, Full professor in Physical and Rehabilitation Medicine, at "Politecnica delle Marche" University. Her research interests include the epidemiology of neurologic disease–related disability, the effectiveness of stroke rehabilitation through action observation, virtual reality, functional electrical stimulation and non-invasive cortical stimulation, the effectiveness of deep brain stimulation in Parkinson's disease and Epilepsy, the instrumental assessment of posture and gait, the neural correlates of financial decisions, and, more recently, the epidemiology of COVID-19–related disability and rehabilitation needs.