Journal of Network and Computer Applications 147 (2019) 102423

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Review

A survey of emotion recognition methods with emphasis on E-Learning environments

Maryam Imani [a,*], Gholam Ali Montazer [b]

[a] Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
[b] Faculty of Information Technology Engineering, Tarbiat Modares University, Tehran, Iran

ARTICLE INFO

Keywords:
Emotion recognition
Facial expression
Body gesture
Speech
Physiological signal
Text
E-learning

ABSTRACT

Emotions play an important role in the learning process. Considering the learner's emotions is essential for electronic learning (e-learning) systems. Some researchers have proposed that system should induce and conduct the learner's emotions to the suitable state. But, at first, the learner's emotions have to be recognized by the system. There are different methods in the context of human emotions recognition. The emotions can be recognized by asking from the user, tracking implicit parameters, voice recognition, facial expression recognition, vital signals and gesture recognition. Moreover, hybrid methods have been also proposed which use two or more of these methods through fusing multi-modal emotional cues. In the e-learning systems, the system's user is the learner. For some reasons, which have been discussed in this study, some of the user emotions recognition methods are more suitable in the e-learning systems and some of them are inappropriate. In this work, different emotion theories are reviewed. Then, various emotions recognition methods have been represented and their advantages and disadvantages of them have been discussed for utilizing in the e-learning systems. According to the findings of this research, the multi-modal emotion recognition systems through information fusion as facial expressions, body gestures and user's messages provide better efficiency than the single-modal ones.

1. Introduction

Individualized tutoring to the learners is provided by computer based learning systems called Intelligent Tutoring Systems (ITS) (Nkambou and Gauthier, 1996). The researches have been shown that the learners who receive one-on-one instructions can learn better and faster than learners in traditional classrooms (Bloom et al., 1956). Providing a real personalized learning environment is beyond the training and education budgets of most organizations (Montazer and Sadegh Rezaei, 2012). But, a virtual and electronic training environment can provide the advantages of one-on-one instruction cost effectively (Caputi and Garrido, 2015). The learner model, as the main component of ITS, consists of the motivational, cognitive, and other affective states that have important effects to performance of learner's learning (Saberi and Montazer, 2012). The tutor model cooperates between the learner model and domain to handle tutoring actions and strategies (Villiger et al., 2019). Detection of learner's emotional states and their reaction for a given situation is one of the most important elements for every electronic learning (e-learning) environment.

1.1. Search items and selection of articles

This work reviews papers written in English and published in various scientific journals and proceedings of international symposia, conferences and workshops from beginning to 2020. The main used electronic database is sciencedirect, i.e., the papers published by Elsevier. Also, several related works published by institute of electrical and electronic engineering (IEEE) have been reported. This review paper is made of two main parts. The researches of the first part substantially are the articles published in journals related to cognitive science and psychology. The second part substantially reports the articles published in journals or proceedings related to machine learning, pattern recognition and artificial intelligence.

For the first part, the keywords 'emotion theories' and 'emotion models' have been searched in sciencedirect and the main works have been selected by scanning the whole paper with emphasis to title and abstract. For the second part, at first a general search is done in sciencedirect and IEEE. The search has been conducted using the keywords 'emotion recognition' and one of the following terms: 'e-learning', 'face recognition', 'gesture recognition', 'physiological (vital) signals', 'speech

* Corresponding author.
E-mail addresses: maryam.imani@modares.ac.ir (M. Imani), montazer@modares.ac.ir (G.A. Montazer).

https://doi.org/10.1016/j.jnca.2019.102423
Received 6 April 2019; Received in revised form 14 July 2019; Accepted 21 August 2019
Available online 27 August 2019

Table 1
Total number of articles identified from sciencedirect database.

	E-learning	Face recognition	Gesture recognition	Physiological signals	Speech recognition	Text	self-report
Emotion recognition	15,890	35,041	6043	10,457	13,562	14,569	32,466

(voice) recognition', 'text' and 'self-reporting'. The total number of articles identified from sciencedirect database is reported in Table 1. The term 'emotion detection' is also searched in IEEE and 1892 papers were found. The identified papers are screened by title and abstract and the main related works for emotion recognition using machine learning methods have been chosen.

1.2. Contents

Education and learning are among the most important bases of each society. They are influential processes in the perpetual enrichment of knowledge and specialization, and also for improvement of communication between individuals and nations of the world. In a real learning environment (school), the human instructor (teacher) has face-to-face relationship with learner (student) and can show appropriate reaction according to action or affection of learner, and also presents the appropriate educational points. But, in an electronic educational system, the structure of learning domain and content are usually presented in the static way, without taking into account the learners' affections and without interactivity and feedbacks from the instructor. However, considering affection and emotion will increase the quality of learning. To deal with this problem, it has been proposed to provide an adaptive electronic learning system associated with the affections of learners. This approach improves the quality of learning, increases the concentration and strengthens the memory. The emotion recognition methods using computers are generally divided into seven groups (see Fig. 1):

- Asking from user
- Tracking implicit parameters
- Voice recognition
- Facial expression recognition
- Vital signals
- Gesture recognition
- Hybrid methods

Each of them has its advantages and disadvantages. For each of them there are some methods and algorithms that are discussed in the following sections. But they are similar from one aspect. They usually do emotion recognition in three main stages: 1- data pre-processing, 2- discovering effective features for recognizing emotions, and 3- classifying based on extracted or selected features, i.e., assigning appropriate emotional states to them. In this paper we have explained each of these research areas and their benefits and difficulties for using them in the e-learning context.

This paper is composed of two parts. In part I, different models and theories about emotions and also the applications of emotion recognition are represented. In part II, a wide range of emotion recognition methods is discussed. This paper is organized as follow. In sections 2-3 in part I, the emotion models and theories, applications of emotion recognition, online learning environments and effects of emotions in learning are represented, respectively. In part II, section 4 discusses emotion recognition using asking from the user. Section 5 is about tracking implicit parameters. Section 6 explains different methods for emotion recognition using voice recognition. Facial expression recognition is discussed in section 7. Different methods using vital signals are reviewed in section 8. Section 9 is about gesture recognition and hybrid methods are discussed in section 10. A Comparison between different emotion recognition methods for e-learning goals is given in section 11. Finally, conclusions and future works are represented in section 12.

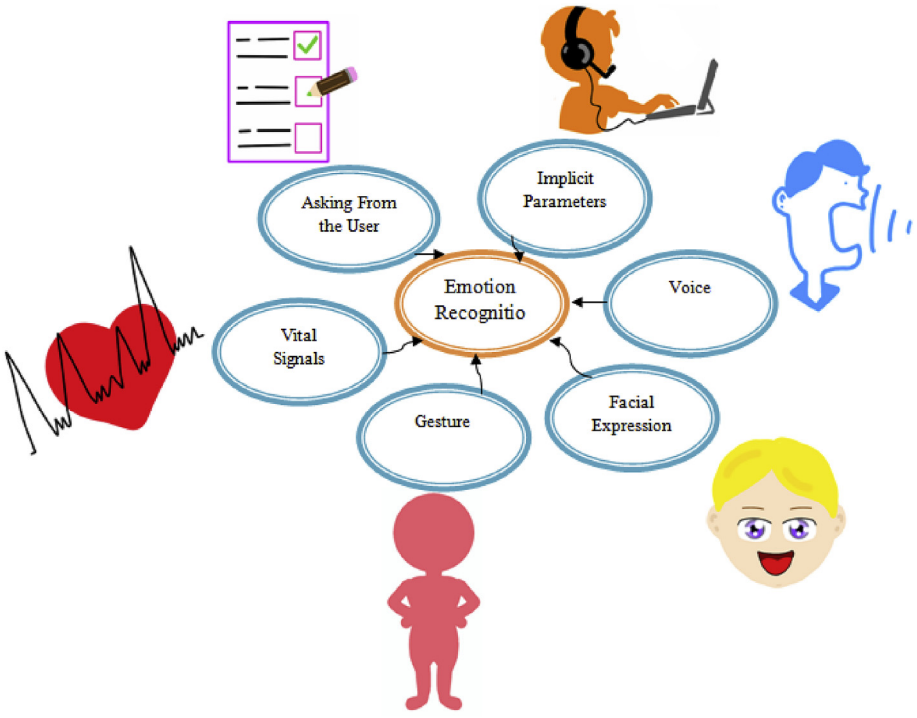


Fig. 1. Different sources for emotion recognition.

2. Emotion models and theories

Emotions have undeniable role in different aspects of life. So far, many different definitions of emotion have been presented (Kleinginna and Kleinginna, 2005). Emotions are feeling states with negative or positive affective valence (Ortony et al., 1988). The ability to reason about emotions supports countless social behaviors, from maintaining healthy relationships to scheming for political power. Not only emotions have a substantial role in human life and social interactions of each person, but also, emotions have important role in human perception and cognition. The neurological researches and studying utilitarian functions within brain human show the obvious role of emotions human in the rational decision making (Picard, 2001; Sander et al., 2005a).

The collection of all cognitive processes for reasoning about others' emotions is called affective cognition. So far, many studies have been done to describe how people obtain accurate and complex attributions about others' psychological and emotional states (Tomasello et al., 2005; Zaki and Ochsner, 2011). Lay theories, sometimes called folk theories or intuitive theories, provide a structured knowledge about the world and an abstract framework for reasoning the other's emotions (Flavell, 1999; Leslie et al., 2004). Similar to scientific theories, lay theories coherently describe how the world works. While a scientific theory describes the world, a lay theory makes sense of that.

In (Ong et al., 2015), a model of lay theory that explains how people inference others' emotions is introduced (see Fig. 2). A reasoner (observer), called agent by the lay theory, infers about target of the reasoning. The existence of emotional stimuli and interaction of outcomes of those with other mental states such as goals produces emotions in agent. The external cues of agent's emotions include speech, body language, facial expressions or further actions. Except goals and emotions, which are internal mental states, all of the variables of model are potentially observable while emotion as a latent variable is unobservable. A probability distribution can be used to represent each of directed causal relationships of model.

The emotion theories are categorized in terms of two different views (Lopatovska and Arapakis, 2011a): manifestation and structure (see Fig. 3). In the first category, there are two groups of emotion theories. The first group is based on cognitive factors while the second group is based on somatic factors. The emotion theories based on cognitive factors consider cognition as a necessary element of emotion and take a form of a thought or judgment. They explain the subjective manifestations of emotional experiences where the cognitive activity can be intentional or unintentional, conscious or unconscious.

According to (Lazarus, 1984), cognitive factors are the most important stimuli. In (Frijda, 1994) emotion is defined as a reaction to affective events. In (Scherer, 2005) the componential theory of emotion is represented where emotion is defined as a synchronization of different cognitive and perceptual processes. Other emotion theories based on cognitive factors can be found in (Bilal and Bachir, 2007; Gwizdka and Lopatovska, 2009; Lopatovska, 2009; Lopatovska and Mokros, 2008).

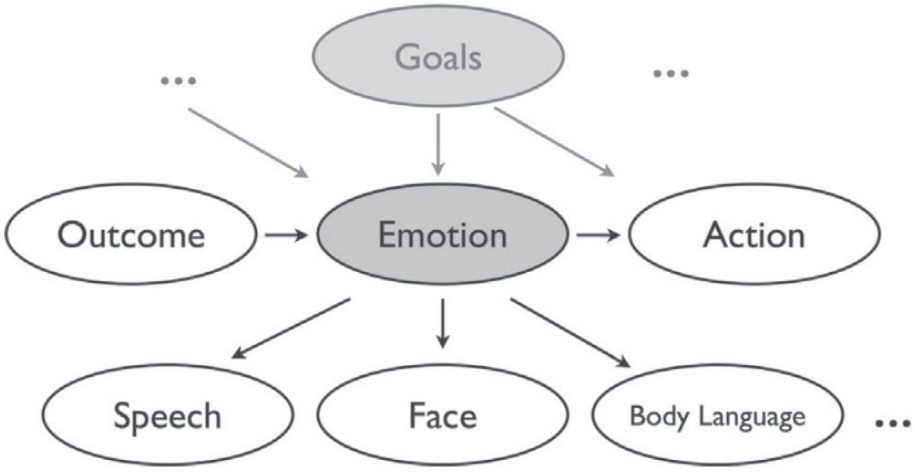


Fig. 2. Model of a lay theory (Ong et al., 2015).

