https://doi.org/10.1007/s42979-022-01116-x

SURVEY ARTICLE

Human Emotion: A Survey focusing on Languages, Ontologies, Datasets, and Systems

Mohammed R. Elkobaisi[1] · Fadi Al Machot[2] · Heinrich C. Mayr[1]

Received: 25 May 2021 / Accepted: 28 March 2022 / Published online: 10 May 2022

Abstract
Emotions are an essential part of a person's mental state and influence her/his behavior accordingly. Consequently, emotion recognition and assessment can play an important role in supporting people with ambient assistance systems or clinical treatments. Automation of human emotion recognition and emotion-aware recommender systems are therefore increasingly being researched. In this paper, we first consider the essential aspects of human emotional functioning from the perspective of cognitive psychology and, based on this, we analyze the state of the art in the whole field of work and research to which automated emotion recognition belongs. In this way, we want to complement the already published surveys, which usually refer to only one aspect, with an overall overview of the languages ontologies, datasets, and systems/interfaces to be found in this area. We briefly introduce each of these subsections and discuss related approaches regarding methodology, technology, and publicly accessible artefacts. This comes with an update to recent findings that could not yet be taken into account in previous surveys. The paper is based on an extensive literature search and analysis, in which we also made a particular effort to locate relevant surveys and reviews. The paper closes with a summary of the results and an outlook on open research questions.

Keywords Human emotion · Emotion markup language · Emotion ontology · Emotion datasets · Emotion models

Introduction

Human emotion recognition refers to the process of discovering a person's state of mind and reactions, that are associated with a specific event or situation. It is performed with sophisticated algorithms that, thanks to technological progress, have become increasingly efficient in tracking emotional changes. The research interest is motivated above all by promising applications in the areas of ambient assistance, decision support, and the prevention of emotional hazards [1]. According to "Research and Markets" report,[1] the global market of emotion recognition is expected to rise about 65 Billion by 2023, with 39% Annual Growth Rate between 2017 and 2023.

Algorithms for detecting emotional changes use different context information and modalities: Facial cues, speech variations, gestures, data from body or brain sensors, and more. Some approaches rely on the person's self-assessment to measure instant emotion. Each approach exhibits advantages and disadvantages, depending on the contexts or settings in which they are used.

This paper aims to provide a comprehensive overview of the current state of important aspects of machine recognition of human emotions to assist the interested community in developing and improving methods, techniques, and systems. In doing so, we not only add recent findings to the body of knowledge established by previously published reviews and original papers. Rather, we bring together different perspectives: (1) Systems that use emotional data, (2) the datasets they use, (3) description languages, and

✉ Mohammed R. Elkobaisi
M3mohammed@edu.aau.at

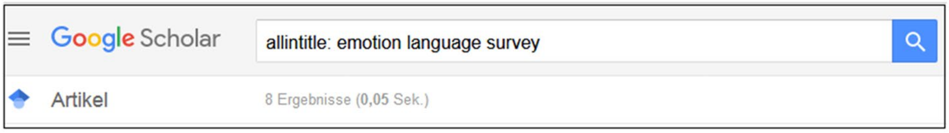
Fadi Al Machot
Fadi.al.machot@nmbu.no

Heinrich C. Mayr
heinrich.mayr@aau.at

[1] Application Engineering, Alpen-Adria-Universität Klagenfurt, 9020 Klagenfurt, Austria

[2] Faculty of Science and Technology, Norwegian University for Life Science (NMBU), Ås, Norway

[1] https://www.marketresearchfuture.com/reports/emotion-detection-recognition-market-3193, Accessed at 01.10.2019.



allintitle: emotion language survey

8 Ergebnisse (0,05 Sek.)

Fig. 1 Google Scholar search result using term "InTitle"

(4) domain-specific models and ontologies as the basis for recognition algorithms.

To this end, we have analyzed a large number of papers from various relevant areas such as emotion interpretation theories, modeling languages, ontologies, data sets, and output interfaces from the perspective of the use of emotion recognition by intelligent systems. In addition, we have tried to get an overview of the survey articles already existing in this area and to evaluate them to be able to present as comprehensive a compendium as possible with our paper.

The paper has the following structure: "Methodology" briefly describes the criteria we used to search literature sources and, in particular, previously published review articles. "Human Emotional Functioning" deals with the human emotional functioning, the manifestations of human emotions, and approaches to their interpretation. "Languages and Ontologies" focuses on ontologies and domain-specific modeling approaches to describe emotions. "Datasets" gives an overview of the data sets used in the literature for emotion recognition, followed by a compilation of systems for emotion recognition "Systems for Emotion Recognition" and of systems that exploit emotion data "Information Systems (IS) Exploiting Emotion Data". The paper concludes with a summary of the results "Summary" and a brief outlook on open research questions "Open Research Questions".

Methodology

The comprehensive overview provided in this paper is based on findings from existing studies and literature available online (e.g., Google scholar[2]). Five aspects play a special role in the context of emotion recognition: models, languages, ontologies, datasets, and systems. In total, we selected 230 relevant literature sources on these aspects and referenced them in the bibliography. In contrast to our approach, which considers all five aspects, most published studies focus on one of these aspects or on one emotional modality like spoken language, video, image, etc. For example, study [2] focuses on speech datasets from very recent years, and [3, 4] overviews miscellaneous facial datasets only based on video, audio, or image input. [5, 6] review different approaches for detecting emotion from text only.

To validate this finding, we systematically searched Google Scholar for articles that had relevant terms in their title (e.g., allintitle: emotion language survey). In detail, the search terms used were combinations of the phrases "emotion model," "emotion language," "emotion ontology," "emotion dataset," and "emotion information system," in conjunction with the keywords "survey," and "review," respectively, as shown in Fig. 1. The keyword "review" has been selected additionally as it seems that authors use it more often than "survey". The quantitative results of these searches are shown in Fig. 2 which also contains the result of a search with all emotion aspects combined in one query: as expected, such a study could not be found.

Human Emotional Functioning

Emotions are specific reactions to an experienced event [7]. In the domain of cognitive psychology, the term emotion refers to "specific sets of physiological and mental dispositions triggered by the brain in response to the perceived significance of a situation or object" [8]. Theories of cognitive psychology suggest that the individual interpretation of a situation or event influences emotions and behaviors [9–11]: guided by their beliefs, different people may interpret the same events differently. Psychologists study how people interpret and understand their worlds and how emotions prepare people for acting and regulating their social behavior [12]. Emotional expression is also an important part of emotional function, as human emotions can influence a person's physical reactions [13]. These emotional reactions are mediated by speech, facial expressions, body gestures, or physiological signals [14].

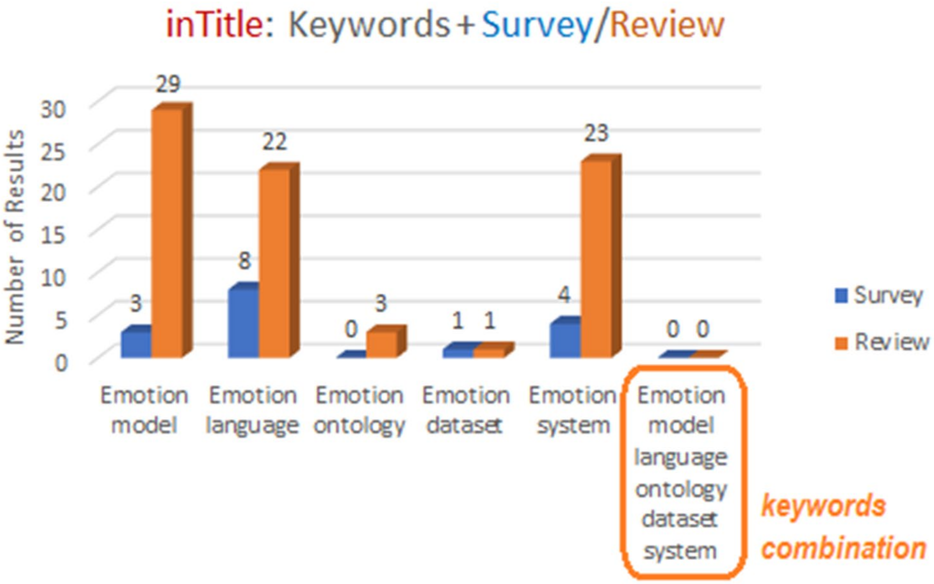
In this section, we present approaches to categorize, model, interpret, and understand emotions.

Categorizing the Manifestations of Emotion

The emotion felt by a person cannot be "grasped" physiologically in everyday life, so that one is dependent on the external manifestations in which emotions express themselves: Emotion controls many modes of human visible behavior like gestures, facial expression, postures, voice tone, respiration, and skin color. All this affects the way people interact with each other. Psychologists and engineers have conducted several studies to understand and categorize the manifestations of emotions.

[2] https://scholar.google.com.

Fig. 2 The graph shows the results of human emotion aspects identified in the literature followed by "Survey" and "Review" keywords



Facial Expressions

Ekman and Friesen [15] conducted a study on the universality of facial expressions and classified them in relation to six basic emotions: anger, happiness, sadness, disgust, surprise, and fear. Based here-on, they developed a taxonomy of facial muscle movements (Facial Action Coding System, FACS) that is general enough to describe a person's basic emotional state through analyzing the relationship between points on the person's face [16, 17]. Many studies use Ekman's results as the basis for the recognition task. A similar approach [18] describes facial expression as a result of the so-called action units (AUs) that capture the possible movements of facial muscles. Such facial movements occur in most people and can reflect certain emotions in combination. Table 1 shows the connection between certain combinations of action units and the basic emotion they express. For instance, happiness is calculated as a combination of action unit 6 (cheek raiser) and 12 (lip corner puller) [19].

Vocal Expressions

The speech signal contains both explicit (linguistic, i.e., the message presented) and implicit (paralinguistic) information such as references to the emotional state of the speaker. Acoustic speech signals are mainly generated by the vibration of the vocal chords, whereby the frequency determines the pitch of the tone. Further parameters of the speech signal are the intensity, duration of the spectral speech properties, contour, melt-frequency cepstral coefficients (MFCCs), tone base, and voice quality [22]. The variation of the pitch and its intensity together form the prosody. Many features of speech signals are used to extract emotions. Table 2 outlines different emotional behavior which are listed in relation to the common vocal parameters.

Physiological Signals

In general, physiological signals are divided into two categories: (1) signals derived from peripheral nervous system phenomena such as heart rate (ECG) and skin conductance (EMG) and (2) signals derived from the central nervous system such as brain signals (EEG) [23]. Physiological signals can be collected via wearable sensors and evaluated for the classification and identification of emotions [24]. For emotion classification, signal features like frequency, amplitudes, minima, and maxima are analyzed. Popular approaches are Support Vector Machine (SVM) [25], Fisher linear discriminant projection [26], Canonical Correlation analysis (CCA) [27], Artificial Neural Network (ANN) [28], K-Nearest Neighbor (KNN) [29], Adaptive Neuro-Fuzzy Interference System (ANFIS) [30], or the Bayesian network method [31]. There are several ways to derive emotions from physiological signals: (1) measure various parameters of the signal and compare the results to a self-assessment Manikin (SAM) questionnaire [32] (see Fig. 3); (2) estimate emotion based on facial expressions and overall impressions by psychological experts; (3) correlate the results to a gold standard, such as facial recognition or EEG; (4) compare the result with well-known dataset to elicit the emotions, such as the International Affective Picture System (IAPS) dataset [33] to generate comparable results [34].

Table 1 Action units (AUs) that correspond to determine basic emotions [20, 21]

Basic emotion	Action units	Facial description
Happiness	6+12	Cheek Raiser, Lip Corner Puller
Sadness	1+4+15	Inner Brow Raiser, Brow Lowerer, Lip Corner Depressor
Surprise	1+2+5+26	Inner Brow Raiser, Outer Brow Raiser, Upper Lid Raiser, Jaw Drop
Fear	1+2+4+5+7+20+26	Inner Brow Raiser, Outer Brow Raiser, Brow Lowerer, Upper Lid Raiser, Lid Tightener, Lip Stretcher, Jaw Drop
Disgust	9+15+16	Nose Wrinkler, Lip Corner Depressor, Lower Lip Depressor
Anger	4+5+7+23	Brow Lowerer, Upper Lid Raiser, Lid Tightener, Lip Tightener

Table 2 The properties of speech signal-based emotion analysis [22]

Basic emotion	Speech rate	Pitch average	Pitch range	Intensity	Voice quality	Pitch changes	Articulation
Happiness	Faster or slower	Much higher	Much wider	Higher	Breathy blaring	Smooth upward	Normal
Sadness	Slightly slower	Slightly lower	Slightly narrower	Lower	Resonant	Downward inflections	Slurring
Surprise	Much faster	Much higher	–	Higher	–	Rising contour	–
Fear	Much faster	Very much higher	Much wider	Normal	Irregular voicing	Normal	Precise
Disgust	Very much slower	Very much lower	Slightly wider	Lower	Chest tone	Wide downward	Normal
Anger	Slightly faster	Very much higher	Much wider	Highest	Breathy chest tone	Abrupt on chest	Tense

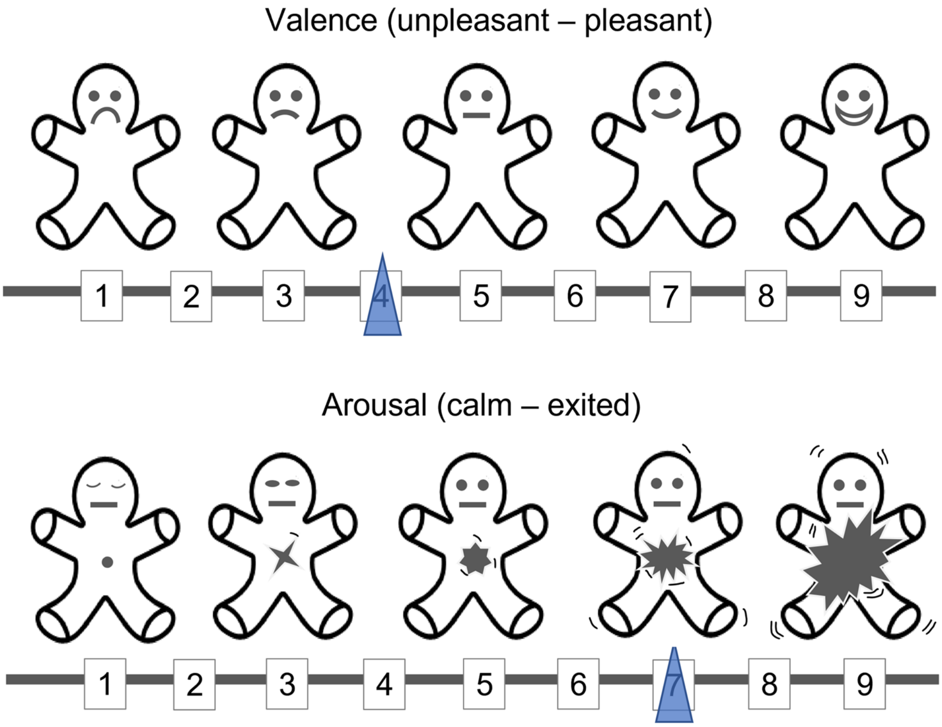


Fig. 3 Self-assessment Manikin to quantify emotion, original see [67]

Body Gesture

Gestures (except sign language) are a form of non-verbal interaction in which a human moves a certain part of the body, especially hands or head. This movement is used to convey a message and additional information such as human emotions [35]. Figure 4 depicts emotion expressions which associated with body pose and motion. In addition, one can deduce emotion parameters from measured movement values such as speed, amplitude, and time expenditure of body parts involved in the various gesture phases (preparation, stroke, and relaxation). Table 3 depicts the frequent arm movements which form a certain emotion.

