Emotion Modelling for Social Robotics Applications: A Review

Filippo Cavallo[1*], Francesco Semeraro[1], Laura Fiorini[1], Gergely Magyar[2], Peter Sinčák[2], Paolo Dario[1]

1. *BioRobotics Institute, Scuola Superiore Sant'Anna, Pontedera 56025, Italy*
2. *Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Kosice 04001, Slovakia*

Abstract

Robots of today are eager to leave constrained industrial environments and embrace unexplored and unstructured areas, for extensive applications in the real world as service and social robots. Hence, in addition to these new physical frontiers, they must face human ones, too. This implies the need to consider a human-robot interaction from the beginning of the design; the possibility for a robot to recognize users' emotions and, in a certain way, to properly react and "behave". This could play a fundamental role in their integration in society. However, this capability is still far from being achieved. Over the past decade, several attempts to implement automata for different applications, outside of the industry, have been pursued. But very few applications have tried to consider the emotional state of users in the behavioural model of the robot, since it raises questions such as: how should human emotions be modelled for a correct representation of their state of mind? Which sensing modalities and which classification methods could be the most feasible to obtain this desired knowledge? Furthermore, which applications are the most suitable for the robot to have such sensitivity? In this context, this paper aims to provide a general overview of recent attempts to enable robots to recognize human emotions and interact properly.

Keywords: social robotics, service robotics, human-robot interaction, emotion recognition, robot learning, robot behavioural model

1 Introduction

According to recent statistics[1], robots have become more common in our daily life. In a not too distant future, robots will concretely share environments with human beings to actively collaborate with them in specific daily tasks, such as, for example, the management of daily activities[2], the promotion of social inclusion[3] and the suggestion of healthy activities[4]. As social robots need to work closely with humans, they are expected to interact and communicate with them by embodying human social behaviours and rules and to have high levels of acceptability in providing useful and efficient services[5]. In this context, the capability to understand and recognize human emotions and intentions plays a fundamental role in the development of acceptable robotic solutions. In fact, emotions are crucial elements of our life, while facilitating the interplay between the organism and the environment. When two or more human beings interact they use verbal and non-verbal cues, such as body, posture, gesture, gaze and can also introduce psychological aspects which can depend on thought, cultural and social background[6].

Social robotics is a growing branch of robotics, where also not technical disciplines, such as psychology and sociology, converge. The importance of social robotics in the design and development of robot abilities in acting and interacting physically, emotionally, socially, and safely with humans has generally been poorly understood so far[7]. Social robotics extends this concept by leading to the unexplored field of Human Robot Interaction (HRI). A robot could modify its behaviour according to a user's humour, or even more, express a related emotion, resulting in an increase of synergy with the owner in the cooperation. In this sense, enabling robots to sense and model emotions will improve their performance across a wide variety of human-robot interaction applications[8]. Consequently, with these improved HRI abilities, robots will not be perceived as cold impersonal entities and will earn the designation of social robots. In this approach, the robot will interact and communicate with humans or other autonomous physi-

---

*Corresponding author: Filippo Cavallo
E-mail: filippo.cavallo@santannapisa.it

cal agents by means of social behaviours and rules.

Other important aspects which will influence the acceptance and the usage of robotic solutions are the appearance and embodiment properties. A social robot does not necessarily have to be a humanoid. In fact, Gray and Wegner[9], through statistical experiments based on the uncanny valley theory, demonstrated that, too humanlike appearance of robots is more likely to unnerve people. This also strongly depends on people cultural factors, nonetheless by age and gender[10]. Through their progressive insertion in social life over time and a technological empowerment, fear of them will eventually fade away, and even humanoids will be accepted in communities. Ongoing studies about this topic have evaluated the appropriate terms with which to measure the comfort of people in interaction with robots[11,12]. However, some humanlike elements should be present in the very first prototypes. This was pointed out by Donald Norman[13], who felt that object design should be pleasing to users, to appear more effective. This surely is related to the concept of human-centred design. A first element, which could also be the main interaction vector between robots and humans, is a humanlike head. Through this, a person would be more naturally convinced to convey his/her thoughts toward the robot and, conversely, the robot could use it to express its own internal state to the person[14].

However, numerous and ambitious challenges remain in achieving such results. The first important question concerns the modalities needed to sense the emotional state of people by the robot. Secondly, there is the problem of modelling the interaction between human and robot, not only on a haptic level, but also on an emotional level. A simple motivation is the lack of a proper definition of emotions. Furthermore, as social robotics involve different disciplines, such as computer science and artificial intelligence, but also psychology and cognitive science, no common methodology has been established for research[15].

Finally, one issue is fundamental: once having achieved this knowledge, how should the robot use it to improve relations with humans? Should it resonate with the owner's state of mind or, for instance, try to relieve him/her from a negative condition[14]? Should the robot use such knowledge only to produce an emotion or should its behaviour be modified as consequence[7]? In this paper, an overview about how engineers and scientists worldwide have recently faced all these puzzling questions is presented.

The rest of this paper is organized as follows. In section **2**, the research methodology for the review is explained. In section **3**, the most common theories to model human emotions are described. In section **4**, a discussion about sensing modalities is presented. In section **5**, dissertations about features extracted from sensors' signals are given. In section **6**, learning models used to build the emotion recognition process are discussed. Finally, section **7** talks about the applications of emotion-sensitive robots and discusses possible developments in this relevant topic of robotics.

2 Methodology: Search strategy and selection criteria

References were included into the article through the research and screening procedures shown below. An electronic database search was performed from February 2017 to May 2017 using IEEE Xplore, Pubmed, Science Direct, Scopus, and Web of Knowledge databases to identify articles concerning aspects of modelling, recognizing, interpreting, and implementing emotions in social robotics applications. Specifically, terms and keywords used for the literature research were emotion and robot and [nothing or (recognition and model) or interaction], emotion and (classification or model) in default settings for each database. Three main topics were identified: how emotions can be modelled from a medical point of view; which signals from the human body are used to reveal emotions; and which types of sensors or artificial intelligence techniques are used to classify, recognize, or predict emotions. Papers were screened from independent reviewers (the authors) and disagreements were solved through meetings and discussions. Data abstracted from the papers and reported in Table 1 are considered: emotional theoretical models, sensing processing and extracted features, classification techniques, experimental setup, and applications.

Concerning selection criteria, first, results were sorted in default order for each database. Second, a maximum of 100 results for each combination of keywords for each database was initially withheld. Then,

Table 1 Schematic synthesis of literature cases for robot design models considered in the review

Title	Theory	Sensing	Processing	Classification	Experimental setup	Application
Ref. [14]	Ekman	Vision (face)	Constrained Local Model (CLM)	HMM	CONBE robot has 2 7-DOF arms and a head with 2 DOF. It copes with emotion expressed by a human while it is approaching an object.	Service robotics (cooperative companion)
Ref. [15]	Ekman	Vision (face), sound	Local Binary Pattern (LBP)	SVM	Free interaction with robot in a stage of I2R exposition	Affective social robots
Ref. [18]	Valence-arousal	Sound, temperature, time, pressure	Not reported	Fuzzy cognitive maps	Simulated crowded environment	Emotion forecasting for a social robot
Ref. [20]	Valence, arousal and control	Vision (face), neural activity	Recursive Feature Elimination (RFE), Independent Component Analysis (ICA), Actions Units	Gaussian Naive Bayes classifier and linear regression combined in weighted fashion	Video and EEG recordings from Manhob HCI dataset, where 30 participants react to 20 clips from Hollywood movies	Emotion recognition
Ref. [21]	Ekman	Vision (face)	CLM and LBP for feature extraction, PCA, Factor analysis and locality preserving projection for feature reduction	NARX	Audio and visual recordings from SEMAINE dataset, where participants interact with four stereotyped personalities performed by actors	Emotion recognition for social robots
Ref. [22]	Arousal, valence and stance	Vision (face)	Gabor LBP	k-NN	Instances from FERET dataset in first experiment, then 15 elder subjects were asked to spend time in the smart environment	Eldercare robot detecting MCI and controlling parameters of a smart environment
Ref. [24]	Hourglass of emotion	Vision (face), sound, typed text	Not reported	Extreme Learning Machine (ELM)	Video clips from USC IEMOCAP dataset	Emotion recognition
Ref. [28]	Ekman	Vision (body postures and gestures)	Laban movement analysis, particularly quantity of motion and expansion index	Not reported	In the first experiment, a dancer performed dance sessions in scenario designed by him to invoke different emotions; in the second one, an actress played distinct roles in improvisation scenarios; recordings through Kinect	Emotion recognition for social robots
Ref. [29]	Ekman, Myers-Briggs	EDA	SAX representation	HMM, iBk	30 participants divided in two groups. First one has emotion elicited by an actor's performance, the second one through movements of a NAO robot put in front of him/her	Service robot (cooperative companion)
Ref. [30]	Ekman	Typed text	Not reported	Unknown; tuning of robot's behaviour according to level of emotions, mood, and attitude	Robot interacted with several people, known and unknown, asking for information, with different attitudes	Service robot (receptionist)
Ref. [32]	Ekman	Vision (face)	Gabor features, then selected through Adaboost algorithm	Error-Correcting Output Codes (ECOC) classifiers	Instances from Cohan-Kanade dataset	Emotion recognition for social robots
Ref. [33]	Ekman	Vision (face)	Not reported	Probabilistic classifier based on new neuroscience theories	Dataset generated through FaceGen software	Emotion recognition
Ref. [34]	Ekman	Vision (body postures and gestures)	Skin colour detection, motion detection, rectangular region marking, tracking	Laban Movement Analysis	Hand-waving experiments within scenarios, where different emotions and mental states have been enacted, recorded through a camera	Service robot (therapist)
Ref. [35]	PAD	Vision (face, body postures and gestures), sound	Mapping in Speed Intensity iRregularity and Extent (SIRE) model	Gaussian Markov Model (GMM)	10 subjects' utterances from EMO-DB dataset	Emotion recognition for social robots