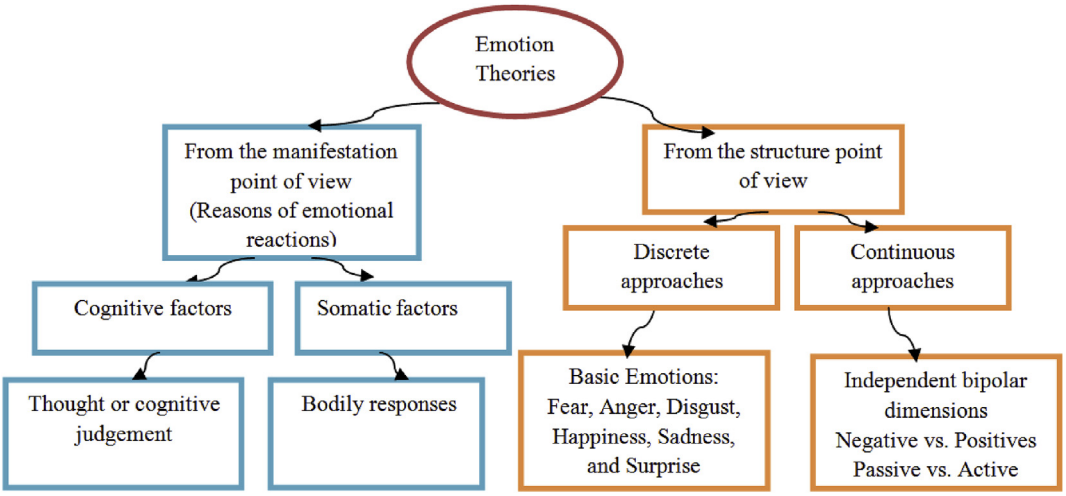


Fig. 3. Categorization of emotion theories.

The emotion theories based on somatic factors, instead of cognitive judgments, consider bodily responses as the reasons of emotional reactions. According to (Ekman, 1984), emotions by mobilizing an organism for quick respond to prototypical events can be considered as psychosomatic states appeared over time. An affect system has amplification effect on bodily and physical functions (Tomkins, 1984). Other somatic based theories can be found in (Plutchik, 1980; Arapakis et al., 2009; Smeaton and Rothwell, 2009).

The second category of emotion theories discuss the structure of emotions in terms of discrete or continuous approaches. Discrete emotion approaches consider the existence of several universally recognized basic emotions (such as fear, anger, disgust, happiness, sadness, and surprise) (Ekman, 1992; Darwin, 2005). All people in the world express and recognize the basic emotions the same way. Other emotions, which do not locate in the six basic emotions, can be considered as combinations of the basic emotions. For instance, guilt is a variant of basic sadness. Many studies focused on separable basic emotions (discrete emotion model). The use of discrete emotions has some disadvantages. The researchers have no consensus on the number of discrete emotions although the most accepted model consists of six basic emotions. From the views of some critics, the discrete model has incapability to capture some human emotions. However, the discrete emotion model is widely used because of its simplicity, interpretability, and high plausibility.

The emotion theories based on continuous structure consider the existence of two or more dimensions for description of different emotions. Russell, instead of a small number of discrete emotions, introduced independent bipolar dimensions of emotion such as pleasure–displeasure (Russell, 1994). In the circumplex model of affect proposed by Russell, distribution of emotions is located in a two dimensional circular space containing arousal and valence (Russell, 1980). Arousal/valence represents the vertical/horizontal axes while the center of the circle is equivalent to a medium level of arousal and a neutral valence. In this model, the emotional expressions can be illustrated at any level of arousal and valence, or at a neutral level of one or both of them. An illustration of six basic emotions in this model is shown in Fig. 4 (Kim and Andre, 2008). A variant on Russell's scheme containing two dimensions of activation in vertical axis and pleasantness in horizontal axis was introduced in (Larsen and Diener, 1992). In (Fernández-Caballero et al., 2016), six basic emotions are added in the border of circle of Russell's circumplex model (see Fig. 5).

In (Scherer, 2002), all kinds of emotions meaning consist of negative versus positives as valence, and passive versus active as activation have been shown. Other examples of continuous based structure approach are reported in (Chan and Jones, 2005; Soleymani et al., 2008). The recalibrated speech affective space model (rSASM) (Scherer, 2005) and the 12-point affective circumplex (12-PAC) (Russell, 1980) are instances of two-dimensional models of emotions. In rSASM, the primitive values of emotions are derived from the four quadrants of emotion space. The 12-PAC model is a geometrical model which provides higher accuracy in emotional profiling compared to rSASM. Some instances of illustration of different emotions based on the rSASM and 12-PAC model are shown in Fig. 6.

With arising an emotional state, the facial expressions, body gestures, and intonation of voice is affected. The normal cognitive operations such as decision-making, attention, and perception are modulated by the emotional significance of perceived environmental stimuli. In an artificial intelligence system, the major goal is development of believable t autonomous agents. In (Rodríguez et al., 2016), authors address the interaction of cognition and emotion in agent architectures for generation of believable and consistent affectional behaviors.

Emotions have a main role in different stages of human life. Emotions effect different aspects of human from childhood to old age. Sometimes, the gender causes key differences in social behaviors of boys and girls. In (KuhnertSander et al., 2017) the gender differences are investigated in the relations between children's prosocial behavior and their theory of mind, emotion understanding, and social preference ratings. The experimental results on both girl and boy children show that the relation between earlier theory of mind and prosocial behavior are the same for all children, irrespective of their gender. But, the relation between emotion understanding and prosocial behavior is effected by gender such that a positive association between emotion understanding and prosocial behavior is seen just for girls.

The control-value theory is proposed to analyze the role of achievement emotions (Pekrun, 2006). According to this theory, two cognitive appraisals, i.e., subjective value and subjective control related to outcomes and achievement activities arouse the achievement emotions. Intrinsic values and extrinsic values are two types of subjective values. Intrinsic values are due to academic tasks such as sense of satisfaction from studying. In other hand, the sense of pleasure due to usefulness of academic outcomes to achieve other goals is source of extrinsic values. The person's perceived causal influence of the self over outcomes and achievement activities is known as subjective control often operationalized as self-efficacy (Pekrun et al., 2011). The combination of subjective value and subjective control reveals various achievement emotions such as enjoyment and boredom. In a learning activity, according to the control-value theory, students feel enjoyment when the subject is interesting or when students have high dominance in that. Students feel pride when they achieve success or when they have a high sense to achieve

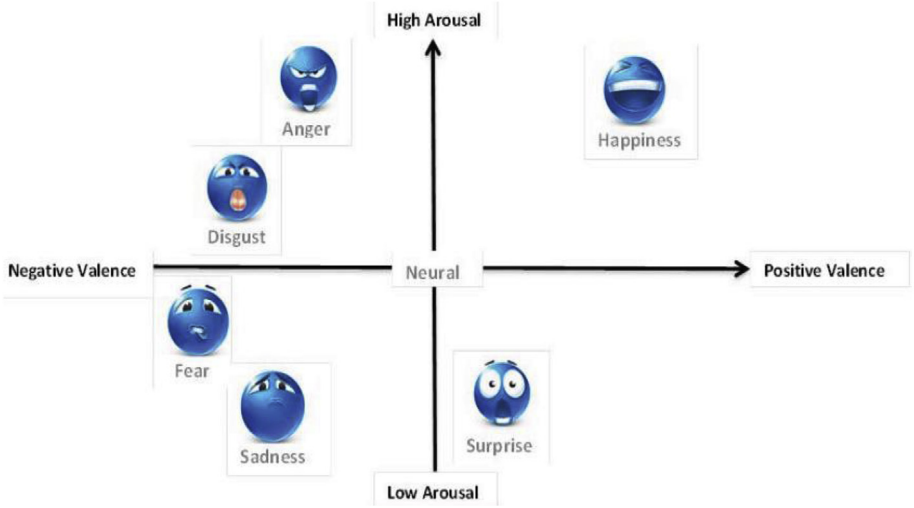


Fig. 4. The basic emotions in a 2-dimensional model of arousal and valence (Kim and Andre, 2008), (Khezri et al., 2015).

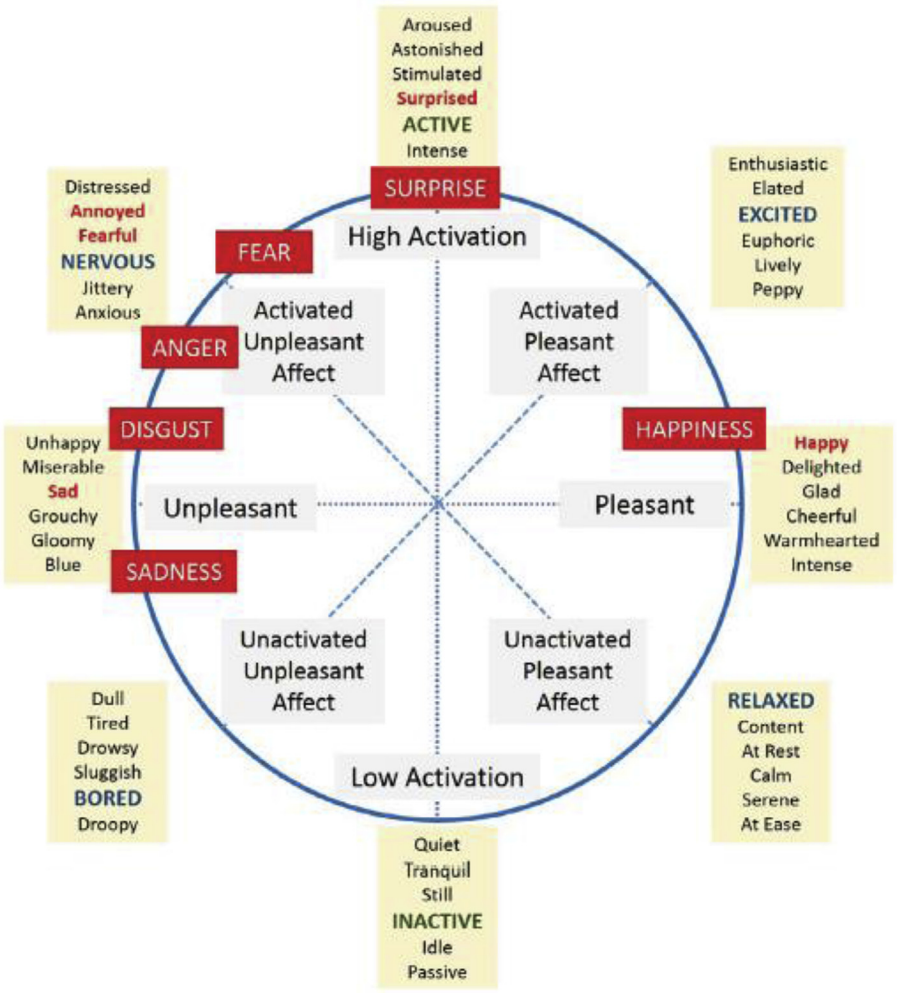


Fig. 5. The circumplex model of affect (Fernández-Caballero et al., 2016).

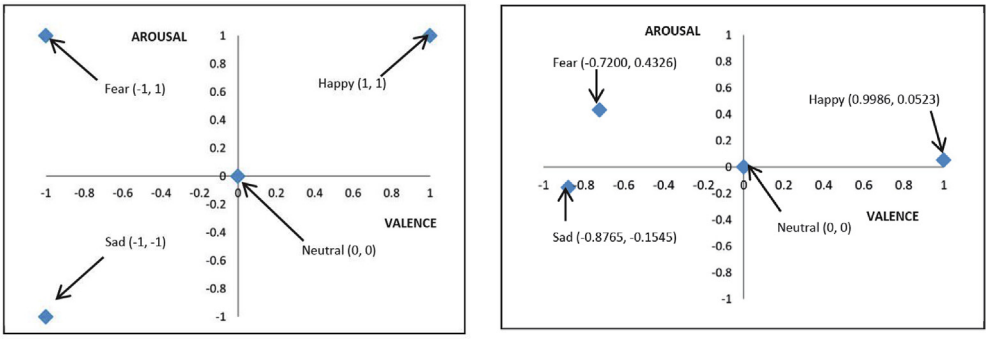


Fig. 6. Some instances of illustration of different emotions based on the rSASM (left) and 12-PAC nodel (right) (Othman et al., 2013).

success. When a learning activity is not a value for students, they experience boredom. When the outcome of a test is important for students, they tend to feel anxiety. The researchers in (Putwain et al., 2013) by doing an examination on 434 undergraduate students produce evidence for control-value theory. They found that there are relations between achievement goals and learning-related emotions. Moreover, the mastery goals are related to both activity-focused emotions and outcome-focused emotions.