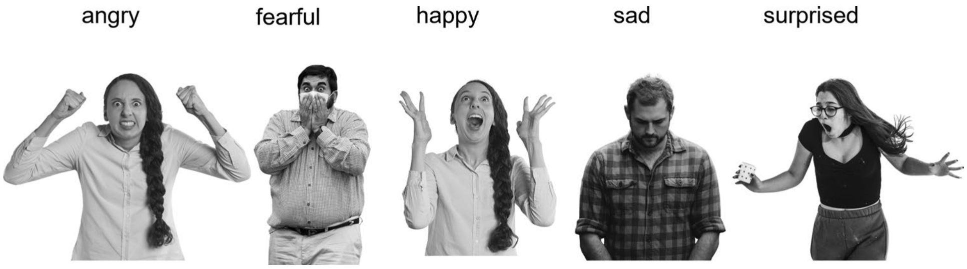


Fig. 4 Emotion interpretation from body movements, original see [208]

Table 3 Gesture-based movement factors that express emotion [36]

Basic emotion	Frequent gesture features
Happiness	High peak flexor and extensor elbow velocities, arms stretched out to the front
Sadness	Longest movement time, smallest amplitude of elbow motion, least elbow extensor velocity
Fear	Arms stretched sideways
Anxiety	Short movement times, constrained torso range of motion
Interest	Lateralized hand/arm movement, arms stretched out to the front
Anger	Lateralized hand/arm movement, arms stretched out to the front, largest amplitude of elbow motion, largest elbow extensor velocity, highest rising arm

Multimodal Emotion

The aforementioned methods of identifying emotions can also be combined, in which case we speak of a multimodal approach to emotion recognition. Fusing modalities together may increase the performance of systems for emotion recognition [14]. For example, the integration of facial expressions and speech signals leads to a new “audiovisual” signal. [37] and the combined evaluation of audio-visual and physiological signals can further reduce the error rate of emotion recognition.

Emotion Theories

In the theoretical discussion, different views on emotions are represented, which are reflected in several theories and descriptive models. A simple approach consists in relating “emotion” to phenomena like anger, fear, or happiness. More sophisticated approaches describe emotions in a multidimensional space with an unlimited number of categories. Brosch et al. [38] distinguish four main directions: the Ekman basic emotion theory [39], the appraisal theory [40], dimensional theories [41–43], and the constructivist theory of emotion [44]. Combining theories leads to further refinements. For instance, [45] combines appraisal and dimensional theory, and [46] claim that there are two main views for emotion classification, namely, basic and dimensional.

Basic Emotion Theory

Ekman [39] assumes emotions to be discrete and related to a fixed number of neural and physically represented states. As already mentioned in Facial Expressions, he proposes a division into six basic emotion categories. Barrett [47] extends this view by associating each emotion with a specific and unmistakable set of bodily and facial responses. An emotion is thus represented by a characteristic syndrome of hormonal, muscular and autonomic responses that are coordinated in time and correlated in intensity [48]. E.g., each emotion comes with different, distinguishable facial movements.

Appraisal Theory

Appraisal theory defines emotions as processes and not as states [49]. It assumes that the interpretation (appraisal, assessment) of a given situation causes a specific emotional reaction in the interpreting person [40]. Consequently, not the situation per se, but the individual assessment of that situation causes the type and intensity of emotional reaction.

Two-Dimensional Theory

The central assumption of dimensional theories is that emotional states can be represented by variation across certain dimensions [42, 43, 50]. In the two-dimensional case, the dimensions valence and arousal are considered (see Fig. 5): Valence denotes the polarity of emotions (positive or negative), and arousal indicates the intensity (high or low).

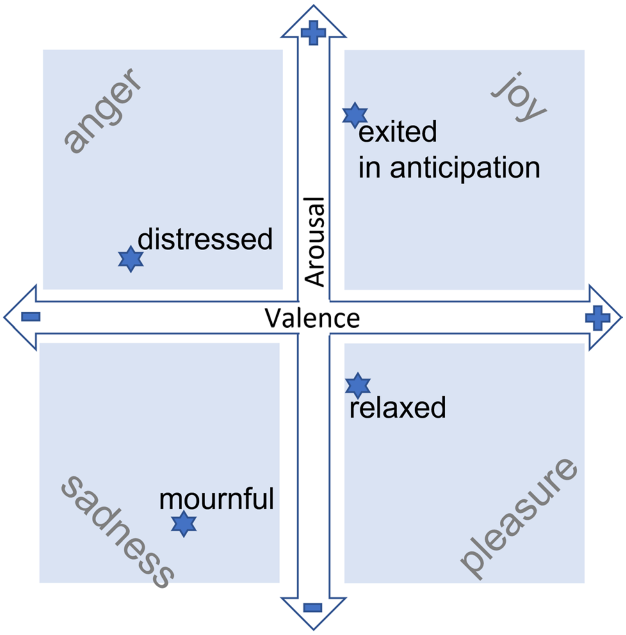


Fig. 5 Two-dimensional arrangement of valence and arousal following [190]

In this way, all emotions can be classified in the arousal-valence coordinate system.

PAD Three-Dimensional Theory

Reference [41] introduce a third dimension and focus on “Pleasure”, “Arousal”, and “Dominance” (*PAD three-dimensional model*; see Fig. 6). Pleasure expresses the range to which the person aware the situation as enjoyable or not, arousal represents the extent to which the situation stimulates the person; dominance describes the extent to which the person is able to control her/his emotional state in the given situation [51]. The PDA model has been used to represent emotions for non-verbal interaction such as body language [52], or to represent emotions of animated characters in virtual worlds [53].

Plutchik’s Wheel of Emotions

Robert Pluchik [54] assumes eight basic emotions, namely, *joy, trust, fear, surprise, sadness, disgust, anger, and anticipation*, the combination of which results into more complex ones. In his “wheel” model (see Fig. 7), these basic emotions form the “spokes” of a wheel and are dyed in different colors from the color spectrum. Related emotions are housed on the same spoke and are gradually colored with the spoke color according to their intensity (arousal). E.g., joy is organized with ecstasy (higher arousal) and serenity (lower arousal) on the same (yellow) spoke. Opposing emotions are arranged on opposite spokes, e.g., joy versus sadness, anger versus fear, trust versus disgust, and surprise versus anticipation [54]. More complex emotions can be represented as combinations of basic emotions, similar to the way primary colors are combined. The wheel rims indicate such combinations. For example, joy and trust combine to be love, boredom, and annoyance combine to contempt.

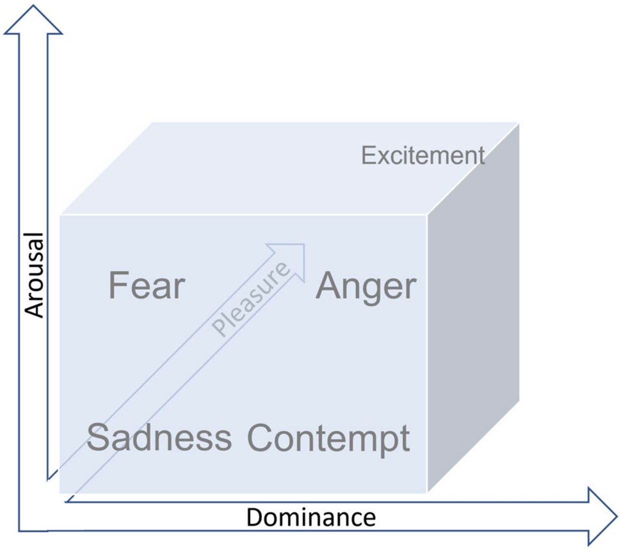


Fig. 6 PAD three-dimensional representation of emotions following [189]

OCC Model of Emotion

Ortony, Clore, & Collins (OCC) [55] assume that emotions arise from assessing observations about events, agents, and objects. Events are manifestations that occur at a certain point in time; agents can be people, animals, machines and similar, or abstract entities such as institutions. Several studies employed the OCC model to reason about emotion or generate emotions in artificial characters. The OCC model classifies emotions into 22 categories (see Fig. 8).

Construction Theory

Man builds up his own knowledge of the world based on his experiences. The constructive model therefore assumes that emotions are psychological connections built from more basic psychological components [44].

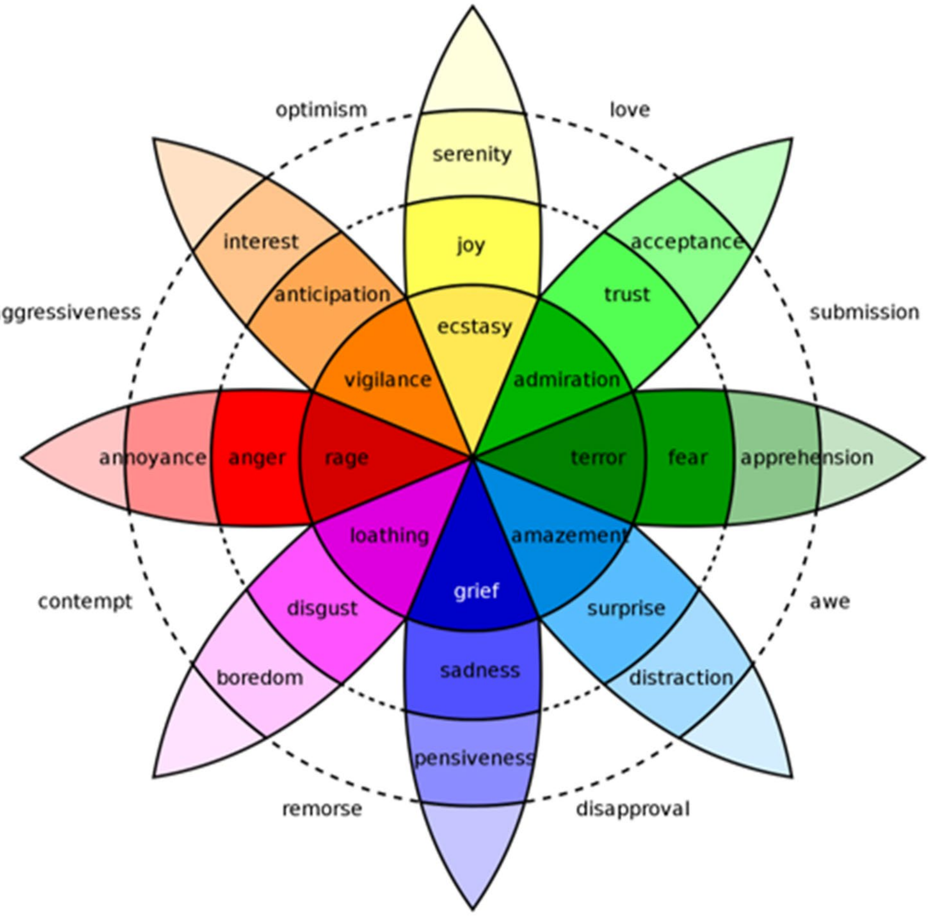


Fig. 7 Plutchik's Wheel of emotions [188], picture from wikimedia.org

Languages and Ontologies

For dealing with emotions in computerized applications, suitable descriptive languages are needed, i.e., languages that allow a representation of, e.g., the models described above. Although, of course, universal modeling languages could be used here—as in all areas—first efforts to develop special domain-specific languages for the field of emotion modeling can be observed. However, these are mainly simply structured markup languages, although the emergence of powerful meta-modeling platforms [41, 43, 50–52] would allow for the definition and use of comprehensive domain-specific conceptual modeling languages [53].

Domain-specific concepts come with various rules, constraints, and semantics [56]. Ontologies are used for their semantic foundation and formal definition [57, 58]. Reference [59, 60] investigate the integration of ontology with domain-specific language at the meta-model level and automated reasoning process. [61] discuss the use of the formal semantics of the Web Ontology Language (OWL)[3] together with reasoning services for addressing constraint definition, suggestions, and debugging. The ontology-based approach presented in [62] allows for integrating different domains and reasoning services. In [63], the authors propose a "User Story Mapping-Based Method" for extracting knowledge from the relevant domain by applying a formal guideline. The authors demonstrate how to establish a full semantic model of a particular domain using ontology concepts.

Languages

Standardizing emotion representations by a closed set of emotion denominators is perceived as being too restrictive. On the other hand, leaving the choice of emotion annotation completely unlimited is considered to be not appropriate [64]. Consequently, there seems to be no specific standard language that covers all aspects of emotions as they appear in the approaches and theories described above. However, markup languages have been presented that provide a set of syntax and semantic description concepts and thus satisfy the demands of some researchers. This section describes popular markup languages and the purpose behind their development.

[3] https://www.w3.org/TR/owl-features.

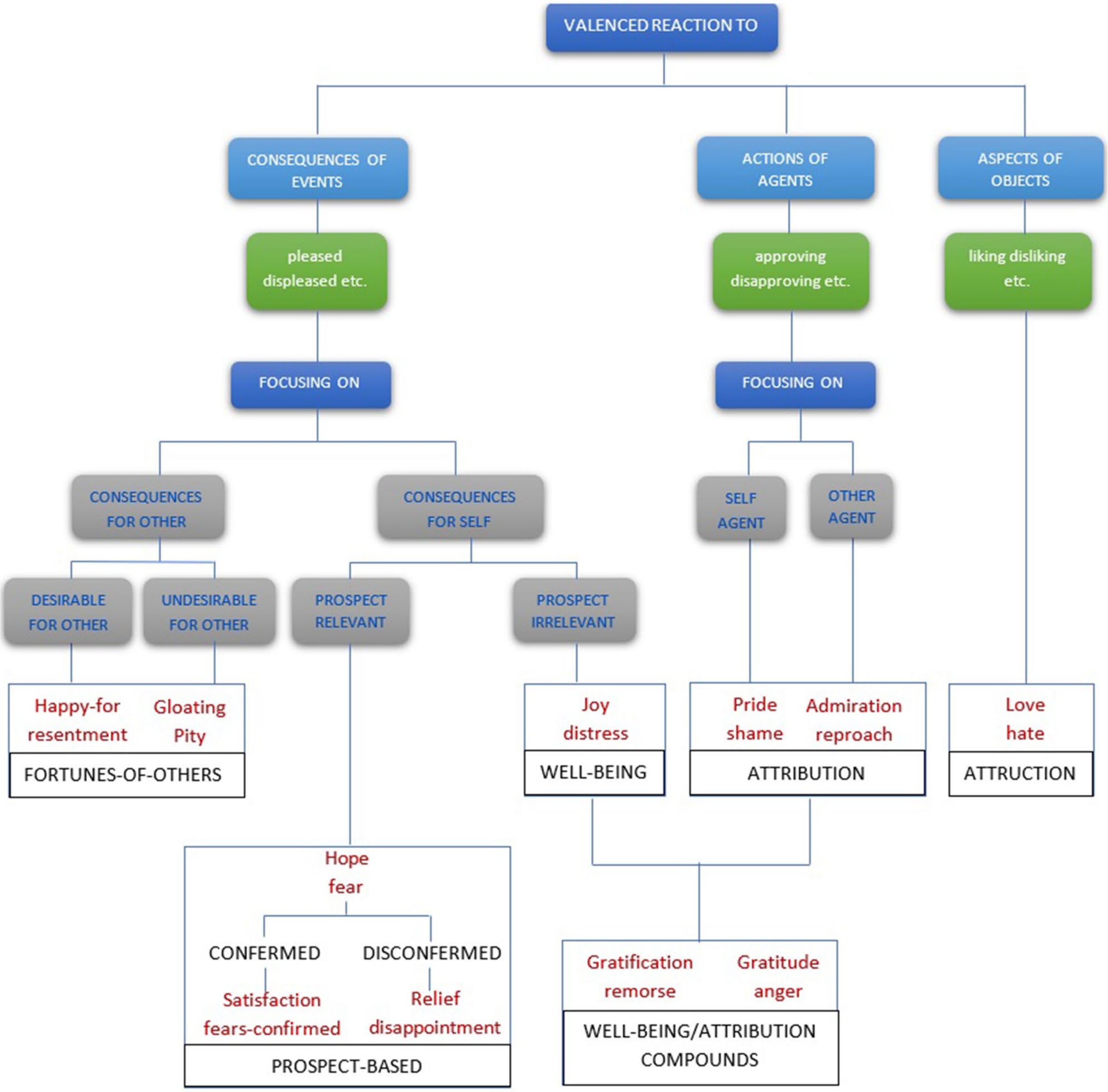


Fig. 8 Structure of OCC model (the 22 categories are dyed with red color)

(EmotionML) Emotion Markup Language [65] has been presented by the W3C Consortium to allow for (1) manual annotation of material involving emotionality, (2) automatic recognition of emotions from sensor data, speech recordings, etc., and (3) generation of emotion-related system responses, which may involve reasoning about the emotional aspects, e.g., in interactive systems.[4] Due to the lack of generally agreed descriptors, EmotionML does not come with a fixed emotion vocabulary, but proposes possible structural elements and allows users to "plug in" their own favorite vocabulary. Concerning the basic structural elements, it is assumed that emotions can be described in terms of categories or a small number of dimensions, and that emotions involve triggers, appraisals, feelings, expressive behavior, and action tendencies [65]. Consequently, EmotionML is a quite general-purpose XML-based language that implicitly realizes Ekman's basic emotion theory and [39] and dimensional theories [41].