Table 1 Continued

Ref. [36]	Ekman	Sound	Multifractal analysis	ANN and SVM, combined through Adaboost algorithm	Recordings processed in MATLAB	Emotion recognition for social robots
Ref. [37]	Ekman	Sound	Classic speech features, mainly coming from intensity, spectrogram, pitch, Mel's coefficients, HNR and LTAS chosen through attribute selection	SVM, Bayes classifier and RBF tested separately	Instances from database of polish emotional speech. The emotion recognition process was divided in speaker-independent and speaker-dependent, to assess improvement of performance in exploiting prior knowledge of speaker; people were asked to express emotions, by reading sentences	Emotion recognition for social robots
Ref. [38]	Ekman	Sound	Optimization of classification through Genetic Algorithms	GMM, SVM, HMM	Instances from EMO-DB dataset	Emotion recognition for social robots
Ref. [39]	Ekman	Vision (face), sound	PCA, optimization through GA	ANN, HMM	First training on purposely made dataset (Mexican dataset); then, dialogue interaction with Bioloid premium robot	Emotion recognition for social robot
Ref. [40]	Valence-Arousal	Heart activity	Standard and nonlinear HRV features	Quadratic Discriminant Classifier (QDC)	27 subjects were elicited in different emotions through sounds from the IADS dataset, classified by level of arousal, separated by neutral sounds	Emotion recognition
Ref. [41]	Plutchik	Heart activity	Features which separately characterize each heart wave interval	Linear Discriminant Classifiers	In the experiment, 60 pictures were shown to subjects individually. Each trial began with a preparation. ECG electrodes were attached on subject's breast. Then, picture to be rated was presented for 6 s, and immediately after the picture left the screen, the subjects made their ratings.	Emotion recognition for social robots
Ref. [42]	Ekman	Heart activity	Noise of acquisition	Decision tree based on Gini index	Participants were asked to watch disgusting, fearful, and neutral movies while recorded through ECG electrode	Emotion recognition
Ref. [43]	Ekman	Neural activity, heart activity, EDA	Several classic features, then selected through sequential forward floating selection algorithm	SVM and KNN, fused in classification	A first pool of subjects was required to tag video clips; then, the second subject viewed clips and had their physiological signals recorded	Emotion recognition
Ref. [44]	Valence-arousal	Neural activity	Basic EEG features from frequency domain and others from multifractal domain	Three different fusion-based classifiers	Instances from DEAP dataset	Emotion recognition
Ref. [45]	Valence-arousal	Thermal state	Smoothing of five ROIs, then feature selection through GA	Linear discriminant analysis on 6 binary classifiers	In a controlled room, subjects were shown different visual stimuli from IAPS dataset, while seated in front of the monitor. Their reactions were recorded through an infrared camera	Emotion recognition
Ref. [46]	Unknown	EDA	SAX representation	HMM	The first group of subjects interacted with a NAO robot, which executed scripted movements and talking; the second group had same type of interaction with a human actor	Emotion recognition for social robots
Ref. [48]	Ekman	Vision (face)	Not reported	CNN	Not reported	Emotion recognition for social robots
Ref. [51]	Ekman	Vision (face)	HOG histograms	SVM	The emotion recognition was initially trained for 7 emotions, by using the Cohn-Kanade dataset; then children were placed in front of the R25 robot and asked to repeat expressions performed by robot, that were only 4, 5 times each	Service robot (therapist)
Ref. [52]	Valence-arousal	Vision (face)	Facial detection, Hue saturation values, PCA	Based on eigen-faces	Instances from a dataset	Emotion recognition

Table 1 Continued

Ref. [55]	Ekman	Vision (face)	Pyramid of LBP	SVM, KNN, Naive Bayesian, Random Forest	First experiment was performed on the extended Cohn–Kanade (CK+) database. Second experiment was performed to test the performance of the proposed framework on low resolution image sequence. Third experiment tests the robustness of the proposed framework when generalizing on a new dataset. Fourth experiment was performed on the MMI facial expression database which contains spontaneous/natural expressions.	Emotion recognition for social robots on low resolution images
Ref. [53]	Ekman	Vision (face)	Active appearance model, optical flow, dynamic face detection	Based on eigenfaces	People are positioned in front of a camera and asked to perform different facial expressions related to different emotions	Emotion recognition for social robots
Ref. [56]	Ekman	Vision (face)	Radon transform, EMD-based feature extraction	SVM, RBF	Instances from CK+ and JAFFE Dataset	Emotion recognition
Ref. [58]	Plutchik	Vision	Not reported	SOM	The robot must grab a ball held by a person	Behavioural model
Ref. [59]	Ekman	Vision (face)	Not reported	Not known; SOM for behaviour production	First part of experiment dealt with simulations in Matlab 2010; in the second part, the robot tuned its behaviour on grasping an object according to emotion felt by the user	Service robot (cooperative companion)
Ref. [60]	Valence-arousal	Neural activity, EDA, muscular activity	Several classic features	Stacked autoencoders	Instances from DEAP dataset	Emotion recognition
Ref. [61]	Ekman	Vision (face)	Gray scale conversion, face detection and cropping, pooling feature extraction	DCNN	Instances from CK+ and JAFFE datasets	Emotion recognition
Ref. [62]	Ekman	Vision (face)	Sobel filters and others	DCNN	Instances from FABO dataset	Emotion recognition for social robots
Ref. [65]	Ekman	Vision (face)	Several techniques of dimensionality reduction	Several classifiers	Cohan-Kanade and JAFFE dataset for training; then the robot wanders around the laboratory, accomplishing tasks and recognizing emotions in people it has never seen before	Service robot
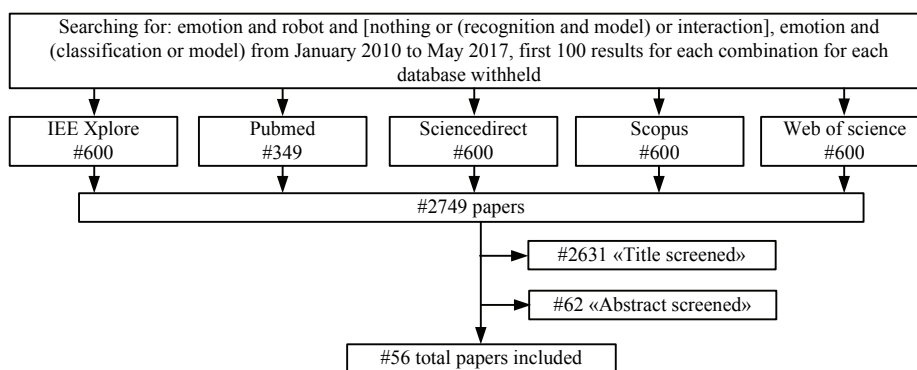
Fig. 1 Research methodology for cited references.

duplicate references were excluded. During the screening procedure, items were excluded if they were an abstract, a chapter from a book, not written in English language and not full access. Then, references were fully assessed during the evaluation procedure, and papers were excluded if they did not appear appropriate for this review after the reading of title and abstract. Particularly, a paper was discarded if: 1) it did not include words that

were strongly related to the purpose of this review, *i.e.* "emotion"," affection" and "social" and their derivatives or synonyms; 2) it focused on cognitive interaction but other aspects than emotion recognition were investigated; 3) it did not include the presence of humans in the experimental setup. However, concerning this third point, few papers were retrieved, if considered helpful in consolidating some specific concepts. Finally, 56 papers were fully evaluated and included in this review (Fig. 1). Only original, full-text articles published in English, dated between January 2010 and May 2017 were included in this review paper. Obtained in the research were 13 references from IEEE Xplore, 3 from PubMed, 18 from Science Direct, 15 from Scopus, and 7 from Web of Knowledge.

