



# Facial Emotion Expressions in Human–Robot Interaction: A Survey

Niyati Rawal<sup>1</sup> · Ruth Maria Stock-Homburg<sup>1</sup>

Accepted: 4 January 2022 / Published online: 24 June 2022  
© The Author(s) 2022

## Abstract

Facial expressions are an ideal means of communicating one's emotions or intentions to others. This overview will focus on human facial expression recognition as well as robotic facial expression generation. In the case of human facial expression recognition, both facial expression recognition on predefined datasets as well as in real-time will be covered. For robotic facial expression generation, hand-coded and automated methods i.e., facial expressions of a robot are generated by moving the features (eyes, mouth) of the robot by hand-coding or automatically using machine learning techniques, will also be covered. There are already plenty of studies that achieve high accuracy for emotion expression recognition on predefined datasets, but the accuracy for facial expression recognition in real-time is comparatively lower. In the case of expression generation in robots, while most of the robots are capable of making basic facial expressions, there are not many studies that enable robots to do so automatically. In this overview, state-of-the-art research in facial emotion expressions during human–robot interaction has been discussed leading to several possible directions for future research.

**Keywords** Facial emotion recognition · Facial emotion expressions · Human–robot interaction · Survey · Overview

## 1 Introduction

Robots are no longer just machines being used in factories and industries. There is a growing need and demand towards robots sharing space with humans as collaborative robotics or assistive robotics [35,63]. Robots are, now, increasingly being deployed in a variety of domains as receptionists [120], educational tutors [49,59], household supporters [111] and caretakers [25,49,67,125]. Thus, there is a need for these social robots to effectively interact with humans, both verbally and non-verbally. Facial expressions are non-verbal signals that can be used to indicate one's current status in a conversation, e.g., via backchanneling or rapport [3,31].

Perceived sociability is an important aspect in human–robot interaction (HRI) and users want robots to behave in a friendly and emotionally intelligent manner [28,48,99,105]. For social robots to be more anthropomorphic and for human–robot interaction to be more like human–human interaction (HHI), robots need to be able to understand human

emotions and appropriately respond to those human emotions. Stock and Merkle show that emotional expressions of anthropomorphic robots become increasingly important in business settings as well [118,121]. The authors of [119] emphasize that robotic emotions are particularly important for the acceptance of a robot by the user. Thus, emotions are pivotal for HRI [122]. In any interaction, 7% of the affective information is conveyed through words, 38% is conveyed through tone, and 55% is conveyed through facial expressions [92]. This makes facial expressions an indispensable mode of affective communication. Accordingly, numerous studies have examined facial expressions of emotions during HRI [e.g.2,8,15,17–19,33,38,50,81,81,91,91,110,116].

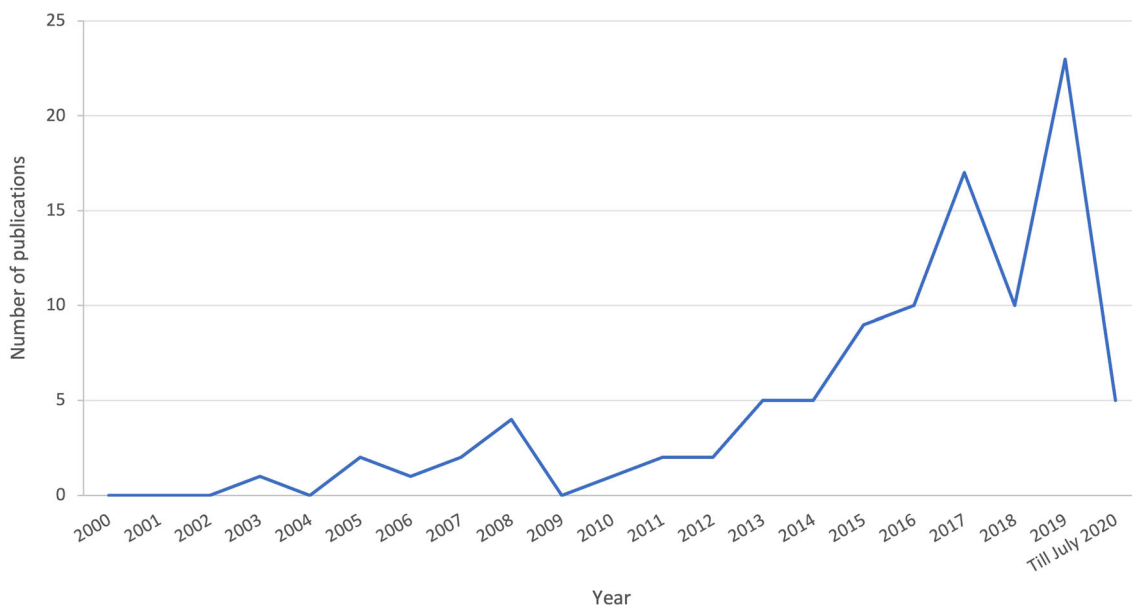
In any HHI, human beings first infer the emotional state of the other person and then accordingly generate facial expressions in response to their peer. The generated emotion could be a result of parallel empathy (generating the same emotion as the peer) or reactive empathy (generating emotion in response to the peer's emotion) [26]. Similarly, in the case of HRI, we would like to study robots recognizing human emotion as well as robots generating their emotion as a response to human emotion.

There has been a growth in the number of papers on facial expressions in HRI in the last decade. Between 2000 and 2020 (see Fig. 1), there has been a gradual increase in

✉ Ruth Maria Stock-Homburg  
RSH@tu-darmstadt.de

Niyati Rawal  
niyati.rawal@tu-darmstadt.de

<sup>1</sup> Technical University Darmstadt, Hochschulstr. 1, 64295 Darmstadt, Germany



**Fig. 1** Publications on emotion recognition of human faces during HRI and generation of facial expressions of robots

the number of publications. Thus, the overarching research question is: What has been done so far on facial emotion expressions in human–robot interaction, and what still needs to be done?

In Sect. 2 the framework of the overview is outlined, followed by the method of selection of studies in Sect. 3. Recognition of human facial expressions and generation of facial expressions by robots are covered in Sects. 4 and 5. The current state of the art and future research are discussed in Sect. 6 with the conclusion in Sect. 7.

## 2 Framework of the Overview

This overview focuses on two aspects: (1) recognition of human facial expressions and (2) generation of facial expressions by robots. The review framework (Fig. 3) is based on these two streams. (1) Recognition of human facial expressions is further subdivided depending on whether the recognition takes place on (a) a predefined dataset or in (b) real-time. (2) Generation of facial expressions by robots is also subdivided depending on whether the facial generation is (a) hand-coded or (b) automated, i.e., facial expressions of a robot are generated by moving the features (eyes, mouth) of the robot by hand-coding or automatically using machine learning techniques.

## 3 Method

Studies with the keywords “facial expression recognition AND human–robot interaction / HRI”, “facial expression

recognition” and “facial expression generation AND human–robot interaction / HRI” between 2000 and 2020 were reviewed on Google Scholar.

In this overview, studies that use voice or body gestures as a modality for emotional expression but do not involve facial expressions are not included. Studies that involve HRI with humans having mental disorders like autism are also not included. Furthermore, studies that work on single emotion such as recognition of smile or facial expression generation of anger are not included. In total, 175 studies of 276 were rejected (Fig. 2).

In Table 3, various studies on facial expression recognition are listed. Here, studies with an accuracy of greater than 90% for facial expression recognition on predefined datasets are selected. For real-time facial expression recognition, all studies that perform facial expression recognition in a human–robot interaction scenario are listed.

## 4 Recognition of Human Facial Expressions

Earlier, facial expression recognition (FER) consisted of the following steps: detection of face, image pre-processing, extraction of important features and classification of expression (Fig. 4). As deep learning algorithms have become popular, the pre-processed image is directly fed into deep networks (like CNN, RNN etc.) to predict an output [71] (Fig. 5).

In the machine learning algorithms, Viola Jones algorithm and OpenCV were popular choices for face detection. However, dlib face detector and ADABOOST algorithm