[4] https://www.w3.org/TR/emotionml/#s1.

(EARL) Emotion Annotation and Representation Language [66] is an XML-based language designed specifically for representing emotions in technological contexts with a focus on emotion annotation, recognition, and generation. EARL can represent emotions as categories, dimensions, or sets of appraisal scale [67]. EARL can be utilized for manual emotion annotation and for generating affective system such as speech synthesizers or embodied conversational agents (ECAs) [68].

(EMMA) Extensible MultiModal Annotation [69] is a markup language intended for representing multimodal user inputs (e.g., speech, pen, keystroke, or gesture) in a standardized way for further processing. As such, it was designed for use in the so-called Multimodal Interaction Frameworks (MMI) to solve uncertainty in user input interpretations. EMMA distinguishes between *instance data* (contained within an EMMA interpretation for an input or an ouput) and the *data model* which is optionally specified as an annotation of that instance. Multiple interpretations of a single input or output are possible. The EMMA structural syntax is quite simple and provides *elements* for the organization of interpretations and instances like *root* (container with version and name space information), *interpretation element*, *container*, or *literal element*. EMMA markup is intended to be generated automatically by components and expected to include speech recognizer, semantic interpreters, and interaction managers. It concentrates on single input/output (e.g., single natural language utterance). EMMA may be used as a "host language" for plugging-in EmotionML for covering emotion interpretation, such that all EmotionML information is encoded in element and attribute structures. As an example,[5] see the following analysis of a non-verbal vocalization where emotion is described as a low-intensity state, maybe “boredom”:

(VHML) *Virtual Human Markup Language* [70] is an XML-based markup language that is intended for use in computer animation of human bodies and facial expressions (e.g., the control of interactive “Talking Heads”). It is created to adapt the various issues of Human–Computer Interaction (HCI) such as Facial or Body Animation, Text-to-Speech production, Dialogue Manager interaction, Emotional Representation plus Hyper and Multi-Media information.[6] For example, a “virtual human” who introduces some bad news to the user *(“I’m sorry, I can’t find that file”)* may speak with a sad voice, a sorry face and with a bowed body gesture. To produce the required vocal, facial and emotional response, tags such as<smile>,<anger>,<surprised> have been defined to make the Virtual Human believable and more realistic.

(SSML) Speech Synthesis Markup Language [71] has been produced by Voice Browser Working Group to provide a rich, XML-based markup language for assisting the generation of synthetic speech in Web and other applications.[7] The essential role of the markup language is to provide authors of synthesizable content a standard way to control aspects of speech such as pronunciation, volume, pitch, rate, etc. across different synthesis-capable platforms. Popular voice assistants (Google assistant, Alexa, and Cortana) are known to use SSML. SSML adds markup elements (or tags) on input text to construct speech waveforms to improve the quality of synthesized content and to sound more natural. For instance,the tag <*amazon:emotion name="excited" intensity="medium"*> tells Alexa to speak a string (e.g., *“Three seconds till lift off”*) in an *“excited”* voice. Speech synthesis can be beneficial in many text-to-speech applications, e.g., reading for the blind, supporting the handicapped, access to the email remotely, and proofreading, etc.

```
<emma:emma version=”1.0” xmlns:emma=”http://www.w3.org/2003/04/emma”
xmlns=”http://www.w3.org/2009/10/emotionml”>
    <emma:interpretation emma:start=”1245790094000”

    emma:end=”1245790095000”
    emma:mode=”voice” emma:verbal=”false”>
        <emotion category-set=”http://www.w3.org/TR/emotion-voc/xml
             #everyday-categories”>
                <category name=”bored”value=”0.1”confidence=”0.1”/>
        </emotion>
    </emma:interpretation>
</emma:emma>
```

[5] https://www.w3.org/TR/emotionml/#s5.2.1.

[6] https://www.vhml.org.

[7] https://www.w3.org/TR/speech-synthesis11.

Emotion Ontologies

In the context of information systems, an ontology defines a set of conceptualizations and representational primitives with which to model a domain of knowledge or discourse [72]. In general, ontologies use primitives like categories, properties, and relationships between categories for organizing information into knowledge using machine readable statements to be computerized, shareable, and extensible. No universal ontology has been proposed for the domain of human emotions so far, but several approaches exist for conceptualizing particular important aspects of human emotions. In this section, we will classify these ontologies in terms of: name, goal, key concept, and underlying emotion theory (according to "Emotion Theories"). As there are many approaches in each class, we present these in table form.

Text Ontologies

Text is a powerful means to interact and transfer information as well to express emotion. When dealing with text ontologies, it has to be taken in mind that in modern social software systems (like Facebook, Twitter etc.) abbreviations of all kinds as well as so-called emoticons, emojies, etc. are used besides of classical language-based text. Thus, establishing such an ontology is challenging. The real meaning of a text, possibly including such abbreviations, depends on the combination of all text components and the particular situation. A given text, therefore, may be ambiguous and lead to different ontological interpretations, if the related situation (the context) is not considered. Table 4 summarizes ontologies that may be used for emotion inference from texts.

Ontologies Conceptualizing Facial Cues, Speech, and Body Expressions

Facial expression synthesis has acquired much interest with the MPEG-4 standard [85]. The MPEG-4 framework is used for modeling facial animation based on facial expressions that implicitly reflect emotion. This, can be beneficial in several domains such as psychology, animation control, or healthcare particularly in the patient–computer interaction field. An appropriate ontological conceptualization may also support the recognition of emotions from speech (e.g., study [86]). The same is true for body movements that, however, are more context dependent and thus harder to synthesize comparing. [87] employed a Virtual Human (VH) to increase realism and reliability of body expressions. Table 5 summarizes the existing emotion ontologies conceptualizing facial cues, speech, and body expressions.

Context-Awareness Ontologies

In our social environment, the context has a significant impact on human's emotions. The relationship between emotion and its cause can be further understood when investigating the context. Based on the observed contextual information, the situation that trigger an emotional state can be better understood. Therefore, appropriate contextual information is required such as: place, time, things in the environment, etc. Today, context-awareness has been introduced as a key feature in emotion recognition projects that demonstrate relevance and effectiveness. Table 6 summarizes emotion ontologies based on contextual information.

High-Level Ontologies

High-level or General-/Upper-/Top-ontologies are very generic and, in general, are defined for providing the ontological foundations about the kind of things. They come with concepts like *object* or *process*, etc. The existing upper emotion ontologies supply the most significant shared concepts to represent human emotions. The developer can extend an upper level ontology by defining lower level concepts (as specializations of higher level concepts) according to the development's purpose. Table 7 summarizes the known high-level emotion ontologies.

Datasets

Datasets may help to accelerate work progress by providing a benchmark resource for analyzing and comparing system performance before tackling a system in real-life settings. This section lists and describes available datasets that are widely used for evaluating emotion recognition systems. The descriptions are summarize in Table 8.

Textual Datasets (Corpora)

Textual datasets, i.e., *corpora*, are widely used in computer linguistics for evaluating natural language processing systems like taggers, parsers, translators, etc. With the increasing number of social-media participants, emotion recognition from unstructured written text is of growing interest. There are several ways to represent emotions in texts depending on the particular emotion model used.

Stack Overflow (Q&A) dataset has been created by annotating manually 4800 posts (questions, answers, and comments) from the platform Stack Overflow (Q&A) [104]. Each post in the dataset was analyzed regarding the presence or absence of emotion. From total 4800 posts, 1959 were labeled with basic emotion.

Table 4 Ontologies used for emotion inference from texts

Ontology	Purpose	Conceptualization	UT
Emotions Ontology [73]	Analyze consumers' emotion regarding electronic products on social websites	Emotion keywords are mined from unstructured text within online consumers' posts	B
Emotive Ontology [74]	Monitor emotional responses in informal postings of social media	Capturing eight types of emotions from sparse, text-based social media streams, such as Twitter	B
Emotion in E-learning [75]	Provide students with a feedback during e-learning	The ontology represents different types of emotions, moods and behaviors	B
Emotions and Feelings [76]	Annotate emotion and navigate automatically through text written in French	Words are bundled together and marked as positive, negative or neutral with regard to their emotional content	B
EmotiNet [77]	Capturing and storing the structure and the semantics of real events and predicting the emotional responses triggered by chains of actions	Store action chains and their corresponding emotional labels from several situations based on real-life self-assessment	B
Visual Sentiment Ontology (VSO) [78]	Detect sentiment from associated text tags of visual content (images and videos in Flickr and YouTube)	The ontology is based on Plutchik's Wheel [79] and consists of 3244 adjective noun pairs (ANP); it comes with a visual concept detector library called SentiBank	PW
Emotion in Chinese [80]	Analyze and classify actors' emotions in Chinese sentences based on HowNet [81]	The ontology contains about 5,500 verb predicates covering 113 emotion categories	B + D
Ontology of SentiMiner [82]	To conduct sentiment analysis for Chinese online reviews from a semantic perspective	Sentiment words are classified into two types: emotional (from eight classes) and evaluation words. The ontology is built based on HowNet to express emotion in a fuzzy way	B + D
Onyx [83]	To annotate and describe the emotions expressed by user-generated content	Onyx consists of sets of classes and properties that, used in schemata, allow to compare emotions coming from different systems (polarity, topics, features)	B + D
Smiley Ontology (SO) [84]	Introduce a better communication between users of social media by representing the structure and semantic of "EmotIcons". Adapt "EmotIcons" visually to the users' moods	SO has a class "Emoticon System" that contains all emoticons' pictures, and "Emotion" class that comprises a set of emotions, each interpreted with appropriate "EmotIcons"	B + D

UT Underlined Theory, *B* Basic emotion, *PW* Plutchik's Wheel, *D* Dimensional emotion

Table 5 Ontologies conceptualizing facial cues, speech, and body expressions

Ontology	Purpose	Conceptualization	UT
Virtual human ontology [88]	Storing, indexing and retrieving prerecorded synthetic facial cues	Facial animation concepts are represented according to the MPEG-4 standard	B + PW
Virtual character's ontology [89]	Create emotional virtual character regarding elements: personality, preferences, goals of the character and the environment events	The virtual character is a non-verbal communication interface between the system and user's emotions. The ontology based on MPEG-4 to obtain emotions with different intensities	B
E-learning ontology [90]	Predict emotions in E-learning environments	The emotion is predicted when students trying to solve Java quiz	B
Face characteristic detection [91]	Detect features of facial expression from images and transferred the detected part into text description	The major detected parts are: face shape, eyes, eyebrows, nose, and mouth	B
Non-verbal communication cues (NVC) [92]	An ontological mapping in virtual environment to support automatic integration of non-verbal communication cues	A complex emotion represented in terms of Ekman's theory by mixing more than one basic emotion with related intensity to gain more accurate level in the virtual environment	B
Visual Emotion Ontology (VEO) [93]	An ontology that links visualizations with specific emotions	Provide a semantic definition of 25 emotions based on established models, as well as visual representations of emotions utilizing shapes, lines, and colors	OCC [94]
Speech Ontology [86]	Recognize emotion from voice using semantic web	Three-dimensional concepts: "evaluation", "activation", and "power.Evaluation". This ontology provides an automatic classification using reasoner and provide automatic hierarchical structure	B
Body Expression Ontology [87]	Create emotional Body expression using virtual human	Standard MPEG-4 used to provide parameters to enhance virtual gestures. The ontology classified animations by associating them to emotions. Gesture classification derived from videos presented in [95]	D

UT Underlined Theory, *B* Basic emotion, *PW* Plutchik's Wheel, *D* Dimensional emotion

Table 6 Ontologies conceptualizing emotions and contextual information

Ontology	Purpose	Conceptualization	UT
EmOCA [96]	Represent emotion in relation to contextual information and body expressions	EmOCA is restricted to phobia and philia; the emotion processing chain is: from sensor data to dimensional model, then via an ontology to a basic model	B + D
Contextual Ontology [97]	Obtain the relationship between emotional states and contextual elements	Ekman's basic model is used to infer three states: positive, negative and neutral	B
Formal context model [98]	Address issues like semantic context, representation, reasoning, classification, dependency and quality of context	Modeling of users' contexts includes: user profiles, EEG data, situation and environment factors	B + D
Mining Minds Context (MMC) [99]	Combine low-level primitives of behavior, locations and emotions to derive meaningful high-level context information	The main concept is the class Context, which represents the context of a user in an interval of time	B

UT Underlined Theory, *B* Basic emotion, *D* Dimensional emotion

Table 7 High-level emotion ontologies

Ontology	Purpose	Conceptualization	UT
EmotionsOnto [100]	Generic ontology to describe emotions based on contextual and multimodal elements	It focused on Emergent Emotion instead of a global taxonomy of emotional states	PW
Human Emotions Ontology (HEO) [101]	Annotate multimedia data about dialogues with associated gesture and emotional state	The structure of ontology includes (face, text, voice and gesture)	B +D
Semantic Human Emotion Ontology (SHEO) [102]	To conceptualize simple emotions, combine them, and work with axioms-based rules to infer complex emotions. SHEO was based on HEO [101]	simple emotions in video and text are identified using computer algorithms in order to infer complex emotions	B + PW
Ontology [103]	Modeling of emotional cues. A simple cue expressed by: facial, gesture and speech; complex emotional cues mix two or more simple cues	Three global modules representing three layers of emotions' detection: the emotion module, the emotional cue module, and the media module	D

UT Underlined Theory, *B* Basic emotion, *PW* Plutchik's Wheel, *D* Dimensional emotion

EmoBank [105] is a large-scale corpus of 10 k English sentences. Data annotation was performed by applying the dimensional Valence-Arousal-Dominance (VAD) scheme [106]. A subset of the dataset has been annotated initially according to Ekmans six basic emotions, such that the mapping between both representation formats (dimensional and basic) becomes possible. Each sentence was rated regarding to both: The writer and the reader.

Emotion intensity dataset of tweets [107] was created to study the impact of word hashtags on emotion intensities in the text. The annotation is performed on 1030 tweets in the form of a hashtag with a query term (#<query term>). The dataset annotates intensities for emotion: anger, joy, sadness, and fear, respectively. The study proved that "emotion-word hash-tags" influence emotion intensity by transferring more emotion.

Social Media Posts-based dataset [108] has been developed for the training of prediction models for valence and arousal that achieve high predictive accuracy. The dataset consists of 2895 Facebook posts that have been rated by two psychologically trained experts on two separate ordinal nine-point scales regarding valence and arousal thus defining each post's position on the circumplex model of affect *CIRCUMPLEX MODEL* of emotion [43].

text_emotion is a CrowdFlower platform [109] dataset consisting of almost 40.000 annotated tweets like<*1966197101 ,"hate","rachaelk_x","why is this english homework so hard i seem to be getting nowhere"*> (csv-format). 13 labels are used for annotation, for example "hate", "sadness", "worry", "boredom", "happiness", "surprise", etc.