According to the control-value theory, the social environments, such as parental expectancy and parental involvement affect students' emotional experiences through the subjective value and the subjective control. The expectations of parents about success of children in academic tasks have high effect on the children's motivation, students' s success beliefs, and academic outcomes. The involvement of parents in the school related activities of children such as doing homework is known as parental involvement. The students' s motivation, learning, and emotion are effected by parental involvement where for example parental involvement has straightforward effect on students' expectancy in their doing homework. In (Luo et al., 2016) a mediation model is hypothesized to investigate four salient achievement emotions in learning mathematics (boredom, anxiety, pride, and enjoyment) from the control-value theory point of view (see Fig. 7). For example, according to

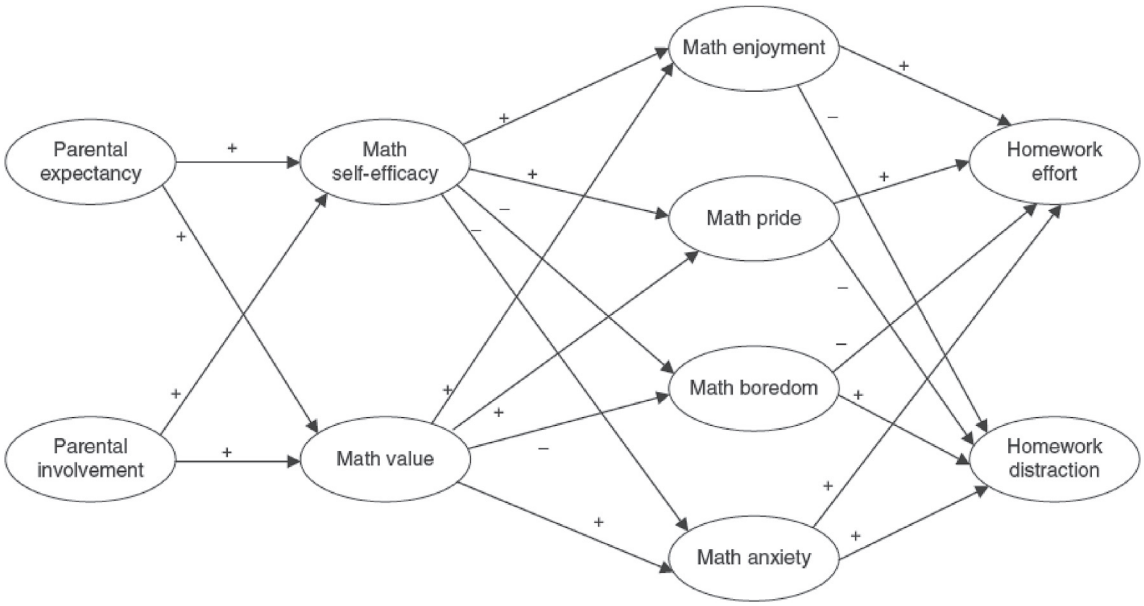


Fig. 7. The mediation model hypothesized in (Luo et al., 2016).

the hypothesized model in Fig. 7, the math enjoyment as a positive activating emotion would be associated positively/negatively with homework effort/homework distraction. For instance, an indication of the math self-efficacy is "I can do almost all mathematical tasks". How students perceive the value of learning mathematics and usefulness of it in their future education is called math value. For instance, an indication of the math value is "learning mathematics is helpful in my routine life" (Luo et al., 2016).

In addition to empirical and theoretical studies for investigation of academic emotions, the implicit theories of intelligence have been proceeded although the existence of a little cross-over of these ideas is possible (Dweck et al., 2011). The possible potential synergies between two mentioned paradigms are studied in (King et al., 2012) where relationship between academic emotions and implicit theories of intelligence is investigated. Academic emotions have different types which can be categorized into activation, i.e., activating versus deactivating and valence, i.e. positive versus negative (Pekrun et al., 2002). Crossing these two emotional dimensions provides positive activating, positive deactivating, negative activating, and negative deactivating emotions. While activating emotions physiologically facilitate excitement, deactivating emotions induce relaxation. Positive and negative emotions can be also differentiated from each other in terms of valence. For example, pleasant happiness and enjoyment felt during learning is versus unpleasant anxiety experienced before an exam. Hope, enjoyment, and pride are some instances of positive activating emotions. Anxiety, shame, and anger are some instances of negative activating emotions. Moreover, while relief can be considered as a positive deactivating emotion, hopelessness and boredom are considered as negative deactivating emotions.

There are two different theories of intelligence about relation between students' s effort and intelligence (Dweck, 1999). The first one, which is called entity theory of intelligence, sees intelligence fixed where effort has marginal effect on it. In contrast, the second theory, called incremental theory of intelligence, sees intelligence malleable where effort and learning change the intelligence. Each of these theories creates different interpreting about failure or success. In other words, entity and incremental theories are located on opposite poles of an uni-dimensional continuum construct (Dweck et al., 2005). The studies showed that students with an entity theory have more need to confirm their smartness through validation of their intelligence. In contrast, the students with an incremental theory have more interests to learn the material and develop their abilities and skills. The studies have been shown the positive advantages of an incremental theory versus the maladaptive results of an entity theory of intelligence. The following hypotheses are considered in (King et al., 2012): Entity theory negatively predicts positive academic emotions while it positively predicts negative academic (activating/deactivating) emotions. As a conclusion about this discussion, there is a straightforward relation between how students think about their intelligence and their feelings in school. It can be useful that students think their intelligence can be improved through more effort and hard work.

Students experience despondent each time they procrastinate their homework. In contrast, they feel well-being when doing their homework. The most of Research on students' motivational processes have mainly focused on interpersonal differences while studying the intrapersonal ups and downs of students' motivation can be also very beneficial to better comprehend the ongoing dynamics that forge students' everyday lives. A five-month diary study with 179 high school participants in a small city in Northern Greece is done in (Mouratidis et al., 2016) for a multilevel analysis about ongoing dynamics of students' thought and motivation. The relation between wellbeing (depressive feelings and subjective vitality) and study manner regulation (procrastination in contrast to study efforts) and autonomous functioning investigated in this study. The results showed that autonomous functioning has positive effects on study efforts, having subjective vitality and negative effects on procrastination and depressive expressions. This pattern was seen particularly true among female participants. Moreover, the studying results showed that the beliefs that ability is malleable have positive effects on study efforts and negative effects on procrastination. In contrast, beliefs that ability is fixed provide lower study efforts, higher homework procrastination, and consequently poorer grades.

3. Applications of emotion recognition

3.1. Varied applications

Affect systems have been designed for various applications such as distance education environments, emotion-sensitive interfaces, increasing drivers'safety, and discovering depressed writers (Quan and Ren, 2016). Life with humanoid robots is one of the future aims of human. By fast growing of technology, the development of robots capable of emotion recognition and understanding people and responding to human needs has noticeable improvement. The service robots with

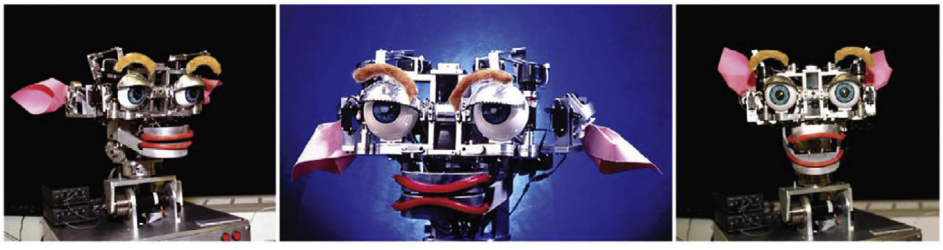
considering the emotional aspect of human can be very helpful for elderly people or people with disabilities (Aceto et al., 2018; PAL, 2015). A humanoid robot tries to emulate the interaction of human people with each other. Robot Kismet (Breazeal, 2003), Jibo (Chambers et al., 2015), Model Of User's Emotions (MOUE) (Lisetti et al., 2003) are some of examples of these robots capable of emotion recognition and interface with human (see Fig. 8 (Perez-Gaspar et al., 2016)). Robot Kismet is social robot which is able to recognize affective intentions (calm, sadness, happiness, surprise, fear, disgust, anger, and tiredness) through the voice. Jibo is an extension of Kismet which acted as a companion robot used for commercial purposes. For instance, it is able to tell stories, to take photographs, and to remind important events. MOUE is an intelligent interface capable of distance monitoring of patient. This system captures emotional gestures and physiological signals through a web-cam and a bracelet connected to a computer. The capture data is sent to a central computer to process. This artifical system can have an interaction by patient through an animated character (avatar), which reflects the facial expression of the patient as a mirror. It can process the emotions of frustration, sadness, fear, anger, and neutral.

The marketers can analyze the textual content of online customer reviews which expressed emotions. Because, theses expressed emotions can affect other customers, this analyze can be helpful for marketers to manage and improve their business (Dai et al., 2015b).

The number of elderly people or patients, which now are treated in home, is increased. These people may require help in critical conditions for example when becoming depressed or facing an accident. So, the health smart homes have been deployed. Advances in the Internet of things and ubiquitous computing have provided cheap and efficient equipment including cameras and wireless communication. Researchers in (Mano et al., 2016) discuss how the use of patient images for their emotion detection can assist them within a healthcare context. The LifeShirt is developed in (Wilhelm et al., 2006) for ambulatory monitoring of a wide variety of experiential responses, motor-behavioral, metabolic, respiratory and cardiovascular. This system consists of a garment with sensors for recording the physiologic data. This device is useful in clinical studies for monitoring effects of emotional and physical stress in naturalistic settings.

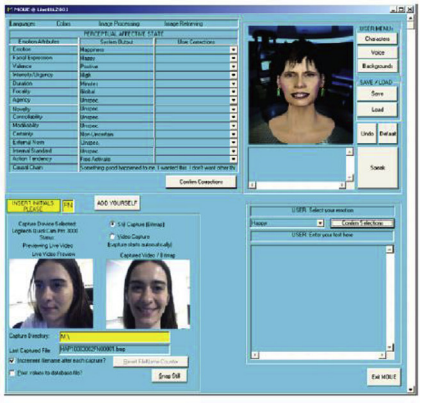
One of the main factors in transportation accidents is driver fatigue. The Driver Fatigue is recognized using analysis of the facial expression in (Zhang and Hua, 2015). In order to have a reasonable choice in decision making problems, different options are considered and compared together from different points of view. Associated with each option, there is emotional response that predicts an aversive or rewarding consequence of that optional choice. How much the emotional response associated to the optional choice is felt positive is considered as the value of that option. This emotional-based value plays an important role in decision making (an and Umair, 2015).

In contrast to real learning environment, the e-learning environments usually lack the advantages of face-to-face interactions between learner and instructor. So, developers try to provide cognition based interactive e-learning environments with considering individual differences, emotions, and personality. These kinds of interactive user interfaces increase learners' motivation, satisfaction and involvement, and also improve learning (Latham et al., 2012). Personality is also a cognitive factor consists of behavioral tendencies, desires, feelings, and feelings of each person (Hartmann, 2006). Many studies consider both personality and emotions for designing an intelligence system for human-computer interaction. The Bayesian Belief Networks are used to model emotions and personality. They consider friendliness and dominance as two dimensions of the personality for implementation of a character-based user interface (Ball and Breese, 1998). In (Egges et al., 2003), a model of emotions, personality, and mood is introduced to implement a conversational virtual human. In (Mehdi et al., 2004), a model comprised from emotions, personality, and mood factors is developed where mood was considered as a filter to moderate the emotions intensity and personality as a threshold of the emotions appearance. For generation of affective behaviors, a combination of emotions, mood, attitudes, and personality traits is presented in (Moshkina, 2006). In (Santos et al., 2011), a combination of emotions, mood, and personality is introduced to provide an artificial intelligence agent for implementation of a decision-support system.



(a) Kismet

(b) Jibo



(c) MOUE

Fig. 8. Examples of emotion recognition by artificial systems (Perez-Gaspar et al., 2016).

3.2. Online learning environments

The fast growth of internet technology have been affected all aspects of human life. Everyone and everything is accessed online. The field of education is also not isolated from involving to the internet. Computer based learning, on-line learning or electronic learning are referred as e-Learning. The substitution of traditional education systems by e-Learning provides several benefits such as improving performance and decreasing costs. E-Learning can be done synchronous or asynchronous (al-shalchi, 2009). Both methods have their advantages and disadvantages. While in asynchronous e-Learning, learners begin and complete their courses each time that is preferred for them, synchronous e-Learning provides a real-time interaction for learners. Several common features of asynchronous and synchronous e-Learning are represented in the following (Hrastinski, 2008; Consulting, 2013). Asynchronous e-Learning are previously recorded or pre-produced, intermittent on-demand access, independent learning, self-paced and individual or poorly collaborative. Synchronous e-learning are live, real-time, scheduled, concurrent learning and collaborative. The conducting ways of asynchronous e-learning are self-paced courses, web-based training, podcasting, computer aided system, discussion groups and message boards. The conducting ways of synchronous e-learning are virtual classrooms, audio and video conferencing, on-line chat, shared whiteboard, application sharing and instant messaging. A synchronous e-Learning is like a virtual classroom to share knowledge among the learners. Different teaching methods such as shared whiteboard, video, slide presentation and application sharing can be chosen by the instructor. The students and the instructor can use instant messaging, voice and chat to communicate together.