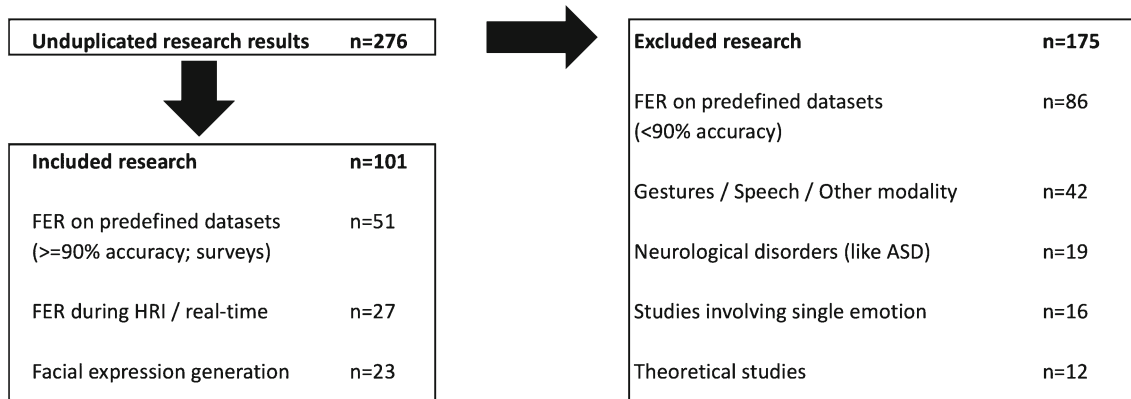


Fig. 2 Flowchart of the literature screening process

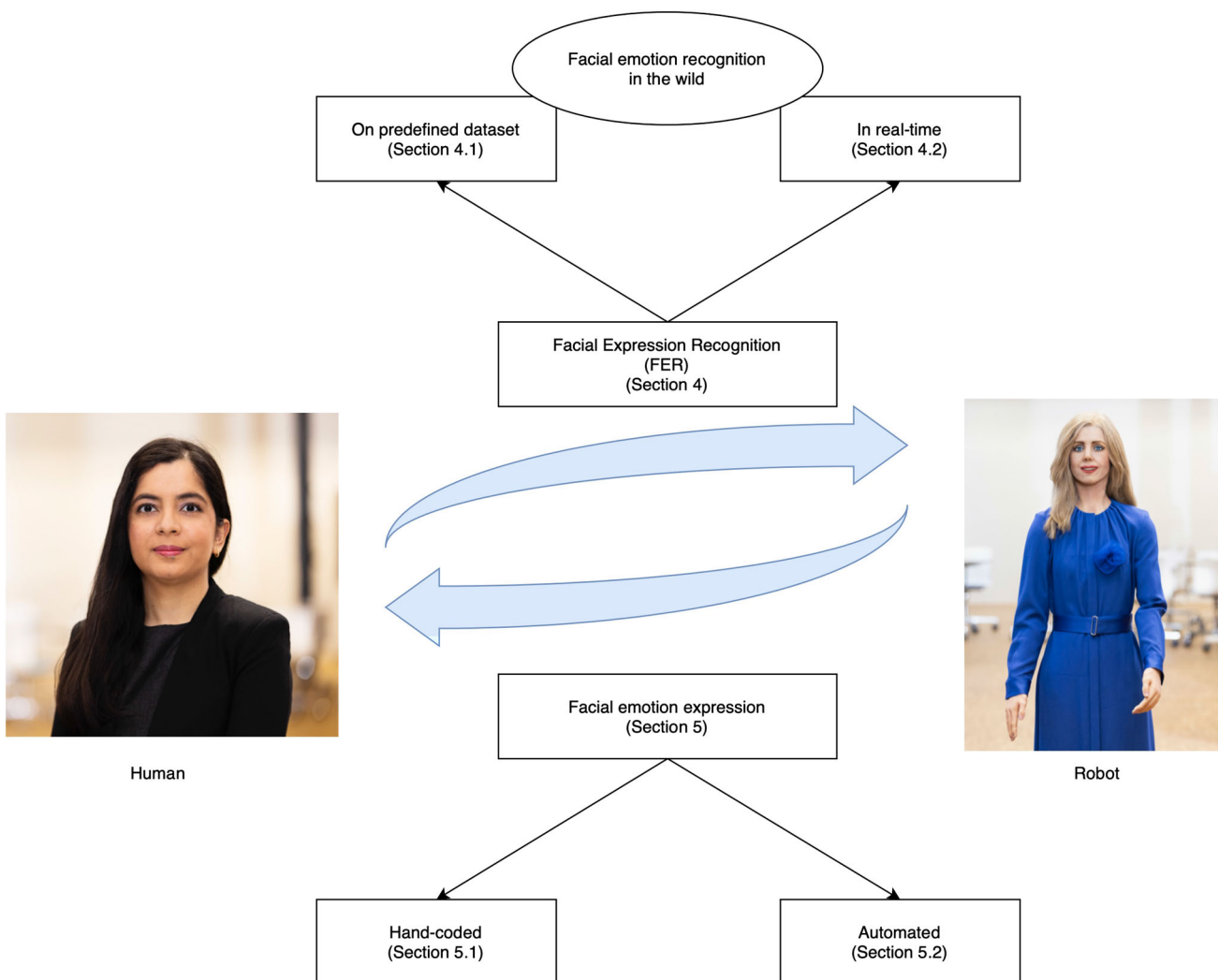
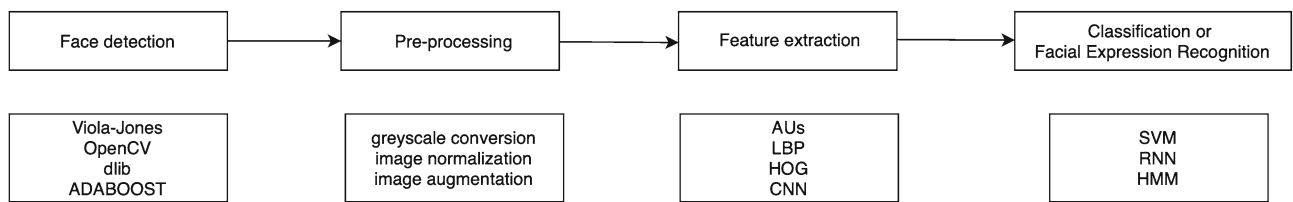
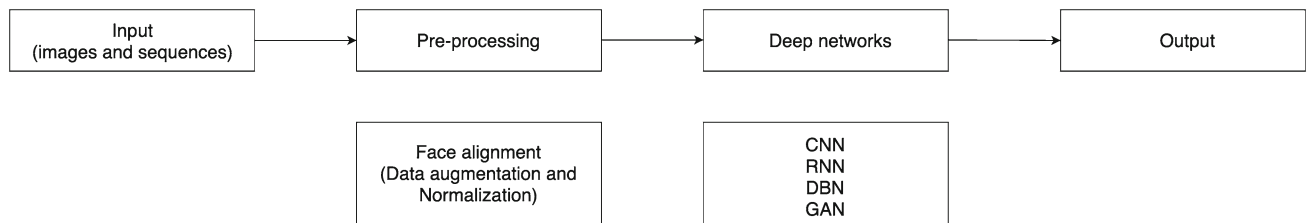


Fig. 3 Framework of the overview



Notes: Action Units (AUs), Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), Convolutional Neural Network (CNN), Support Vector Machine (SVM), Recurrent Neural Network (RNN), Hidden Markov Model (HMM)

**Fig. 4** Process of facial expression recognition in machine learning (adapted from Canedo and Neves [13])



Notes: Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Deep Belief Network (DBN), Generative Adversarial Network (GAN)

**Fig. 5** Process of facial expression recognition in deep learning (adapted from Li and Deng [71])

were also used. To pre-process the images, greyscale conversion, image normalization, image augmentation (such as flip, zoom, rotate etc.) were usually applied. Further, some studies extract the important regions in faces like eyebrows, eyes, nose and mouth (also known as the acting units or AUs) that play an important role in FER. Others use local binary pattern (LBP) or histogram of oriented gradients (HOG) to extract the featural information. Finally, the classification is performed. Most of the studies perform classification for the six universally known emotions (happy, sad, disgust, anger, fear and surprise) and sometimes include a neutral expression. For final classification, k-Nearest Neighbor (KNN), Hidden Markov Model (HMM), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) are used.

In the deep learning algorithms, the input images are first pre-processed by performing face alignment, data augmentation and normalization. Then the images are directly fed into deep networks like CNN, RNN etc. which predict the emotion of the images. The most commonly used classification methods are explained in more detail below. They are arranged in the order in which they were invented.

**KNN:** Nearest neighbor based classifier was first invented in the 1950s [37]. In KNN [57], given the training instances and the class labels, the class label of an unknown instance is predicted. KNN is based on a distance function that measures the difference between two instances. While the Euclidean distance formula is mostly used, there are also other distance formulae such as Hamming distance which can be used.

**HMM:** An HMM [104] was introduced in the late 1960s. It is a doubly embedded stochastic process, bearing a hidden stochastic process (a Markov chain) that is only visible through another stochastic process, producing a sequence of observations. The state sequence can be learned using Viterbi algorithm or Expectation-Modification (EM) algorithm.

**RNN:** RNN [78] was introduced in the 1980s. RNN is a feed-forward neural network that has an edge over adjacent time steps, introducing a notion of time. Hence, RNN is mainly used for a dynamic data input that has a temporal sequence. In RNN, a state depends upon the current input as well as the state of the network at the previous time step, making it possible to contain information from a long time window.

**CNN:** Convolutional Networks [70] were invented in 1989 [69]. CNNs are trainable multistage architectures composed of multiple stages. The input and output of each stage are sets of arrays called feature maps. Each stage of CNN is composed of three layers- a filter bank layer, a non-linearity layer and a feature pooling layer. The network is trained using the backpropagation method. They are used for end-to-end recognition wherein given the input image, the output is predicted by CNNs. They are even used as feature extractors which are further connected with neural networks layers like LSTM or RNN for the prediction.

**SVM:** SVM was invented by Vapnik [128]. In SVM [129], the training data can be separated by a hyperplane.

**LSTM:** LSTM [39] was invented by Hochreiter and Schmidhuber [47]. It also has recurrent connections but

unlike RNN, it is capable of learning long-term dependencies.

Table 1 summarizes the major purpose, application areas, advantages, disadvantages and frequency of use for commonly used algorithms. For the frequency of use, only the number of papers that implement facial expression recognition during HRI or in real-time scenarios were counted. Although RNN has not been used for facial expression recognition during HRI or in real-time, some studies perform facial expression recognition on predefined datasets using RNN.

27 studies on facial expression recognition during HRI were reviewed. Some of the studies have not been performed on a robot platform. These studies perform emotion recognition in real-time and mention HRI as their intended application. The studies are summarized in Table 2. Here, studies that perform facial expression recognition on predefined datasets or studies that perform facial expression recognition but not in real-time were not included.

#### 4.1 FER on Predefined Dataset

Although the goal of this study is to perform FER in real-time and during HRI, the studies on real-time FER are compared with FER on predefined datasets. FER has been carried out on static human images as well as on dynamic human video clips. While some studies, perform facial recognition on still images, others perform facial recognition on videos. In Datcu and Rothkrantz [24], they show that there is an advantage in using data from video frames over still images. This is because videos contain temporal information that is absent in still images.

Results of studies with above 90% accuracy in FER on still images are summarized in Table 3a and on videos are summarized in Table 3b. Table 3a, b are for comparison with Table 3c. Studies are arranged according to their accuracy level. It should be noted that these studies are carried out on predefined datasets consisting of human images and videos and do not involve robots. There are a considerable number of studies that achieve accuracy greater than 90% on CK+, Jaffe and Oulu-Casia datasets on both still images and videos.

#### 4.2 FER in Real-Time

It is easier to achieve high accuracy while performing emotion recognition on predefined datasets as they are recorded under controlled environmental conditions. On the other hand, it is difficult to achieve the same level of accuracy when performing emotion recognition in real-time when the movements are spontaneous. It should be noted that studies that perform facial expression recognition in real-time were carried out under controlled laboratory conditions with little variation in lighting conditions and head poses.

As this study is about facial expressions in HRI, for a robot to be able to recognize emotion, emotion recognition has to be performed in real-time. Table 3c provides studies with facial expression recognition in real-time for HRI. Here, the accuracies are comparatively lower than the accuracies for predefined datasets. As can be seen in Table 3c, only two studies have an accuracy greater than 90%. The robots that are used in the studies are either robotic heads or humanoid robots such as Pepper, Nao, iCub etc. Many studies that perform facial expression recognition in real-time use CNNs, making it a popular choice for facial expression recognition [2,2,8,15,133]. However, the highest accuracy is achieved by Bayesian and Artificial Neural Network (ANN) methods for facial expression recognition in real-time.