News Headlines dataset consists of 1.000 annotated news headlines extracted from news websites [110], such as: New York Times, CNN, and BBC News, as well as News search engines. The reason for focusing on headlines was that they usually have a "high load of emotional content". The annotation was performed by six independently working persons using a web-based interface that displayed one headline at a time, together with six slide bars for emotions

Table 8 Emotion datasets

Dataset	Type	Underlined theory	Accessibility
Stack Overflow (Q&A) [104]	Text	B	Pub
EMOBANK [105]	Text	B + D	Pub
Emotion intensity for tweets [107]	Text	Anger, joy, sadness, fear	Pub
Emotion in posts [108]	Text	D	Pub
Emotion in Text [109]	Text	Happiness, sadness, anger	Pub
Emotion within news headlines [110]	Text	Any lexical of emotion + Basic (optional)	Pub
Twitter, grounded Emotions [111]	Text	Happy, sad	Pub
SemEval [112]	Text	Anger, fear, joy, sadness	Pub
GoEmotions [113]	Text	27 emotions+ N	Pub
EmoEvent [114]	Text	B+ N	Pub
UIBVFED [126]	Image	B+ “None” emotion	Pub
FFHQ [127]	Image	B	OnReq
Google facial [128]	Image	B	OnReq
Yale Face Database [129]	Image	B	OnReq
SFEW [132]	Image	B	OnReq
SPOS [133]	Image	B	OnReq
CMU Multi-PIE [134]	Image	B	–
CAS-PEAL [135]	Image	B	OnReq
TsinghuaL [139]	Image	B	Pub
iCV-MEFED [144]	Image	B	OnReq
MUG [140]	Image	B	Pub
RAF-DB [141]	Image	D	Partially Pub
FERG-DB [142]	Image	B	OnReq
FACES [143]	Image	D	Pub
GFT [136]	Video	B	OnReq
ADFES [137]	Video	B	OnReq
Angled Posed Facial Expression [138]	Video	B	OnReq
CAS(ME)2 [130]	Video	Positive, negative, surprise, "other" emotion	OnReq
AFEW [131]	Video	B	OnReq
EMODB [115]	Speech	B + N	Pub
DES [116]	Speech	B + N	Pub
MASC [117]	Speech	B + N	OnPay
VERBO [118]	Speech	B + N	OnPay
MSP-Podcast corpus [119]	Speech	B+ D	OnPay
URDU-Dataset [120]	Speech	N + angry, happy, sad	Pub
DEMoS [121]	Speech	B	OnReg
AESDD [122]	Speech	B	Pub
Emov-DB [123]	Speech	N+ sleepiness, anger, disgust, amused	Pub
JL corpus [124]	Speech	5 primary and 5 sary	Pub
EmoFilm [125]	Speech	B	Pub
HUMAINE [145]	Speech-video	D	Pub
Belfast [146]	Speech-video	B+ D	OnReg
SEMAINE [147]	Speech-video	B + D	OnReg
IEMOCAP [148]	Speech-video	B+ D	OnReq
GEMEP-FERA [149]	Speech-video	B	OnReq
SAVEE [150]	Speech-video	B + N	OnReq
Biwi [151]	Speech-video	15 emotions	OnReq
OMG [152]	Speech-video	D	Pub
RAVDESS [153]	Speech-video	B + N	Pub
CEAR [154]	Speech-video	B + N	Pub

Table 8 (continued)

Dataset	Type	Underlined theory	Accessibility
SEWA [155]	Speech-video	D	OnReq
CMU-MOSEI [156]	Speech-video	B+ Likert scale	Pub
VAMGS [157]	Speech-video	B	OnReq
Cohn-Kanade [158]	Image–video	B	OnReq
JAFFE [159]	Image–video	B	OnReq
BU-4DFE [161]	Image–video	B	OnReq
MMI [162]	Image–video	B	OnReg
NVIE [163]	Image–video	B	OnReg
EMOTIC [164]	Image–video	D	Pub
Affective-MIT [165]	Image–video	B	OnReq
Affective [166]	Image–video	D	–
DISFA [167]	Image–video	D	OnReq
LIRIS ACCEDE [168]	Image–video	D	OnReq
FABO [169]	Image–video	10 facial and body gestures	OnReg
Kinect FaceDB [170]	Image–video	9 facial expressions	OnReq
YouTube [171]	Image–video	B	Pub
MAHNOB [172]	Audio, video, physiological signals	B	OnReg
DEAP [173]	Audio, video, physiological signals	B	OnReg
RECOLA [174]	Audio, video, physiological signals	D	OnReq
EMDB [175]	Video, physiological signals	D	OnReq
MMSE [176]	Video, physiological signals	D	OnReq
DECAF [177]	Audio, physiological signals	D	OnReq
emoFBVP [178]	Audio, video, body, physiological signals	B	OnReq
DREAMER [179]	Audio, video, physiological signals	D	OnReq
MEISD [180]	Text, audio, video	B + N+ ("Acceptance" from Plutchik wheel)	OnReq
ASCERTAIN [181]	Physiological signals	D	OnReq

Pub Public, *OnReg* On registration, *OnReq* On request, *OnPay* On payment, *N* Neutral, *B* Basic emotion, *D* Dimensional, *PW* Plutchik's Wheel

and one slide bar for valence. For rating the emotional load, a fine-grained scale has been used.

Twitter microblogs dataset [111] was established for exploring the impact and correlation of "external factors" like weather, news events, or time with a user's emotional state. The dataset consists of 2.557 tweets that have been collected early in 2017. The tweets are self-tagged by their respective author with a #happy or #sad hashtag and come with metadata such as author, time, and location. They originate from 20 large US metropolitan areas.

SemEval-2018 Task 1 [112] provides a collection of datasets for English, Arabic, and Spanish tweets. In particular, an *Affect in Tweets* dataset of more than 22.000 annotated tweets has been established. For each emotion dimension (anger, fear, joy, and sadness), the data are annotated for fine-grained real-valued scores indicating the intensity of emotion.

GoEmotions [113] is the currently largest available manually annotated dataset of 58 k English Reddit comments, labeled for 27 emotion categories or Neutral.

EmoEvent [114] is a multilingual dataset collected from Twitter and based on different emotional events. A total of 8409 tweets in Spanish and 7303 in English were labeled by the six Ekman's basic emotions plus the "neutral or other emotions".

Emotional Speech Datasets

Speech datasets are utilized as a foundation for different researches in the field of emotional speech analysis. An overview is given in Table 8.

EMODB [115] contains German sentences of emotional utterances. Ten German actors simulate emotions and produce 10 utterances used in daily communication. The data were labeled by 20 annotators with a total of six emotion types plus neutral.

DES [116] Danish emotional voice data: Four Danish actors (two male and two female) convey emotional language as they each consider it realistic. The recordings of each actor consist of two isolated words, nine sentences and two passages. The data set contains emotional speech expressions in five emotional states, neutrally annotated by 20 people.

MASC [117] Mandarin Affective Speech Corpus is a database of emotional speech consisting of 25,636 audio recordings of utterances and corresponding transcripts. It serves as a tool for linguistic and prosodic feature investigation of emotional expression in Mandarin Chinese, and for research in speaker recognition with affective speech. Five emotional states were recorded: Neutral, Anger, Elation, Panic, and Sadness. The recordings are from 68 speakers (23 females, 45 males); information about the speakers is available.

VERBO [118] is a database of speech in the Portuguese language of Brazil, called Voice Emotion Recognition dataBase. It consists of 14 sentences each spoken by 12 professionals (6 actors and 6 actresses) for each of the six basic emotions plus neutral. This meant that 1176 audio recordings were analyzed. This leads to a total number of 1176 audio recordings, which were annotated by three people.

MSP-Podcast corpus [119] contains speech data from online audio-sharing websites (100 h by over 100 speakers) annotated with sentence-level emotional score regarding four basic emotions (general, joy, anger, and sadness).

URDU-Dataset [120] is an Urdu-language speech emotion database that includes 400 utterances by 38 speakers (27 male and 11 female). Four basic emotions are annotated: anger, happiness, neutral, and sadness.

DEMoS [121] is an Italian emotional speech corpus. It contains 9365 emotional and 332 neutral samples produced by 68 native speakers (23 females, 45 males).

AESDD [122] The Acted Emotional Speech Dynamic Database contains around 500 Greek utterances by a diverse group of actors simulating five emotions.

Emov-DB [123] The Emotional Voices Database contains English recordings from male and female actors and French utterances form a male actor. The emotional states annotated are neutral, sleepiness, anger, disgust, and amused.

JL corpus [124] is a strictly guided simulated emotional speech corpus of four long vowels in New Zealand English. It contains 2400 recording of 240 sentences by 4 actors (2 males and 2 females). Five primary emotions (angry, sad, neutral, happy, and excited) and five secondary emotions (anxious, apologetic, pensive, worried, and enthusiastic) are annotated.

EmoFilm [125] is a multilingual emotional speech corpus that consists of 1115 English, Spanish, and Italian emotional utterances extracted from 43 films and 207 speakers. Five emotions are recognized: anger, contempt, happiness, fear, and sadness.

Facial Expression Datasets

Recognizing emotions from facial expressions is an important area of research. Classifying facial expressions requires large amounts of data to reflect the diversity of conditions in the real world. Table 8 provides an overview.

UIBVFED [126] provides sequenced semi-blurry facial images with different head poses, orientations, and movement. Over 3000 facial images were extracted from the daily news and weather forecast of the public tv-station PHOENIX. Seven basic emotions are categorized, namely: sad, surprise, fear, angry, neutral, disgust, and happy as well as “None” if the facial expression could not be recognized.

FFHQ [127]: Flickr faces HQ includes 70,000 high-quality. png images in high resolution (1024*1024) and contains considerable variation in age, ethnicity, and image background. It also has a good coverage of accessories like glasses, sunglasses, hats, etc. The images were crawled from Flickr.

Google Facial Expression Comparison Dataset [128] is a large-scale dataset consisting of facial image triplets along with human annotations. The latter indicate which two faces in each triplet form the most similar pair in terms of facial expression. The dataset was annotated by six or more human raters, which is quite different from existing expression datasets that focus mainly on basic emotion classification or action unit recognition.

Yale Face Database [129] is a dataset containing 165 GIF images of 15 different people in varying lighting conditions. The people in the images show distinct emotions and expressions (happy, normal, sad, sleepy, surprised, and winking).

CAS (ME) ^{2} Chinese Academy of Sciences Macro-Expressions and Micro-Expressions [130] contain both macro- and microexpressions in long videos, which the authors say facilitates the development of algorithms to detect microexpressions in long video streams. The database consists of two parts, one of which contains 87 long videos containing both macroexpressions and microexpressions from a total of 22 subjects, all filmed in the same setting. The other part includes 357 cropped expression samples containing 300 macroexpressions and 57 microexpressions. The facial expression samples were coded with facial action units marked and emotions labeled. In addition, participants were asked to review each recorded facial movement and indicate their emotional experience of it. Emotion is labeled using four types (negative, positive, surprise, and others). Happiness and sadness are classified as positive and negative, respectively. Surprise refers to an emotion that can be positive or negative. The "others" category represents ambiguous emotions that cannot be categorized to the mentioned categories.

AFEW Acted Facial Expression in Wild [131] AFEW was created in a semi-automatic process from 37 DVD movies: First, the subtitles were parsed and searched for keywords, and then, the relevant clips found were assessed and annotated by a human observer. In total, AFEW version 4.0 includes 1268 clips annotated with the basic emotions (anger, disgust, fear, happiness, neutral, sadness, and surprise). The AFEW dataset was used several times as the basis for the "Emotion Recognition In The Wild Challenge".

SFEW Static Facial Expressions in the Wild [132] have been created by selecting frames from the AFEW [131] dataset. It comprises 700 images labeled by basic six emotions. It presents real-world images with variety of properties of facial cues (e.g., head poses, age range, and illumination variation).

SPOS Spontaneous vs. Posed dataset [133] contains spontaneous and posed facial expressions from seven subjects who had been shown emotional movie clips to produce spontaneous facial expressions. Six categories of basic emotions were considered (happy, sad, anger, surprise, fear, and disgust). Subjects were also asked to pose these six types of facial expressions after watching the movie clips. Data were recorded with both the visual and near-infrared cameras. A total of 84 posed and 147 spontaneous facial expression clips were labeled.

CMU Multi-PIE [134] X contains approximately 750 K images from 337 exposures taken in up to four sessions over a 5-month period under 19 different illumination conditions. The images were captured from 15 different viewpoints to ensure data diversity. Six basic emotion categories are used to label the result (neutral, smile, surprise, squint, disgust, and scream).

CAS-PEAL [135] includes about 99K images taken from 1040 persons (595 males and 445 females), some of them with glasses and/or hats. Five posed expressions were captured in different head poses and different lighting proprieties. Images' backgrounds are adapted with different colors to simulate the real world. The subjects were asked to represent (neutral, smile, fear, and surprise) emotional states.

GFT Facial Expression Database [136] comprises about 172K video frames taken from 96 subjects in 32 three-person groups. To analyze facial expression automatically, GFT includes expert annotations of Facial Action Coding System (FACS) occurrence and intensity, facial landmark tracking, and baseline results for linear Support Vector Machine (SVM), deep learning, active patch learning, and personalized classification.

ADFES Amsterdam Dynamic Facial Expression Set [137] is a rich stimulus set of 648 emotional expression movies taken from 22 persons (10 female, 12 male). It includes the six "basic emotions" as well as the emotion states of contempt, pride, and embarrassment. Active turning of the head is used to indicate the direction of the expressions.

Angled Posed Facial Expression Dataset [138] contains facial expressions videos shot from different angles and poses. The different observation angles are intended to facilitate emotion recognition.

Tsinghua facial expression dataset [139] contains 110 images of young and old Chinese showing eight facial expressions (Neutral, Happiness, Anger, Disgust, Surprise, Fear, Content, and Sadness). Each image in the dataset was labeled on the basis of perceived facial expressions, emotion intensity, and age by two different age groups.

MUG Facial Expression Database [140] contains image sequences taken from 86 people (35 women and 51 men, all of Caucasian origin) making facial expressions against a blue screen background. The database consists of two parts: The first contains images where subjects were asked to show the six basic expressions (anger, disgust, fear, joy, sadness, and surprise). The second part contains emotions generated in the laboratory: Subjects were recorded while watching a video created to generate emotions. The dataset was created to improve some issues such as high resolution, uniform lighting, and others.

RAF-DB Real-world Affective Faces Database [141] contains approximately 30 K different face images downloaded from the Internet, each labeled by approximately 40 people based on crowdsourcing annotation. The images vary in age, gender and ethnicity of the persons, head pose, lighting conditions, occlusions, post-processing procedures, etc. Each image is labeled with a seven-dimensional (basic) expression distribution vector, landmark locations, race, age and gender attributes, and other features.

FERG-DB [142] Facial Expression Research Group 2D Database. It contains 2D images of six stylized characters (3 males and 3 females) with annotated facial expressions. The database contains 55,767 annotated face images grouped in seven basic expressions—anger, disgust, fear, joy, neutral, sadness, and surprise.

FACES [143] is a set of images of naturalistic faces of 171 young ($n = 58$), middle-aged ($n = 56$), and older ($n = 57$) women and men, each showing one of six facial expressions: Neutrality, Sadness, Disgust, Fear, Anger, and Happiness. The database includes two sets of images per person and per facial expression, resulting in a total of 2052 images.

iCV-MEFED iCV-Multi-Emotion Facial Expression Dataset [144] was designed for multiemotion recognition and includes about 31 K facial expressions with different emotions from 125 people with almost uniform gender distribution. Each person shows 50 different emotions, and for each of these emotions, 5 samples were taken under uniform illumination conditions with relatively uniform backgrounds. All emotion expressions are labeled with seven basic emotional states (anger, contempt, disgust, fear, happiness, sadness, surprise, and neutral). The images were taken and labeled under the supervision of psychologists, and the subjects were trained on the emotions they posed.

Hybrid Emotion Datasets

Recognition results can be more accurate when data are collected from different modalities, such as: text, audio, video, body and physiological data, etc. Table 8 lists the available data sets in the area of emotion recognition based on hybrid data sources. The table has been extended by a column titled "Type" to illustrate the modularity used to extract emotions.

Speech-Video Datasets

The records listed below combine voice and video. Table 8 provides an overview.

HUMAINE [145] data set contains naturalistic clip samples of emotional behavior in relation to the context (static, dynamic, indoor, outdoor, monologue, and dialogue). The emotional state is commented on in each case by a series of annotations associated with the clips: these include core signs in speech and language as well as gestures and facial features related to different genders and cultures. Six annotators have been used for a wide range of emotions (intensity, activation/sounds, valence, and power).

The **Belfast** [146] dataset contains clips extracted from television programs (chat shows and religious programs). These are recordings of people discussing emotional issues. 100 clips were annotated.

SEMAINE [147] dataset was created as part of an iterative approach to the creation of automatic agents called Sensitive Artificial Listener (SAL). SAL involves a person in an emotional conversation. The participants comment on each clip in five emotional dimensions (valence, activation, power, expectation/expectation, and intensity).

IEMOCAP [148] (interactive emotional and dyadic motion capture) is a multimodal and multilingual dataset that includes video, speech, facial motion capture, and text transcriptions over a period of 12 h. Ten actors execute sketched scenarios specially selected to evoke emotional expression. IEMOCAP was annotated by six annotators both in basic terms and dimensions.

GEMEP-FERA [149] comprises ten recordings of actors displaying expressions with different intensities. Five independent discrete emotions are labeled per video.