The human users in virtual classes with synchronous e-learning have an Avatar to represent them where the Avatar acts like a puppet. An Avatar is an electronic image manipulated by a computer user to represent him/her. The real user facial expressions are represented through the face of chosen Avatar by using the mirror approach (Chapman, 2013).

The online learning environments and traditional (face-to-face) classrooms have similarities and differences. The most of researchers (instructors) have tried to maximize the similarities between face-to-face and online courses by providing similar syllabi, providing the same access to the teacher via message or email, and posting the lecture notes. In addition, the video technologies cause more similarity to face-to-face classroom. Based on these similarities, it can be seen that emotions elicited in online environments are similar to face-to-face classrooms. However, because of the certain characteristics of online environments, which are different from face-to-face classrooms, the students' appraisals of control and value and therefore the experienced emotions may be influenced. The varied characteristics of traditional and online learning environments are discussed in terms of control-value theory in (Daniels and Stupnisky, 2012). Emergence of some emotions in traditional classrooms is more than online environments and vice versa. For an example in the face-to-face classrooms, the researches show that pleasant emotions are usually endorsed more than unpleasant (Daniels et al., 2009). In contrast, in the online learning environment, almost all major human emotions are reported by the participants and also the description of positive emotions is as much as negative emotions.

Noting to this point is worthwhile that pleasant emotions should not be necessarily inferred good for students. In duality, unpleasant emotions should not be necessarily inferred bad. For example, according to results of some studies, enjoyment is not always due to achievement (Pekrun et al., 2006). Alternatively, according to report of students, the negative emotions like boredom can have helpful outcomes such as providing time for reflection, relaxation, or thought (Harris, 2000). For another example, the emotion of guilt can move learners into action.

In several studies, it has been shown that learning can be facilitated through simulations or virtual worlds and the performance of learning is increased compared to traditional forms of learning processes (Ahmad and Montazer, 2009). The virtual learning environments can facilitate meaningful and deep learning of team-work, empathy, self-confidence and active problem-solving skills (Devlin et al., 2012). Such platforms help to build trust and increase a sense of belonging. These positive emotions have effective role in increasing the people's leaning capacity and enhance their innovative and creative problem-solving capacities.

The effects of multimedia materials in learner emotions and so, learner performance are investigated in (Chen and Wang, 2011). To this end, the learner emotions are recognized by employing the emWave system. This system developed by the Institute of HeartMath utilizes human pulse physiological signals for recognition of three emotional states: negative, peaceful and positive. The correlation between learning performance and learner emotions when three different types of multimedia learning materials are presented to the learners is studied. The most frequently multimedia learning materials used in modern education are: 1- multimedia materials containing static text and image 2- multimedia materials based on video including moving images with audio, 3- multimedia materials based on animated interaction including text and animated images with interactive features. The most positive emotion and the best learning performance are generated by video-based multimedia material among three types of assessed multimedia materials. Different males and females show significant differences in their emotional states while different multimedia materials are used for learning: female learners are more easily affected than male ones by different multimedia material.

3.3. Effects of emotions in learning

Learning is a cognitive and emotional experience according to the "attention-to-affect" model (Critcher and Ferguson, 2011; Satpute et al., 2013). Emotions have a direct effect on human functions such as learning process. Therefore, it is very important to understand how emotions change the students' learning process. This understanding is significant to design appropriate learning models and consequently improve the learning gains and achievements.

According to (Doulik et al., 2017), the style consists of five strands called stimuli: 1-physiological including auditory, visual, tactile preferences and kinaesthetuc; 2-sociological; 3-environmental; 4-psychological and 5-emotional. The mentioned factors impact how much a person learns. The human behavior characteristics such as mood, personality and emotion should be considered and modeled in the e-learning environments (Fatahi et al., 2016). According to the psychological studies, people have many fundamental and individual differences in problem solving, decision making and learning process. The differences in the learning process consisting of information processing, understanding and assessment are called learning styles (Li et al., 2007). In other words, a learning style includes physiological, cognitive and emotional features used to recognize how the learner interacts with the learning elements and environments and understands the concepts (LoganKThomas, 2002).

Emotions and personality have been modeled in many intelligent agents to provide an automated agent for responding to user's interactions. The researcher results have been shown that the learners prefer user interfaces and learning environments designed based on the learning styles. For example, an intelligent agent based learning environment has been proposed in (Fatahi et al., 2010). Two agents are defined in the proposed learning environment: virtual classmate agent (VCA) and virtual teacher agent (VTA). Based on the learning style of the learner, an appropriate teaching style is used by VTA. Also, based on the present situation, a suitable VCA is proposed by VTA to the learner. VCA has its specific learning style. The results show that the interaction of intelligent agents having human behavior factors with the learner improves the satisfaction level of the learners and improves the learning rate. Another emotional intelligent agent has been introduced in (Chaffar and Frasson, 2004) that includes three modules: perception, control and action. At first, the current emotion of learner is recognized by the agent according to the learner choice of a sequence of colors. Then, the optimal emotional state of learning is identified according to the learner's

personality. Identification of learner's personality is done by the personality questionnaire and the Bayes classifier is used for optimum emotion prediction.

In a tradition learning environment, in a classroom, the instructor (teacher) acts as a facilitator between the learning course and the student. The instructor constantly adjusts the instruction (teaching) process to close to the students' goals and needs as much as possible. But, in the online learning environments, teachers have no easy way to analyze the students' emotions and behavior. To deal with this problem, an appropriate solution is to develop mechanisms that use computers with capability of automatically emotions detection of learners and adapt the learning process to close to learners' real needs. The researchers in (Faria et al., 2017) have been described an emotional learning model and have been developed a software prototype. They use the affective computing techniques to perform adjustments to the learning processes of students. The proposed emotion test platform proposed takes into account the learner's emotional state, learning preferences, and personality traits. The entire teaching process from theoretical learning to exercises, test, and assessment is simulated by the developed emotion test platform. Four main models are used to form this prototype (Faria et al., 2015): 1- the student model, 2- the emotional model, 3- the application model and 4- the emotive pedagogical model. When a learner feels an emotion that needs to be repressed (such as disgust, sadness, anger, and confusion), the emotional interaction mechanism is trigged to deal with undesired expression and facilitate the learning process. The interaction depends on the learner's learning style and personality.

The Broaden-and-Build theory represents that some positive emotions such as interest, love, contentment, and joy can widen individuals' awareness and prompt exploratory and also novel thoughts, ideas, and actions (Fredrickson, 1998; Fredrickson, 2001). In addition, according to this theory, the positive emotions lead to wellbeing and human flourishing. This theory argued that positive emotions can widen learners' scope of cognition, attention, action, and tend and also increase the activity engagement. In contrast, negative emotions narrow the learners' capacity for learning and reduce the activity engagement. The role of Broaden-and-Build theory to facilitate the second language learning is assessed in (Ali and Askari Bigdeli, 2014). The study results show that positive emotions have significant effects on language learners' motivation and learning.

Emotions have a dominant role in different aspects and activities of human. They have an important influence on individuals' beliefs, assessments judgment and decision-making (Gratch and Marsella, 2004). Recently, an emotional dimension has been suggested to individual perceptions (Stein et al., 2015). Authors in (Darban and Polites, 2016) have assessed that how the perception of learners about the radicalness of a new technology relates the emotion of individuals to their willingness to learn that technology. With improving the educators' understandings, the opportunity of students to learn the new technology is increased. Radicalness mediates the relationship between willingness to learn and individuals' emotions. Four emotions: deterrence, challenge, loss, and achievement are considered. The obtained results in the simulated learning environment show that anxiety is positively associated with perceived radicalness while anger and excitement are negatively associated with perceived radicalness. The positive effect of happiness is insignificant. Finally, the effects of negative emotions are greater than positive emotions on radicalness perceptions.

The effects of emotions in the second language learning are assessed in (López and Cárdenas, 2014). According to obtained results, motivation levels, the social context where learning occurs, and self-regulation mediate the effect of emotions. As another result, it is seen that some learners have capability of turning negative feelings into motivational energy. The study in (Zhou, 2016) represents that the performance of the second language learning in addition to learners' motivation and emotion is dependent to the classroom learning orientation. The results show that learners who experienced an emotion of social anxiety in language learning, in particular, having a fear for speaking in public environments, feel less autonomous, have weaker collaborative learning, and obtain less success in their language learning. The role of emotions in the second language acquisition is also assessed in (Lockwood, 2015).

One of the important factors in success of online learning is self-regulated learning. The learners' academic emotions and perceived academic control are the main antecedents of self-regulated learning. Because of interrelating cognition and emotions, to better understand the self-regulated learning process, it is valuable that the joint relationship between academic emotions and perceived academic control on self-regulated learning is investigated. The importance of academic emotions in the relationship between self-regulated learning and perceived academic control in online learning is examined in (You and Kang, 2014). The effects of social interaction and emotions in strengthening the learning process are conceptually and computationally discussed in (Jan and Arlette van Wissen, 2013).

The presence of a teacher in the learning environment not only assists conducting but also helps student to learn the taught contents (Lazarus, 2001). The teacher's behavioral cues such as eye contact, facial expression, postures and gestures convey important and serious communicative contents and messages especially where speech is not available; and so, they have significant influence on student performance. The role of emotional design of multimedia learning environment is discussed in (HeidigJulia Müller and Reichelt, 2015). In order to investigate the emotional design features, some concepts of web design are taken account. The learners were assigned into different conditions provided by two design factors (expressive aesthetics vs. classical), each with two levels (low vs. high), usability factor (low usability vs. high), and control group (gray scale/no color). The results indicated that the learners' emotional states are positively affected by perceived aesthetics and usability but not by objective differences in aesthetics or usability. Moreover, the results showed that the emotional states of learners have not significant impact on the learning outcomes but have a large effect on the learners' motivation to continue working with the designed environment.

As a result from the above discussions, there is a high correlated relation between the ability of learning and the emotions of learner. Our main focus in this paper is the role of emotions in the learning process.

4. Part II. Emotion recognition methods

The emotion recognition methods using computers are generally divided into seven groups:

- Asking from user
- Tracking implicit parameters
- Voice recognition
- Facial expression recognition
- Vital signals
- Gesture recognition
- Hybrid methods

Each of them has its advantages and disadvantages. For each of them there are some methods that are discussed in the following sections. In this paper we have explained each of these research areas and their benefits and difficulties for using them in the e-learning context.

5. Asking from the user

Asking from the user is the simplest one. In this way, system asks questions about user's emotions from himself. For example, a question can be as follow:

Question: how are you?

Answers: 1) happy 2) sad 3) bored 4) angry.

This method is absolutely unusable for some aims such as lie recognition because he/she attempts to cover his/her actual emotions. This method is simply applicable in e-learning systems (Kardan and Einavypour, 2008a). Although this method is very simple but it has its

disadvantages for e-learning systems too. For example, asking questions that are not related to learning content may be annoying for learners. Furthermore the emotion is a temporary state. The learner may be happy at first, but answering to difficult questions may make him/her sad and bored. Thus, emotions must be asked continuously during the learning process. Some recognition methods using asking the user are briefed in Table 2.

The self-report methods simply ask individuals to describe their nature and emotions. The basic assumption in these methods is that participants are willing and able to recognize their emotions and report them (Lopatovska and Arapakis, 2011b). There is a high correlation between the self-report results and the neurological activities evolved from brain. Although the use of self-report method is an easy and efficient technique, it is biased by participants. The self-report methods are generally divided based on two major discrete and continues approaches in emotion theories described in the previous section. In self-reports with the discrete approach, discrete emotion terms such as sadness, happiness and boredom are provided for participant individuals. The participants determine which affective term better describes their emotional experience. The use of discrete emotion terms has several disadvantages. The participants may be biased by some pre-determined emotional responses. The associated respond of participant may not be provided on the response list, participant may be unfamiliar with some terms determined by a researcher. In addition, the results obtained by different researches that employed different lists of emotion labels may be not comparable (Scherer, 2005), (Klein et al., 1999; Scheirer et al., 2002).