## 5 Facial Emotion Expression by Robots

For robots to be empathic, it is necessary that the robots not only be able to recognize human emotions but also be able to generate emotions using facial expressions. Several studies enable robots to generate facial expressions either in a hand-coded or an automated manner (Fig. 6). By hand-coded, we mean that the facial expressions are coded by moving the eyes and mouth of the robot in a desirous manner, and automated is when the emotions are learned automatically using machine learning techniques.

16 studies on facial emotion expression in robots were reviewed. These studies are summarized in Table 4.

### 5.1 Facial Expression Generation is Hand-Coded

Earlier studies started by hand-coding the facial expressions in robots. There is a static as well as dynamic generation of facial expressions on robots.

Among the static methods, there is a humanoid social robot “Alice” that imitates human facial expressions in real-time [91]. Kim et al. [61] introduced an artificial facial expression imitation system using a robot head, Ulkni. As Ulkni is composed of 12 RC servos, with four Degrees of Freedom (DoFs) to control its gaze direction, two DoFs for its neck, and six DoFs for its eyelids and lips, it is capable of making the basic facial expressions after the position commands for actuators are sent from the PC. Bennett and Sabanovic [7] identified minimal features, i.e. movement of eyes, eyebrows, mouth and neck, which are sufficient to identify the facial expression.

In this study, the main program called functions that specified facial expressions according to the direction (used to make or undo an expression) and degree (strength of the expression—i.e. smaller vs. larger). The facial expression functions would in turn call lower functions that moved specific facial components given a direction and degree, fol-

**Table 1** Details about the commonly used algorithms

Algorithm (year)	Major Purpose	Application areas	Advantages	Disadvantages	Frequency of use
KNN (1950s)	Storing of all available cases and classifies new instances by measuring the similarity or difference between two instances using a distance function [57]	Classification of facial expressions	No training required before making predictions	Great computational complexity, because the distance between every sample should be calculated in order to classify [132]	1
HMM (1960s)	Given a sequence of observations, decode the hidden states [104]	Perform FER based on dynamic data input i.e. videos	Capture the dependencies between consecutive measurements	Information from states in the preceding time steps (not the previous time step) cannot be captured	1
RNN (1980s)	Learn temporal dependencies [78]	Perform FER based on dynamic data input i.e. videos	Useful in modelling sequential data	Difficult to learn long-term temporal dependencies as gradients explode or vanish over many time steps [6]	–
CNN (1989)	Given a set of images, extracts featural information such as edges and performs the classification task [70]	Learn the spatial features in an image i.e. perform FER based on static data input	Easier to train and generalizes much better than networks with full connectivity between adjacent layers [68]	Requires a lot of data to prevent over-fitting [138]	13
SVM (1995)	A supervised learning algorithm that can be used for classification or regression on small datasets [129]	Classification of facial expressions	Robust to over-fitting [124]	Does not perform well with large training samples [80]	3
LSTM (1997)	Learn long-term temporal dependencies [39]	Perform FER based on dynamic data input i.e. videos	Ability to learn long-term temporal dependencies [47]	Computationally expensive to train	1

*KNN* k-nearest neighbor, *HMM* hidden markov model, *RNN* recurrent neural network, *CNN* convolutional neural network, *SVM* support vector machine, *LSTM* long short-term memory

**Table 2** Detailed information about studies on emotion recognition and Human Robot Interaction (HRI)

References	Recognition mode	Recognized emotions	Robot	Algorithm type	Major findings
Ahmed et al. [1]	Face	Angry, disgust, fear, happy, neutral, sad, surprise	–	CNN with data augmentation	The model achieved an accuracy of more than 90% for each emotion as it could classify geometrically displaced facial images
Barros et al. [2]	Face	Positive, negative, neutral	iCub	CNN	The network is able to recognize emotions from different environments, different subjects performing spontaneous expressions, and in real-time
Bera et al. [8]	Face	Happy, sad, angry, neutral	Pepper	Bayesian inference and CNN	A multi-channel model to classify pedestrian features into four categories of emotions. Emotional detection accuracy of 85.33% was observed in the validation results
Byeon and Kwak [12]	Face	Happiness, sadness, anger, surprise, disgust, fear	–	3D-CNN	The experimental results on video-based facial expression database revealed that the method showed a good performance in comparison to the conventional methods such as PCA and TMPCA
Chen et al. [15]	Face	Angry, disgust, fear, happy, sad, surprise, neutral	XiaoBao	CNN (VGG-16)	The facial expression recognition in the wild (FERW) model can recognize facial expressions in the real world with an accuracy of 79% and a real-time positive emotion incentive system (PEIS) was able to enhance user experience
Cid et al. [18]	Face, voice	Sad, happy, fear, anger, neutral	Muecas	Dynamic Bayesian Network	The Bayesian approach to the emotion recognition problem presents good results for real-time applications with untrained users in an uncontrolled environment
Dandil and Özdemir [23]	Face	Anger, fear, happy, surprise, sad, neutral	–	CNN	Successful results obtained on real-time video, in changing light and environment conditions with 80% accuracy

Table 2 continued

References	Recognition mode	Recognized emotions	Robot	Algorithm type	Major findings
Deng et al. [27]	Face	Surprise, sad, neutral, happy, fear, disgust, anger	–	cGAN	Can learn not only the regions related to expression but also maximally capturing nuanced characteristics relevant to expression and then transforming the original expression to another expression with identity and other factors preserved
Faria et al. [34]	Face	Afraid, angry, happy, neutral, disgusting, sad, surprised	Nao	Dynamic Bayesian Mixture Model (DBMM)	An overall accuracy around 85% on KDEP dataset and 80% on tests on-the-fly during human–robot interaction
Ge et al. [38]	Face	Happy, sad, fear, disgust, anger, surprise	Robotic head	SVM	The experimental results showed that the proposed nonlinear facial mass-spring model coupled with the SVM classifier is effective to recognize the facial expressions compared with the linear mass-spring model
Inthiam et al. [56]	Face and bodily movement	Positive mood and negative mood	–	HMM	Facial expression alone may be misleading, since it may not be a true expression of inner feeling. By including bodily expression in the analysis, the estimation model gave a better result with estimation accuracy over 70%
Li et al. [72]	Face	Happiness, anger, disgust, fear, sadness, surprise	Harley	CNN-LSTM	CNN and LSTM are combined to exploit their advantages in the proposed model and the system is applied to a humanoid robot to demonstrate its practicability for improving the HRI
Liu et al. [81]	Face	Happy, angry, surprise, fear, disgust, sad, and neutral	Mobile robot	ELM classifier	Recognized human emotions with 80% accuracy
Liu et al. [79]	Face	Anger, scared, sad, surprise, neutral, disgust	–	CNN	An average weighting method was proposed to avoid potential errors in real-time facial expression recognition based on the traditional convolutional neural network



Table 2 continued

References	Recognition mode	Recognized emotions	Robot	Algorithm type	Major findings
Lopez-Rincon [82]	Face	Sadness, happiness, surprise, anger, disgust, fear	Nao	AFFDEX SDK and CNN	The global emotion classification accuracy of the CNN is better than the AFFDEX system, but the face detection in AFFDEX is slightly better
Martin et al. [87]	Face	Anger, disgust, fear, surprise, neutral, happy, sad	–	Active Appearance Model (AAM), Multi Layer Perceptron (MLP) and SVM	Compared three different facial expression classifiers (AAM classifier set, MLP and SVM)
Meghdari et al. [91]	Face	Happiness, sadness, fear, anger, surprise	Alice	ANN	Can recognize emotions with 92.52% accuracy in real-time
Nunes [100]	Face and upper body	Anger, disgust, fear, happiness, sadness, surprise and neutral	–	CNN	Bimodal approach (86.6% accuracy) based on emotion recognition from both face and upper body produced better results than monomodal approach.
Romero et al. [108]	Face, voice, body gestures	Neutral, fear, angry, sad, happy	–	Dynamic Bayesian Network	The standardized variables associated to the AUs of the user were obtained, allowing real-time recognition of each facial expression with different factors, such as lighting conditions, gender, unusual facial features (like injuries or scars), among others
Ruiz-Garcia et al. [110]	Face	Surprise, happy, disgust, angry, fear, sad	–	CNN	Images from 28 participants were collected in an uncontrolled environment to test our CNN emotion recognition model, resulting in a classification rate of 73.55%
Shi et al. [114]	Face and body	Interested, distracted, confused	Pepper	k-Nearest Neighbour (KNN)	A multi-student affect recognition system which, starting from eight basic emotions detected from facial expressions, can infer higher emotional states relevant to a learning context, such as “interested”, “distracted” and “confused”

Table 2 continued

References	Recognition mode	Recognized emotions	Robot	Algorithm type	Major findings
Simul et al. [116]	Face	Neutral, surprise, sad, smile, angry	Ribo	SVM	Robot Ribo recognizes human facial expression, facial gesture movement and detects human gender in real-time
Vithanawasam and Madhusanka [130]	Face and upper body	Anger, fear, bored	–	Fisherface algorithm / position of arms	Anger was correctly predicted 81.54% of the times, followed by bored (72.20%) and fear (68.37%). The results were sensitive to the lighting conditions
Webb et al. [133]	Face	Angry, disgust, fear, happy, neutral, sad, surprise, neutral	Nao	CNN	When evaluated on novel data with nonuniform conditions taken by a Nao robot an accuracy of 79.75% was achieved
Wimmer et al. [134]	Face	Anger, disgust, fear, happy, sad, surprise	B21 robot	Binary Decision Tree	Experimental evaluation reports a recognition rate of 70% on the Cohn–Kanade facial expression database, and 67% in a robot scenario
Wu et al. [136]	Face	Angry, sad, disgust, fear, happy, neutral, surprise	–	Weight-Adapted Convolutional Neural Network (WACNN)	The recognition accuracies of the proposed algorithm were higher than the deep CNN without HGA, indicating a better global optimization ability
Yu and Tapus [142]	Face, body gesture	Neutral, happy, angry, sad	Pepper	Random Forests (RF)	Developed a multimodal emotion recognition model with gait and thermal facial data, which is based on RF model and the modified confusion matrices of two individual models