SAVEE [150] (Surrey Audio-Visual Expressed Emotion) dataset comes with six basic emotions plus neutral. It has been created as a pre-requisite for the development of an automatic emotion recognition system. SAVEE consists of recordings from four English actors in seven different emotions with a total of 480 British English utterances.

Biwi 3D-Audiovisual Corpus [151] contains 1109 dynamic 3D face scans taken while uttering an English sentence. The information was extracted by tracking the frames using a simple face template, by splitting the speech signal into phonemes, and by evaluating the emotions using an online survey. The data set can be used in areas such as audio-visual emotion recognition, emotion-independent lip reading, or angle-independent facial expression recognition.

OMGEmotion (One-Minute-Gradual Emotion) [152] dataset is composed of 567 emotion videos with an average length of 1 min, collected from a variety of YouTube channels using the search term "monologue". The videos were separated into clips based on utterances, and each utterance was annotated by at least five independent subjects using an arousal/valence scale and a categorical emotion based on the universal emotions from Ekman.

RAVDESS [153] (Ryerson Audio-Visual Database of Emotional Speech and Song) is a set of multimodal, dynamic expressions of basic emotions. The data set includes 24 professional actors (12 female, 12 male), vocalizing two lexically matched statements by providing audio-visual recordings of vocal communication in North American English.

CEAR [154] (Context-Aware Emotion Recognition) dataset contains 13,201 video clips (with audio and visual tracks) and about 1.1 M frames that were extracted from 79 TV shows. Each clip is manually annotated with six emotion categories, including “anger”, “disgust”, “fear”, “happy”, “sad”, and “surprise“, as well as “neutral”. The clips range from short (around 30 frames) to longer ones (more than 120 frames) with an average length of 90 frames. A static image subset contains about 70,000 images. The dataset is randomly split into training, validation, and testing sets.

SEWA [155] is an audio-visual, multilingual dataset with recordings of facial, vocal, and speech behaviors made “in the wild”. It includes>2000 min of audio-visual data from 398 individuals (201 males and 197 females) and a total of six different languages. The recordings are annotated with face landmarks, facial action unit (FAU) intensities, different vocalizations, verbal cues, mirroring and rapport, continuous rated valence, arousal, liking, and prototypical examples (templates) of (dis)liking and mood.

CMU-MOSEI Multimodal Opinion Sentiment and Emotion Intensity dataset [156] contains>23 K sentence utterances in>3300 video clips from >1000 online YouTube speakers. The dataset is gender balanced. The sentences utterance is randomly chosen from various topics and monologue videos. The videos are annotated with the basic six emotion categories (happiness, sadness, anger, fear, disgust, and surprise).

VAMGS The Vera Am Mittag German Audio-Visual Emotional Speech Database [157] consists of 12 h of recordings of the German TV talk show “Vera am Mittag” (Vera at Noon). They are divided into broadcasts, dialogue acts, and utterances, and contain spontaneous and highly emotional speech recorded from unscripted, authentic discussions between talk show guests. The video clips were annotated by a large number of human raters on a continuous scale for three emotion primitives: Valence (negative vs. positive), Activation (calm vs. excited), and Dominance (weak vs. strong). The video section contains 1421 segmented utterances from 104 different speakers, the audio section contains 1018 utterances, and the facial image section contains 1872 facial images labeled with emotions.

Hybrid Facial Expression Datasets

Several image and video datasets have been introduced for supporting the analysis and prediction of human emotional reactions based on facial expressions. An overview is given in Table 8.

Cohn-Kanade [158] CK dataset was made available to the research community in 2000. The image data consisted of about 500 image sequences of 100 subjects and were FACS (Facial Action Coding System, see above) annotated. An extended data set called CK+ was published in 2010. In CK+, the number of sequences is increased by 22% and the number of subjects by 27%. The target expression for each sequence is fully FACS coded; the emotion labels have been revised. In addition, non-positive sequences for different types of smiles and the associated metadata have been added.

JAFFE [159] *Japanese Female Facial Expression* dataset includes 213 annotated images of 7 facial expressions (6 basic expressions + neutral) that have been posed by Japanese female models. Each image rated by 60 Japanese annotators on 6 emotion classes. The images are in.tiff format with no compression (see also [159]). Semantic ratings on emotion adjectives, averaged over 60 subjects, are provided in a text file. The JAFFE images may be used for non-commercial scientific research.

BU-4DFE [160] is a 3D facial expression data set comprising 606 3D facial expression sequences posed by 101 persons of different ethnic origin. For each person, there are six model sequences showing six prototypic facial expressions (anger, disgust, happiness, fear, sadness, and surprise), respectively. Each sequence consists of about 100 frames, the resolution is about 1040×1329 pixels per frame. BU-4DFE is an extension of BU-3DFE [161] dataset (3D + time).

MMI [162] has been created as a resource for building and evaluating recognition algorithms of facial expression. It comprises over 2900 videos and high-resolution images of 75 subjects. Action Units (AU) in videos were fully annotated and partially coded on frame level. **NVIE** [163] *Natural Visible and Infrared Facial Expressions* dataset contains both spontaneous and posed expressions of more than 100 subjects. The images were taken synchronously with a visible and an infrared thermal imaging camera, with illumination from three different angles. The data set also allows a statistical analysis of the relationship between face temperature and emotions.

EMOTIC [164] *EMOTions In Context* is a database of images with people in real environments, annotated with their apparent emotions. The images are annotated with an extended list of 26 emotion categories combined with the three common dimensions (valence, arousal, and dominance). The dataset contains 23,571 images and 34,320 annotated people.

Affective-MIT [165] is a labeled dataset of spontaneous facial responses recorded in natural settings over the

Internet: online viewers watched one of three intentionally amusing Super Bowl commercials and were simultaneously filmed using their webcam. They answered three self-report questions about their experience. The dataset consists of 242 facial videos (168,359 frames).

Affectiva [166] is described as the largest emotion dataset growing to nearly 6 million faces analyzed in 75 countries, representing about 2 billion face frames analyzed. Affectiva includes spontaneous emotional responses to consumers while doing a variety of activities. The data set consists of viewers watching media content (ads, movie clips, TV shows, and viral campaigns online). The dataset has been expanded to include other contexts such as videos of people driving cars, people in conversation interactions, and animated GIFs.

DISFA [167] *Denver Intensity of Spontaneous Facial Action* The dataset contains high-resolution stereo videos (1024× 768) of 27 people (12 women and 15 men) that capture the spontaneous (non-posed) emotions of the persons while watching video clips. Each record frame was manually coded for presence, absence, and intensity of facial action units according to the facial action unit coding system (FACS). An extension, **DISFA+**, comprises also posed facial expressions data, more detailed annotations, and meta data in the form of facial landmark points (in addition to the self-report of each individual regarding every posed facial expression).

LIRIS-ACCEDE [168] comprises 9800 video segments with a large content diversity. Affective annotations along the valence and arousal axes were achieved using crowdsourcing through a pair-wise video comparison protocol. The videos were selected from 160 diversified movies.

FABO [169] *Bimodal Face and Body Gesture* contains 1900 videos of face and body expressions recorded simultaneously by two cameras. The dataset combines facial cues and body in an organized bimodal manner.

Kinect FaceDB [170] includes facial images of 52 persons acquired by Kinect sensors. The data were captured in different time periods involving 9 different facial expressions under different conditions: neutral, smile, open mouth, left profile, right profile, occlusion eyes, occlusion mouth, occlusion paper, and light on.

YouTube emotion datasets [171] contain 1101 videos annotated with 8 basic emotions using Plutchik's Wheel of Emotions [54]. The research efforts was focused on recognizing emotion-related semantics.

Multimodal Emotion Datasets

Multimodal datasets combine video data with synchronously recorded physiological signals. The following examples are also listed at the bottom of Table 8.

MAHNOB-HCI [172] consists of multimodal recordings of participants in their response to excerpts from movies, images, and videos. The modalities include multicamera video of face, head, speech, eye gaze, pupil size, ECG, GSR, respiration amplitude, and skin temperature. The recordings for all excerpts were annotated by the 27 participants immediately after each excerpt using a form asking five questions about their own emotive state through self-assessment manikins (SAMs) [32]. A precise synchronization permits researchers to study the simultaneous emotional responses using different channels.

DEAP [173] *Database for Emotion Analysis using Physiological Signals* is a multimodal dataset combining face videos with electroencephalogram (EEG) and peripheral physiological signals. 32 participants were recorded while watching 40 1-min-long excerpts of music videos. The participants rated each video in terms of the levels of arousal, valence, like/dislike, dominance, and familiarity.

RECOLA [174] *Remote Collaborative and Affective Interaction* dataset consists of audio, visual, and physiological signal (ECG and EDA) recordings. Video conference interactions between 46 French-speaking participants while solving a cooperation task were recorded synchronously. The Emotions expressed by the participants were reported by themselves using the Self-Assessment Manikin (SAM) [32].

EMDB [175], the *Emotional Movie Database*, consists of 52 affective movie clips from different emotional categories without auditory content. Recorded signals are kin conductance level (SCL) and heart rate (HR). Subjective scores for annotation by the participants were arousal, valence, and dominance (all on a scale from 1 to 9).

MMSE [176] Multimodal Spontaneous Emotion Database. The data captured from different sensors, such as 3D models, 2D videos, thermal, facial expressions, and FACS codes. The physiological signals are also recorded such as heart rate, blood pressure, electrical conductivity of the skin, and respiration rate. The datast was captured from 140 individuals from various nationalities. Ten emotions are recorded per person including surprise, sadness, fear, anger, disgust, happiness, embarrassment, startle, sceptical, and pain.

DECAF [177] *Multimodal Dataset for Decoding Affective Physiological Responses* combines brain signals collected

Table 9 Current emotion interfaces

Application	Input	Underlined theory	Accessibility
IBM Watson	Text	Anger, fear, joy, sadness, analytical, confident, tentative	OnReg
ToneAPI	Text	Joy, Trust, Interest, Surprise, Sadness, Disgust, Anger, Fear	OnPay
Receptiviti	Text	B	OnPay
Synesketch	Text	B	Pub
EmoTxt	Text	B	Pub
EMOSpeech	Audio	Wide spectrum of acoustic emotions	OnReq
Vokaturi	Audio	Happy, sad, afraid, angry, or neutral	OnReq
FaceReader	Video	B + neutral and contempt	OnReq
Realayes	Video	Happy, confusion, Disgust, sad, scared, surprise, engagement	OnReq
CrowdEmotion	Video	B	OnReg
Face++	Image	B + neutral	OnPay
SkyBiometry	Image	B + neutral and contempt	OnPay
Google-Cloud Vision	Video, image	Joy, Sorrow, Anger, Surprise, Exposed, Blurred	OnPay
Microsoft Cognitive Services	Video, image	Anger, contempt, disgust, fear, happiness, sadness, surprise + neutral	OnPay
CLMtrackr	Video, image	Angry, sad, surprised, happy	Pub
Kairos	Video, image	B	OnPay
Amazon	Video, image	Happy, sad, angry, confused, disgusted, surprised, calm, unknown	OnPay
Sightcorp	Video, image	B	OnPay
SHORE	Video, image	Anger, happiness, sadness, surprise	OnPay
nViso	Video, image, text	B	OnPay
iMotions	Video, body, physiological signals	Joy, Anger, Surprise, Fear, Contempt, Sadness, Disgust	OnReq
Affectiva	Image, audio, video, physiological signals	Anger, Contempt, Disgust, Fear, Joy, Sadness, Surprise	OnPay

Pub Public, *OnReg* On registration, *OnReq* On request, *OnPay* On payment, *B* Basic emotion

using the Magnetoencephalogram (MEG) sensor with explicit and implicit emotional responses of 30 participants to 40 1-min music video segments (used in DEAP) and 36 movie clips. This allows for comparisons between the EEG vs. MEG modalities as well as movie vs. music stimuli for affect recognition. The Recorded Signals are: MEG data, horizontal electrooculogram (hEOG), electrocardiogram (ECG), electromyogram of the Trapezius muscle (tEMG), and near-infrared face video.

emoF-BVP [178] is a multimodal dataset of face, body gesture, voice, and physiological signals recordings. It consists of audio and video sequences of actors displaying three different intensities of expressions of 23 different emotions and the corresponding physiological data.

DREAMER [179] is a multimodal database consisting of electroencephalogram (EEG) and electrocardiogram (ECG) signals together with 23 participants' self-assessments of their emotion in terms of valence, arousal, and dominance. The signals were captured using portable, wearable, and wireless equipment during affect elicitation by means of audio-visual stimuli.

MEISD [180] a large-scale balanced Multimodal Multi-label Emotion, Intensity, and Sentiment Dialogue dataset (MEISD) collected from different TV series that has textual, audio, and visual features. For annotating the dataset, six basic emotions are used. Emotion annotation list is extended to incorporate two more labels, namely, Acceptance and neutral. The "acceptance" emotion has been taken from the Plutchik's [54] wheel of emotions.

ASCERTAIN [181] contains emotional self-assessments (Arousal, Valence, Engagement, Liking, and Familiarity) from 58 users along with synchronously recorded electroencephalogram (EEG), electrocardiogram (ECG), galvanic skin response (GSR), and facial activity data recorded with commercially available sensors while watching affective movie clips. This multimodal database can be used to detect personality traits and emotional states via physiological responses.

Systems for Emotion Recognition

This chapter sketches a number of systems or Application Programmable Interfaces (APIs) for emotion recognition, which are used in numerous areas such as health care, education, and entertainment. These systems are based on the various methods discussed above, namely face analysis, speech processing, physiological signs, recognition, and analysis of emotional phrases in social media, body language, and gesture expressions. An overview of these systems is given in Table 9.

Text-Based Interfaces for Emotion Detection

IBM Watson[8] is an analyzer of emotions in written text. It detects emotional tones, social tendencies and writing styles from simple texts of any length. Currently, the tool can analyze online texts such as tweets, online reviews, email messages, product reviews, or user texts for emotional content.

ToneAPI[9] was created for marketing people to evaluate (and potentially improve) the emotional impact of their advertising texts quantitatively and qualitatively. For this purpose, input texts are analyzed, compared with other texts from a corpus, and emotions and their intensity are derived. A total of 8 emotions are identified and their intensity is evaluated with a value between 1 and 100.

Receptiviti[10] is a computational language psychology platform that aims at helping to understand the emotions, drives, and traits that affect human behavior. A set of algorithms uncovers signals from everyday human language, e.g., stress, depression, etc. The analysis is performed in real time without needing self-reports or surveys.

Synesketch [182] is an open source tool for analyzing the emotional content of text sentences and transforming the emotional tone into some visualizations. It is a dynamic text representation in animated visual patterns to reveal the underlying emotion.

EmoTxt [183] is an open-source toolkit for emotion detection from text. It was trained and tested on two large gold standard datasets mined from Stack Overflow and Jira. It provides supporting both emotion recognition from text and training of custom emotion classification models.

Audio-Based Interfaces for Emotion Detection

EMOSpeech[11] is an interface based on an end-to-end psychological model. It is designed to help automated call agents analyze recorded customer calls and then send real-time feedback to supervisors. It uses a three-dimensional emotion representation model and recognizes ten emotions from acoustic features in the voice.

Vokaturi[12] was developed to detect the emotions “happy,” “sad,” “scared,” “angry,” or “neutral” from a speaker’s voice. The open-source version of the software selects between these five emotions with high accuracy, even when hearing the speaker for the first time, according to the manufacturer. The “plus” version is said to reach the performance level of a dedicated human listener.

Video-Based Interfaces for Emotion Detection

FaceReader[13] created by Noldus [184] is a professional tool for automatic analysis of facial expressions. More than 10,000 manually annotated images were used to train the recognition component: emotion, gaze direction, head orientation, and personal characteristics, such as gender and age.

RealEyes[14] is a platform for emotion recognition using Webcams [185], used to measure people’s feeling when they watch video content online. Computer vision and machine learning techniques were used to analyze signals from physiological sensors, voice, and posture.

CrowdEmotion[15] used to explore facial points in real-time video, to detect the time series of Ekman six basic emotions. The Webcam tracks eye and what they are paying attention, as well as facial coding to understand emotion.

[8] https://www.ibm.com/watson/services/tone-analyzer.

[9] https://adoreboard.com.

[10] https://www.receptiviti.com/platform.

[11] https://emospeech.net.

[12] https://vokaturi.com.

[13] https://www.noldus.com/facereader.

[14] https://www.realeyesit.com.