The self-report method based on continues emotion theory is established in (Sander et al., 2005b). The proposed method for description of emotions is the use of a three-dimensional space formed by the tension (tense-relaxed), arousal (excited-calm), and valence (negative-positive). In this approach, the participants report their emotional experience by simply choosing associated coordinates in the three-dimensional emotional space. Since identifying a third dimension from excitation or arousal is difficult, often only two of three dimensions (arousal and valence) are used to form adimensional emotional space. This simple and straightforward approach provides interval data that is useful for statistical processing. However, this method has also some disadvantages. There is limitation in degrees of negative or positive valence or arousal. Moreover, the simple perception and the intuitiveness of the discrete basic emotions are lost in this approach (Partala and Surakka, 2004; Peter and Herbon, 2006). The free response report is used to degrade the disadvantages of self-report methods where participants are allowed to represent their emotions using their expressions with each associated words or terms. Although the statistical analyze of free-response data is difficult, but because it provides a high level of personality specification, this technique is an appropriate choice where accuracy is important (Scherer, 2005), (Lopatovska and Arapakis, 2011b).

Interviews, journals, and think-aloud protocols are some other popular techniques for data collection in library and information science research (Bilal and Kirby, 2002). In (Kuhlthau, 1991), in order to collect the emotion data, the participants are asked to record their thoughts and feelings during information searching in a journal. Think-aloud methods have been used in several library and information science researches by recording their think-aloud reports. The screen logging software is used to record the interactions of users with the study monitor and their search activities (Nahl and Tenopir, 1996). The participants are asked to think aloud when interacting with system (Wang and Soergel, 1998; Tenopir et al., 2008). Interview is another popular approach of studying emotion in library and information science researches. Interviews with individuals are done before and after individuals' engagement in a seek activity. Both the one-on-one interviews and group interviews can be used to collect emotion data. In an interview, participants must recall specific experimental assignments and represents their thoughts and emotions. Questionnaire is other technique of self-report methods. The participants are asked to complete a questionnaire that is designed for collection of emotional states of participants (Mentis, 2007; Lopatovska, 2009). Achievement Emotions Questionnaire (AEQ) is used for measurement of emotions in many identified studies where most of them conceptualize emotions using both dimensions of activation (activating/deactivating) and valence (pleasant/unpleasant) (Pekrun et al., 2011). Because the reliability for the scales have been adequate in the reported studies, AEQ can be used an appropriate measurement tool.

Table 2
Recognition methods by asking from the user.

Method	References
Self-report methods- discrete approaches	(Klein et al., 1999; Sander et al., 2005b)
Self-report methods- continues approaches	(Partala and Surakka, 2004; Peter and Herbon, 2006), (Kuhlthau, 1991)
Self-report methods- free response report	(Scherer, 2005), (Lopatovska and Arapakis, 2011b)
Self-report methods- questionnaire	(Kuhlthau, 1991), (Mentis, 2007; Lopatovska, 2009), (Pekrun et al., 2011)
Interviews	(Bilal and Bachir, 2007), (Bilal and Kirby, 2002)
Journals	(Bilal and Kirby, 2002; Kuhlthau, 1991)
Think-aloud protocols	(Nahl and Tenopir, 1996; Tenopir et al., 2008)
Outside-observers- assessments	(Truong et al., 2012; Truong et al., 2008)

In (Truong et al., 2012), the differences between self-annotations and observer-annotations are explored. The emotion recognizers trained on these obtained annotations differ from each other. According to obtained results, recognition of the self-reported emotion is much harder than the observed emotion. As another result, the averaging ratings obtained from multiple observers increase the recognizer performance. This hypothesis that individuals can better recognize their own emotions are investigated in several researches. The people were asked to represent and label what they felt compared to what they expressed. The comparison between self-assessments of emotions and outside-observers- assessments showed a mismatch between the perception and expression of emotion (Busso and Narayanan, 2008; Truong et al., 2008).

The objective tests and self-report questionnaires are two types of assessment methods in the subject of emotional intelligence. They correspond to ability and mixed emotional intelligence models (Mayer et al., 2000). Similar to other types of intelligence such as spatial or verbal, emotional intelligence is considered as a set of cognitive competencies and abilities in ability models. A broad definition of emotional intelligence as an array of motivational traits, personality, and cognitive is considered in mixed models, sometimes called trait models. While the proponents of ability models utilize the objective tests with right or wrong as answers, the proponents of mixed models use the self-report questionnaires like personality inventories. Because of known limitations of self-report approach (Lyusin and Ovsyannikova, 2016), focuses on objective tests. In other words, the emotion recognition ability is evaluated independent of beliefs about individual's behavior and self-concept. One of the hardest problems is scoring because there is not any logical foundation to establish the correct answers. Three major approaches for scoring are expert, consensus, and target scoring. Expert scoring uses the opinions of experts about the correct answer or the best choice among candidate answers (Legree et al., 2005). Consensus scoring uses the opinion of the majority of the participant individuals about the correct answers. Target scoring considers the opinion of the target person who creates the stimuli.

6. Tracking implicit parameters

Generally in emotion recognition methods using implicit parameters, the learner's activities are observed. System can make inferences about learner's emotional state by tracking implicit parameters such as his/her number of mistakes, time of answering questions, his/her written sentences and so on. Although this method is not very precise, but because it has no need to additional means, it is simply applicable for e-learning applications. Some of recognition methods by tracking implicit

Table 3
Recognition methods by tracking implicit parameters.

Method	References
mean time of solving of problems and the number of mistakes	Kardan and Einavypour (2008b)
scrolls up, left mouse clicks	Lopatovska (2009a)
Interactive behaviors in log files	Kapoor et al. (2007)
Elapsed time on the result search page and the number of visited search pages	Fox et al. (2005)
Keystroke data	(Hernandez et al., 2014; Syed et al., 2014)
The behavior of users with touch-screen advices	(Johnson, 1965; Syed et al., 2014)
Short text- a multi-label maximum entropy model	Li et al. (2016a)
Short text- topic-level maximum entropy model	Bao et al. (2012)
Text- HMM	Quan and Ren (2010)
Text- mutual information	yasmina et al. (2016)
Text- Semi-supervised (based on adaptive multi-view selection)	Yang et al. (2014)
Text- ensemble classifier	Perikos and Hatzilygeroudis (2016)
Text- sentiment of topics	(Bao et al., 2012; Mano et al., 2016)
Text- hidden de-noising classification model	Wang et al. (2016)

parameters are represented in Table 3. In (Kardan and Einavypour, 2008b) researchers have used two implicit parameters: mean time of solving of problems and the number of mistakes. They have introduced Agility Factor for comparing it in different sessions. Then, they have made deductions about learners' emotional state (positive or negative) by comparing the sessions' agility factor.

It was found that the specific behaviors during search in web environment, such as wheel scrolls up or left mouse clicks are cues and patterns of emotional states preceding or following the behaviors of person (Lopatovska, 2009a). The human–computer interaction researches can infer emotions from interactive behaviors captured in log files of computers. In (Kapoor et al., 2007), frustration is predicted using log data. There is a correlation between certain behaviors of user during search, such as elapsed time on the result search page and the number of visited search pages with satisfaction and dissatisfaction of users (Fox et al., 2005). The relation between happy expressions of users, the number of search activities, certain clicks, and elapsed time in a search session is also shown in (Lopatovska, 2009a).

The behavior of users with touch-screen advices can be used for recognition of their emotions. Touch-screen technology based human-computer interaction firstly introduced in 1965 (Johnson, 1965). Recently, with rapid growth of mobile technologies such as smart phone, touch as an interface methodology is available anywhere. Similarly, the mouse and keystroke data have been used for prediction of emotional states such as stress (Hernandez et al., 2014). The keystroke data of users' profiles are analyzed in (Syed et al., 2014). This information is used for prediction of whether the user is male or female, uses one or two hands to type, is right- or left-handed, and age category. Two consecutive keystrokes and SVM classifiers are used to this end. Soft-biometrics use the latest generation of touch-screen technologies which utilize the touch input data for prediction of user information. Swipe gestures can be used as a more secured means for information recognition of users. The swipe gesture data combined with accelerometer data is used for user identification in (Bo et al., 2013). Three gestures are considered for this purpose: Scroll, Tap and Fling. A touch-based feature set consists of touch pressure, touch coordinates on the screen, and duration across three applications are used. The Finger-gestures Authentication System using Touch screen (FAST) is also proposed in (Feng et al., 2012), which uses six swipe gestures: zoom out, zoom in, right-to-left swipe, left-to-right swipe, up-to down swipe and down-to-up swipe. In (Miguel-Hurtado et al., 2016), the user's sex is predicted using swipe gesture data taken from a smart phone. 14 features are extracted from coordinate position, pressure, speed, acceleration, arc distance and thickness. The BestFirst (Dash and Liu, 1997) is used for Feature selection. Decision trees, naïve Bayes, support vector machine (SVM), and logistic regression are used for classification. Although this method is proposed for sex prediction, it can be used for prediction of single or-two handed usage, age category, handedness or even emotion recognition.

Some researchers have proposed recognizing emotions from text in e-learning environments. It means that system can investigate the learner's written text and then it can deduce the emotional state of learner using his/her utilized words. Many individuals express their emotions in online social media. Emotion recognition, i.e., emotion classification, of social media reveals the preferences and opinions of the general public. Most of the suggested social emotion recognition methods use the long documents. Since the short text is vastly prevalent on the Web, a multi-label maximum entropy model for emotion recognition of users, i.e., writers of short texts, has been proposed in (Li et al., 2016a). The rich features are generated by the modeled multiple emotional labels and valence scale of 0–100 for six emotion labels (surprise, sad, joy, fear, disgust, and anger) scored by many users jointly. A maximum entropy model has been also proposed for classification of social emotions embedded in users' comments (short texts) at the social web (Rao et al., 2014).

The textual emotion recognition methods generally are divided into three main groups: machine learning based methods (Mishne, 2005; Xia et al., 2011), conceptual based methods (Grassi et al., 2011; Olsher, 2012), and emotional keywords potting (Chunling et al., 2005; Hancock et al., 2007). Hidden Markov model (HMM), which is an efficient statistical framework to solve the time series problems, has been used to model the typical time-ordering contained in sentence emotions. The statistical features of emotion transfer are revealed from a blog article in (Quan and Ren, 2010). The analysis of emotion transfer in adjacent contexts of blogs reveals the emotion continuity.

The effects of preprocessing, classifier type and ensemble methods in analysis of emotions over twitter texts are evaluated in (Troussas et al., 2019). The conclusions show the positive effect of feature selection, superiority of SVM and Naïve Bayes classifiers and improvement of sentiment analysis by utilizing the ensemble of multiple classifiers. As a case study, the application of sentiment analysis is studied in an e-learning context for enhancement of adaptivity in the learning process. Improvement of delivering recommendations about student activities and providing personalized assistance are two main benefits of using emotion analysis in the e-learning system. Different classifiers and datasets are compared for sentiment analysis of public messages in twitter using confusion matrix in (Krouska et al., 2017). The results showed that the performance of the sentiment analysis is not dependent to the used dataset.

The emotion of YouTube comments are also recognized based on their textual exchanges in (yasmina et al., 2016). A text entry is classified into a particular emotion class by computing its similarity to each of target emotions. To this end, the point wise mutual information measure is used. To classify a text entry into a particular emotion category, its similarity to each target emotion is computed using the point wise mutual information measure. The six basic emotions of Ekman are used for emotion classification. A list of emotional words is used to represent the each emotion class. For determination of emotion expressed in a sentence of text, at first its component words are classified. The nouns, adverbs, adjectives, and verbs are extracted. The other words such as interjections, prepositions, and pronouns are neglected because of their neutral substance. The probability of belong of each word to each emotion class is computed. This probability is the normalized value of Point wise Mutual Information between the corresponding word and the representative words of the each emotion class. To obtain the probability of whole sentence, the achieved probabilities of the classified words are averaged.

A semi-supervised text emotion recognition method based on adaptive multi-view selection is proposed in (Yang et al., 2014). Typically, the quality of the training samples is degraded by some mislabeled samples. The adaptive multi-view selection (AMVS) method, proposed in (Yang

et al., 2014), introduces two distributions for construction of multiple discriminative feature views: distribution of feature emotional strengths and distribution of view dimensionality. This work is able to improve the accuracy of labeling process of unlabeled samples.

An ensemble classifier is developed for emotion recognition from text through analyzing the natural language in (Perikos and Hatzilygeroudis, 2016). This analysis is conducted at sentence level. So, each given document must be firstly split in sentences. The used ensemble is based on three classifiers: two statistical-based and one knowledge-based. The ensemble framework integrates the results of three classifiers using a majority voting rule. The used statistical classifiers consists of a naïve Bayes and a maximum entropy learner which are trained using Affective text datasets and the International Survey on Emotion Antecedents and Reaction (ISEAR). The Naïve Bayes classifier treats each document as a bag-of-words, with this assumption that the document has no internal structure, and there are no relationships between the words. The knowledge-based tool tries to extract knowledge from each sentence to specify its sentimental status (Perikos et al., 2013). It determines word dependencies and specifies how the connections between words convey emotional content. For detection of words containing emotional content, the lexical resources, such as Word Net Affect (Strapparava and Valitutti, 2004) are used. A Tree Tagger is used to determine the grammatical role of each word in the sentence (Schmid, 1994). Tools such as Stanford parser are utilized to analyze the sentence's structure (de Marneffe et al., 2006).