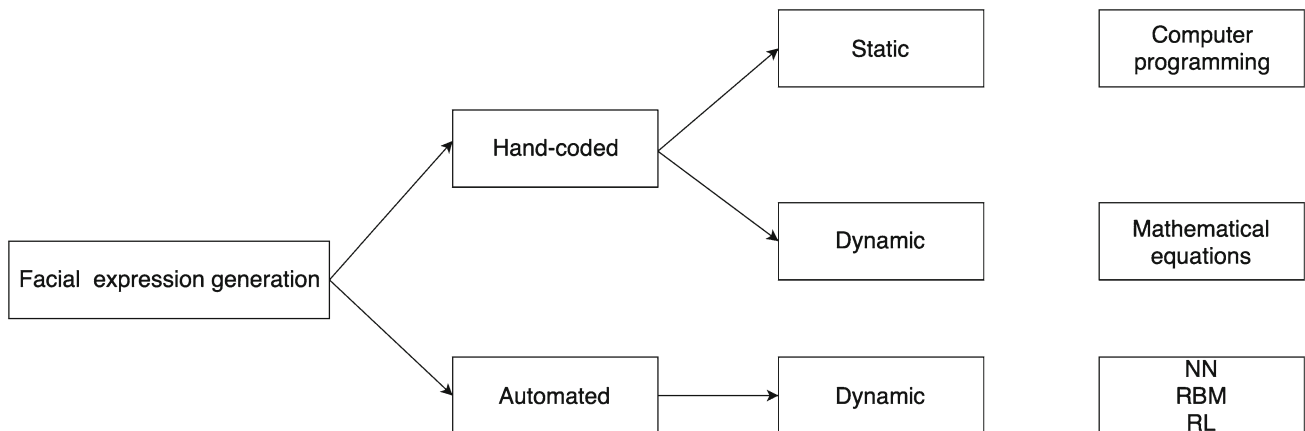
**Table 3** Studies on FER; *Note:* Studies listed according to accuracy level

Study	Dataset	Algorithm	Classes	Accuracy	
(a) With accuracy greater than 90% on static input i.e., human images					
Mistry et al. [95]	CK+/MMI	SVM	7	100%/94.66%	
Kotsia and Pitas [66]	CK	SVM	6	99.7%	
Hossain et al. [51]	Jaffe/CK	GMM	7	99.8%/99.7%	
Kar et al. [60]	CK+	BPNN	6	99.51%	
Mliki et al. [45]	CK/Jaffe	SVM	7	99.24%/96.50%	
Chen et al. [16]	CK+/Jaffe	CNN	7	99.1597%/87.7350%	
Zhang et al. [146]	CK+	CNN	6	98.9%	
Mayya et al. [89]	Jaffe/CK+	CNN	7	98.12%/96.02%	
Minaee and Abdolrashidi [94]	CK+/Jaffe	CNN	7	98.0%/92.8%	
Nwosu et al. [101]	Jaffe/CK+	CNN	7	97.71%/95.72%	
Yang et al. [140]	CK+/Oulu-Casia/Jaffe	CNN	6	97.02%/92.89%/92.21%	
Yang et al. [139]	CK+/Oulu-Casia/Jaffe	WMDNN	6	97.0%/92.3%/92.2%	
Ding et al. [30]	CK+/Oulu-Casia	CNN	8/6	96.8%/87.71%	
Gogić et al. [40]	CK+/Jaffe/ MMI	NN	7	96.48%/85.88%/73.73%	
Kim et al. [62]	CK+/Jaffe	CNN	6	96.46%/91.27%	
Hua et al. [52]	Jaffe	CNN	7	96.44%	
Mannan et al. [85]	CK+	SVM	7	96.36%	
Ruiz-Garcia et al. [110]	KDEF/CK+	CNN-SVM	7	96.26%/95.87%	
Hamester et al. [44]	Jaffe	CNN	7	95.8%	
Meng et al. [93]	CK+/MMI	CNN	6	95.27%/71.55%	
Liliana et al. [77]	CK+	SVM	7	93.93%	
Ferreira et al. [36]	CK+/Jaffe	CNN	8/6	93.64%/89.01%	
Mollahosseini et al. [97]	CK+	DNN	7	93.2%	
Yaddadenet al. [137]	Jaffe/KDEF	KNN	7	92.29%/79.69%	
(b) With accuracy greater than 90% on dynamic input i.e., human videos					
Liang et al. [76]	CK+/Oulu-Casia/MMI	CNN-BiLSTM	6	99.6%/91.07%/80.71%	
Carcagni et al. [14]	CK+	SVM	7	98.5%	
Wu et al. [135]	CK+	HMM	7	98.54%	
Zhang et al. [145]	CK+/Oulu-Casia/MMI	CNN-RNN	6	98.5%/86.25%/81.18%	
Kotsia et al. [65]	CK	SVM	6	98.2%	
Uddin et al. [126]	Depth	DBN	6	96.67%	
Zhao et al. [147]	CK+/Oulu-Casia/MMI	SVM	7	95.8%/74.37%/71.92%	
Elaiwat et al. [32]	CK+/MMI	RBM	7	95.66%/81.63%	
Uddin et al. [127]	CK	CNN	6	95.42%	
Sikka et al. [115]	CK+/Oulu-Csia	HMM	7	94.60%/75.62%	
Kabir et al. [58]	Depth	HMM	6	94.17%	
Study	Robot	Sensor	Algorithm	Classes	Accuracy
(c) FER in real-time i.e., on dynamic input during HRI					
Cid et al. [18]	Muecas	camera	Bayesian	5	94%
Meghdari et al. [91]	Alice	Kinect	ANN	6	92.52%
Simul et al. [116]	Ribo	Webcam	SVM	5	86%
Bera et al. [8]	Pepper	Camera	CNN	4	85.33%
Liu et al. [81]	Mobile robot	Kinect	ELM	7	Above 80%

**Table 3** continued

Study	Robot	Sensor	Algorithm	Classes	Accuracy
Yu and Tapus [142]	Pepper	camera	RF	4	78.125%
Webb et al. [133]	Nao	Camera	CNN	8	79.75%
Chen et al. [15]	XiaoBao	Camera	CNN	7	79%
Barros et al. [2]	iCub	RGB camera	CNN	3	74.2%
Ruiz-Garcia et al. [110]	Nao	Built-in camera	CNN-SVM	7	68.75%
Wimmer et al. [134]	B21 robot	Camera	Binary decision tree	6	67%

*CK* Cohn–Kanade, *Jaffe* Japanese Female Facial Expression, *GAN* Generative Adversarial Network, *KNN* k-Nearest Neighbor, *HMM* Hidden Markov Model, *RNN* Recurrent Neural Network, *CNN* Convolutional Neural Network, *SVM* Support Vector Machine, *LSTM* Long Short-Term Memory, *WMDNN* Weighted Mixture Deep Neural Network, *NN* Neural Network, *ANN* Artificial Neural Network, *ELM* Extreme Learning Machine, *BPNN* Back Propagation Neural Network, *DBN* Deep Belief Network, *RF* Random Forests



Notes: Neural Network (NN), Restricted Boltzmann Machine (RBM), Reinforcement Learning (RL)

**Fig. 6** Facial expression generation techniques

lowing the movement related to specific AUs in the facial acting coding system (FACS).

Breazeal's [9] robot Kismet generated emotions using an interpolation-based technique over a 3-D space, where the three dimensions correspond to valence, arousal and stance. The expressions become intense as the affect state moves to extreme values in the affect space. Park et al. [102] made diverse facial expressions by changing their dynamics and increased the lifelikeness of a robot by adding secondary actions such as physiological movements (eye blinking and sinusoidal motions concerning respiration). A second-order differential equation based on the linear affect-expressions space model is used to achieve the dynamic motion for expressions. Prajapati et al. [103] used a dynamic emotion generation model to convert the facial expressions derived from the human face into a more natural form before rendering them on the robotic face. The model is provided with the facial expression of the person interacting with the system and corresponding synthetic emotions generated are fed to the robotic face.

*Summary of findings* The robot faces are capable of making basic facial expressions as they contain enough DoFs in

the eyes and mouth. They are able to generate static emotions [7,61,91]. Additionally, the robot faces are able to generate dynamic emotions [9,102,103].

## 5.2 Facial Expression Generation is Automated

Some of the studies automatically generate facial expressions on robots. Unlike hand-coded techniques where the commands for the position of features like eyes and mouth are sent from the computer, here, the facial expressions are generated using machine learning techniques such as neural networks and RL.

Breazeal et al. [10] presented a robot Leonardo that can imitate human facial expressions. They use neural networks to learn the direct mapping of a human's facial expressions onto Leonardo's own joint space. In Horii et al. [50], the robot does not directly imitate the human but estimates the correct emotion and generates the estimated emotion using RBM. RBM [46] is a generative model that represents the generative process of data distribution and latent representation, and can generate data from latent signals [98,117,123].