3 Emotional models

To embed the capability of recognizing emotions into a robot, a preliminary discussion about how they should be represented is compulsory. In most works regarding social robotics, representation of emotions has two main theories. The most famous theory was proposed by Ekman and Friesen[16]. According to them, every emotive expression can be classified into one of six main categories: anger, disgust, fear, happiness, sadness, and surprise, plus a neutral one (Fig. 2). Later, he added other emotional states (fun, contempt, contentment, embarrassment, excitement, guilt, pride, relief, satisfaction, pleasure, shame), called secondary emotions. However, in emotion recognition, only the six historical emotions are considered for two main reasons. The first one is strictly related to the effort to instantiate an efficient multi-class recognition problem of such complexity. As concern the second reason, taking a closer look at the most recent ones, they can be related to the six historical ones, at least in computational terms. Ekman *et al*. stated that they should be considered as independent states of mind. In final expressions of emotion, features like sets of microfacial expressions, related to each different emotion, can of course overlap.

The other school of thought is related to Russell's theory[19]. He asserted that emotions can be retrieved in empirical sections of a plane of two main coordinates: valence, which gives positive or negative connotation to the emotional state, and arousal, which measures the
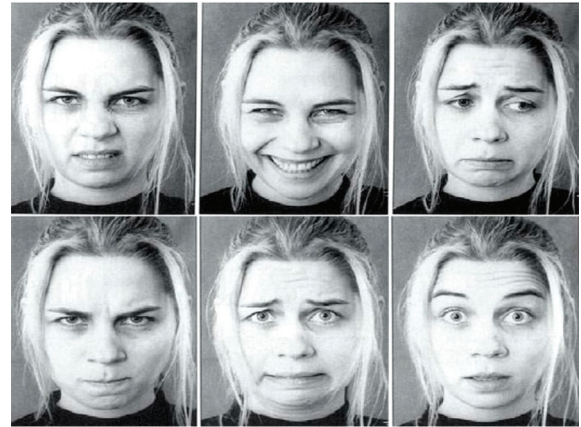

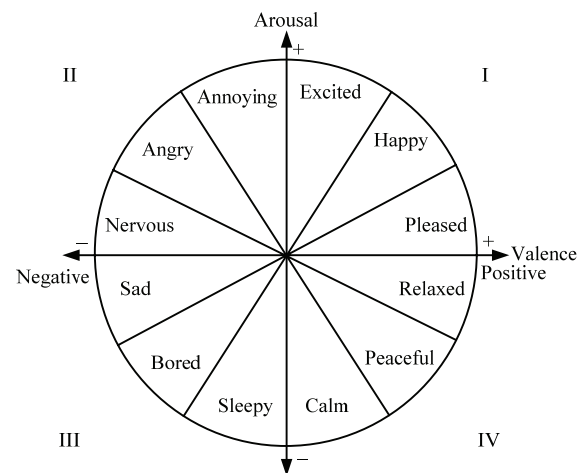Fig. 2 The historical 6 emotions pointed out by Ekman[17].

Fig. 3 The valence-arousal model[18].

intensity of the feeling (Fig. 3). He ended up defining several emotional states, as well. Hence, to avoid a disproportionate amount of emotions to handle, for the previously mentioned reasons, the six historical emotions initially mentioned by Ekman are considered. A reason to prefer Russell's theory to Ekman's is its better suitability for a classification problem, thanks to its implicit reference to a hyper-plane separation problem.

A few research groups decided to follow different schematizations, inspired by the two described here. For instance, Koelstra and Patras[20] decided to model the space of emotion, by adding, besides arousal and valence, a third dimension, called control, which refers to the capability of the person to handle the emotion. Banda *et al*.[21] used a third additional dimension to Russell's model, called stance, in the behavioural model of the Kismet robot. This is inspired by Mehrabian's theory of

the PAD model (pleasure, arousal, dominance)[22]. Moreover, other emotional models in literature can be used as inspiration to build the emotion recognition problem in robots. These include the hourglass of emotions, also known as Plutchik's model[23–25]. Furthermore, looking toward neuroscience to search for new models should be encouraged[26,27]. However, the use of categorical emotions in the classification process has begun to be revised, as this approach has several limitations. Lewis and Canamero[28] said that this categorization can create mistakes in interactional cases in which very different movements represent the same emotion. This is mediated by the situation in which the emotion recognition takes place. It might also depend on people's personalities. This last aspect was considered by Henriques et al.[29], by tuning interaction of their robots based on the user's profile according to Myers-Briggs theory. This theory states that 16 personalities exist in the world, and these personalities differentiate from each other by attitude, perceptual and judgement functions, and lifestyle.

However, even if a robot could perfectly understand humans' emotions, this would not be enough to achieve a complete social robot. As it should be a companion, it strongly needs to build long-term relationships[30], which are not considered by the models described so far. Having internal models of robots that account for such relationships would strongly impact its social navigation. For instance, if a robot continuously relates the same person with a positive emotion, it could consider the situation as less risky to be closer to him/her, prior to visualizing an emotion from him/her. Furthermore, differentiated behaviours can be imposed to the robot, according to the allowed distance from users. For this purpose, detailed studies about proxemity should be considered. If the emotion related was disgust, the robot could try to avoid him/her. Kirby et al.[30] built a robot in which concepts of mood and attitude were modelled, together with emotions. In their work, mood was intended as a modulator of intensity of emotions felt by the robot, triggered by a certain event. Depending on the event, this mood fades over time at different speeds. Attitude was modelled as a long-term mood related to a person. It does not have to be extinguished so easily, for the attitude to be consistent.

This leads to a more general concept: the awareness of the context. It is a key concept: the outcome of the decision-making process implemented in robots can radically change, according to the outbound conditions of the environment. For instance, if a person is scared by watching a horror movie, a robot should not be alarmed in recognizing an expression of fear. This concept was pointed out by Lewis and Canamero[28] and was accurately considered by Raymundo et al.[26]. Particularly, they discerned between context and situation. Context-aware systems focus on characterizing the environment, whereas situation-aware systems try to determine when some entities are in a relevant state, to intervene. This suggests a good approach to model the representation of the context.

Recommendations and Trends

The majority of works regarding emotion recognition settle their theoretical background in Ekman's[16] or Russell's[19] theories. The first is a categorical model, based on six independent categories (anger, fear disgust, joy, sadness and surprise); the second is a two-dimension model, mapping emotions in terms of valence and arousal, which represent, respectively, the connotation and dimension of feeling. In some extensions of this model, a third dimension called stance, which is related to the capability of handling the feeling, is considered[21,22].

The common definition of emotion says that it is an instantaneous affective response to an experienced event. On the contrary, mood is a longer lasting affective response, which has not a specific input and clear starting point and has low intensity. If it is triggered not only by current events, but also by memory of past events, it is called attitude[30]. In this review, studies investigating long-term affective models were not identified, confirming that mood modelling and theories applied to social robotics represents an interesting research challenge.

Most of the works used Ekman and Russell models, while few of them used alternative theories to model their work. For instance, Plutchik's model[23] uses hybrid models that merge positive aspects of both theories. In fact, that model firstly identifies emotions within eight specific sectors; then, according to the feeling's intensity, the emotion can be categorized in three additional sub-

levels in each sector.

Another research challenge in emotion recognition concerns the use of context awareness elements, since environments, situations and people can consistently influence the perception and production of emotions. In this sense, only two papers have showed a preliminary modelling of emotions by considering context awareness inputs. The work of Banda et al.[21] represented a preliminary tentative to use appraisal-based models in emotion recognition mentioned, by accounting individual variances of emotional reactions to the same events. Jing et al.[22], instead, used the appraisal approach of the Ortony, Clore & Collins (OCC) model that considers context treats the key aspects of situations as triggers that cause an emotion.

4 Sensing modalities

Social robotics is supported in research by affective computing, which is a branch of the data mining field whose aim is the instantiation of processes of emotion recognition[31]. This type of service is generally used in economic surveys to obtain feedback on products from customers, in commercials, and in medical applications, particularly in the monitoring of Autism Spectrum Disorder (ASD) patients during therapeutic sessions. Recently, data mining has begun to address the problem of human robot interaction, as well.

Most affective computing applications primarily use vision[32], as the informative channel. Detection of face expressions, in fact, is the most natural way to perceive emotions of people. Vitale et al.[33] used vision to try to emulate the human brain's representations of a perceived emotion via facial expressions.

Gestures and body movements in general, also have a significant role in determining emotions, but only a small number of studies, to date, have examined the use of gestures. An interesting work on the topic is by Lourens et al.[34]. They developed an emotion recognition process based on gestures described through Laban movement analysis; it can discern from human segments of different people entering the field of view of the robot and discard them in the classification. Lim and Okuno[35] used a rudimental gait analysis to extract part of their selected features. They mainly focused on cadence, peak acceleration, step timing variance and maximum stride length.

However, it recently has become evident that the typical use of vision in affective computing applications sometimes may not be as feasible in social robotics. In fact, the standard procedure to acquire information through vision is to make a person sit in front of a camera with properly placed lights and ask him/her to perform an emotion[20]. In other cases, publicly available datasets of emotional faces, such as JAFFE or CK+, or those generated via software[33], are used. In a real scenario, the person often would not stand in front of the robot at a fixed distance during the interaction. Furthermore, in an environment such as a house, the presence of irregular light sources would invalidate the training process previously performed. Additionally, an elicited emotion is quite different from a spontaneous one[7].