The tag-based search is personalized in collaborative tagging systems in (Xie et al., 2016). It incorporates available sentiment information, embedding user profiles and resource ones. In (Liang et al., 2018), a universal effective model is used for classification of readers' emotions. The model consists of two sub-models: topic level and term level. By reading negative comments in online social networks, users may not buy the advertised products. The sentiment analysis of online reviewers about services and products in online shopping is important from two main views: help to users to make proper decisions when purchasing and improvement of the quality of services and products (Yang et al., 2019). The social emotion mining methods are classified into two main groups: word-level model (Katz et al., 2007; Lin et al., 2008) where the sentiment of individual words is focused and topic-level models (Bao et al., 2012) where the sentiment of topics is focused. Two sentiment topic models called multi-label supervised topic model and sentiment latent topic model have been proposed in (Rao et al., 2014) which can be applied for two tasks: generation of social emotion lexicons and classification of social emotion.

In social media, the ground truth for emotions and sentiments is often constructed through hashtags, surveys or emoticons, and thus, the existence of error in the labels is possible. To reconcile this noise, a hidden de-noising classification model (HDCM) has been proposed in (Wang et al., 2016) that does not need any lexicons or outsourcing systems for estimation of the actual emotional or sentimental category from data with noisy labels. The HDCM method firstly uses the logistic function to model the likelihoods of visible noisy labels equal to a hidden actual category. Then, the expectation maximization algorithm is used for estimation of the above objective function. Finally, SVM is used for classification of the optimized hidden variables. HDCM can classify both sentiment (binary) and emotion (multi-labels) datasets with noisy labels.

Extraction of relevant emotions from texts is a classical and challenging problem. A sparse feature space can typically characterize the online comments where it makes the emotion classification a hard task. Deep neural networks have been shown high ability in production of dense high level features from the sparse low level features. A hybrid deep neural network is proposed in (Li et al., 2017) that uses the unsupervised teaching models for increase of neural network power in emotion classification.

The emotion recognition methods using text analysis can be generally divided into following categorizes (Shivhare and Khethawat, 2008):

- Keyword Spotting Technique:

This technique is based on certain keywords which are predefined and classified into categories such as angry, surprised, fearful, happy, sad, disgusted, etc. In the Keyword spotting technique, a text document is considered as input and an emotion class is generated as output. At this technique, at first, the text data is transformed into tokens and emotion words are detected and identified from these tokens. Then, the intensity of emotion words is analyzed.

It is checked whether sentence contains negation or not. Finally, an emotion class is assigned as the output.

- Lexical Affinity technique:

This technique, which is an extension version of keyword spotting technique, assigns a probabilistic affinity associated with each particular emotion to arbitrary words not just emotional keywords. As a disadvantage, this method misses out the emotional contents of words.

- Learning-based techniques:

Initially, the problem was emotion detection from the text data, but now the problem is classification of the text data into different emotions. Learning-based techniques recognize emotions using a classifier, which is previously trained by applying various machine learning methods such as conditional random fields and SVM, to assign an appropriate emotion category to the text data.

- Hybrid Methods:

Some of above methods can be used together. For example, the lexical affinity methods and learning-based methods are combined in (Wu et al., 2006) to achieve more precise results.

The most disadvantage of emotion recognition using implicit parameters is its complexity. In some cases, finding a relationship between specific emotional states and implicit parameters seems to be difficult. These parameters may be numerable. Another problem is its low preciseness for emotion recognition. By this method we can guess that the emotional state is positive or negative but it is difficult to determining emotional states like stress or sadness.

7. Voice recognition

The verbal communication is one of the observer methods to study emotions emerged in the early 20th century (Mendoza and Carballo, 1999). There are some acoustic cues leads to emotion recognition. The most often auditory features extracted from the voice signal are pitch (frequency of acoustic signal), intensity (vocal energy), speech rate (number of words spoken in a time interval), pitch contour (geometric patterns for description of pitch variations), and phonetic features (consonants, and pronunciation) (Pantic and Rothkrantz, 2003).

Speech, which has an important role in the human communication, exhibits the cognitive states, intentions, and emotions of individuals (MorrisonLiyanage and De Silva, 2007), (Sadoughi and Busso, 2019). A Speaker Emotion Recognition (SER) system consists of two main parts: speech feature extraction and classification. A standard SER system that consists of acoustic feature extraction and emotion identification is shown in Fig. 9. Table 4 shows some of emotion recognition methods using voice signals.

According to how to construct an acoustic emotion model, the speech emotion recognition methods can be categorized into three groups: speaker-independent (SI), speaker-dependent (SD), and speaker-adapted (SA) models. The speaker-independent approach uses the training samples acquired from a number of specific speakers who are not relevant to real users. Speaker-independent have simplicity and efficiency in common applications. But, because of acoustic characteristics, which are

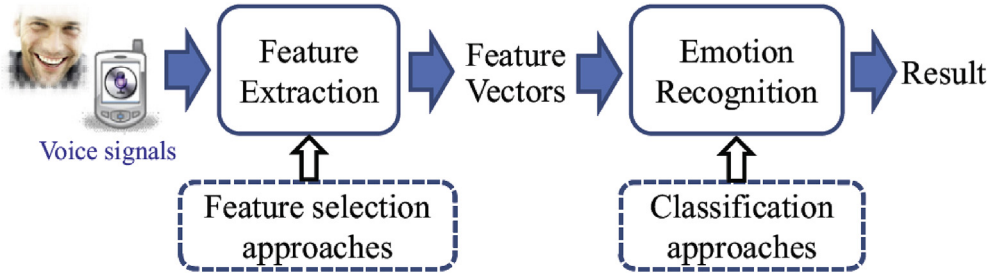


Fig. 9. A standard speech emotion recognition system (Kim and Park, 2016).

unmatched between speakers in real users and training data, cannot guarantee a stable performance. The speaker-dependent approach uses the system's user data to build the acoustic model. So, it can deal with the variations between speakers. Nevertheless, because collection of a sufficient emotional data from individual users is difficult, are significant limitations for using this approach in commercial applications. The speaker-adapted approach is a model transformed from speaker-independent model associate with speaker adaptation procedures. Although the adaptation just needs a relatively small set of data, but, the produced user-characterized acoustic model achieves the performance near to the speaker-dependent model (Matsui and Furui, 1998; Choi et al., 2015). A representative speaker adaptation method by utilizing the maximum likelihood linear regression has been proposed in (Kim and Park, 2016). This personalized SER method takes the advantage of personal devices such as smart phones.

The distinct emotions and some other environmental factors may degrade robustness and recognition rate of SER systems. The extraction of salient features from the speech signals can solve these problems (Moataz et al., 2011). A survey of different methods for SER systems is given in (El Ayadi et al., 2011). Four types of speech features are intended for emotion recognition: acoustic features, linguistic features, context information, and hybrid features. Some spectral features of speech signal are tonal power ratio, spectral flux, pitch chroma and Mel Frequency Cepstral Coefficient (MFCC). Tonal power ratio, which is calculated by ratio of tonal power of the spectrum components and overall power, measures tonalness of the speech signal (Alexander Lerch, 2012). The spectral flux extracts the spectral components of speech signal (Subramanian, 2004). By changing the spectral content of the speech signal over the time, the recognition performance is degraded. So, the spectral components are important. The squared difference between the normalized magnitudes of consecutive spectral distributions of signal associated with consecutive frames of that is calculated to find out the input speech signal. A powerful method for categorizing the pitches and approximating to equal tempered pitch scale is extraction of the pitch chroma. It reduces the noise and transients present in the signals. For extraction of features using pitch chroma, after calculation of Fourier transform, the frequency transform is mapped with the 12 semi-tones pitch classes (Peeters, 2006). MFCC, which provides the perceptual frequency bands for the cepstral analysis, extracts significant features from the speech signal (On et al., 2006). MFCC, which is based on the human ear's frequency bandwidth, utilizes a non-linear Mel scale frequency to emulate the auditory system of human. To obtain the MFCC coefficients, the triangular filters linearly spaced on the Mel frequency scale are applied to the speech signal. Then, a discrete cosine transform is applied to the output log energies of filter banks. The MFCC and HMM are used for feature extraction and classification of speech signal, respectively in (Nanavare and Jagtap, 2015). A two level hierarchical ensemble of classifiers is proposed in (Vasuki and Aravindan, 2012) for fusion of speech data in decision-level. The MFCCs of speech signal are classified independently by Gaussian Mixer Model (GMM) and SVM classifiers. Then, discriminate function values of SVM and the posterior probabilities of GMM are fed to the second level SVM classifier for emotion recognition. A speech emotion recognition method has been proposed in (Mannepalli et al., 2017) which uses four spectral features of speech signal. Four features extracted from speech signal are concatenated into a single vector as $F = \{T; P; C; M\}$ where1×1 vectors of T, P, and C denote tonal power ratio, spectral flux, and chroma, respectively and $1 \times m$ vector of M contains MFCC components. The extracted features are fed into the classifier for emotion recognition.

In (Chenchah and Lachiri, 2017), some different methods are used for feature extraction, namely, MFCC coefficients, Perceptual Linear Prediction (PLP) coefficients (Hermansky, 1990) and some simple variants of Power Normalized Cepstral Coefficients (PNCC) (Kim and Stern, 2016). PLP coefficients are derived similar to MFCC. The difference between them is due to filter-banks, the application of linear prediction, the intensity-to-loudness conversion, and the equal-loudness pre-emphasis. PLP, which uses techniques similar to human auditory system, applies linear prediction to the hearing spectrum to achieve predictor coefficient. To increase the robustness of speech recognition systems under noisy environment, several feature extraction methods have been proposed in literature (Ogunfunmi et al., 2015). PNCC, which uses relevant physiological phenomena, is an alternative to MFCC. PNCC features are robust against reverberations and noise. Some differences of PNCC with MFCC are represented in the following. PNCC replaces triangular filters by Gammatone auditory filters to simulate the behavior of the cochlea. PNCC subtracts the medium-duration power bias and replaces the nonlinear logarithm of MFCC by nonlinear power function with exponent 1/15. Four different simplified versions of PNCC are used in (Chenchah and Lachiri, 2017). Moreover, HMM is used as classifier for speech emotion recognition. The HMM classifiers are widely used for speech emotion recognition. This is due to strength of HMM to implement segmentation and classification within an integrated formalism. HMM is a stochastic process that contains a first order Markov chain with hidden states. Each state, which is associated with a random process, generates the sequence of observations. The temporal structure of the speech data are represented by hidden states. A HMM model is determined by five characteristics: 1- number of hidden states, 2- number of observation symbols per state, 3- initial state probability distribution, 4- observation symbol probability distribution in each state, and 5- state transition probability distribution.

A hierarchical binary decision tree classifier motivated by the appraisal theory of emotions, applied to acoustic features, is proposed for speech emotion recognition in (Lee et al., 2011). According to the appraisal theory, emotion perception is a multi-stage unconscious and conscious process. In an appraisal process, several decisions may be thought such as how novel is the stimulus, how positive is the stimulus, and what is the cause of the stimulus. A human person, at each stage, appraises the situation, reacts to the situation, and reappraises. The proposed hierarchical binary decision tree in (Lee et al., 2011), based on the idea of appraisal and reappraisal processes in human decision making, splits the single multi-class emotion classification problem into several binary emotion classification problems. In the hierarchical decision tree classifier, the most distinguishable classes are early recognized in tree. Recognition of ambiguous emotional classes is done at the bottom of the tree. This process degrades error propagation. Different types of binary classifiers such as SVM, logistic regression, and Fisher

Table 4
Emotion recognition methods by voice recognition.