**Table 4** Detailed information about studies on emotion expression and Human Robot Interaction (HRI)

References	Expression mode	Expressed emotions	Robot	Coding	Major findings
Bennett and Sabanovic [7]	Face	Neutral, sadness, happiness, anger, fear, surprise	MiRAE	Hand-coded	Identified minimal features, i.e. movement of eyes, eyebrows, mouth and neck are sufficient to identify the facial expression
Breazeal [9]	Face	Anger, calm, disgust, fear, content, interest, sorrow, surprise, tired	Kismet	Hand-coded	Kismet can generate emotions using an interpolation-based technique over a 3-D space where the three dimensions correspond to valence, arousal and stance
Breazeal et al. [10]	Face		Leonardo	Automated (Neural Network)	Built a robot capable of learning how to imitate facial expressions from simple imitative games played with a human, using biologically inspired mechanisms
Churamani et al. [17]	Face	Anger, happiness, neutral, sadness, surprise	Nico	Automated (RL)	Explored a continuous representation of expression on the NICO robot using the complete face LED matrix to generate expressions
Cid et al. [19]	Face	Sad, happy, fear, anger, neutral	Muecas	Hand-coded	The output of the Bayesian classifier is imitated by the robotic head Muecas
Esfandbod et al. [33]	Face	Neutral, happy, sad, surprised, angry	RASA	Hand-coded	From the subjects' viewpoint, the system's performance was fairly promising with a score of 4.1 out of 5
Ge et al. [38]	Face	Happy, fear, surprise, disgust, sadness, anger	Robotic head	Hand-coded	The robot head are triggered to imitate human facial expressions by the emotion generator engine and can generate a vivid imitation according to the tester's facial expression
Horii et al. [50]	Face, hands, voice	Happiness, neutral, anger, sadness	iCub	Automated (RBM)	The robot does not copy the human's expressions directly but generates expressions on its own after estimating the human's emotion
Ilic et al. [54]	Face	Anger, disgust, fear, happiness, sadness, surprise, indifference	Aisoy	Hand-coded	Learned a model of the emotional value of the robot's facial expressions without humans' explicit feedback
Kim et al. [61]	Face	Neutral, anger, happiness, fear, sadness, surprise	Ulkn	Hand-coded	Introduced an artificial facial expression imitation system using a robot head

Table 4 continued

References	Expression mode	Expressed emotions	Robot	Coding	Major findings
Kishi et al. [64]	Face	Anger, sadness, fear, disgust, surprise, happy	KOBIAN	Hand-coded	Developed a new head for biped walking robot KOBIAN that could express the 6 basic emotions
Liu et al. [81]	Face	Happy, angry, surprise, fear, disgust, sad, neutral	Mobile robot	Hand-coded	Generated facial expressions adapting to human emotions using a four-layer framework designed for the system to recognize human emotion based on HRI
Maeda and Geshi [84]	Face	Anger, contempt, disgust, fear, happiness, neutral, sadness, surprise	TAPIA	Hand-coded	An interactive communication method of a human and a robot based on the Markovian emotional model (MEM) by using the facial expression, was significantly better than using an identical, symmetric or random emotion interaction methods
Meghdari et al. [91]	Face	Happy, sad, angry, surprised, disgusted, neutral	Alice	Hand-coded	A humanoid social robot "Alice" imitates human facial expressions in real-time
Park et al. [102]	Face	Anger, disgust, fear, happiness, sadness, surprise, dislike	Robotic head	Hand-coded	Made diverse facial expressions by changing their dynamics and increased the lifelikeness of a robot by adding secondary actions such as physiological movements (eye blinking and sinusoidal motions concerning respiration)
Yoo et al. [141]	Face	Angry, disgust, fear, happiness, sadness, surprise	Robotic head	Hand-coded	A fuzzy integral-based generation method of composite facial expressions was proposed and demonstrated its effectiveness through the experiment with the developed robotic head

Li and Hashimoto [73] developed a KANSEI communication system based on emotional synchronization. KANSEI is a Japanese term that means emotions, feeling, sensitivity etc. The KANSEI communication system first recognizes human emotion and maps the recognized emotion to the emotion generation space. Finally, the robot expresses its emotion synchronized with the human's emotion in the emotion generation space. When the human changes his/her emotion, the robot also synchronizes its emotion with the human's emotion, establishing a continuous communication between the human and the robot. It was found that the subjects became more comfortable with the robot and communicated more with the robot when there was emotional synchronization.

In Churamani et al. [17], the robot Nico learned the correct combination of eyebrow and mouth wavelet parameters to express its mood using RL. The learned expressions looked slightly distorted but were sufficient to distinguish between various expressions. The robot could also generate expressions that were not limited to the basic five expressions that were learned. For a mixed emotional state (for example, anger mixed with sadness), the model was able to generate novel expression representations representing the mixed state of the mood.

*Summary of findings* In all of the above studies, the robots learn to generate facial expressions automatically using machine learning techniques. While Breazeal [10], Li and Hashimoto [73] used direct mapping of human facial expressions, Horii et al. [50] generated the estimated human's emotion on the robot. In Churamani et al. [17], the robot was able to associate the learned expressions with the context of the conversation.

## 6 Discussion

### 6.1 Summary of the State of the Art

There are already studies having high accuracy (greater than 90%) in facial expression recognition on CK+, Jaffe and Oulu-Casia datasets. (see Table 3a, b). The accuracies on CK+, Jaffe and Oulu-Casia datasets have been as high as 100%, 99.8% and 92.89% respectively. In comparison to this, the accuracy for facial expression recognition in real-time is not as high.

Zhang et al. [146] used a deep convolutional network (DCN) that had an accuracy of 98.9% on CK+ dataset and 55.27% on Static Facial Expressions in the Wild (SFEW) dataset. Here, the same network produced very different results for two different datasets. SFEW [29] consists of close to a real-world environment extracted from movies. The database covers unconstrained facial expressions, varied head poses, large age range, occlusions, varied focus, different resolution of faces, and close to real-world illumination.

In Zhang et al. [146] the accuracy for "in the wild" settings was considerably lower than on CK+ dataset, implying that the expression recognition algorithms can still not handle the variations in environment, head poses etc. in real-world settings.

Table 5 provides possible categories for facial recognition in the wild. It contains the basic emotional facial expressions, situation-specific face occlusions, permanent face features, face movements, situation-specific expressions and side activities during facial expressions.

Most of the current research in facial expression recognition relates to the first category of basic emotional facial expression. Survey articles on facial expression recognition have been cited in the Table 5 [11,13,21,42,43,71,88,109,112]. For more details on individual studies, refer to Table 3. Facial expression recognition in the presence of situation-specific face occlusions like a mouth–nose mask, glasses, hand in front of face etc. has also been studied [74,75,131]. Pose invariant facial expression recognition when the face is moving or turned sideways has also been partially studied [96,113,143,144].

For the facial expression generation, robots can make certain basic facial expressions by moving their eyes, mouth and neck. However, they cannot make as many expressions as human beings due to the limited number of DoFs present in a robot's face. There are relatively fewer studies for automated facial expression generation in robots [10,17,50,73]. While the robots are capable of displaying their facial expressions by manually coding the movement of the eyes and mouth, there are fewer studies that would make a robot learn to display its facial expressions automatically.

Most of the studies on facial expression generation have been carried out on robotic heads or humanoid robots like iCub and Nico [e.g.9,10,17,50]. In Becker-Asano and Ishiguro [5], Geminoid F's facial actuators are tuned such that the readability of its facial expressions is comparable to a real person's static display of emotional expression. It was found that the android's emotional expressions were more ambiguous than that of a real person and 'fear' was often confused with 'surprise'.

An advantage of automated facial expression generation over hand-coded facial expression generation is that in automated facial expression generation, a robot could learn mixed expressions than simply the learned expressions. Unlike in hand-coded facial expression generation, where a robot can only express the emotions that it has learned, in Churamani et al. [17], the robot could express complex emotions that were made up of a combination of emotions.

### 6.2 Future Research

Although facial expression recognition under specific settings has high accuracy and robots can express basic emotions

**Table 5** Possible categories for facial recognition in the wild

Category	Description	Examples	Related studies	Applied algorithm
<b>Basic emotional facial expressions</b>	<b>Expression of basic emotions, such as FACS by Ekman (2001)</b>	<b>Happiness, sadness, anger, disgust, fear, surprise, neutral</b>	<b>[11,13,21,42,43,71,88,109,112]</b>	<b>KNN, HMM, RNN, CNN, SVM and LSTM</b>
<i>Face movements</i>	<i>Movements of the face itself (as a whole)</i>	<i>Moving forward, backward, turning around, turning sideways, jerking the head forward, spinning</i>	<b>[96,113,143,144]</b>	<i>GAN, CNN</i>
<i>Situation-specific face occlusions</i>	<i>Occlusions of the face due to situational requirements</i>	<i>Mouth-nose mask, glasses, hand in front of face, resting hand on mouth, headset</i>	<b>[74,75,131]</b>	<i>CNN</i>
<i>Permanent face features</i>	<i>Permanently installed features of the face</i>	<i>Artificial eye, beard</i>	<b>[74,75,131]</b>	<i>CNN</i>
<i>Situation-specific expressions</i>	<i>Facial expressions of the face due to situational requirements</i>	<i>Nodding, blinking, looking down, yawning, shake head in agreement, eye roll, closing eyes, lip biting, pursed lips, sticking tongue out, winking</i>		
<i>Side activities during facial expression</i>	<i>Additional activities during the expressions of emotions</i>	<i>Talking, eating, drinking, brushing teeth, fixing hair, combing hair, biting nails, cleaning eyes with a hand, coughing, supporting the face with a hand, itching on face, blowing nose, sneezing, rubbing eyes, sipping through a straw, applying cream</i>		

(1) examples were generated based on 50 life observations and 50 video-based observations by the authors; (2) bold = very well understood in research; italic = partly understood in research; bolditalic = hardly understood



through facial expressions, there are several possible directions for future research in this area.

*Suggestion 1: Performing facial expression recognition in the wild needs to be emphasized upon.*

To efficiently recognize facial expressions in real-time and in a real-world environment, the robot should be able to perform facial expression recognition with varied head poses, varied focus, presence of occlusions, different resolutions of the face and varied illumination conditions. The studies that perform facial expression recognition in real-time are limited to a laboratory environment which is far different from a real-world scenario. A good study would be the one where facial expression recognition in the wild is performed.

Some studies perform facial expression recognition in the wild, but their accuracy is much less than the accuracy on predefined datasets like CK+, Jaffe, MMI etc. To increase the efficiency of facial expression recognition in real-world scenarios, the performance of facial expression recognition in the wild needs to be improved. This can also be used to recognize facial expressions in real-time. Based on this, a direct adaptation of emotions would make HRI smoother.

*Suggestion 2: Facial expressions during activities like talking, nodding etc. need to be studied.*

Situation-specific expressions (nodding, yawning, blinking, looking down) and side activities during facial expressions (talking, eating, drinking, sneezing) in Table 5 have not been studied. To understand vivid expressions, it is required to be able to recognize facial expressions for all categories. Humans also express emotions while interacting with someone verbally, such as smiling while speaking when they are happy. In this case, it should be possible to recognize a smile during speech.