[15] https://www.crowdemotion.co.uk.

Image-Based Interfaces for Emotion Detection

Face++[16] detects faces within images and gain high-precision face location rectangles. Each detected face can be stored for future usage and analysis. Detected face is compared with stored faces to return a confidence score and thresholds to evaluate the similarity. It also can determine, if a subject is smiling or not.

SkyBiometry[17] is created by biometric company to detect faces and emotion in photos with a percentage rate for each emotion. The application also determines gender, smile, eyeglasses and sunglasses presence, age, roll and yaw, eyes, nose and mouth position, checks if lips are parted or sealed, eyes open or closed.

Multimodal Interfaces for Emotion Detection

Cloud Vision[18] is a tool created by Google, which understands faces, signs, landmarks, objects, text, as well as emotions by detecting facial features within image or video. The Cloud platform takes an image as an input, and then returns the expected percentage of each emotion for each face in that image.

Microsoft cognitive services[19] or Microsoft Project Oxford is a set of tools that make it possible for a computer to identify emotions in photographs using facial recognition technology. It detects emotional depth for each face using the core seven emotional states as well as "neutral". Each scanned image is bounded with a box for the face, and then assigned a score between zero to one, where zero corresponds to a complete absence of the emotion in question and one is a strong emotional response.

CLMtrackr[20] created by MIT is a javascript library for fitting facial models to faces in videos or images [186]. It is an open and free to use JavaScript library for precise tracking of facial features. This library recognizes four emotional states: angry, sad, surprised, and happy.

Kairos[21] integrates both face detection and important demographics data (Age, Gender, and Ethnicity). It detects real-time emotion from face as well as ethnicity to understand the diversity of human face.

Amazon[22] recognition is used to identify the objects, people, text, scenes, and activities, as well as emotion. This tool accepts two sources of data input: image and video. The level of confidence in the determination is ranging: (zero) minimum value to (100) maximum value. This application based on the same learning technology is developed by Amazon's computer vision to analyze billions of images and videos daily.

Sightcorp[23] Platform provides face analysis and face recognition software using computer vision and deep learning techniques. It allows for emotion recognition, age detection, gender detection, attention time, and eye gaze tracking in images, videos, and real-life environments.

SHORE[24] is used to detect the emotion, age, and gender of a person from a standard webcam. The special feature of this tool is its ability to analyze and recognize the respective emotion from a video input with multiple faces simultaneously.

nViso[25] analyzes real-time emotions from facial expressions in video using 3D facial imaging technology. nViso can monitor many different facial data points to produce likelihoods for main emotion categories.

The **iMotions**[26] Facial Expression Analysis (FEA) module provides 20 facial expression measures, seven core emotions (joy, anger, fear, disgust, contempt, sadness, and surprise), facial landmarks, and behavioral indices such as head orientation and attention. These output measures are assigned probability values to represent the likelihood that the expected emotion will be expressed. Summary values for engagement and valence are also provided.

Affectiva[27] is designed to detect facial cues or physiological responses based on emotions. It tracks a person's heart rate from the human face using the webcam without any other sensors being worn, depending on the color change in the person's face, which pulses every time the heart beats.

[16] https://www.faceplusplus.com.

[17] https://skybiometry.com/tag/emotions.

[18] https://cloud.google.com/vision.

[19] https://azure.microsoft.com.

[20] https://github.com/auduno.

[21] https://www.kairos.com.

[22] https://aws.amazon.com.

[23] https://sightcorp.com.

[24] https://www.iis.fraunhofer.de/de/ff/sse/imaging-and-analysis/ils/tech/shore-facedetection.html.

[25] https://www.nviso.ai/en.

[26] https://imotions.com.

[27] https://www.affectiva.com.

Table 10 Exploiting emotion recognition systems in different domains

Domain	Sub-domain
Health/medical	Measure emotional reaction about a kind of treatment [189, 190]
	Help to decide when patients necessitate medicine [191]
	Help monitoring (rehabilitation/therapy) [192]
	Monitor the autism spectrum disorders (ASD) [193, 194]
E-learning	Emotion feedback of the students [195]
	Monitor the level of students concentration [196]
	Detect the emotion of the learner [197]
Hiring/interview	Track candidate's emotions during interview [198]
	Analyze the stress level of employees [199, 200]
Entertainment	Adapt games according to the player's mood [201, 202]
	Test game success according to user's experience [203, 204]
HCI/Robotics	Improve Human-Computer Interaction (HCI) [205, 206]
	Effective human-to-robot communication [207, 208]
Marketing	Monitor the impact of advertisements [209, 210]
	Monitor emotion in purchasing decisions to improve Sales [211, 212]
Automotive industry	Alert the driver when his/her looks sleepy or drowsy [213, 214]
	Improve driving experience [215, 216]

Information Systems (IS) Exploiting Emotion Data

Emotion recognition applications such as those mentioned in the previous chapter are used in various fields and embedded in domain-specific information systems. Examples of such domains include medicine, e-learning, human resources, marketing, entertainment, and automotive; these are briefly outlined below. Table 10 lists some such systems, organized by application domain.

Health/Medical: Stress and various psychological problems require proper psychometric analysis of the patient. A healthcare system that focuses on emotional aspects may improve the quality of life. Such systems automatically monitor both the environment and the person to provide help and services. Also, coupling emotion recognition systems with activity recognition systems [187, 188] can generate important insights for supporting older people in the AAL context.

E-learning: Student emotions play an important role in any learning environment, whether in a classroom or in e-learning. In face-to-face classes, it is easier to observe student behavior, because the instructor interacts with students face-to-face in the same environment. In the virtual classroom, this is more difficult, especially since course participants often turn off their cameras. This is where techniques based on multimodal recognition come in to assist the lecturer.

Hiring/interview: Companies have begun to integrate emotion recognition technology into hiring processes to capture employees' emotions. Employers argue that the technology can help eliminate hiring biases and reject candidates with undesirable traits. During interviews, employers can track candidates' emotional responses to each question in the interview to determine how honest the applicant is. In addition, analyzing employee stress levels can impact productivity and career success. The discussion of whether one can ethically justify such an approach has to take place elsewhere.

Entertainment: Modern computer games use information about the player's emotions and emotional reactions to dynamically adjust the game's difficulty level and audiovisual features. Monitoring is done with multimedia tools, for example in video games: this involves measuring player behavior and emotional states to improve player engagement, challenge, immersion, and excitement, in addition to adjusting game features. In addition, emotion recognition is also used to evaluate the success of a game according to the experience of the player interacting with the game.

HCI/Robotics: Human–computer interaction, or the interaction between information technologies and humans, has become similar to human–human interaction. Detecting the emotional state during human–computer interaction aims to make this interaction more comfortable, effective, and easy. Similarly, the integration of emotions in robots is being researched to make robots more "social" and "human-like".

Marketing: Marketing has moved to the study of consumer attitudes and measures the factors that influence consumer decisions. Emotions have become an important aspect of this. They help companies better understand the opinions expressed about a consultation or product.

Automotive industry: In the automotive industry, emotional information can be used to respond appropriately to recognized driver emotions to enhance safety and support the driving experience, e.g., by suggesting useful information or a conversation.

Summary

As shown in the previous sections, there are a variety of methods and tools that can be used to model, analyze, and predict human emotions. We have reviewed the relevant literature on this as far as it was available to us. The extensive bibliography substantiates this.

Overall, we can summarize the contribution of our study as follows:

- It combines emotion models, languages, ontologies, datasets, and systems.
- It provides a comprehensive and systematic overview of the field of human emotion recognition, organized into tables to give the reader an easy and coherent way to find the data they need.
- It provides a comprehensive benchmark for users to create human support in a specific environment by exploiting the available semantic and contextual information; there is no need to reinvent the wheels.
- It can serve as a starting point for anyone studying human emotions, especially budding researchers.
- It introduces a resource for many applications in various fields, for instance, computer science, psychologists, machine learning, human–computer interaction, e-learning, information systems and cognitive science.

It has been noted that there are numerous trends in emotion research with different goals and models. Facial expressions tend to be the most promising approaches to emotion measurement, but they are easier to fake in different situations compared to other recognition methods (e.g., voice or biosignals).

Emotion recognition accuracy can be increased by combining multiple modalities with information about the context, preferences, and situation of the observed person. Multimodal emotion recognition has been shown to be superior to unimodal, so there are several datasets that serve as benchmarks for emotion modalities that have recently been made available. The next generation of recognition tools will capture emotions from multiple input devices using both recent technological advances and integration methods.

The context in which emotions are experienced is another important topic currently being discussed by researchers. Some work has implemented ontology-based methods to better interpret contextual human emotion manifestation [96–99].

Other developments include the ability to recognize and interpret not only current physical activities and reactions, but also relevant mental states beyond basic emotions, such as shame and pride.

Open Research Questions

As an additional outcome of this literature review, we summarize a list of limitations and open research questions that we would like to share with the research community:

- Current emotion representation languages such as [65, 66, 68–71,] are XML extensions that do not provide more advanced knowledge representation or automated reasoning capabilities. None of these languages is general enough to cover all emotion vocabularies.
- Known recognition frameworks (listed in “Datasets”) have a limited concept space by ignoring some real-world details. This limits their ability to reflect the user’s intentions in unpredictable situations.
- So far, several emotion ontologies have been proposed that share many similarities in terms of their concepts, classes, and underlying psychological theory. For example, Ekman’s basic theory [39] is adapted in most ontologies. However, there does not yet seem to be a universally shared ontology.
- The majority of emotion recognition interfaces are based on recognizing and analyzing facial data (i.e., using image or video data) to measure emotions. However, studies [217–223] have shown that integrating different modalities provides better accuracy than a single model for emotion recognition. Therefore, it stands to reason that future research will focus on multimodal emotion recognition [172–175, 177–179].
- A dataset that integrates emotion recognition with context is lacking; such data could be useful to answer theoretical questions about the causes of emotional responses.
- We identify the need for a conceptual human emotion framework, in particular, a domain-specific modeling language [224] that can describe human emotions in a comprehensive, reliable, and flexible way. The main challenge in developing such a language is the diversity of theoretical models and the complexity of human emotions. In addition, there is the influence of temporal structure and human context in interpreting emotions. For example, a person may express the same emotion in different situations depending on the context, some-

times with different intensity. We have presented a first approach to such a language, our "Human Emotion Modeling Language", in [225].

Finally, we take the opportunity to thank the anonymous reviewers of this paper for their valuable comments.

Funding Open access funding provided by University of Klagenfurt. We did not apply for external funding for this research.

Declarations

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

References

1. Bechara A. Brain and cognition, the role of emotion in decision-making: evidence from neurological patients with orbitofrontal damage, 55, 1, 30–40, Elsevier, 2004
2. Lieskovská E, Jakubec M, Jarina R, Chmulík M. A review on speech emotion recognition using deep learning and attention mechanism. Electronics, 2010.
3. Krumhuber EG, Skora L, Küster D, Fou L. A review of dynamic datasets for facial expression research, 2017.
4. Haamer RE, Rusadze E, Lusi I, Ahmed T, Escalera S, Anbarjafari G. Review on emotion recognition databases, 2017.
5. Murthy AR and Anil Kumar KA. A review of different approaches for detecting emotion from text. In: IOP Conference Series: Materials Science and Engineering. 2021.
6. Francisca Adoma Acheampong and Chen Wenyu and Henry Nunoo-Mensah, Text–based emotion detection: Advances, challenges, and opportunities, 2020
7. Verma S, Prakashan OB. Personality development and soft skills, 2013.
8. Brader T. Campaigning for hearts and minds: How emotional appeals in political ads work. Chicago: University of Chicago Press; 2006.
9. Alan Baddeley MM, Gruneberg MPE, Sykes RN. But what the hell is it for? Practical Aspects of Memory, John Wiley, 1988.
10. Crocker LD, Heller W, Warren SL, O'Hare AJ, Infantolino ZP and Miller GA. Relationships among cognition, emotion, and motivation: implications for intervention and neuroplasticity in psychopathology. Front Hum Neurosci Front. 2013.
11. Lench HC, Darbor KE, Berg LA. Functional perspectives on emotion, behavior, and cognition. Multidisciplinary Digital Publishing Institute; 2013.
12. Stangor C. Principles of social psychology—1st International Edition. Psychology. 2014
13. Nummenmaa L, Glerean E, Hari R, Hietanen JK. Bodily maps of emotions. Proc Natl Acad Sci. 2014;111(2):646–51.
14. Sebe N, Cohen I, Gevers T, Huang TS. Multimodal approaches for emotion recognition: a survey, book Internet Imaging VI, 5670. Int Soc Opt Photon. 2005; pp. 56–68.
15. Ekman P, Friesen WV. Facial action coding system: investigator's guide. Consulting Psychologists Press; 1978.
16. Chang Y, Hu C, Feris R, Turk M. Manifold based analysis of facial expression. Image Vis Comput. 2006;24(6):605–14 (**Elsevier**)
17. Zhang Y, Ji Q: Active and dynamic information fusion for facial expression understanding from image sequences. In: IEEE Transactions on pattern analysis and machine intelligence, 27, 5, 699–714, IEEE, 2005
18. Bennett CC, Sabanovic S. Deriving minimal features for human-like facial expressions in robotic faces. Int J Soc Robot. 2014;6(3):367–81 (**Springer**)
19. Farnsworth B. Facial Action Coding System (FACS)—a visual guidebook, August 18th, 2019.
20. Ekman P and Friesen WV. Unmasking the face: a guide to recognizing emotionsfrom facial clues, Malor Books, 2003.
21. Friesen WV and Ekman P, Alto P. Facial action coding system: a technique for themeasurement of facial movement. 1978
22. Kaminska D, Sapnski T, Pelikant A. Recognition of emotional states in natural speech. In: Signal Processing Symposium (SPS), 2013, IEEE, 2013; pp. 1–4.
23. Valdma J. Art installation from brain waves for tedxtartu 2012 report, 2012.
24. Ragot M, Martin N, Em S, Pallamin N, Diverrez J-M. Emotion recognition using physiological signals: laboratory vs. wearable sensors. In: International Conference on Applied Human Factors and Ergonomics. Springer, 2017; pp. 15–22.
25. Naji M, Firoozabadi M and Azadfallah P. Classification of music-induced emotions based on information fusion of forehead biosignals and electrocardiogram. Cogn Comput. 2014;6:241–52.
26. Healey J and Picard R. Digital processing of affective signals, 1998; pp. 3749–52.
27. Li L and Chen J. Emotion recognition using physiological signals from multiple subjects. In: International Conference on Intelligent Information Hiding and Multimedia, 2006; pp. 355–8.
28. Uyl MJD, Kuilenburg HV. The FaceReader: online facial expression recognition, book Psychology. 2005; pp. 589–90.
29. Rani P, Liu C, Sarkar N and Vanman E. An empirical study of machine learning techniques for affect recognition in human-robot interaction. Pattern Anal Appl. 2006
30. Jang JR. ANFIS, adaptive-network-based fuzzy inference system. In: IEEE Transactions on Systems, 1993; pp. 665–85.
31. Dai K, Fell HJ and MacAuslan J. Recognizing emotion in speech using neural networks. In: Proceedings of the 4th IASTED International Conference on Telehealth and Assistive Technologies, Telehealth AT 2008, 2008; pp. 31–6.
32. Bradley MM, Lang PJ. Measuring emotion: the self-assessment manikin and the semantic differential. J Behav Therapy Exp Psychiatry. 1994 (**Elsevier**)
33. Lang PJ, Bradley MM and Cuthbert BN. International affective picture system (IAPS): affective ratings of pictures and instruction manual. In: Technical Report A-8, 2008
34. Maria E, Matthias L, Sten H. Emotion recognition from physiological signal analysis: a review. Electron Notes Theor Comput Sci. 2019;343:35–55.
35. Schindler K, Van Gool L, de Gelder B. Recognizing emotions expressed by body pose: a biologically inspired neural model. Neural Netw. 2008;21(9):1238–46 (**Elsevier**)