This large amount of problems would induce the search for other sensing modalities, more feasible in real-world applications. In fact, speech and writings are used in affective computing, as well. In Ref. [36], authors implemented a recognition process based on multifractal theory to analyse recordings from Mandarin people. Juszkiewicz[37] employed speech for the emotion recognition system of a social robot, taking advantage of the well-founded literature about relevant features of speech, such as pitch and Mel's cepstrum coefficients. Le and Lee[38] applied genetic algorithms to optimize their classification process, based on mixtures of Support Vector Machine (SVM) and Hidden Markov Models (HMM). Perez-Gaspar et al.[39] also used genetic algorithms for improving performances in recognizing emotions by a robot during a dialogue with the user. Roboceptionist of Carnegie Mellon Institute[30] has a keyboard to receive inputs from a user. By means of the words used, the system can discern different emotions. However, this last modality of receiving inputs could be less affordable in a social robot, because in this case an excessive amount of attention should be paid in transmitting the emotion to the robot by the user, resulting in a subsequent decrease of spontaneity.

The use of vision and speech for emotion recognition both present significant difficulties. Since human beings have a large amount of control of these information channels, they could distort the provided informa-

tion, invalidating the emotional analysis. As social robots are intended to be our personal companions, they should be able of discerning these situations.

This could be achieved by exploiting different informative channels that are not directly controlled by humans. These can be their physiological parameters. Many applications of affective computing have employed these signal sources, as well. Nardelli *et al.*[40] used electrocardiogram (ECG) signals to classify emotions in subjects by means of auditory stimuli. Rattanryu *et al.*[41] used the same modality with the declared intent to subsequently embed this process into service robots. Another interesting case is the one of Ferreira *et al.*[42] They, also taking inspiration from Nardelli's team, elaborated an emotion recognition based on ECG, but which analysed the noise and extracted information through its entropy.

Khezri *et al.*[43] and Wang *et al.*[20,44] used electroencephalograms (EEGs) to acquire signals from the human brain for the emotion recognition process. The problem of using EEG is its unfeasibility in the real world. In worst cases, electrode cuffs can take up to 2 hours to be properly installed, so it is not possible to use them for daily-life applications.

In Ref. [29], measurements of electrodermal activity (EDA) in human-robot interaction scenarios were performed. To do this, they explored variations and classification accuracy for a human-human interaction.

Nonetheless, there is an interesting case in literature in the choice of the type of information to analyse. Nhan and Chau[45] decided to build an emotion classification process by using thermal infrared imaging of the face.

A general problem of physiological parameters is that they vary according to the emotional state of the person with a slower speed than vision and speech. Additionally, they undergo variability to age and possible diseases of people[46]. Thus, the feasibility of these types of signals must be carefully assessed for a real-time application. In this context, Henriques *et al.* performed a deep survey about problems of the up-to-date analysis of physiological signals in general, by defining seven key principles in the definition of a data mining process, including this type of signals, and illustrating their own procedure.

A promising way to obtain relevant performances in a real-world scenario is through a sensor fusion approach, which is the joint use of sensing modalities coming from diverse sources. In fact, we can determine how another person feels by seeing his/her movements and expressions and by hearing his/her voice tone at the same time[20,24,43,47]. Conversely, the use of various sources of information can make the process more resilient against irregularities from the environment (*e.g.*, a noisy room or excessive source of light in the field of view). Albert Mehrabian stated that 7% of the communication information is transferred by language, 38% by paralanguage, and 55% by facial expressions[48]. Through different informative channels, a social robot could be able to receive additional information needed to make its emotion recognition process more sensitive to hidden states of mind. In several situations during daily life, we are not able to listen to a person's words during an interaction, but we can detect or at least have an idea of the emotion the person is feeling, by only looking at his/her face. In this case, the robot could have an even more enhanced capability, by exploiting physiological parameters, as well. Koelstra and Patras[20] included visual features together with EEG recordings to achieve better classification accuracy. Khezri *et al.*[43] performed a detailed analysis by examining multiple parameters. They recorded EEG, electrooculography (EOG), electromyography (EMG), Blood Volume Pressure (BVP), Galvanic Skin Response (GSR), and Heart Rate Variability (HRV) from subjects simultaneously while building an emotion recognition process based on a weighted linear fusion model. Poria *et al.*[24] and Banda *et al.*[21] combined visual and audio features for their multimodal sentiment analysis. The Probo robot[47] is provided with visual, audio, and tactile sensors to infer the emotions of children during therapy sessions.

Recommendations and trends

Several sensing approaches for emotion recognition have been used in literature. Vision-based sensors appear to be the most widely used, focusing on facial expressions and body gesture analysis; however they present limitations in case of irregular light sources and target person not in front of the camera or in movement, thus preventing their use in real social robotics applica-

tions. Physiological sensors provide direct information at autonomic nervous system level, however they suffer from possible bias due to age-related and disease conditions; additionally they are still obtrusive in terms of wearability, thus resulting not usable and acceptable for daily living activities. Recently, wearable devices, such as wrist, ear[49] and ring[50] sensors, appear to be promising solutions for monitoring physiological parameters, but more improvements are required in terms of miniaturization and accuracy of the measurements.

Sensor fusion seems to be a promising approach to overcome limitations of different sensors, combining vision sensors on the robot and wearable sensors on users. Additionally, an extraordinary empowerment of robot capabilities to recognize emotion could surely come from installation of sensors in indoor and outdoor environments. They, in fact, could provide unlimited additional information to robots to improve their behaviours. This information is not simply in terms of additional informative channels; the information could be used by the robot to evaluate the situation and change its position or behaviour to put itself in a better condition to acquire signals from the user. For instance, if light sensors recognize a large amount of luminous intensity from a certain direction, they could tell the robot how to move to adjust its position to better acquire the image. Plus, they can surely provide the context awareness needed for a complete handling of the interaction with the user. Still, too few research studies have been conducted to exploit the environment to this purpose. One such study by Salmeron[18] involved a fuzzy system that was intended to receive information from sound, temperature, and pressure sensors to obtain a model of artificial robot emotion forecasting in crowded environments.

5 Features

Following the selection of informative channels to exploit, subsequent research is conducted to find useful information in acquired signals for the emotion recognition (Fig. 4).

Mainly regarding vision, several image processing filters to extract salient features distinguish between component and global approaches in feature extraction of the face[51]. In the first approach, facial components

that are used to classify emotions are detected. This is done, for example, by matching the corresponding components as in Ref. [52] or in Ref. [53]. Global approaches, instead, try to extract a representation of the expressions through the appearance of the global face. This is the case, for instance, of Histogram of Gradients (HOG) descriptors. For instance, Lourens et al.[34] used Laban movement analysis to detect human body segments in a real-time process.

The main features of speech come from intensity, spectrogram, pitch, Mel's cepstrum coefficients, harmonics-to-noise ratio, and long-term average spectrum[37]. Regarding physiological parameters, there is less literature. The frequency domain is generally preferred for feature extraction. For ECG, Rattanryu et al.[41] pre-processed fast and continuous wavelet transform, while Nardelli et al.[40] employed the power spectral density. This last approach was followed by Koelstra and Patras[20] for EEG which was combined with video analysis at different levels to investigate how the accuracy variations.

In a real-time application, the extraction of all common features by this type of data, particularly regarding vision, can be excessively time consuming and have a high computational cost. Therefore, it is necessary to find ways to reduce the numbers of features to use in the upcoming classification process. A solution is the selection of features extracted by signal that carries the greatest information content. Regarding the selection of signals to obtain features, Blais et al.[54], by analysing random portions of face images, demonstrated that better performance in emotion classification through partial images comes from analysing the mouth. Reduction of the considered image was used by Jing et al.[22] for their application. Khan et al.[55] developed a framework for real-time facial expression recognition which receives low-resolution images as input. Having a smaller number of pixels than typical robot cameras, feature extraction is speeded up. In addition, this method could be used in the case of low-resolution images due to different distances of the user from the robot. De et al.[52] performed Principal Component Analysis (PCA) to obtain a singular value decomposition, generalization of the more known value using eigenvalue, to reduce dimensionality of expression mapping. Juszkiewicz[37]
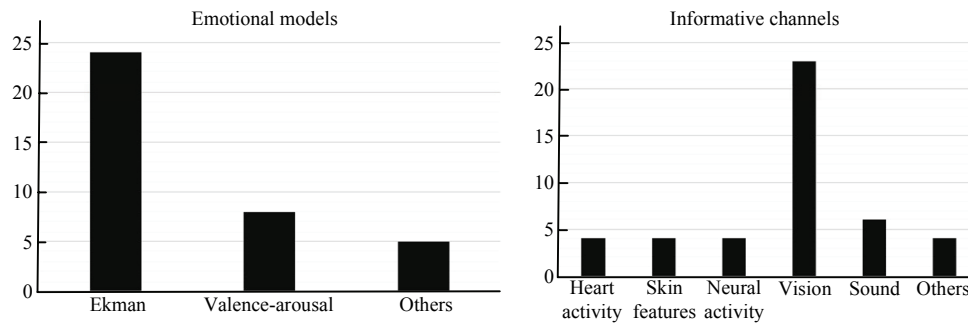
Fig. 4 Model preferences among cited articles (left) and statistics of informative channels exploited throughout the papers (right).

used feature selection algorithms provided by WEKA software to reduce dimensionality of his feature space.