*Suggestion 3: Combine facial expression recognition with the data from other modalities such as voice, text, body gestures and physiological data to improve the emotion recognition rate.*

Although this overview focuses on facial expression recognition, it may be possible to control one's face and not express the emotion one is truly experiencing. Some studies combine facial expression recognition with audio data, body gestures or physiological data for an improved emotion recognition [41,53,83]. There are very few studies that combine facial data with both audio and physiological data [106,107] and studies that analyze all modalities (face, voice, text, body gestures and physiological signals) have not been found. Humans can recognize the emotion of a person quickly and effectively by taking into account their facial expression, body gestures, voice and words. Combining facial, audio, text and body gestures with physiological data could lead to a higher emotion recognition rate by machine learning algorithms than by humans.

*Suggestion 4: How should a robot react towards a given human emotion?*

In HHI, a human's reaction to a given emotion is either a result of parallel empathy or reactive empathy [26]. It should be studied with which emotion should a robot appropriately react to a given human emotion. Moreover, it needs to be studied if a robot should be able to express negative emotions. Most of the existing studies allow a robot to be able to express basic emotions (anger, fear, happiness, neutral, sadness, surprise). It may be reasonable for a robot to react with a sad expression when a human being expresses anger. But, should a robot be able to express extreme emotions such as anger?

For facial expression generation, while robots are capable of displaying facial expressions both static and dynamic, they are unable to generate facial expressions when they are speaking. For example, robots could smile while talking to express their happiness or they could speak with a frown when angry. Robots could also express their emotions through partial facial or bodily gestures instead of showing a full face expression. For example, tilting head down to express sadness, frowning to express anger, eyes wide open to express surprise and raising eyebrows.

*Suggestion 5: Robots should be able to recognize and generate facial expressions with various intensities.*

Emotions form a continuous range and can have various intensities. If one is less happy, one would smile less. Similarly, if someone is very happy, the smile would also be big. It should be possible to recognize not just the emotion but the intensity of emotion. Moreover, in most of the existing studies, robots express their emotions with only one configuration per emotion. Robots should also be able to express their emotions with different intensities. Finally, it needs to be studied whether the intensity of emotion with which a robot reacts to a given human emotion has any effects on the human and whether these effects are positive or negative.

*Suggestion 6: Robots should be able to express their emotions through a combination of body gestures and facial expressions.*

While in this overview, we focus on robotic facial expressions, there are other articles where emotional expression is performed through the robot's body postures [4,20,22,55,86,90]. A potential future study could be to compare the robot's facial expressions with robot's bodily expressions and also with the combination of facial and bodily expressions to see if there is any difference in the recognition of these.

*Suggestion 7: Robots should be able to both recognize and generate complex emotions such as that of thinking, calm and bored states.*

For both facial expression recognition and generation, there is a need to go beyond the basic seven emotions to recognizing and generating more complex emotions such as calm, fatigued, bored etc. It might be difficult to generate complex emotions given the hardware limitations of the robot, but if

this is made possible, robots could express a wider range of emotions similar to human beings.

## 7 Conclusion

This overview emphasizes the recognition of human facial expressions and the generation of robotic facial expressions. There are already plenty of studies having high accuracy for facial expression recognition on pre-existing datasets. Accuracy on facial expression recognition in the wild is considerably lower than the experiments which have been conducted under controlled laboratory conditions. For human facial emotion recognition, future work would be to improve emotion recognition for non-frontal head poses in presence of occlusions (i.e. emotion recognition in the wild). It should be made possible to recognize emotions during speech as well emotions with varying intensities. In the case of facial expression generation in robots, robots are capable of making the basic facial expressions. Few studies perform autonomous facial generation in robots. In the future, there could be studies comparing robotic facial expressions with the robot's bodily expressions and also with a combination of facial and bodily expressions to see if there is any difference in recognizing these. Robots should be able to express their emotion with partial bodily or facial gestures while speaking. They should also be express their emotions with various intensities instead of a single configuration per emotion. Lastly, there is a need to go beyond the basic seven expressions for both facial expression recognition and generation.

**Acknowledgements** The authors thank Vignesh Prasad for his insightful comments.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This research was funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft). The authors also thank the ZEVEDI Hessen and the leap in time foundation for the grateful funding of the project.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** The authors declare that there are no compliance issues with this research.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence,

unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahmed TU, Hossain S, Hossain MS, ul Islam R, Andersson K (2019) Facial expression recognition using convolutional neural network with data augmentation. In: 2019 joint 8th international conference on informatics, electronics vision (ICIEV) and 2019 3rd international conference on imaging, vision pattern recognition (icIVPR), pp 336–341
- Barros P, Weber C, Wermter S (2015) Emotional expression recognition with a cross-channel convolutional neural network for human–robot interaction. In: 2015 IEEE-RAS 15th international conference on humanoid robots (humanoids), pp 582–587
- Bavelas J, Gerwing J (2011) The listener as addressee in face-to-face dialogue. *Int J Listen* 25:178–198
- Beck A, Cañamero L, Hiolle A, Damiano L, Cosi P, Tesser F, Sommavilla G (2013) Interpretation of emotional body language displayed by a humanoid robot: a case study with children. *Int J Soc Robot* 5(3):325–334
- Becker-Asano C, Ishiguro H (2011) Evaluating facial displays of emotion for the android robot geminoid f, pp 1–8
- Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. *IEEE Trans Neural Netw* 5(2):157–166
- Bennett CC, Sabanovic S (2014) Deriving minimal features for human-like facial expressions in robotic faces. *Int J Soc Robot* 6:367–381
- Bera A, Randhavane T, Prinja R, Kapsaskis K, Wang A, Gray K, Manocha D (2019) The emotionally intelligent robot: improving social navigation in crowded environments. *ArXiv arXiv:1903.03217*
- Breazeal C (2003) Emotion and sociable humanoid robots. *Int J Hum Comput Stud* 59(1–2):119–155
- Breazeal C, Buchsbaum D, Gray J, Gatenby D, Blumberg B (2005) Learning from and about others: towards using imitation to bootstrap the social understanding of others by robots. *Artif Life* 11:31–62. <https://doi.org/10.1162/1064546053278955>
- Buciu I, Kotsia I, Pitas I (2005) Facial expression analysis under partial occlusion, pp v/453–v/456, vol 5
- Byeon YH, Kwak KC (2014) Facial expression recognition using 3d convolutional neural network. *Int J Adva Comput Sci Appl* 5(12)
- Canedo D (2019) Facial expression recognition using computer vision: a systematic review. *Appl Sci*. <https://doi.org/10.3390/app9214678>
- Carcagni P, Del Coco M, Leo M, Distanto C (2015) Facial expression recognition and histograms of oriented gradients: a comprehensive study. *Springerplus* 4(1):645
- Chen H, Gu Y, Wang F, Sheng W (2018) Facial expression recognition and positive emotion incentive system for human–robot interaction. In: 2018 13th world congress on intelligent control and automation (WCICA), pp 407–412
- Chen X, Yang X, Wang M, Zou J (2017) Convolution neural network for automatic facial expression recognition. In: 2017 international conference on applied system innovation (ICASI), pp 814–817
- Churamani N, Barros P, Strahl E, Wermter S (2018) Learning empathy-driven emotion expressions using affective modulations