36. Lhommet M and Marsella S. Expressing emotion through posture and gesture, 2015
37. Avots E, Sapinski T, Bachmann M, Kaminska D. Audiovisual emotion recognition in wild. Mach Vis Appl. 2018;1–11 (**Springer**)
38. Brosch T, Pourtois G, Sander D. The perception and categorisation of emotional stimuli: a review. Cogn Emotion. 2010;24(3):377–400 (**Taylor Francis**)
39. Ekman P. An argument for basic emotions. Cogn Emotion 1992;6(3–4):169–200 (**Taylor Francis**)
40. Scherer KR, Schorr A, Johnstone T. Appraisal processes in emotion: theory, methods, research. Oxford University Press, 2001.
41. Mehrabian A, Russell JA. An approach to environmental psychology. The MIT Press, 1974
42. Russell JA, Barrett LF. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. J Personal Soc Psychol. 1999;76(5):805 (**American Psychological Association**)
43. Russell JA. A circumplex model of affect. J Personal Soc Psychol. 1980;39(6):1161 (**American Psychological Association**)
44. Gendron M, Lisa BF. Reconstructing the past: a century of ideas about emotion in psychology. Emotion Rev. 2009;1(4):316–39. (**London, England: Sage Publications Sage UK**)
45. Yin D, Bond S. Zhang H. Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews, 2013.
46. Gregor S, Lin ACH, Gedeon T, Riaz A, Zhu D. Neuroscience and a nomological network for the understanding and assessment of emotions in information systems research. J Manag Inf Syst. 2014;30(4):13–48 (**Taylor Francis**)
47. Barrett LF. Are emotions natural kinds?, Perspectives on psychological science, 1, 28–58. Los Angeles: SAGE Publications Sage CA; 2006.
48. Barrett LF. Are emotions natural kinds? 2006
49. Moors A, Ellsworth PC, Scherer KR, Frijda NH. Appraisal theories of emotion: state of the art and future development. Emotion Rev. 2013;5(2):119–24. (**London, England: Sage Publications Sage UK**).
50. Kim J, Andre E. Emotion recognition based on physiological changes in music listening. IEEE Trans Pattern Anal Mach Intell. 2008;30(12):2067–83.
51. Lance B, Marsella S. Glances, glares, and glowering: how should a virtual human express emotion through gaze?, Auton Agents Multi-Agent Syst. 2010;20(1):50 (**Springer**)
52. Albert M. Nonverbal ccation. Taylor and Francis; 2007.
53. Becker C, Kopp S, Wachsmuth I. Why emotions should be integrated into conversational agents, conversational informatics: an engineering approach, 49–68. Chichester: John Wiley & Sons, Ltd; 2007.
54. Plutchik R. Emotion: a psychoevolutionary synthesis, stress and emotion recognition: an internet experiment using stress induction, January 1, 1980
55. Ortony A, Clore GL, Collins A. The cognitive structure of emotions, New York: Cambridge University Press; 1988.
56. Gross JJ. The emerging field of emotion regulation: an integrative review. Review of general psychology, 2, 3, 271, Educational Publishing Foundation, 1998
57. Blackburn MR, Denno PO. Using semantic web technologies for integrating domain specific modeling and analytical tools. Procedia Comput Sci. 2015;61:141–6 (**Elsevier**)
58. Terkaj W, Pedrielli G, Sacco M. Virtual factory data model. In: Proceedings of the Workshop on Ontology and Semantic Web for Manufacturing, Graz, Austria, 2012; pp. 29–43.
59. Liao C, Lin P-H, Quinlan DJ, Zhao Y, Shen X. Enhancing domain specific language implementations through ontology. In: Proceedings of the 5th International Workshop on Domain-Specific Languages and High-Level Frameworks for High Performance Computing, 3, ACM, 2015
60. Walter T, Parreiras FS, Staab S. Ontodsl: an ontology-based framework for domain-specific languages, book International Conference on Model Driven Engineering Languages and Systems. Springer, 2009; pp. 408–22.
61. Antunes G, Bakhshandeh M, Mayer R, Borbinha JL, Caetano A. Using ontologies for enterprise architecture integration and analysis. CSIMQ. 2014;1:1–23.
62. Ekman P. Body position, facial expression, and verbal behavior during interviews. Psychol Sci Public Interest. 1964;68(3):295–301.
63. Kiritsis D, Milicic A, Perdikakis A. User story mapping-based method for domain semantic modeling. In: Domain-Specific conceptual modeling. Springer, 2016; pp. 439–54.
64. Elkobaisi MR, Maatuk AM, Aljawarneh SA. A proposed method to recognize the research trends using web-based search engines. In: ICEMIS ’15, September 24–26, 2015, Istanbul, Turkey, ACM; 2015.
65. Burkhardt F, Schroder M, Baggia P, Pelachaud C, Peter C, Zovato E. W3C Emotion Markup Language (EmotionML), W3C Recommendation 22 May 2014.
66. Schroder M, Pirker H, Lamolle M. First suggestions for an emotion annotation and representation language. In: Proceedings of LREC, vol. 6. 2006; pp. 88–92.
67. Schuller B, Karpouzis K, Pelachaud C. What should a generic emotion markup language be able to represent?, 2007
68. Prendinger H, Ishizuka M. Life-like characters: tools, affective functions, and applications. Springer Science Business Media, 2013
69. Froumentin, Max, Extensible multimodal annotation markup language (EMMA): invited talk, book Proceeedings of the Workshop on NLP and XML (NLPXML-2004): RDF/RDFS and OWL in Language Technology, 33–33, Association for Computational Linguistics, 2004
70. Marriott A. VHML–virtual human markup language. In: Talking head technology workshop, at OzCHI Conference, 2001; pp. 252–64.
71. Bagshaw P, et al. Speech synthesis markup language (SSML) version 1.1, 2007
72. Gruber T. Ontology. In: Liu L and Tamer Özsu M (eds). Springer-Verlag, 2009.
73. Sam KM and Chatwin CR. Ontology-based text-mining model for social network analysis. In: Management of innovation and technology (ICMIT), 2012 IEEE International Conference on, IEEE, 2012; pp. 226–31.
74. Sykora MD, Jackson T, O’Brien A, Elayan S. Emotive ontology: extracting fine-grained emotions from terse, informal messages, IADIS-International Association for Development of the Information Society, 2013
75. Arguedas M, Xhafa F, Daradoumis T, Caballe S. An ontology about emotion awareness and affective feedback in E-learning. In: Intelligent networking and collaborative systems (INCOS). 2015 International Conference on, IEEE, 2015; pp. 156–63.
76. Mathieu YY. Annotation of emotions and feelings in texts. In: International conference on affective computing and intelligent interaction, Springer, 2005; pp. 350–7.
77. Balahur A, et al. Emotinet: a knowledge base for emotion detection in text built on the appraisal theories. In: International Conference on Application of Natural Language to Information Systems. Springer, 2011; pp. 27–39.
78. Borth D, Ji R, Chen T, Breuel T, Chang S-F. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In: Proceedings of the 21st ACM international conference on Multimedia, 223–232, ACM, 2013

79. Williams Y. Robert Plutchik's Wheel of Emotions, Education Portal, Retrieved from http://education-portal.com/academy/lesson/robert-plutchiks-wheel-of-emotionslesson-quiz.html, 2013
80. Yan J, Bracewell DB Ren F, Kuroiwa S. The creation of a Chinese emotion ontology based on HowNet. Eng. Lett. 2008;16:1.
81. Dong Z, Dong Q, Hao, Hownet and its computation of meaning, changling. In: Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations, 53–56, Association for Computational Linguistics, 2010.
82. Shi W, Wang H, He S. EOSentiMiner: an opinion-aware system based on emotion ontology for sentiment analysis of Chinese online reviews. J Exp Theor Artif Intell. 2015;27(4):423–8 (**Taylor Francis**)
83. Rada S, Fernando J, Fernandez I. Carlos Angel. Onyx: Describing emotions on the web of data, Telecomunicacion; 2013.
84. Radulovic F, Milikic N. Smiley ontology. In: Proceedings of The 1st International Workshop On Social Networks Interoperability, 2009
85. Raouzaiou A, Tsapatsoulis N, Karpouzis K, Kollias S. Parameterized facial expression synthesis based on MPEG-4. EURASIP J Adv Signal Process. 2002;10:521048 (**Springer**)
86. Francisco V, Gervas P, Peinado F. Ontological reasoning to configure emotional voice synthesis. In: International conference on web reasoning and rule systems. Springer, 2007; pp. 88–102.
87. Rojas A, et al. Emotional body expression parameters in virtual human ontology. In: 1st International Workshop on Shapes and Semantics, 2006; pp. 63–70.
88. Garc Rojas A et al. Emotional face expression profiles supported by virtual human ontology. Comput Anim Virtual Worlds. 2006;17(3–4):259–69 (**Wiley Online Library**)
89. Lera I, Arellano D, Varona J, Juiz C, Puigjaner R. Semantic model for facial emotion to improve the human computer interaction in ami. In: 3rd Symposium of Ubiquitous Computing and Ambient Intelligence 2008. Springer, 2009; pp. 139–48.
90. Eyharabide V, Amandi A, Courgeon M, Clavel C, Zakaria C, Martin J-C. An ontology for predicting students' emotions during a quiz. Comparison with self-reported emotions, book Affective Computational Intelligence (WACI), 2011 IEEE Workshop on, 1–8, IEEE, 2011
91. Khoonnaret C, Nitsuwat S. A face characteristic detection system using ontology and supervised learning. Int J Comput Internet Manag 2017;25(1):62–9.
92. Honold F, Schussel F, Panayotova K, Weber M. The nonverbal toolkit: towards a framework for automatic integration of nonverbal communication into virtual environments. In: Intelligent environments (IE), 2012 8th International Conference on, IEEE, 2012; 243–50.
93. Lin L, Amith M, Liang C, Duan R, Chen Y, Tao C. Visualized Emotion Ontology: a model for representing visual cues of emotions. In: BMC Medical Informatics and Decision Making, 2018
94. Ortony A, Clore GL, Collins A. The cognitive structure of emotions. 1988.
95. Caridakis G, Raouzaiou A, Karpouzis K, Kollias S. Synthesizing gesture expressivity based on real sequences. In: Workshop on Multimodal Corpora. From Multimodal Behaviour Theories to Usable Models. 5th International Conference on Language Resources and Evaluation (LREC'2006), 2006; pp. 19–23.
96. Berthelon F, Sander P. Emotion ontology for context awareness. In: Cognitive Infocommunications (CogInfoCom), 2013 IEEE 4th International Conference on, IEEE, 2013; pp. 59–64.
97. Benta K-I, Rarau A, Cremene M. Ontology based affective context representation. In: Proceedings of the 2007 Euro American conference on Telematics and information systems, 46, ACM, 2007.
98. Zhang X, Hu B, Chen J, Moore P. Ontology-based context modeling for emotion recognition in an intelligent web. World Wide Web, 16, 4, Springer, 2013; pp. 497–513.
99. Villalonga C, Razzaq MA, Khan WA, Pomares H, Rojas I, Lee S, Banos O. Ontology-based high-level context inference for human behavior identification. Sensors 2016;16(10):1617 (**Multidisciplinary Digital Publishing Institute**)
100. Gil L, Miguel J and Garc Gonzalez R, Gil Iranzo RM, Ordonez C, Cesar A. EmotionsOnto: an ontology for developing affective applications. J Univ Comput Sci. 2014;13(20):1813–28 (**Graz University of Technology**)
101. Grassi M. Developing HEO human emotions ontology. In: European workshop on biometrics and identity management. Springer, 2009; pp. 244–51.
102. Tapia SAA, Gomez AHF, Corbacho JB, Ratt S, Torres-Diaz J, Torres-Carrion PV, Garcia JM. A contribution to the method of automatic identification of human emotions by using semantic structures, Interactive Collaborative Learning (ICL), 2014 International Conference on, IEEE, 2014; pp. 60–70.
103. Obrenovic Z, Garay N, Lopez JM, Fajardo I, Cearreta I. An ontology for description of emotional cues. In: International Conference on Affective Computing and Intelligent Interaction, Springer, 2005; pp. 505–12.
104. Novielli N, Calefato F, Lanubile F. A gold standard for emotion annotation in stack overflow, arXiv preprint arXiv:1803.02300, 2018.
105. Buechel S, Hahn U. EMOBANK: studying the impact of annotation perspective and representation format on dimensional emotion analysis. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2. Short Papers. 2017;2:578–85.
106. Russell JA, Mehrabian A. Evidence for a three-factor theory of emotions. J Res Personal. 1977;11(3):273–94 (**Elsevier**)
107. Saif M. Mohammad and Felipe BravoMarquez. CoRR: Emotion Intensities in Tweets; 2017.
108. Preotiuc-Pietro, Daniel and Schwartz, H Andrew and Park, Gregory and Eichstaedt, Johannes and Kern, Margaret and Ungar, Lyle and Shulman, Elisabeth, Modelling valence and arousal in facebook posts, Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 9–15, 2016
109. Sentiment analysis: emotion in text. https://data.world/crowdflower/sentiment-analysis-in-text. Accessed: 01.10.2018
110. Carlo, Mihalcea R. Semeval-2007 task 14: Affective text, Strapparava. In: Proceedings of the 4th international workshop on semantic evaluations, 70–74, Association for Computational Linguistics, 2007
111. Liu V, Banea C, Mihalcea R. Grounded emotions. In: 2017 Seventh international conference on affective computing and intelligent interaction (ACII). IEEE, 2017; pp. 477–83.
112. Mohammad SM, Bravo-Marquez F, Salameh M, Kiritchenko S. SemEval-2018 task 1: affect in Tweets. In: Proceedings of International Workshop on Semantic Evaluation (SemEval-2018), New Orleans, LA, USA, 2018
113. Demszky D, et al. GoEmotions: a dataset of fine-grained emotions, 2020
114. Plaza-del-Arco FR, et al. EmoEvent: a multilingual emotion corpus based on different events. 2020.
115. Burkhardt F, Paeschke A, Rolfes M, Sendlmeier WF, Weiss B. A database of German emotional speech. In: Ninth European Conference on Speech Communication and Technology, 2005.
116. Engberg IS, Hansen AV, Andersen O, Dalsgaard P. Design, recording and verification of a Danish emotional speech database. In: Fifth European Conference on Speech Communication and Technology, 1997.