Method		References
Features	Classifier	
MFCC	HMM	Nanavare and Jagtap (2015)
MFCC	A two level hierarchical ensemble of classifiers with GMM and SVM classifiers	Vasuki and Aravindan (2012)
Tonal power ratio, spectral flux, chroma and MFCC	Adaptive Fractional Deep Belief Network (AFDBN)	Mannepalli et al., 2017
MFCC, PLP and PNCC	HMM	Chenchah and Lachiri (2017)
Acoustic features	Binary decision tree classifier consists of SVM, logistic regression, and Fisher discriminant analysis	Lee et al. (2011)
Acoustic features: spectral features (LPC, MFCC, wavelet, wavelet, PLP) and prosodic features (average value of pitches, pitch variability, intensity, pause duration, speech rate, and voice quality)	Ranking SVM	Cao et al. (2015)
Acoustic information, contextual information and elapsed time of spoken utterances	SVM, ANN, KNN	Chakraborty et al. (2016)
MFCCs, short time energy and pitch	ANN, SVM	Rajisha et al. (2016)
Acoustic parameters (spectral, prosody and sound quality parameters) such as voice speed, MFCC and pitch	LS-SVR	Dai et al. (2015a)
RASTA-PLP acoustic features	Probabilistic echo state network (π-ESN)	Trentin et al. (2015)
Acoustic features (spectral and prosodic features)	SVM	Mariooryad and Busso (2014)
Pitch contour, amplitude and bandwidth of formants extracted by LPC	LDA	Mencattini et al. (2014)
Long-term modulation spectral features (MSFs)	SVM	Wu et al., 2011
Phonetic interpretation of concept of syllable	SVR	Origlia et al. (2014)
AR parameters of different orders	An ensemble of classifiers (KNN, ANN, SVM, GMM and discriminant classifiers)	Milton and Tamil Selvi (2014)
Entropy and wavelet packet energy features applied to ERB, Bark and Mel scale	HMM	Chenchah and Lachiri (2014)
Prosodic features (TEO, ZCR, log-energy and pitch) and spectral features (BDPCA + LDA + MFCC)	RBF neural network	Chien et al. (2014)
Weighted spectral features based on Local Hu moments	SVM	Sun et al. (2015)
The instantaneous features, the first and second derivative and the statistic of the instantaneous features + Fisher discriminator and PCA	SVM, ANN	ChenXia and XueLee (2012)
Pitch related features, Bark Band Energy and spectral features	GMM	Clavel et al. (2008)

discriminant analysis can be used as binary classifiers in the hierarchical tree structure. In real interactions, the emotion of neutral is the most ambiguous and dominant emotional class.

Almost all speech emotion recognition methods extract acoustic features from the speech signal and then train a classifier using these representations to able determine emotion of a new utterance. So far, a variety of pattern recognition methods have been proposed to this end such as HMM (Palaz et al., 2019; Meng et al., 2007), neural network (Nicholson et al., 2000a), Gaussian mixture models (Luengo et al., 2005; Vondra and Vich, 2009), regression (Grimm et al., 2007), and SVM (Sun et al., 2019; Wang et al., 2008). All of these methods are designed for emotion prediction of a single test utterance in isolation while in real applications, emotion recognition task is usually needed to do on a complete recorded conversations. For example, in recordings of broadcast and telephone conversations, speeches, meetings, and political debates, the demand is emotion recognition of parts where feelings and affects are expressed. In these cases multiple utterances from the same speaker are available and the emotion detection system should utilize this information. In such cases, a classification score is assigned to each test utterance without consideration of beneficial information contained in other utterances. Deciding for an utterance expresses is easier than deciding on a larger set of utterances which conveys a variety of emotions. All utterances are ranked according to their degree in conveying a particular emotion. All possible emotions expressed in the utterance are considered where each utterance is characterized by computing its distance from the hyperplane for several binary emotion classifiers. Therefore, in ranking approaches such as ranking SVM (Joachims, 2002), there is the same speaker in both testing and training. So, with beneficial of speaker information incorporation, the precision of emotion recognition is increased. A comprehensive set of standard acoustic features together a ranking SVM are used for speech emotion recognition in (Cao et al., 2015). It introduces a novel learning approach for ranking SVM and uses the common acoustic features including spectral features and prosodic features. Spectral features characterize short segments of the utterance. The most popular spectral features are Linear Prediction Code (LPC) coefficients (Nicholson et al., 2000b), MFCC coefficients, wavelet features (Neiberg et al., 2006), and PLP coefficients. Prosodic features are usually computed for the entire utterance and provide global characteristics of speech signal. They usually include average value of pitches, pitch variability in given parts of the utterance, intensity and changes in intensity, pause duration, speech rate, and voice quality (Fernandez and Picard, 2003; Busso et al., 2011). The openSMILE feature extraction library can extract most of these features (Eyben et al., 2010). The most acoustic features are divided into two groups: prosodic and spectral. There are the important emotional cues of speech signal in prosodic features. Even in cases that there is not knowledge about the best discriminative emotional features, the prosodic features are the most commonly choices for SER systems. The spectral features including cepstral features contain the complementary information to prosodic ones. The spectral features contain the frequency content of the speech signal. Extraction of spectral features is usually done over a short frame for example in a 20–30 ms duration. The longer temporal information is incorporated through obtaining the local derivatives. The short-term spectral features such as MFCCs have some limitations. Even by incorporation of local derivatives, short-term characteristics of features are obtained and important temporal behavior information is omitted. In other words, the short-term spectral features discard the long-term temporal cues perceived by human listeners. To deal with these shortcomings, long-term modulation spectral features (MSFs) is proposed in (Wu et al., 2011) for emotion recognition. These features are obtained by frequency analysis of amplitude modulations, i.e., temporal envelopes of several acoustic frequency bins by using a modulation filter bank and an auditory filter bank for speech analysis. Therefore, they contain both temporal and spectral characteristics of the speech signal. The proposed features of (Wu et al., 2011) are used for recognition of both discrete and continuous emotions by using SVM classifier.

The all-pole transfer function of a system producing the observed signal is estimated using Autoregressive (AR) modeling where the denominator coefficients of it are LPCs. In many cases, the spectra, formants, pitch and vocal tract area function can be estimated using the AR model of speech signal. The speech producing system from the vocal cords to the lips can be accurately modeled as an all-pole linear system. The voiced, fricative and nasal of speech signals can be represented by

the all-pole model if model has poles with a high enough number (Rabiner and Schafer, 2004). The results of emotion recognition from speech signal using features of the AR parameters show that the use of a specific set of the AR features, coupled with a specific classifier, can accurately recognize a specific emotion. According to this result, a classification method is proposed in (Milton and Tamil Selvi, 2014), which utilizes the potential of the specific emotion recognition using a specific feature set of AR parameters with a specific classifier.

To improve the performance of spontaneous speech emotion recognition system, two main categories of information are used in addition to acoustic information in (Chakraborty et al., 2016): contextual information and elapsed time of spoken utterances. SVM, artificial neural network (ANN), and K nearest neighbor (KNN) are used in the experiments. There is significant disagreement among human individuals when they annotate spontaneous utterances. This disagreement is largely degraded when additional knowledge related to the conversation, i.e., contextual information is provided for them. The contexts are derived from linguistic contents. Moreover, time lapse of the utterances in the context of a spontaneous audio conversation improves the current emotion recognition.

Most speech emotion recognition studies have been focused on acted context (Lopatovska and Arapakis, 2011a), and some others do emotion recognition in spontaneous context. Because of ideal recording conditions and high ambiguous ground truth labels, emotion recognition from spontaneous speech is a challenging task. In the acted databases, many recording conditions can be systematically and carefully controlled. The differences between spontaneous and acted emotional speech is discussed in (Chenchah and Lachiri, 2014).

The vocal based emotion recognition method proposed in (Dai et al., 2015a) estimates the three dimensional Position-Arousal-Dominance (PAD) space of vocal emotion using Least Squares-Support Vector Regression (LS-SVR) model (Suykens and Vandewalle, 1999). The PAD values of speech signal are estimated as the affective computing results, and finally, the PAD values are expressed as the percentage distributions of the candidate emotions based on their converted Euclidean distances in the PAD space (Sun and Tao, 2008). Generally three categories of acoustic feature parameters consist of spectral parameters, prosody parameters and sound quality parameters determine the affective characteristics on vocal signal. The following 25 parameters from the above categories are chosen for PAD value estimation: number of voice breaks, voice speed, 12 MFCC coefficients, pitch (max, min, mean), short-time zero crossing rate (max, min, mean), short-time energy (max, min, mean), first formant, and second formant.

The probabilistic echo state network (π-ESN) is proposed for emotion recognition from speech signals in (Trentin et al., 2015). A set of 21 acoustic features are extracted and used as input of π-ESN. The π-ESN classifier is a hybrid method constituted from echo state network (ESN) and radial basis function (RBF). While ESNs are a particular subclass of recurrent neural networks (RNN), RBFs are feed-forward networks containing linear combinations of Gaussian kernels.

Voice is one of the richest communicative channels for revealing human emotions. But, this channel in addition to emotions conveys other communicative aspects such as idiosyncratic characteristics of the speaker and the lexical content. These additional factors cause variability in the acoustic signal. To have a robust emotion recognition system, the underlying speaker- and lexical-dependent variability should be compensated. The emotional and lexical dependencies on the frame level acoustic features are quantified in (Mariooryad and Busso, 2014). The proposed metric quantifies the dependency between communication traits (i.e., emotional, speaker, and lexical factors) and acoustic features, which is motivated by the mutual information framework. The variability due to underlying lexical content and speaker dependency are mitigated by using a normalization scheme. The implementation of the compensation schemes are done at feature or model level. Both approaches have their advantages and disadvantages.

Most of standard SER approaches provide low recognition rate, probably because of considering distinct emotional model where emotions be considered as independent affective states. Some recently SER systems assume the dimensional circumplex model of emotions where arousal and valence are predicted in a two-dimensional emotional domain. A SER system which labels corpus in terms of arousal and valence is suggested in (Mencattini et al., 2014).

Although many researches adopted in prosody have indicated the importance of concentration on specific regions of speech for studying the intonation phenomena, several works have shown the importance of syllables for conveying of emotions. A feature extraction method is introduced in (Origlia et al., 2014) that uses the phonetic interpretation of concept of syllable. The feature weighting is done based on syllabic prominence. The method is applied on a three-dimensional continuous emotional model on the axes of activation, valence and dominance. The Support Vector Regressors (SVRs) is used to perform emotion regression.

An audio emotion recognition method is proposed in (Chien et al., 2014) that uses both prosodic and spectral features. The emotion recognition is done in two paths. Path 1 does the intensive analysis of prosodic features such as Teager Energy Operator (TEO), zero-crossing rate (ZCR), log-energy and pitch while path 2 does analysis on spectral features. In path 2 Bi-directional Principle Component Analysis (BDPCA), and LDA are applied to MFCC features and RBF neural network is used for classification. This path contains three parallel BDPCA + LDA + RBF sub-paths where each sub-path handles two emotions. Then, the decision-level fusion is applied to fuse information from both paths 1 and 2.

Although MFCC is the most widely used in speech emotion, but, it does not consider both the relationship among coefficients of Mel filters of neighbor frames and the relationship among neighbor coefficients of Mel filters of a frame. So, many useful features from spectrogram may be lost. To deal with disadvantages of MFCC, the use of weighted spectral features based on Local Hu moments is proposed in (Sun et al., 2015). The idea of this method is motivated by that some emotions such as happy and angry cause drastic variation in energy of spectrogram while some other emotions such as fear and sadness causes slight variation. So, the local energy distribution of spectrogram is affected by this phenomenon in both time and frequency axes.

An emotion recognition system from Malayalam language speech is investigated in (Rajisha et al., 2016). It uses the MFCCs, short time energy and pitch as audio features and utilizes the ANN and SVM as classifier. ANN and SVM achieve 88.4% and 78.2% recognition accuracy, respectively. The Bark Band Energy and spectral features as the most useful unvoiced content are used with GMM for recognition of emotion of fear occurring during abnormal situations (Clavel et al., 2008).

A three-level speech emotion recognition framework is proposed in (ChenXia and XueLee, 2012) where each level of it consists of one or two classifiers for pairwise classification. While the lower levels do rough classification, higher levels complete precise classification. In each level, Fisher discriminator and PCA are used for dimensionality reduction and SVM and ANN are used for classification. The instantaneous features, the first and second derivative and the statistic of the instantaneous features are extracted. The results of this work show better performance of Fisher compared to PCA for dimensionality reduction and show that SVM has more expansibility than ANN.

As a conclusion of above discussions, one of the main issues of SER is extraction of proper features. There are two important issues in feature extraction which appropriately should be determined: 1- the use of global or local features, 2- the feature types such as energy and pitch. Because of non-stationary nature of speech signals, dividing the speech signals into small segments, called frames, is common where each frame can be considered approximately stationary. Prosodic features such as energy and pitch are extracted from each frame of speech and known as local features. In contrast to local features, the global statics are extracted from the whole speech utterance. There is not agreement about which of global or local features better choice for speech emotion recognition task. However, many researchers claim that global features have better

performance than local ones in the recognition rate and classification time point of view (Picard et al., 2001; Shami and Kamel, 2005). As a disadvantage of global features, the researchers claim that the good performance of global features is only in discrimination between low-arousal emotions versus high-arousal ones.