18. Cid F, Moreno J, Bustos P, Núñez P (2014) Muecas: a multi-sensor robotic head for affective human robot interaction and imitation. *Sensors* (Basel, Switzerland) 14:7711–7737
19. Cid F, Prado JA, Bustos P, Núñez P (2013) A real time and robust facial expression recognition and imitation approach for affective human–robot interaction using gabor filtering. In: 2013 IEEE/RSJ international conference on intelligent robots and systems, pp 2188–2193 . <https://doi.org/10.1109/IROS.2013.6696662>
20. Cohen I (2010) Recognizing robotic emotions: facial versus body posture expression and the effects of context and learning. Master's thesis
21. Corneanu CA, Simón MO, Cohn JF, Guerrero SE (2016) Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition: history, trends, and affect-related applications. *IEEE Trans Pattern Anal Mach Intell* 38(8):1548–1568
22. Costa S, Soares F, Santos C (2013) Facial expressions and gestures to convey emotions with a humanoid robot. In: International conference on social robotics. Springer, pp 542–551
23. Dandil E, Özdemir R (2019) Real-time facial emotion classification using deep learning. *Data Sci Appl* 2(1):13–17
24. Datu C, Rothkrantz L (2007) Facial expression recognition in still pictures and videos using active appearance models. A comparison approach, p 112. <https://doi.org/10.1145/1330598.1330717>
25. Dautenhahn K (2007) Methodology & themes of human–robot interaction: a growing research field. *Int J Adv Robot Syst* 4:15
26. Davis M (2018) Empathy: a social psychological approach
27. Deng J, Pang G, Zhang Z, Pang Z, Yang H, Yang G (2019) cgan based facial expression recognition for human–robot interaction. *IEEE Access* 7:9848–9859
28. de Graaf M, Allouch S, Van Dijk JA (2016) Long-term acceptance of social robots in domestic environments: insights from a user's perspective
29. Dhall A, Goecke R, Lucey S, Gedeon T (2011) Static facial expression analysis in tough conditions: data, evaluation protocol and benchmark. In: 2011 IEEE international conference on computer vision workshops (ICCV Workshops), pp 2106–2112
30. Ding H, Zhou SK, Chellappa R (2017) Facenet2expnet: regularizing a deep face recognition net for expression recognition. In: 2017 12th IEEE international conference on automatic face gesture recognition (FG 2017), pp 118–126
31. Drolet A, Morris MW (2000) Rapport in conflict resolution: accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts. *J Exp Soc Psychol* 36:26–50
32. Elaiwat S, Bennamoun M, Boussaïd F (2016) A spatio-temporal rbm-based model for facial expression recognition. *Pattern Recogn* 49:152–161
33. Esfandbod A, Rokhi Z, Taheri A, Alemi M, Meghdari A (2019) Human–robot interaction based on facial expression imitation. In: 2019 7th international conference on robotics and Mechatronics (ICRoM), pp 69–73
34. Faria DR, Vieira M, Faria FCC, Premebida C (2017) Affective facial expressions recognition for human–robot interaction. In: 2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 805–810
35. Feil-Seifer D, Matarić MJ (2011) Socially assistive robotics. *IEEE Robot Autom Mag* 18(1):24–31
36. Ferreira PM, Marques F, Cardoso JS, Rebelo A (2018) Physiological inspired deep neural networks for emotion recognition. *IEEE Access* 6:53930–53943
37. Fix E (1951) Discriminatory analysis: nonparametric discrimination, consistency properties. USAF School of Aviation Medicine
38. Ge S, Wang C, Hang C (2008) A facial expression imitation system in human robot interaction
39. Gers FA, Schmidhuber J (2000) Recurrent nets that time and count. In: Proceedings of the IEEE-INNS-ENNS international joint conference on neural networks. IJCNN 2000. Neural computing: new challenges and perspectives for the new millennium. IEEE, vol 3, pp 189–194
40. Gogić I, Manhart M, Pandžić I, Ahlberg J (2018) Fast facial expression recognition using local binary features and shallow neural networks. *Vis Comput*. <https://doi.org/10.1007/s00371-018-1585-8>
41. Gunes H, Piccardi M (2007) Bi-modal emotion recognition from expressive face and body gestures. *J Netw Comput Appl* 30(4):1334–1345
42. Gunes H, Schuller B (2013) Categorical and dimensional affect analysis in continuous input: current trends and future directions. *Image Vis Comput* 31:120–136
43. Gunes H, Schuller B, Pantic M, Cowie R (2011) Emotion representation, analysis and synthesis in continuous space: a survey. *Face Gesture* 2011:827–834
44. Hamester D, Barros P, Wermter S (2015) Face expression recognition with a 2-channel convolutional neural network, pp 1–8 . <https://doi.org/10.1109/IJCNN.2015.7280539>
45. Hazar M, Fendri E, Hammami M (2015) Face recognition through different facial expressions. *J Signal Process Syst*. <https://doi.org/10.1007/s11265-014-0967-z>
46. Hinton G, Salakhutdinov R (2006) Reducing the dimensionality of data with neural networks. *Science* (New York, NY) 313:504–7. <https://doi.org/10.1126/science.1127647>
47. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
48. Hoffman G, Breazeal C (2006) Robotic partners' bodies and minds: an embodied approach to fluid human–robot collaboration. In: AAAI workshop—technical report
49. Hoffman G, Zuckerman O, Hirschberger G, Luria M, Shani Sherman T (2015) Design and evaluation of a peripheral robotic conversation companion. In: Proceedings of the Tenth Annual ACM/IEEE international conference on human–robot interaction, HRI '15, pp 3–10. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/2696454.2696495>
50. Horii T, Nagai Y, Asada M (2016) Imitation of human expressions based on emotion estimation by mental simulation. *Paladyn J Behav Robot*. <https://doi.org/10.1515/pjbr-2016-0004>
51. Hossain MS, Muhammad G (2017) An emotion recognition system for mobile applications. *IEEE Access* 5:2281–2287
52. Hua W, Dai F, Huang L, Xiong J, Gui G (2019) Hero: human emotions recognition for realizing intelligent internet of things. *IEEE Access* 7:24321–24332
53. Huang Y, Yang J, Liao P, Pan J (2017) Fusion of facial expressions and eeg for multimodal emotion recognition. *Comput Intell Neurosci* 2017:1–8. <https://doi.org/10.1155/2017/2107451>
54. Ilic, D., Žužić, I., Brscic, D.: Calibrate my smile: robot learning its facial expressions through interactive play with humans, pp 68–75 (2019)
55. Inthiam J, Hayashi E, Jitviriyi W, Mowshowitz A (2019) Mood estimation for human–robot interaction based on facial and bodily expression using a hidden Markov model. In: 2019 IEEE/SICE international symposium on system integration (SII). IEEE, pp 352–356
56. Inthiam J, Mowshowitz A, Hayashi E (2019) Mood perception model for social robot based on facial and bodily expression using a hidden Markov model. *J Robot Mechatron* 31:629–638
57. Jiang L, Cai Z, Wang D, Jiang S (2007) Survey of improving k-nearest-neighbor for classification. In: Fourth international conference on fuzzy systems and knowledge discovery (FSKD 2007), vol 1, pp 679–683
58. Kabir MH, Salekin MS, Uddin MZ, Abdullah-Al-Wadud M (2017) Facial expression recognition from depth video with patterns of oriented motion flow. *IEEE Access* 5:8880–8889

59. Kanda T, Hirano T, Eaton D, Ishiguro H (2004) Interactive robots as social partners and peer tutors for children: a field trial. *Hum Comput Interact (Special issues on human–robot interaction)* 19:61–84
60. Kar NB, Babu KS, Jena SK (2017) Face expression recognition using histograms of oriented gradients with reduced features. In: Raman B, Kumar S, Roy PP, Sen D (eds) *Proceedings of international conference on computer vision and image processing*. Springer, Singapore, pp 209–219
61. Kim DH, Jung S, An K, Lee H, Chung M (2006) Development of a facial expression imitation system, pp 3107–3112
62. Kim J, Kim B, Roy PP, Jeong D (2019) Efficient facial expression recognition algorithm based on hierarchical deep neural network structure. *IEEE Access* 7:41273–41285
63. Kirgis FP, Katsos P, Kohlmaier M (2016) *Collaborative robotics*. Springer, Cham, pp 448–453
64. Kishi T, Otani T, Endo N, Kryczka P, Hashimoto K, Nakata K, Takanishi A (2012) Development of expressive robotic head for bipedal humanoid robot. In: 2012 IEEE/RSJ international conference on intelligent robots and systems, pp 4584–4589
65. Kotsia I, Nikolaidis N, Pitas I (2007) Facial expression recognition in videos using a novel multi-class support vector machines variant. In: 2007 IEEE international conference on acoustics, speech and signal processing–ICASSP '07, vol 2, pp II-585–II-588
66. Kotsia I, Pitas I (2007) Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Trans Image Process* 16(1):172–187
67. Kozima H, Nakagawa C, Yasuda Y (2005) Interactive robots for communication-care: a case-study in autism therapy. In: ROMAN 2005. In: *IEEE international workshop on robot and human interactive communication*, pp 341–346
68. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
69. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD (1989) Backpropagation applied to handwritten zip code recognition. *Neural Comput* 1(4):541–551
70. LeCun Y, Kavukcuoglu K, Farabet C (2010) Convolutional networks and applications in vision. In: *Proceedings of 2010 IEEE international symposium on circuits and systems*, pp 253–256
71. Li S, Deng W (2020) Deep facial expression recognition: a survey. *IEEE Trans Affect Comput*, pp 1–1
72. Li TS, Kuo P, Tsai T, Luan P (2019) Cnn and lstm based facial expression analysis model for a humanoid robot. *IEEE Access* 7:93998–94011
73. Li Y, Hashimoto M (2011) Effect of emotional synchronization using facial expression recognition in human–robot communication
74. Li Y, Zeng J, Shan S, Chen X (2018) Occlusion aware facial expression recognition using cnn with attention mechanism. *IEEE Trans Image Process*, pp 1–1 (2018)
75. Li Y, Zeng J, Shan S, Chen X (2018) Patch-gated cnn for occlusion-aware facial expression recognition. In: 2018 24th international conference on pattern recognition (ICPR), pp 2209–2214
76. Liang D, Liang H, Yu Z, Zhang Y (2019) Deep convolutional bilstm fusion network for facial expression recognition. *Vis Comput* 36:499–508
77. Liliana DY, Basaruddin C, Widyanto MR (2017) Mix emotion recognition from facial expression using svm-crf sequence classifier. In: *Proceedings of the international conference on algorithms, computing and systems, ICACS '17*. Association for Computing Machinery, New York, NY, USA, pp 27–31
78. Lipton ZC, Berkowitz J, Elkan C (2015) A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*
79. Liu K, Hsu C, Wang W, Chiang H (2019) Real-time facial expression recognition based on cnn. In: 2019 international conference on system science and engineering (ICSSE), pp 120–123 . <https://doi.org/10.1109/ICSSE.2019.8823409>
80. Liu P, Choo KKR, Wang L, Huang F (2017) Svm or deep learning? A comparative study on remote sensing image classification. *Soft Comput* 21(23):7053–7065
81. Liu ZT, Wu M, Cao W, Chen LF, Xu J, Zhang R, Zhou M, Mao J (2017) A facial expression emotion recognition based human–robot interaction system. *IEEE/CAA J Autom Sin* 4:668–676. <https://doi.org/10.1109/JAS.2017.7510622>
82. Lopez-Rincon A (2019) Emotion recognition using facial expressions in children using the nao robot. In: 2019 international conference on electronics, communications and computers (CONIELECOMP), pp 146–153
83. Ma F, Zhang W, Li Y, Huang SL, Zhang L (2020) Learning better representations for audio-visual emotion recognition with common information. *Appl Sci* 10:7239. <https://doi.org/10.3390/app10207239>
84. Maeda Y, Geshi S (2018) Human–robot interaction using Markovian emotional model based on facial recognition. In: 2018 Joint 10th international conference on soft computing and intelligent systems (SCIS) and 19th international symposium on advanced intelligent systems (ISIS). IEEE, pp 209–214
85. Mannan MA, Lam A, Kobayashi Y, Kuno Y (2015) Facial expression recognition based on hybrid approach. In: Huang DS, Han K (eds) *Adv Intell Comput Theor Appl*. Springer, Cham, pp 304–310
86. Marmpena M, Lim A, Dahl TS, Hemion N (2019) Generating robotic emotional body language with variational autoencoders. In: 2019 8th international conference on affective computing and intelligent interaction (ACII). IEEE, pp 545–551
87. Martin C, Werner U, Gross H (2008) A real-time facial expression recognition system based on active appearance models using gray images and edge images. In: 2008 8th IEEE international conference on automatic face gesture recognition, pp 1–6 . <https://doi.org/10.1109/AFGR.2008.4813412>
88. Martinez B, Valstar MF, Jiang B, Pantic M (2019) Automatic analysis of facial actions: a survey. *IEEE Trans Affect Comput* 10(3):325–347
89. Mayya V, Pai RM, Pai MMM (2016) Automatic facial expression recognition using DCNN. *Proc Comput Sci* 93:453–461. <https://doi.org/10.1016/j.procs.2016.07.233>
90. McColl D, Nejat G (2014) Recognizing emotional body language displayed by a human-like social robot. *Int J Soc Robot* 6(2):261–280
91. Meghdari A, Shouraki S, Siamy A, Shariati A (2016) The real-time facial imitation by a social humanoid robot
92. Mehrabian A (1968) Communication without words. *Psychol Today* 2:53–56
93. Meng Z, Liu P, Cai J, Han S, Tong Y (2017) Identity-aware convolutional neural network for facial expression recognition. In: 2017 12th IEEE international conference on automatic face gesture recognition (FG 2017), pp 558–565 (2017)
94. Minaee S, Abdolrashidi A (2019) Deep-emotion: facial expression recognition using attentional convolutional network. *arXiv preprint arXiv:1902.01019*
95. Mistry K, Zhang L, Neoh SC, Lim CP, Fielding B (2017) A micro-GA embedded PSO feature selection approach to intelligent facial emotion recognition. *IEEE Trans Cybern* 47(6):1496–1509
96. Moeini A, Moeini H, Faez K (2014) Pose-invariant facial expression recognition based on 3d face reconstruction and synthesis from a single 2d image. In: 2014 22nd international conference on pattern recognition, pp 1746–1751. <https://doi.org/10.1109/ICPR.2014.307>
97. Mollahosseini A, Chan D, Mahoor MH (2016) Going deeper in facial expression recognition using deep neural networks. In: 2016 IEEE winter conference on applications of computer vision (WACV), pp 1–10 (2016)