Another way, less commonly used, is to find new features from acquisitions, with the expectation that they will be less time consuming[36,42].

Recommendations and trends

A consolidate set of features which can recognize users' emotions with a high level of accuracy is lacking. Most of the works are using video analysis, less studies involve the analysis of other parameters. By using video related features, there could be differences between positive and negative expressions due to the small changes between disgust, sadness and fear and high changes in positive expressions[52,53]. Future research efforts should tend to use more points on the face to capture even these little changes, finding a compromise between computational efforts required or use a global approach which will also analyse the body posture[34]. The sound feature can add important information on the emotion, nevertheless they can be gender-dependent, indeed the design of speakers must deal with the difference in masculine and feminine voice (*i.e.* tone and frequencies)[37].

Another important aspect to consider when you have to deal with psychological parameters or brain activity is related to the methodology used[20]. Having participants taking part in multiple experimental sessions can degrade the performance as differences in brain activity and physiological status can occur between different mood and electrodes can be placed in slightly different positions which can affect the measures. In this context, the implementation of a precise stimulus methodology should be recommended and investigated promoting the creation of labelled datasets with different data sources. Finally, research efforts should investigate which stimulus is appropriate to generate a specific emotion in the users and which kind of feature is more significant to reveal a particular emotion or mood.

6 Learning models

Techniques inherited by data mining theory allow the construction of well-known types of classifiers (Fig. 5). De *et al.*[52] applied a k-NN algorithm to determine the classification of what they termed "eigenfaces", similarly to the process employed by Luo *et al.*[53]. Furthermore, Khan *et al.*[55] used Naïve Bayesian and random forest classifiers. Ferreira *et al.*[42] employed decision tree algorithms.

Another way to instantiate the classification is a learning process achieved through neural networks. In the fashion of a Multi-Layer Perceptron (MLP), this process automatically performs a remapping of extracted features in a new feature space, by using multiple layers, without needing any previous knowledge about the system. If the neural network has been correctly instantiated, it results in an improved classification through the new feature space.

Training of neural networks is generally performed through supervised learning. Labelled instances of a dataset are presented to the network, which modifies its internal conformation according to the mismatch between the output produced and the expected result, which is the correct classification of the instances. Salmeron[18] used fuzzy cognitive maps, based on the implementation of fuzzy rules, for his system of emotion forecasting.

In opposition to neural networks is the commonly used method of the Support Vector Machine (SVM). It is a supervised learning process that solves a hyper-plane

separation problem by maximizing the distance from the closest instances of each different class. It has a more efficient algorithm, capable of representing nonlinear relationships, with whom the multi-layer perceptron can achieve by increasing the number of layers and, by that, the computational effort. Moreover, as this model is based on a constrained nonlinear optimization problem, it guarantees the absolute minimum of the error function, while multi-layer perceptron, being unconstrained, suffers the problem of local minima. But, of course, this model strongly requires prior hypotheses. Ali *et al.*[21,36,37,55,56] used neural networks in the support vector machine and Radial Basis Function (RBF) network fashions to make linear separation of different regions of the feature space to associate to defined classes.

In a supervised learning of neural networks, another methodology is to process sequences of data related to one another through time[24,37,57]. In this case, we can have dependency of the output from all previous histories of inputs. To overcome these limitations, Recurrent Neural Network (RNN) was used. For instance, Juszkiewicz[37] used a particular type of RNN, called NARX, based on non-autoregressive exogenous inputs, for his classification of acoustic signals. Kushiro *et al.*[57] exploited RNN in what is known as the MoNAD fashion. It is a complex architecture, based on redundant circuit of multi-layered RNNs. This structure was built with the intent of growing dynamically at the detection of unknown inputs. A slight drawback in training neural networks is the computational time required by the backpropagation training algorithm. This inconvenience is increased in back propagation through time of recurrent neural networks. In this context, Poria *et al.*[24] realized an extreme learning machine, as a substitutive classification process. They compared it with support vector machines and Artificial Neural Networks (ANNs), demonstrating a decrease of time needed for training with the former.

Another approach in training neural networks is given by unsupervised learning. In this case, learning is not based on adjustment of the structure for errors committed in classification. Traditional versions of unsupervised learning deal with a Self-Organizing Map (SOM)[58]. They are networks in which a neuron of the output layer fires when similar inputs are presented to it,

consequently inhibiting firing of other neurons. Mei and Liu[59] implemented a SOM to solve cognitive problems of service robots in unstructured environments.

Today, unsupervised learning is conveyed in deep learning networks. They are structures of neurons, in which each layer performs different internal representations of instances of the dataset, with an increasing level of hierarchy with propagation of input along the network. Yin *et al.*[60] exploited deep learning models in analysing their features from physiological signals. In image analysis, a deep learning structure frequently used is Convolutional Neural Networks (CNN). Mayya *et al.*[61] and Barros *et al.*[62] used a Discrete CNN (DCNN) for extraction of features and classification of emotion from face images.

Another more intriguing and untamed method is the Adaptive Resonance Theory (ART). In this topic, neural networks are created in such a way that when an input very discordant from categories accounted by the network is presented, a new input is generated to include the new instance and similar next instances. This dynamic growth of the representation of a phenomenon could consider an aspect discarded so far, that is, variability among different subjects and in a same subject over time in emotion expression. This reveals an interesting potentiality for the robot: the capability of getting to know a user increasingly better by adding new internal states when a new state of mind, totally different from states seen so far, is presented to the robot. Chumkamon *et al.*[14] used an ART-like network to build a behavioural model of a robot capable of sympathizing with emotions recognized in the user and to learn which movements are good to perform. This theory still needs to be explored in detail to understand its potential in a robotic application.

Recommendations and future trends

The choice of the appropriate learning methods is a challenge in the emotion learning process and it is related to different aspects. As highlighted from literature, supervised approach is the most used (see Fig. 5), indeed a total of 9 work used this method. Analysing hystogram depicted in Fig. 5, it is evident that there is not a predominant choice, indeed, among these tecniques, SVM are the most used in this application; but also HMM and ANN classifiers received attention. Other learning
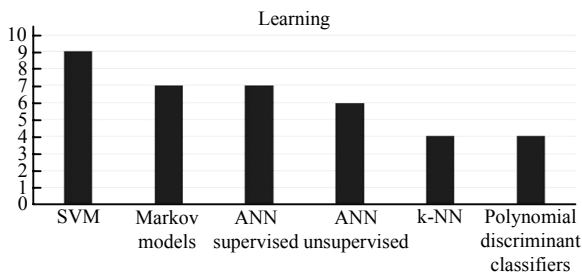
Fig. 5 Classification methods reported in cited papers.

methods used are: k-NN and polynomial discriminant classifiers. The remain of the papers used custom algorithms.

However, supervised learning has the inconvenience that each instance has to be preemptively labeled before the training. This implies a hard categorization of each possible emotional display, which may not fit well the problem of emotion recognition in a real scenario[63]. Recently, unsupervised learning received attention in this field to overcome this limitation, since they do not require the set of labels. Among unsupervised techniques ANN is the most used classifiers (Fig. 5).

All learning models and classifiers can be generally enhanced in their proficiency of a correct emotion recognition by exploiting fusion modalities mentioned so far. There is, in fact, a cunning way to organize the analysis of features coming from distinct types of signals. Features coming from different informative channels can be used separately to perform different classifications or they can be fused together to produce the final output of classification. For instance, features from facial expression can be combined at different levels (feature or decision level) with other parameters such as GSR, ECG, EEG or sound thus to increase the level of accuracy of more complex emotions as presented in Ref. [35] where vision signals (face, body posture, gesture) were fused with sound signals.

Additionally, the variation of weights of different classifiers in the result could be a very efficient way to handle reliability of different sensing modalities according to the current situation. For instance, Khezri *et al*.[43] performed a weighted linear fusion model. The features extracted from EEG signals and muscular (EMG, EOG) and cardiovascular activity was separately exposed to the same type of feature selection and classification and conclusive results were achieved by a weighted sum of the three outcomes. Whereas, Koelstra and Patras[20] modelled the weighting through a Bayesian approach, based on probability of different classifiers to recognize the correct emotion.

Another important aspect which should be considered in this kind of application is the time-dependency, especially in real-time scenario. In fact, if a time window is sampled to collect the instances through which to perform the classification, the number of collected samples will never be the same in same length windows, due to slight differences of sampling frequencies[37] and to the synchronization between different devices. The advantage of using RNN, is that we do not have to make a prior hypothesis about how much the history affects the outcome. Indeed, they can express the outcome of classification without considering the variable number of samples to process.