117. Wu T, Yang Y, Wu Z, Li D. MASC: a speech corpus in mandarin for emotion analysis and affective speaker recognition, bookSpeaker and language recognition workshop, 2006; pp. 1–5.
118. Torres Neto JR, Filho GPR, Mano LY, Ueyama J. VERBO: voice emotion recognition database in Portuguese Language, 2018
119. Martinez-Lucas L, Abdelwahab M, Busso C. The MSP-conversation corpus; 2020.
120. Siddique L, et al. Cross lingual speech emotion recognition: Urduvs. Western Languages, 2020
121. Parada CE, et al. DEMoS: an Italian emotional speech corpus Elicitation methods, machine learning, and perception, 2019.
122. Nikolaos V, et al. Speech emotion recognition for performance interaction, 2018.
123. Adaeze A, et al. The emotional voices database: towardsControlling the emotion dimension in voice generation systems. 2018.
124. James J, Tian L, Watson CI. An open source emotional speech corpus for human robot interaction applications. 2018.
125. Parada-Cabaleiro E, et al. Categorical vs dimensional perception of Italian emotional speech, 2018.
126. Oliver MM. Esperança Amengual Alcover. UIBVFED: Virtual facial expression dataset; 2020.
127. Karras T, Laine S and Aila T. A style-based generator architecture for generative adversarial networks, 2018.
128. Vemulapalli R, Agarwala A. A compact embedding for facial expression similarity, 2019
129. Yale Face Database. http://vision.ucsd.edu/content/yale-face-database
130. Qu F, et al. CAS(ME)2: a database of spontaneous macro-expressions and micro-expressions. HCI, 2016.
131. Dhall A, Goecke R, Lucey S, Gedeon T. Collecting large, richly annotated facial-expression databases from movies. IEEE Multimed. 2012;19:34–41.
132. Dhall A, Goecke R, Lucey S, Gedeon T. Static facial expression analysis in tough conditions: data, evaluation protocol and benchmark. In: Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCVWorkshops), Barcelona, Spain, 6–13 November, 2011; pp. 2106–12.
133. Pfister T, Li X, Zhao G, Pietikainen M. Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework, Computer Vision Workshops (ICCV Workshops). In: 2011 IEEE International Conference, vol., no., 2011; pp. 868–75.
134. Sim T, Baker S, Bsat M. The CMU pose, illumination, and expression (PIE) database. In: Proceedings of the CMU pose, illumination, and expression (PIE) database. The Fifth IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA, 20–21 May, p. 53, 2002
135. Gao W, Cao B, Shan S, Chen X, Zhou D, Zhang X, Zhao D. The CAS-PEAL large-scale Chinese face database and baseline evaluations. IEEE Trans Syst Man Cybern Part A Syst Hum. 2008;38:149–61.
136. Girard JM, et al. GFT facial expression database. OSF, 23 June 2021.
137. Van der Schalk J, Hawk ST, Fischer AH, Doosje B. Moving faces, looking places: validation of the Amsterdam Dynamic Facial Expression Set (ADFES). Emotion. 2011;11(4):907–20.
138. Nizar EZ. Angled posed facial expression dataset. October 15, IEEE Dataport, 2020.
139. Yang T, Yang Z, Xu G, Gao D, Zhang Z, Wang H, et al. Tsinghua facial expression database—a database of facial expressions in Chinese young and older women and men: Development and validation. PLoS One. 2020;15(4): e0231304.
140. Aifanti N, Papachristou C and Delopoulos A. The MUG facial expression database. In: Proceeding of 11th international workshop on image analysis for multimedia interactive services (WIAMIS), Desenzano, Italy, April 12-14 2010.
141. Shan L, Weihong D, JunPing D. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017
142. Aneja D, Colburn A, Faigin G, Shapiro L, Mones B. Modeling stylized character expressions via deep learning, Asian Conference on Computer Vision (ACCV), 2016
143. Ebner NC, Riediger M, Lindenberger U. FACES—a database of facial expressions in young, middle-aged, and older women and men: development and validation. Behav Res Methods. 2010;42(1):351–62.
144. Lusi I, Jacques Junior JCS, Gorbova J, Baro X, Escalera S, Demirel H, Allik J, Ozcinar C and Anbarjafari G-l. Joint challenge on dominant and complementary emotion recognition using micro emotion features and head-pose estimation: Databases. In: Automatic Face and Gesture Recognition, Proceedings. 12th IEEE International Conference on. IEEE, 2017
145. Douglas-Cowie E, Cowie R, Sneddon I, Cox C, Lowry O, Mcrorie M, Martin J-C, Devillers L, Abrilian S, Batliner A, et al. The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data. In: International conference on affective computing and intelligent interaction. Springer, 2007; pp. 488–500.
146. Douglas-Cowie E, Cowie R, Schroder M. A new emotion database: considerations, sources and scope. In: ISCA tutorial and research workshop (ITRW) on speech and emotion, 2000.
147. Douglas-Cowie E, Cowie R, Cox C, Amir N, Heylen D. The sensitive artificial listener: an induction technique for generating emotionally coloured conversation. In: LREC Workshop on Corpora for Research on Emotion and Affect, 1–4, ELRA, 2008.
148. Busso C, Bulut M, Lee C-C, Kazemzadeh A, Mower E, Kim S, Chang JN, Lee S, Narayanan SS. IEMOCAP: interactive emotional dyadic motion capture database. Language Resour Evaluat. 2008;42(4):335 (**Springer**)
149. Valstar MF, Jiang B, Mehu M, Pantic M, Scherer K. The first facial expression recognition and analysis challenge, bookAutomatic Face & Gesture Recognition and Workshops (FG 2011). In: 2011 IEEE International Conference on, IEEE, 2011. pp. 921–6.
150. Haq S, Jackson PJB and Edge J. Speaker-dependent audio-visual emotion recognition. AVSP, 2009; pp. 53–8.
151. Fanelli G, Gall J, Romsdorfer H, Weise T, Van Gool L. A 3-d audio-visual corpus of affective communication. IEEE Trans Multimed. 2010;12(6):591–8.
152. Barros P, Churamani N, Lakomkin E, Siqueira H, Sutherland A, Wermter S. The OMG-emotion behavior dataset, 2018
153. Livingstone SR, Russo FA. The ryerson audio-visual database of emotional speech and song. PLoS One. 2018;13(5): e0196391.
154. Lee J, Kim S, Kim S, Park J, Sohn K. Context-aware emotion recognition networks. In: IEEE International Conference on Computer Vision (ICCV), 2019.
155. Jean K, et al. SEWA DB: a rich database for audio-visual emotion and sentiment research in the wild, 2020.
156. Zadeh A, et al. Multi-attention recurrent network for human communication comprehension. 2018.
157. Grimm M, Kroschel K, Narayanan S. The Vera am Mittag German audio-visual emotional speech database. In: Proceedings of the 2008 IEEE International Conference on Multimedia and Expo, Hannover, Germany, 2008.
158. Kanade T, Tian Y, Cohn JF. Comprehensive database for facial expression analysis, IEEE, 2000
159. Lyons M, Akamatsu S, Kamachi M, Gyoba J. Coding facial expressions with gabor wavelets, Automatic Face and Gesture

Recognition, 1998. Proceedings. Third IEEE International Conference, IEEE, 1998

160. Yin L, et al. A high-resolution 3d dynamic facial expression database, Automatic Face and Gesture Recognition (FGR08). In: 8th international conference, 2008.
161. Yin L, Wei X, Sun Y, Wang J, Rosato MJ. A 3D facial expression database for facial behavior research. Automatic face and gesture recognition, 2006. FGR 2006. 7th international conference on, IEEE, 2006
162. Pantic M, Valstar M, Rademaker R, Maat L. Web-based database for facial expression analysis. In: 2005 IEEE international conference on multimedia and Expo, IEEE, 2005
163. Patel VM, Gopalan R, Li R, Chellappa R. Visual domain adaptation: a survey of recent advances. IEEE Signal Process Magazine. 2015;32(3):53–69.
164. Kosti R, Alvarez JM, Recasens A, Lapedriza A. Emotion recognition in context. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017
165. McDuff D, Kaliouby R, Senechal T, Amr M, Cohn J, Picard R. Affectiva-mit facial expression dataset (am-fed): naturalistic and spontaneous facial expressions collected. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2013; pp. 881–888.
166. Zijderveld G. The World's largest emotion database: 5.3 million faces and counting. https://blog.affectiva.com/the-worlds-largest-emotion-database-5.3-million-faces-and-counting, 2017, accessed 10-March-2020.
167. Mavadati, SM, Mahoor MH, Bartlett K, Trinh P, Cohn JF. Disfa: a spontaneous facial action intensity database. In: IEEE Transactions on Affective Computing, IEEE, 2013
168. Baveye Y, Dellandrea E, Chamaret C, Chen L. Liris-accede: a video database for affective content analysis. IEEE Transactions on Affective Computing; 2015.
169. Gunes H, Piccardi M. A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior. ICPR: Pattern Recognition; 2006.
170. Min R, Kose N, Dugelay J-L. Kinectfacedb: a kinect database for face recognition. IEEE Trans Syst Man Cybernet: Syst. 2014.
171. Jiang Y-G, Xu B and Xue X. Predicting emotions in user-generated videos, AAAI, 2014.
172. Soleymani M, Lichtenauer J, Pun T, Pantic M. A multimodal database for affect recognition and implicit tagging. IEEE Trans Affect Comput. 2012.
173. Koelstra S, Muhl C, Soleymani M, Lee J-S, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I. Deap: a database for emotion analysis; using physiological signals. In: IEEE Transaction on Affective Computing; 2012.
174. Ringeval F, Sonderegger A, Sauer J, Lalanne D. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In: Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, IEEE, 2013
175. Carvalho S, et al. The emotional movie database (EMDB): a self-report and psychophysiological study. Appl Psychophysiol Biofeedback. 2012. Springer.
176. Zhang Z, Girard JM, Wu Y, Zhang X, Liu P, Ciftci U, Canavan S, Reale M, Horowitz A, Yang H et al. Multimodal spontaneous emotion corpus for human behavior analysis. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),Washington, DC, USA, 27–30 June, 2016; pp. 3438–46.
177. Abadi MK, Subramanian R, Kia SM, Avesani P, Patras I, Sebe N. DECAF: MEG-based multimodal database for decoding affective physiological responses. IEEE Trans Affect Comput. 2015;6(3):209–22.
178. Ranganathan H. Chakraborty S. Panchanathan S. Multimodal emotion recognition using deep learning architectures, Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on, IEEE, 2016
179. Stamos K, Naeem R. DREAMER: a database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. IEEE J Biomed Health Inform. 2018;98–107.
180. Firdaus M, et al. MEISD: a multimodal multi-label emotion, intensity and sentiment dialogue dataset for emotion recognition and sentiment analysis in conversations, 2020
181. Subramanian R, Wache J, Abadi M, Vieriu R, Winkler S, Sebe N. ASCERTAIN: emotion and personality recognition using commercial sensors. IEEE Trans Affect Comput. 2016
182. Krcadinac U, Pasquier P, Jovanovic J, Devedzic V. Synesketch: an open source library for sentence-based emotion recognition. IEEE Computer Society Press, 2013
183. Calefato F, Lanubile F, Novielli N. EmoTxt: a toolkit for emotion recognition from text. In: 2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), 2017.
184. Den Uyl MJ, Van Kuilenburg H. The FaceReader: online facial expression recognition. Proc Measur Behav. 2005;30(2):589–90.
185. Schultz R, Peter C, Blech M, Voskamp J, Urban B. Towards detecting cognitive load and emotions in usability studies using the RealEYES framework. In: International Conference on Usability and Internationalization, 2007
186. Javascript library for precise tracking of facial features via Constrained Local Models (CLM), avialable at https://github.com/auduno/clmtrackr
187. Al Machot F, Elkobaisi MR, Kyamakya K. Zero-shot human activity recognition using non-visual sensors. Sensors (Basel, Switzerland). 2020.
188. Elkobaisi MR and Al Machot F. Human emotion modeling (HEM): an interface for IoT systems. J Ambient Intell Humaniz Comput. 2021.
189. Mano LY, Faical BS, Nakamura LHV, Gomes PH, Libralon GL, Meneguete RI, Filho GPR, Giancristofaro GT, Pessin G, Krishnamachari B, Ueyama. Exploiting IoT technologies for enhancing health smart homes through patient identification and emotion recognition. Comput Commun. 2016
190. Sako A, Saiki S, Nakamura M, Yasuda K. Developing face emotion tracker for quantitative evaluation of care effects, Lecture Notes in Computer Science, 10917, 2018, Springer
191. Hermann H, Trachsel M, Elger BS, Biller-Andorno N. Emotion and value in the evaluation of medical decision-making capacity: a narrative review of arguments. Front Psychol. 2016.
192. Ilyas CMA, Haque MA, Rehm M, Nasrollahi K, Moeslund TB. Effective facial expression recognition through multimodal imaging for traumatic brain injured patient's rehabilitation. In: Imaging and computer graphics theory and applications, computer vision; 2018.
193. Mohanapriya N, Malathi L, Revathi B. A survey on emotion recognition from EEG signals for autism spectrum disorder, 2018.
194. Taj-Eldin M, Ryan C, O'Flynn B, Galvin P. A review of wearable solutions for physiological and emotional monitoring for use by people with Autism Spectrum Disorder and their caregivers, 2018
195. Akputu OK, Seng KP, Lee Y and Ang LM. Emotion recognition using multiple kernel learning toward E-learning applications. ACM Trans. Multimedia Comput. Commun; 2018.
196. Garcia-Garcia JM, Penichet VMR, Lozano MD, Garrido JE, Law EL-C. Multimodal affective computing to enhance the user experience of educational software applications. Mobile Inf Syst. 2018.

197. Krithika LB, Lakshmi Priya GG. Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric, International Conference on Computational Modeling and Security (CMS 2016). 2016; pp. 767–76.
198. Ryan A, Cohn JF, Lucey S, Saragih J, Lucey P, De la Torre F, Rossi A. Automated facial expression recognition system. In: 43rd Annual 2009 International Carnahan Conference on Security Technology, 2009, pp. 172–7.
199. Garcia-Ceja E, Osmani V, Mayora O, Automatic stress detection in working environments from smartphones' accelerometer data: a first step. IEEE J Biomed Health Inform. 2016.
200. Hänggi Y. Stress and emotion recognition: an internet experiment using stress induction. Swiss J Psychol. 2004;63:113–25.
201. Mishra P. HMM based emotion detection in games. In: 3rd International Conference for Convergence in Technology (I2CT), 2018.
202. Scott HH, Bowman ND. Video games, emotion, and emotion regulation: expanding the scope. Ann Int Commun Assoc. 2018;42(2):125–43.
203. Hussain J, Khan WA, Hur T, et al. A multimodal deep log-based user experience (UX) platform for UX evaluation. Sensors (Basel). 2018.
204. Wiklund M, Rudenmalm W, Norberg L, Mozelius P. Evaluating educational games using facial expression recognition software – measurement of gaming emotion. In: The 9th European Conference on Games Based Learning, Steinkjer, Norway, 2015
205. Mukeshimana M, Ban X, Karani N, Liu R. Multimodal emotion recognition for human-computer interaction: a survey, 2017.
206. Palm G, Glodek M, Apolloni B, Bassis S, Esposito A, Morabito FC. Towards emotion recognition in human computer interaction, "Neural Nets and Surroundings: 22nd Italian Workshop on Neural Nets, WIRN 2012, May 17–19, Vietri sul Mare, Salerno, Italy", 2013
207. Faria DR, Vieira M, Faria FCC and Premebida C. Affective facial expressions recognition for human-robot interaction. In: 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 2017; pp. 805–10.
208. Burgos FC, Manso L and Trujillo P. A novel multimodal emotion recognition approach for affective human robot interaction. 2015.
209. Shukla A, Gullapuram SS, Katti H, Yadati K, Kankanhalli MS, Subramanian R. Affect recognition in Ads with application to computational advertising. CoRR, 2017.
210. Shukla A, Gullapuram SS, Katti H, Kankanhalli MS, Winkler S, Subramanian R. Recognition of advertisement emotions with application to computational advertising. CoRR, 2019.
211. Consoli D. Annales universitatis apulensis series oeconomica, emotions that influence purchase decisions and their electronic processing, Faculty of Sciences, 2009; pp. 1–45.
212. Kidwell B, Hardesty DM, Murtha BR, Sheng S. Emotional intelligence in marketing exchanges. J Market. 2011;78–95.
213. Hachisuka S, Ishida K, Enya T, Kamijo M, Harris D. Facial expression measurement for detecting driver drowsiness. Eng Psychol Cogn Ergon. 2011.
214. Assari MA, Mohammad R. Driver drowsiness detection using face expression recognition. In: 2011 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2011, Kuala Lumpur, Malaysia. 2011
215. Bosch E, Oehl M, Jeon M, Alvarez IJ, Healey J, Ju W, Jallais C. Emotional GaRage: a workshop on in-car emotion recognition and regulation. AutomotiveUI, 2018.
216. Fridman L, et al. MIT autonomous vehicle technology study: large-scale deep learning based analysis of driver behavior and interaction with automation. CoRR. 2017
217. Schirmer A, Adolphs R. Emotion perception from face, voice, and touch: comparisons and convergence. 2017.
218. Bänziger T, Grandjean D, Scherer K. Emotion recognition from expressions in face, voice, and body: the multimodal emotion recognition test (MERT), 2009
219. Chen LS, Huang TS, Miyasato T, Nakatsu R. Multimodal human emotion/expression recognition. In: Proc. of Int. Conf. on Automatic Face and Gesture Recognition, 1998; pp. 366–71.
220. D'Mello K, Sidney and Graesser A. Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features, 2010
221. Emerich S, Lupu E, Apatean A. Bimodal approach in emotion recognition using speech and facial expressions. In: 2009 International Symposium on Signals, Circuits and Systems, 2009; pp. 1–4.
222. Kapoor A, Picard RW. Multimodal affect recognition in learning environments. In: Proceedings of the 13th Annual ACM International Conference on Multimedia, series = MULTIMEDIA '05, 2005
223. Scherer K. Multimodal expression of emotion: affect programs or componential appraisal patterns? Ellgring Heiner, 2007.
224. Michael J, Mayr HC. Creating a domain specific modelling method for ambient assistance. In: Karagiannis D, Mayr HC, Mylopoulos J (eds.) Proc. Int. Conf. on Advances in ICT for Emerging Regions ICTer2015, Colombo, August 2015. Domain-Specific Conceptual Modeling - Concepts, Methods and Tools. Springer, 2015; 119–24.
225. Elkobaisi MR, Mayr HC, Shekhovtsov V. Conceptual human emotion modeling. In: Advances in Conceptual 768 Modeling, ER Workshops 2020, Springer LNCS, 2020; pp. 71–81.

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