98. Ngiam J, Khosla A, Kim M, Nam J, Lee H, Ng AY (2011) Multimodal deep learning. In: ICML
99. Nicolescu MN, Mataric MJ (2001) Learning and interacting in human–robot domains. *IEEE Trans Syst Man Cybern Part A Syst Hum* 31(5):419–430
100. Nunes ARV (2019) Deep emotion recognition through upper body movements and facial expression. Master's thesis, Aalborg University
101. Nwosu L, Wang H, Lu J, Unwala I, Yang X, Zhang T (2017) Deep convolutional neural network for facial expression recognition using facial parts. In: 2017 IEEE 15th international conference on dependable, autonomic and secure computing, 15th international conference on pervasive intelligence and computing, 3rd international conference on big data intelligence and computing and cyber science and technology congress, pp 1318–1321
102. Park JW, Lee H, Chung M (2014) Generation of realistic robot facial expressions for human robot interaction. *J Intell Robot Syst* 78:443–462
103. Prajapati S, Shrinivasa Naika CL, Jha S, Nair S (2013) On rendering emotions on a robotic face, pp 1–7
104. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE* 77(2):257–286
105. Ray C, Mondada F, Siegwart R (2008) What do people expect from robots? pp 3816–3821
106. Ringeval F, Eyben F, Kroupi E, Yuce A, Thiran JP, Ebrahimi T, Lalanne D, Schuller B (2015) Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data. *Pattern Recogn Lett* 66:22–30. <https://doi.org/10.1016/j.patrec.2014.11.007> (**Pattern Recognition in Human Computer Interaction**)
107. Ringeval F, Schuller B, Valstar M, Jaiswal S, Marchi E, Lalanne D, Cowie R, Pantic M (2015) Av+ec 2015—the first affect recognition challenge bridging across audio, video, and physiological data
108. Romero P, Cid F, Núñez P (2013) A novel real time facial expression recognition system based on candid-3 reconstruction model. In: Proceedings of the XIV workshop on physical agents (WAF 2013), Madrid, Spain, pp 18–19
109. Rouast PV, Adam MTP, Chiong R (2019) Deep learning for human affect recognition: insights and new developments. *ArXiv arXiv:1901.02884*
110. Ruiz-Garcia A, Elshaw M, Altahhan A, Palade V (2018) A hybrid deep learning neural approach for emotion recognition from facial expressions for socially assistive robots. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-018-3358-8>
111. Saerbeck M, Bartneck C (2010) Perception of affect elicited by robot motion, pp. 53–60
112. Sariyanidi E, Gunes H, Cavallaro A (2015) Automatic analysis of facial affect: a survey of registration, representation, and recognition. *IEEE Trans Pattern Anal Mach Intell* 37(6):1113–1133
113. Saxena S, Tripathi S, Sudarshan TSB (2019) Deep dive into faces: Pose illumination invariant multi-face emotion recognition system. In: 2019 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp. 1088–1093. <https://doi.org/10.1109/IROS40897.2019.8967874>
114. Shi Y, Chen Y, Ardila LR, Venture G, Bourguet ML (2019) A visual sensing platform for robot teachers. In: Proceedings of the 7th international conference on human–agent interaction, pp 200–201
115. Sikka K, Dhall A, Bartlett M (2015) Exemplar hidden Markov models for classification of facial expressions in videos. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 18–25
116. Simul NS, Ara NM, Islam MS (2016) A support vector machine approach for real time vision based human robot interaction. In: 2016 19th international conference on computer and information technology (ICCIT), pp 496–500
117. Srivastava N, Salakhutdinov RR (2012) Multimodal learning with deep Boltzmann machines. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ (eds) *Advances in neural information processing systems*, vol 25. Curran Associates Inc, Red Hook, pp 2222–2230
118. Stock R, Merkle M (2018) Can humanoid service robots perform better than service employees? a comparison of innovative behavior cues. <https://doi.org/10.24251/HICSS.2018.133>
119. Stock R, Nguyen MA (2019) Robotic psychology what do we know about human–robot interaction and what do we still need to learn?
120. Stock RM (2016) Emotion transfer from frontline social robots to human customers during service encounters: testing an artificial contagion model. In: ICIS
121. Stock RM, Merkle M (2017) A service robot acceptance model: user acceptance of humanoid robots during service encounters. In: IEEE international conference on pervasive computing and communications workshops (PerCom Workshops), pp 339–344. <https://doi.org/10.1109/PERCOMW.2017.7917585>
122. Stock-Homburg R (2021) Survey of emotions in human–robot interaction—after 20 years of research: What do we know and what have we still to learn? *Int J Soc Robot*
123. Sukhbaatar S, Makino T, Aihara K, Chikayama T (2011) Robust generation of dynamical patterns in human motion by a deep belief nets. In: Asian conference on machine learning, pp 231–246
124. Taira H, Haruno M (1999) Feature selection in SVM text categorization. In: AAAI/IAAI, pp 480–486
125. Tanaka F, Cicourel A, Movellan J (2007) Socialization between toddlers and robots at an early childhood education center. *Proc Natl Acad Sci USA* 104:17954–8
126. Uddin MZ, Hassan MM, Almogren A, Alamri A, Alrubaian M, Fortino G (2017) Facial expression recognition utilizing local direction-based robust features and deep belief network. *IEEE Access* 5:4525–4536
127. Uddin MZ, Khaksar W, Torresen J (2017) Facial expression recognition using salient features and convolutional neural network. *IEEE Access* 5:26146–26161
128. Vapnik VN (1995) *The nature of statistical learning theory*. Springer, New York
129. Vapnik VN (1999) An overview of statistical learning theory. *IEEE Trans Neural Netw* 10(5):988–999
130. Vithanawasam T, Madhusanka A (2019) Face and upper-body emotion recognition using service robot's eyes in a domestic environment, pp 44–50
131. Wang K, Peng X, Yang J, Meng D, Qiao Y (2020) Region attention networks for pose and occlusion robust facial expression recognition. *IEEE Trans Image Process* 29:4057–4069
132. Wang Q, Ju S (2008) A mixed classifier based on combination of HMM and KNN. In: 2008 fourth international conference on natural computation, vol 4, pp 38–42
133. Webb N, Ruiz-Garcia A, Elshaw M, Palade V (2020) Emotion recognition from face images in an unconstrained environment for usage on social robots. In: 2020 international joint conference on neural networks (IJCNN), pp. 1–8
134. Wimmer M, MacDonald BA, Jayamani D, Yadav A (2008) Facial expression recognition for human–robot interaction—a prototype. In: Sommer G, Klette R (eds) *Robot Vis*. Springer, Berlin, pp 139–152
135. Wu C, Wang S, Ji Q (2015) Multi-instance hidden Markov model for facial expression recognition. In: 2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG), vol 1, pp 1–6
136. Wu M, Su W, Chen L, Liu Z, Cao W, Hirota K (2019) Weight-adapted convolution neural network for facial expression recog-

- inition in human–robot interaction. *IEEE Trans Syst Man Cybern Syst*
137. Yaddaden Y, Bouzouane A, Adda M, Bouchard B (2016) A new approach of facial expression recognition for ambient assisted living. In: Proceedings of the 9th ACM international conference on PErvasive technologies related to assistive environments, PETRA '16. Association for Computing Machinery, New York, NY, USA
  138. Yamashita R, Nishio M, Do RKG, Togashi K (2018) Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 9(4):611–629
  139. Yang B, Cao J, Ni R, Zhang Y (2018) Facial expression recognition using weighted mixture deep neural network based on double-channel facial images. *IEEE Access* 6:4630–4640
  140. Yang H, Yin L (2017) CNN based 3d facial expression recognition using masking and landmark features. In: 2017 seventh international conference on affective computing and intelligent interaction (ACII), pp 556–560
  141. Yoo B, Cho S, Kim J (2011) Fuzzy integral-based composite facial expression generation for a robotic head. In: 2011 IEEE international conference on fuzzy systems (FUZZ-IEEE 2011), pp 917–923
  142. Yu C, Tapus A (2019) Interactive robot learning for multimodal emotion recognition. In: Salichs MA, Ge SS, Barakova EI, Cabibihan JJ, Wagner AR, Castro-González Á, He H (eds) Social robotics. Springer, Cham, pp 633–642
  143. Zhang F, Zhang T, Mao Q, Xu C (2018) Joint pose and expression modeling for facial expression recognition. In: 2018 IEEE/CVF conference on computer vision and pattern recognition, pp 3359–3368. <https://doi.org/10.1109/CVPR.2018.00354>
  144. Zhang F, Zhang T, Mao Q, Xu C (2020) Geometry guided pose-invariant facial expression recognition. *IEEE Trans Image Process* 29:4445–4460. <https://doi.org/10.1109/TIP.2020.2972114>
  145. Zhang K, Huang Y, Du Y, Wang L (2017) Facial expression recognition based on deep evolutionary spatial-temporal networks. *IEEE Trans Image Process* 26(9):4193–4203
  146. Zhang Z, Luo P, Loy CC, Tang X (2016) From facial expression recognition to interpersonal relation prediction. *Int J Comput Vis*
  147. Zhao L, Wang Z, Zhang G (2017) Facial expression recognition from video sequences based on spatial-temporal motion local binary pattern and Gabor multiorientation fusion histogram. *Math Probl Eng*

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.