Future research efforts should also investigate what happens to the system when a "new emotion" will occur. If you desired to exploit robots in real world, they must deal with different emotions that should be managed in the correct way with the appropriate algorithm.

7 Emotion recognition in service robotics applications

Emotion recognition capabilities are expected to be present in actual social service robots, which move themselves in an open world, strictly related to a group of persons to fulfil specific tasks. To prove this point and the results of this paper here we provide a summary of already existing robotic applications using affective computing.

Some researchers feel that emotion recognition in robots should be used for collecting feedback on robots' behaviour from the user and then adjusting it consequently. Mei and Liu[59] built a robotic grasper with two cameras, one for detecting objects and the other to observe the user's face. According to the emotion expressed by the user during the process, the robot decided to catch or avoid the visualized object. Chumkamon *et al*.[14] developed a robot capable of sympathizing with the emotion expressed by the user. It is interesting to point out that in both these examples unsupervised learning was used. Nho *et al*.[64] evaluated the behaviour of a cleaning robot by using hand gestures to tune its

internal emotional model. In these cases, the user showed an emotion caused directly by the robot's behaviour, so the emotion was related to a specific action and not to the general context. Hence, a robot could also be taught to act directly based on the emotional feeling of the user, so not related to the robot's behaviour. The robot could use this as a tuning of internal parameters for example in order not to act too invasively in the task performed. Röning et al.[65] developed a remarkable example of an autonomous agent behaving in the real world; their Minotaurus robot could recognize people through vision, detect their mood, and move inside smart environments, and receive information from it. Similarly, the CuDDler robot[15] was used in entertainment and socially interactive robot applications. Their polar bear-like robot produced scripted movements according to the recognized emotion.

Another group of researchers think that robots should have an internal model for emotions to regulate their behaviour. Jitviriya et al.[58] used a SOM to categorize eight inner emotions of a ConBe robot. Based on the difficulty of grabbing the ball, the robot tuned its behaviour and emotion expression when grabbing the ball held by a person. Van Chien et al.[66] developed a humanoid robotic platform expressing its internal emotions through changes in its walking pattern.

The design of a social robot strongly depends on the application and set of contexts it will handle. The presence of a head in the robot body can surely constitute an effective way to express emotions for robots, although its design may be quite demanding and constrained. Hence, some alternative interactions could also be developed. Terada et al.[25] designed a robotic emotional interface based on selection of parameters of a dynamic coloured light, realizing a detailed set of lights for a subtle type of emotions, such as pensiveness.

Interestingly, emotion recognition in robotics is important due to several reasons, not only for pure human robot interaction, but also for bioengineering applications (see Fig. 6). A substantial part of research work in this area deals with treatment of children with ASD or elderly with mild cognitive impairments[67]. In this case, robots are used as tools to show emotions for kids to perform during the therapy session, since they can communicate with robots more easily than with
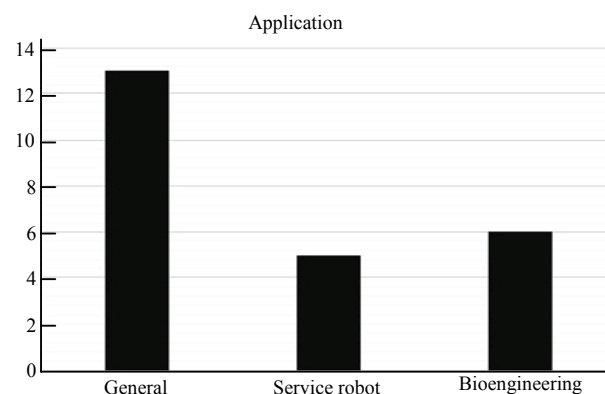

Fig. 6 Paper distribution considering their final application fields. Category "General" includes those works in which the practical application of emotion recognition process was not discussed. In the other two, the scope of the correspondent work was specifically declared. Among those ones, "BioEngineering" category applications, related to assistive therapies for children or elders.

humans. Then, they detect the emotion expressed by the patients and try to categorize them based on the degree of correct performance[36,47,51,68–70]. A physiotherapist is always present to monitor the work and judge classifications made by robots, since these solutions were not intended to replace human work[70]. Indeed, in these applications robots behave as mere tools for helping workers, and are better known as supportive robots. In such applications, a robot is expected to only provide an objective evaluation that integrates the subjective one of the physiotherapist. It does not require emotion recognition nor context awareness to model its own behaviour since it does not have an actual emotion. Alternatively, the robot could be used to monitor the mood of a person inside of his/her own house and try to relieve them from a stressful condition. An interesting example comes from Jing et al.[22]. Their robot recognized emotions in older people and modified parameters of a smart ambient environment to try to induce happiness. In a second instance, it could be used to detect the beginning of dementia or mild cognitive impairment.

Recommendation and trends

The majority of works did not consider the final application in the design of the emotion recognition process, and this concept becomes evident in the histogram of Fig. 6. Although the aim of emotion recognition is the enhancement in human-robot interaction, the way it will modify robot's behaviour is not so often discussed.

If a core objective was considered in the realization of such design from the very beginning, the consequent translational process of the achieved piece of technology would surely become more immediate.

However, other works clearly state how emotion recognition would be used in their application. For instance, in Chumkamon *et al.*[14] the objective was to modulate a physical interaction between a robot and a human in passing objects. They also suggested that, for a better interaction, an online learning in naturally social situations would be needed.

A consistent part of aim-guided works used the robot as an alternative tool to deliver medical and assistive therapies. In these applications, robots are tested in singular sessions with patients. Leo *et al.*[51] suggested that multiple evaluation sessions with the same patients should be consequent to fully evaluate the efficacy of this alternative therapeutic mean.

The complexity and variety of applications often require strong adaptation and flexibility in the used emotion recognition and modelling approach, thus highlighting that several challenges have not yet been appropriately investigated. For example, something poorly explored, but very promising so far is the issue of context awareness. In fact, Röning *et al.*[65] suggested that environmental sensors and use of the Internet surely could enhance the capability of robots to model context, and by that, to use emotion recognition more properly.

8 Discussion and conclusion

This review provides an overview of current solutions, methodologies and approaches implemented for emotion recognition in social robotics applications. By doing so, it discusses various theories to model human emotions, different sensing modalities and the features extracted from sensors' signals. It also covers the learning models used to build the emotion recognition process. Table 2 summarizes recommendations for the reader about which emotion models, sensing modalities and learning models are the most used in social robotics.

The high number of papers identified in the electronic databases demonstrates that emotion modelling and recognition is a tremendously growing research topic. As robots are required to be more socially believable and to operate closer to humans, human robot interaction has become one of the most important research challenges, characterized by multi-channel data sensor integration and methods. Indeed, in order to socially interact with humans, a robotic system should be able not only to understand users' behaviour and intentions, but also to estimate their emotional state with a high level of acceptability and usability[71].

Furthermore, in considering new enabling technologies such as Cloud computing, major issues on legality, privacy, and safety are brought to attention[72]. This paradigm has had significant impact on various fields, and robotics is not an exception. In 2010, James

Table 2 Limitations and Challenges

Type	Barrier/Limitation	Challenge/Opportunity
Enabling Technology	Limited computing resources	A cloud-based solution should be implemented to leverage the robot from computation intensive tasks
	Sensors should be portable and unobtrusive	New devices should be wireless, portable and usable in indoor and outdoor environments. They should be able to detect the emotions in an unobtrusive way.
	Acceptability and usability	The system should be designed considering the acceptability and usability of the end-users to be exploited in real applications
Data Management	Privacy and legal issues	Ethical and legal issues related to the data acquisition and data management should be further investigated.
	Data Storage	The system should be able to storage and manage large quantity of data.
Methodology	Identify the correspondence stimulus/emotion	Strict methodology should be defined to evoke a set of specific emotions in human beings.
	Experimental sessions	The system should be tested in a single session for each user since differences in brain activity and physiological status can occur between different moods and electrodes can be placed in slightly different positions which can affect the measurements.
	Experimental sessions	The system should be tested in real settings, humans show very complex emotion which can differ according to the socio-cultural background.
	Dataset	Available datasets focused mainly on images. Future datasets should include also other signals.
Learning Methods	Data Fusion	Learning methods should investigate different fusion methods to improve the accuracy.

Kuffner at the international conference on humanoid robots introduced the concept of "cloud robotics" and the first formal definition was established by Mell and Grance[73]. Their cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable resources (*e.g.*, servers, storage, networks, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. Since then, many studies have made use of the advantages of this technology[74,75] respectively. However, it is worth to mention that many of the applications focused on navigation and manipulation tasks. Similarly, social "cloud" robots could benefit from publicly available facial emotion assessment cloud services. These are offered by companies such as Microsoft (Face API), Google (Vision API), Emovu, Affectiva, and others. By using the concept of cloud robotics, they could offload the task of emotion recognition to remote servers and exploit computational and storage capabilities and intelligence sharing for tasks more related to the robot's immediate actions, where no latency is acceptable. The use of a cloud architecture also enables robots to share capabilities one to each other, once one of them has acquired it. Conversely, Chen *et al.*[76] highlight Cloud computing can improve emotion recognition for human robot interaction by embedding this design paradigm into Smart Clothing.

Today, indoor and outdoor environments are becoming increasingly smarter, thanks to new Internet of Things (IoT) technologies. Using sensor data fusion coming from environments or in/on-body sensors seems to represent a promising way to structure the acquisition and analysis of data from the user. There is a big opportunity for robots to obtain environmental factors, which can influence the decision-making process in recognizing emotions or for the improvement and empowerment of robotic sensing.

Furthermore, learning models, independently from the chosen ones, should be exploited in such a way that it enhances not only the emotion recognition, but also its expression by the robot[77]. However, in the current state of the art, these two models are independently implemented, so there is no real improvement of the interaction mediated by a user's emotion. In this context, Lim and Okuno[77] developed an emotional model for a robot, inspired by concepts of native language, which used the recognition of emotion to modulate the facial expression of the robot.

Although eventually someone will develop the perfect classifier for a specific situation, great questions and demands still will be posed about the potential best use of this capability for a robot in the world. Such ability surely presents ethical and legal issues. By simply considering Asimov's three laws of robotics, a complex pool of scenarios in which these empirical statements show inconsistencies rapidly emerges. Most of them are due to the lack of a real definition of interaction between human and robot and methodology research, as social robotics involve several disciplines, including engineering, computer science, and ethical sciences.

Finally, much more attention should be paid to the assessment of the feasibility of a social robot in real-life scenarios among different cultures and people[10]. Most works related to this topic found interesting solutions in emotion recognition which could miserably fail when encountering a real-life situation, because they perform evaluations through experiments in ideal conditions. Real environments in which the robot could be inserted have numerous cases in far from ideal conditions, invalidating the results of the process[78]. Hence, the experimental paradigm must be carefully evaluated. More specifically, emotion recognition of a robot should be tested in the real world, instead of only using publicly available datasets.

References

[1]   International Federation of Robotics. *World Robotics Report 2016: European Union Occupies Top Position in the Global Automation Race*, Report, Frankfurt, Germany, 2016.

[2]   Fiorini L, Esposito R, Bonaccorsi M, Petrazzuolo C, Saponara F, Giannantonio R, De Petris G, Dario P, Cavallo F. Enabling personalised medical support for chronic disease management through a hybrid robot-cloud approach. *Autonomous Robots*, 2017, **41**, 1263–1276.

[3]   Reppou S, Karagiannis G. Social inclusion with robots: A RAPP case study using NAO for technology illiterate elderly at ormylia foundation. In: Szewczyk R, Zieliński C, Kaliczyńska M eds., *Progress in Automation, Robotics and Measuring Techniques. Advances in Intelligent Systems and Computing*, Springer, Cham, Switzerland, 2015, 233–241.

[4] Fasola J, Matarić M J. A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction*, 2013, **2**, 3–32.

[5] García-Soler Á, Facal D, Díaz-Orueta U, Pigini L, Blasi L, Qiu R X. Inclusion of service robots in the daily lives of frail older users: A step-by-step definition procedure on users' requirements. *Archives of Gerontology and Geriatrics*, 2018, **74**, 191–196.

[6] Lisetti C L. Affective computing. *Pattern Analysis and Applications*, 1998, **1**, 71–73.

[7] Castellano G, Leite I, Pereira A, Martinho C, Paiva A, McOwan P W. Affect recognition for interactive companions: Challenges and design in real world scenarios. *Journal on Multimodal User Interfaces*, 2010, **3**, 89–98.

[8] Spaulding S, Breazeal C. Towards affect-awareness for social robots. 2015 *AAAI Fall Symposium Series*, USA, 2015, 128–130.

[9] Gray K, Wegner D M. Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 2012, **125**, 125–130.

[10] Rau P L P, Li Y, Li D. A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics*, 2010, **2**, 175–186.

[11] Ho C C, MacDorman K F. Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, 2010, **26**, 1508–1518.

[12] Mathur M B, Reichling D B. Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition*, 2016, **146**, 22–32.

[13] Hassenzahl M, Tractinsky N. User experience - A research agenda. *Behaviour & Information Technology*, 2006, **25**, 91–97.

[14] Chumkamon S, Hayashi E, Masato K. Intelligent emotion and behavior based on topological consciousness and adaptive resonance theory in a companion robot. *Biologically Inspired Cognitive Architectur*, 2016, **18**, 51–67.

[15] Limbu D K, Anthony W C Y, Adrian T H J, Dung T A, Kee T Y, Dat T H, Alvin W H Y, Terence N W Z, Jiang R D, Li J. Affective social interaction with CuDDler robot. 2013 *6th IEEE Conference on Robotics, Automation and Mechatronics* (*RAM*), IEEE, Manila, Philippines, 2013, 179–184.

[16] Ekman P, Friesen W V. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 1971, **17**, 124–129.

[17] Bennett C C, Sabanovic S. Deriving minimal features for human-like facial expressions in robotic faces. *International Journal of Social Robotics*, 2014, **6**, 367–381.

[18] Salmeron J L. Fuzzy cognitive maps for artificial emotions forecasting. *Applied Soft Computing*, 2012, **12**, 3704–3710

[19] Russell J A. A circumplex model of affect. *Journal of Personality and Social Psychology*, 1980, **39**, 1161–1178.

[20] Koelstra S, Patras I. Fusion of facial expressions and EEG for implicit affective tagging. *Image & Vision Computing*, 2013, **31**, 164–174.

[21] Banda N, Engelbrecht A, Robinson P. Feature reduction for dimensional emotion recognition in human-robot interaction. 2015 *IEEE Symposium Series on Computational Intelligence*, Cape Town, South Africa, 2015, 803–810.

[22] Han J, Xie L, Li D, He Z J, Wang Z L. Cognitive emotion model for eldercare robot in smart home. *China Communications*, 2015, **12**, 32–41.

[23] Cambria E, Livingstone A, Hussain A. The hourglass of emotions. In: Esposito A, Esposito A M, Vinciarelli A, Hoffmann R, Müller V C, eds, *Cognitive Behavioural Systems*. *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, Germany, 2012.

[24] Poria S, Peng H, Hussain A, Howard N, Cambria E. Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis. *Neurocomputing*, 2017, **261**, 217–230.

[25] Terada K, Yamauchi A, Ito A. Artificial emotion expression for a robot by dynamic color change. 2012 *IEEE RO-MAN*, Paris, France, 2012, 314–321.

[26] Raymundo C R, Johnson C G, Vargas P A. An architecture for emotional and context-aware associative learning for robot companions. 2015 *24th IEEE International Symposium on Robot and Human Interactive Communication* (*RO-MAN*), Kobe, Japan, 2015, 8799–8804.

[27] Shi X-F, Wang Z L, Ping A, Zhang L-K. Artificial emotion model based on reinforcement learning mechanism of neural network. *The Journal of China Universities of Posts Telecommunications*, 2011, **18**, 105–109.

[28] Lewis M, Canamero L. Are discrete emotions useful in human-robot interaction? Feedback from motion capture analysis. 2013 *Humaine Association Conference on Affective Computing and Intelligent Interaction*, 2013, 97–102.

[29] Henriques R, Paiva A, Antunes C. Accessing emotion patterns from affective interactions using electrodermal activity. 2013 *Humaine Association Conference on Affective Computing and Intelligent Interaction* (*ACII*), Geneva, Switzerland, 2013, 43–48.

[30] Kirby R, Forlizzi J, Simmons R. Affective social robots. *Robotics and Autonomous Systems*, 2010, **58**, 322–332.

[31] Picard R W. Affective computing. *Pattern Recognition*, 1995, **20**.

[32] Giorgana G, Ploeger P G. *Facial Expression Recognition for Domestic Service Robots*, Springer, Berlin, Heidelberg, Germany, 2012.

[33] Vitale J, Williams M A, Johnston B, Boccignone G. Affective facial expression processing via simulation: A probabilistic model. *Biologically Inspired Cognitive Architectures*, 2014, **10**, 30–41.

[34] Lourens T, Van Berkel R, Barakova E. Communicating emotions and mental states to robots in a real time parallel framework using Laban movement analysis. *Robotics & Autonomous Systems*, 2010, **58**, 1256–1265.

[35] Lim A, Okuno H G. The MEI robot: Towards using motherese to develop multimodal emotional intelligence. *IEEE Transactions on Autonomous Mental Development*, 2014, **6**, 126–138.

[36] Liu H, Zhang W. Mandarin emotion recognition based on multifractal theory towards human-robot interaction. 2013 *IEEE International Conference on Robotics and Biomimetics* (*ROBIO*), Shenzhen, China, 2013, 593–598.

[37] Juszkiewicz Ł. Improving speech emotion recognition system for a social robot with speaker recognition. 2014 19*th International Conference on Methods and Models in Automation and Robotics* (*MMAR*), Miedzyzdroje, Poland, 2014.

[38] Le B V, Lee S. Adaptive hierarchical emotion recognition from speech signal for human-robot communication. 2014 *Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing* (*IIH-MSP*), Kitakyushu, Japan, 2014, 807–810.

[39] Perez-Gaspar L A, Caballero-Morales S O, Trujillo-Romero F. Multimodal emotion recognition with evolutionary computation for human-robot interaction. *Expert Systems with Applications*, 2016, **66**, 42–61.

[40] Nardelli M, Valenza G, Greco A, Lanata A, Scilingo E P. Arousal recognition system based on heartbeat dynamics during auditory elicitation. 2015 37*th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (*EMBC*), Milan, Italy, 2015.

[41] Rattanyu K, Ohkura M, Mizukawa M. Emotion monitoring from physiological signals for service robots in the living space. 2010 *International Conference on Control Automation and Systems* (*ICCAS*), Gyeonggi-do, South Korea, 2010, 580–583.

[42] Ferreira J, Br As S, Silva C F, Soares S C. An automatic classifier of emotions built from entropy of noise. *Psychophysiology*, 2017, **54**, 620–627.

[43] Khezri M, Firoozabadi M, Sharafat A R. Reliable emotion recognition system based on dynamic adaptive fusion of forehead biopotentials and physiological signals. *Computer Methods and Programs in Biomedicine*, 2015, **122**, 149–164.

[44] Wang S, Du J, Xu R. Decision fusion for EEG-based emotion recognition. 2015 *International Conference on Machine Learning and Cybernetics* (*ICMLC*), Guangzhou, China, 2015, 883–889.

[45] Nhan B R, Chau T. Classifying affective states using thermal infrared imaging of the human face. *IEEE Transactions on Bio-medical Engineering*, 2010, **57**, 979–987.

[46] Henriques R, Paiva A. Learning effective models of emotions from physiological signals: The seven principles. *Lecture Notes in Computer Science*, 2014, **8908**, 137–155.

[47] Saldien J, Goris K, Vanderborght B, Vanderfaeillie J, Lefeber D. Expressing emotions with the social robot probo. *International Journal of Social Robotics*, 2010, **2**, 377–389.

[48] Ghayoumi M, Bansal A K. Multimodal architecture for emotion in robots using deep learning. *Future Technologies Conference* (*FTC*), San Francisco, USA, 2016.

[49] Esposito D, Cavallo F. Preliminary design issues for inertial rings in Ambient Assisted Living applications. 2015 *IEEE Instrumentation and Measurement Technology Conference*, Pisa, Italy, 2015, 250–255.

[50] Park J-H, Jang D-G, Park J, Youm S-K. Wearable sensing of in-ear pressure for heart rate monitoring with a piezoelectric Sensor. *Sensors*, 2015, **15**, 23402–23417.

[51] Leo M, Del Coco M, Carcagnì P, Distante C, Bernava M, Pioggia G, Palestra G. Automatic emotion recognition in robot-children interaction for ASD treatment. 2015 *IEEE International Conference on Computer Vision Workshop* (*ICCVW*), Santiago, Chile, 2015, 537–545.

[52] De A, Saha A, Pal M C. A human facial expression recognition model based on eigen face approach. *Procedia Computer Science*, 2015, **45**, 282–289.

[53] Luo R C, Lin P H, Wu Y C, Huang C Y. Dynamic face recognition system in recognizing facial expressions for service robotics. 2012 *IEEE/ASME International Conference on Advanced Intelligent Mechatronics* (*AIM*), Kachsiung, Taiwan, 2012, 879–884.

[54] Blais C, Roy C, Fiset D, Arguin M, Gosselin F. The eyes are not the window to basic emotions. *Neuropsychologia*, 2012, **50**, 2830–2838.

[55] Khan R A, Meyer A, Konik H, Bouakaz S. Framework for reliable, real-time facial expression recognition for low resolution images. *Pattern Recognition Letters*, 2013, **34**,

1159–1168.

[56] Ali H, Hariharan M, Yaacob S, Adom A H. Facial emotion recognition using empirical mode decomposition. *Expert Systems with Applications*, 2015, **42**, 1261–1277.

[57] Kushiro K, Harada Y, Takeno J. Robot uses emotions to detect and learn the unknown. *Biologically Inspired Cognitive Architectures*, 2013, **4**, 69–78

[58] Jitviriya W, Koike M, Hayashi E. Behavior selection system based on emotional variations. 2015 24*th IEEE International Symposium on Robot and Human Interactive Communication* (*RO-MAN*), Kobe, Japan, 2015, 462–467.

[59] Mei Y, Liu Z T. An emotion-driven attention model for service robot. 2016 12*th World Congress on Intelligent Control and Automation* (*WCICA*), Guilin, China, 2016, 1526–1531.

[60] Yin Z, Zhao M, Wang Y, Yang J, Zhang J. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Computer Methods & Programs in Biomedicine*, 2017, **140**, 93–110.

[61] Mayya V, Pai R M, Pai M. Automatic facial expression recognition using DCNN. *Procedia Computer Science*, 2016, **93**, 453–461.

[62] Barros P, Jirak D, Weber C, Wermter S. Multimodal emotional state recognition using sequence-dependent deep hierarchical features. *Neural Networks*, 2015, **72**, 140–151.

[63] Mower E, Matarić M J, Narayanan S. A framework for automatic human emotion classification using emotion profiles. *IEEE Transactions on Audio*, *Speech*, *and Language Processing*, 2011, **19**, 1057–1070.

[64] Nho Y-H, Seo J-W, Seol W-J, Kwon D-S. Emotional interaction with a mobile robot using hand gestures. 2014 11*th International Conference on Ubiquitous Robots and Ambient Intelligence*, Kuala Lumpur, Malaysia, 2014, 506–509.

[65] Röning J, Holappa J, Kellokumpu V, Tikanmäki A, Pietikäinen M. Minotaurus: A system for affective human–robot interaction in smart environments. *Cognitive Computation*, 2014, **6**, 940–953.

[66] Van Chien D, Sung K J, Trung P X, Kim J-W. Emotion expression of humanoid robot by modification of biped walking pattern. 2015 15*th International Conference on Control, Automation and Systems* (*ICCAS*), Busan, South Korea, 2015, 741–743.

[67] Sinčák P, Novotná E, Cádrik T Magyar G, Mach M, Cavallo F. Bonaccorsi M. Cloud-based Wizard of Oz as a service. 2015 *IEEE* 19*th International Conference on Intelligent Engineering Systems* (*INES*), Bratislava, Slovakia, 2015, 445–448.

[68] Mazzei D, Zaraki A, Lazzeri N, De Rossi D. Recognition and expression of emotions by a symbiotic android head. 2014 14*th IEEE-RAS International Conference on Humanoid Robots* (*Humanoids*), Madrid, Spain, 2014, 134–139.

[69] Boccanfuso L, Barney E, Foster C, Ahn Y A, Chawarska K, Scassellati B, Shic F. Emotional robot to examine differences in play patterns and affective response of Children with and without ASD. 2016 11*th ACM/IEEE International Conference on Human-Robot Interaction* (*HRI*), Christchurch, New Zealand, 2016, 19–26.

[70] Cao H L, Esteban P G, De Beir A, Simut R, Van De Perre G, Lefeber D, Vanderborght B. ROBEE: A homeostatic-based social behavior controller for robots in Human-Robot Interaction experiments. 2014 *IEEE International Conference on Robotics and Biomimetics* (*ROBIO*), 2014, 516–521.

[71] Cavallo F, Limosani R, Manzi A, Bonaccorsi M, Esposito R, Di Rocco M, Pecora F, Teti G, Saffiotti A, Dario P. Development of a socially believable multi-robot solution from town to home. *Cognitive Computation*, 2014, **6**, 954–967.

[72] Kopacek P. Ethical and social aspects of robots. *IFAC Proceedings Volumes*, 2014, **47**, 11425–11430.

[73] Mell P, Grance T, Grance T. *The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology*, National Institute of Standards and Technology-Special Publication 800-145, 2011.

[74] Goldberg K, Kehoe B. *Cloud Robotics and Automation*: *A Survey of Related Work*, Technical Report, USA, 2013.

[75] Kehoe B, Patil S, Abbeel P. Goldberg K. A survey of research on cloud robotics and automation. *IEEE Transactions on Automation Science & Engineering*, 2015, **12**, 398–409.

[76] Chen M, Ma Y, Song J, Lai C-F, Hu B. Smart clothing: Connecting human with clouds and big data for sustainable health monitoring. *Mobile Networks & Applications*, 2016, **21**, 825–845.

[77] Lim A, Okuno H G. The MEI robot: Towards using motherese to develop multimodal emotional intelligence. *IEEE Transactions on Autonomous Mental Development*, 2014, **6**, 126–138.

[78] Yan H, Ang Jr M H, Neow Poo A. A survey on perception methods for human–robot interaction in social robots. *International Journal of Social Robotics*, 2014, **6**, 85–119.