The speech features can be divided into four categories in general: continuous features, spectral features, TEO-based features, and qualitative features. According to results of many researches the prosody continuous features such as energy and pitch contain much of the emotional content of a spoken utterance (Busso et al., 2009). The arousal state of the speaker, i.e., high or low activation affects the energy distribution across the frequency spectrum, overall energy, duration and frequency of pauses of speech signal. These acoustic features are categorized into: formants, pitch-related, energy-related, articulation and timing features. Formants, energy, duration and fundamental frequency are some commonly global features used in speech emotion recognition. Beside time-dependent acoustic features such as energy and pitch, there are spectral features as a short-time representation for speech signal. The distribution of the spectral energy is affected by the emotional content of an utterance. For instance, while happiness emotion has high energy at high frequency range, sadness emotion has small energy at the same range.

According to studies of Teager, the speech signal is generated by nonlinear flow of air in the vocal system (Teager and Teager, 1990). The air flow in the vocal system is affected by the muscle tension under stressful conditions. So, detection of nonlinear speech features is important. The main idea for development of TEO is based on this evidence that hearing is the process of detecting energy. According to (Zhou et al., 2001), TEO of multi-frequency signal, in addition to revelation of individual frequency components, represents the interaction between them. So, TEO-based features can be useful for detection of stress in speech. The normalized TEO autocorrelation envelope area (TEO-Auto-Env) and TEO-decomposed FM variation (TEO-FM-Var) are other TEO-based features.

The voice quality such as breathy, harsh and tense dramatically affects the emotional content of an utterance such that there is a strong relation between the perceived emotion and voice quality (Gobl and Chasaide, 2003). The features related to voice quality are categorized into: 1- temporal structures, 2- word, phoneme, phrase and feature boundaries, 3- voice pitch, and 4- voice level: duration, energy, and amplitude of signal.

There are several challenges in a SER system. There is no agreement about the most discriminative features to separate different emotional classes. The acoustic features are varied with different speakers, sentence and speaking styles. Speaking rates affect many features extracted from speech such as energy contours and pitch. Moreover, it is possible that there is more than one perceived emotion in the same utterance where each different portion of the utterance contains an emotion type. Furthermore, determination of the boundaries among these portions is difficult. Another challenge is depending among expressed emotions, and speakers' environment and culture. In other words, the difficulties of SER systems are large variations between speakers and ambiguity between emotions. The large variations are found in acoustic characteristics of different speakers even if they express the same emotion. Moreover, some pairs of emotions contain similar acoustic characteristics. For example, emotions of boredom and sadness have similar vocal characteristics. Thus, there is a large overlap in acoustic feature space. Some studies reported that emotion recognition of other persons is not an easy task, even for humans (Kim et al., 2009). Some other limitations of emotion recognition using voice signal analysis are represented in (Jaimes and Sebe, 2007) as follows: classification of speech signals into a few discrete labels of emotion expressions, analysis of the audio signal without considering the present context, analysis of emotion information of vocal signals on short time scales which may not provide inferences about attitudes and mood of speech, non-accurate assumptions about the quality of the voice signal such as clear and noise-free speech, and intermediary pauses within short sentences.

Emotion recognition by voice can be suitable in some contexts such as lie detection or crime detection, but it is not significantly appropriate for e-learning except some applications such as teaching languages. In many e-learning systems such as systems for learning mathematics, programing and so on, there is no need for speaking of the learner. Therefore, there is no voice for emotion recognition. Furthermore, voice recognition has difficulties to make difference between some emotions such as happiness and angriness. Thus, even in the languages learning systems an additional method should be applied with the voice recognition method.

8. Facial expression recognition

Among different ways for body emotions expression, the facial expressions are the most important way. It consists of an excellent source of non-verbal cues in communication and interaction among human individuals (Quraishi et al., 2012). Sometimes, a facial expression is more efficient than speaking a lot of words. Several different methods for facial emotion recognition are listed in Table 5. By movement of muscles beneath the skin of the face, a facial expression is revealed. For example, with evoking the emotion of disgust, the muscles of mouth and nose move such that provides facial disfigurement. The severity of facial disfigurement is increased by feeling greater levels of disgust (Shanmugarajah et al., 2012). Facial expression can be recognized from the static images or a sequence of images or videos. The purpose of facial expression recognition usually is categorization of facial expression into some specific classes of expression labels. In other words, the facial expression recognition can distinguish between different facial gestures and also interpret the states of mind. The face deformation due to emotional states is similar for all human individuals despite of the variation gender, ethnicity, and age. Darwin was the first person that stated this theory in 1872 (Darwin, 1872) and Ekman & Friesen (Ekman and Friesen, 1971), after almost 100 years suggested six basic expressions that are similarly known among all people in the world named as fear, anger, disgust, happiness, sadness, and surprise. In the field of facial expression recognition, detection, recognition and tracking of face are the related research topics. Facial expression recognition is a multidisciplinary research used in various applications such as psychological studies, face animations, image understanding, video games, robotics, interactive devices, cognitive science, neuroscience, computer vision, and machine learning. A general survey on different facial expression methods such as RGB, thermal, 3D and multimodal approaches is presented in (Corneanu et al., 2016) where a taxonomy including all steps of facial expression recognition is provided.

In the early 1990s, a number of facial emotion recognition researches were done to enable machines more emotionally intelligent. It has lasted over 15 years since (Pantic and Rothkrantz, 2000) was published. The Facial Action Coding System (FACS) developed by Ekman and Friesen has been usually used for facial expression analysis in computer vision problems (Ekman and Friesen, 1978). An unequivocal and accurate set of isolated and atomic facial movements is defined by FACS. A set of 46 main action units with regard to their intensity and location are particularly used by FACS. FACS recognizes the six universally distinguished facial emotions: fear, anger, disgust, happiness, sadness, and surprise, and their combinations. Each special facial emotion is result of muscle contractions. The facial expressions are due to temporary deformations of facial elements, such as mouth, eyebrows, eye, and nose. The degree of changes in all facial regions indirectly determines the intensity of the emotion. The FACS method, which can automatically recognizes the user emotions by computer programs without utilizing the non-obtrusive equipment and just by a type of camera such as webcam, has high accuracy rates. FACS and neural networks are used for facial expression analysis and classification, respectively in (Zhang et al., 2013). According to FACS, each Facial action unit (AU) is related to the contraction of a specific set of facial muscles. Some examples of AUs with their corresponding face regions, interpretations, and facial muscles are shown in

Table 5
Emotion recognition methods by facial expression recognition.

Type of features	Method		References
	Features	Classifier	
Geometric features	FACS	ANN	Zhang et al. (2013)
	Facial landmarks + Muti-class AdaBoost	SVM	Ghimire and Lee (2013)
	Point distribution model (PDM)	SVM	Saeed et al. (2014)
	Elastic bunch graph matching (EBGM)+ Multi-class AdaBoost	SVM	Ghimire et al. (2015)
	Three-dimensional geometric features and transient features	SVM, ANN	Niese et al. (2012)
	Active Shape model (ASM)	Machine learning-based classifiers	Loconsole et al. (2014)
	Facial movements detection	ANN	Kurihara et al. (2009)
	Surface electromyography (sEMG)	Elman neural network (ENN)	Chen et al. (2015)
	Geometric feature points comprising eyebrow, eye and lip	Kohonen self-organizing map	Majumder et al. (2014)
	Facial AUs	Fuzzy c-means (FCM) clustering	Zhang et al. (2015b)
Appearance features	LDA	KNN	Long et al. (2012)
	PCA	KNN	Long et al. (2012)
	ICA	KNN	Long et al. (2012)
	PCA + LBP	SVM	Loa et al. (2013)
	ICA	SVM	Long et al. (2012)
	LBP	KNN	Khan et al. (2013)
	PLBP	SVM	Khan et al. (2013)
	LBP- TOP	SVM, KNN	Zhao and Pietikäinen (2007)
	LBP and LBP-TOP	SVM	Jiang et al. (2011)
	STLMBP	SVM, KNN	Huang et al. (2012)
	LOCP-TOP	Normalized Correlation, Chi-Squared Histogram Distance	Chan et al. (2012)
	HOG	SVM	Li et al. (2009)
	Gabor wavelets	SVM	Almaev and Valstar (2013)
	LBP and HOG	Sparse representation based classifier	Ouyang et al. (2015)
	AWELBPP	Minimum distance classifier	Goa et al. (2013)
	ES-LBP	SVM	Chao et al. (2015)
	WLD and LBP	Minimum distance classifier	Muhammad et al. (2012)
	Gabor wavelets and LBP-TOP	SVM	Almaev and Valstar (2013)
	(LBP, LGBP, LBPV)+ firefly optimization algorithm	Probabilistic neural network and SVM	Zhang et al. (2016)
	LGBP + similarity and Pareto-based feature selection	Ensemble classifier with ANN	Chin Neoh et al. (2015)
	LBP, contourlet transform	Nearest neighbor	Hemprasad et al. (2016)
	Deep Convolution Neural Network (DCNN)	SVM	Mayya et al. (2016)
Geometric and appearance features	ASM + SIFT + FAP	SVM	Zhang et al. (2014a)
	Geometric features with Gabor wavelets	ANN	Tian et al. (2002)
	Descriptor the histogram of action units and LBP-TOP	SVM	Sanchez-Mendoza et al. (2015)
	Geometric features + mRMR	SVR, ANN	Zhang et al. (2015a)
	Multiscale morphological applied to regions of mouth and eye	Auto associative neural network	Sreenivasa Rao et al. (2011)
	3D facial geometry + Free-Form Deformations	GentleBoost and HMM	Sandbach et al. (2012)
	A combination of appearance, geometric and surface deformation features	KNN	Tsalakanidou and Malassiotis (2010c)
	AdaBoost, Haar and AAM features	SVM and Bayesian network (BN)	Wang et al. (2014a)
	mouth and the forehead/eyes regions + Gabor filters + PCA	Fuzzy logic and clustering based classifier	Hernandez-Matamoros et al. (2016)

Fig. 10 (Li et al., 2016b; Ekman et al., 2002). According to FACS, every facial emotion can be represented by one AU or decomposed by a combination of AUs. A five level scoring (trace, slight, marked/pronounced, severe/extreme, maximum) is used to determine the intensity of each AU (Zhang et al., 2015a). However, because of existence of many subtle facial expressions and many defined rules in FACS, annotation of AUs is a time consuming and tedious task which needs to certified human annotators.

There are two main categories of facial features: geometric features and appearance features. Localizing and tracking of a set of facial points is the main task of the geometric-based techniques. In other words, the shape of face determined by the location and displacement of several landmark points is represented by geometric features. The use of known landmarks locations for estimation of landmark locations is proposed in (Ghimire and Lee, 2013) where for searching landmarks at the center of two eyes, the Haar-like features are used. The median of 52 facial landmarks forms each expression. Then, the Muti-class AdaBoost is used for selection of discriminative features. Finally, SVM is used for classification. In (Saeed et al., 2014), the facial points are manually detected. Then, the point distribution model (PDM) is used to deal with the point localization deficiencies with projection of facial points onto the facial points subspace. The elastic bunch graph matching (EBGM) is proposed for determining the facial landmarks initially at the first frame and then tracking to other image frames over time (Ghimire et al., 2015). The most discriminative geometric features are selected by Multi-class AdaBoost. In (Niese et al., 2012), the photogrammetric techniques are used to extract three-dimensional geometric features. Then, the optical flow-based motion is implemented between consecutive images to obtain transient features. SVM and ANN are used for classification. The geometric-based technique in (Loconsole et al., 2014) uses the Active Shape model (ASM) for determination of facial landmarks locations.

The facial expression recognition method in (Kurihara et al., 2009) detects movements of important face points such as eyebrows, eyes, mouth, and cheeks and use a neural network for recognition of five emotions: happiness, sadness, anger, surprise, and no emotion. An extended Kohonen self-organizing map is developed for recognition of six basic emotions in (Majumder et al., 2014) where the facial geometric feature points comprising eyebrow, eye and lip are used as input.

Generally, in the geometric-based techniques, accurately locating and tracking of facial features is mandatory. In many real applications, locating and detection of facial features is a time consuming, complicated, and error prone task which sometimes needs to a manual labor.

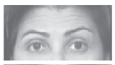
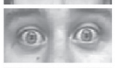
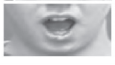
AUs	Picture	Interpretation	Facial Muscles	AUs	Picture	Interpretation	Facial Muscles
AU1		Inner Brow Raiser	Frontalis, pars medialis	AU2		Outer Brow Raiser	Frontalis, pars lateralis
AU4		Brow Lowerer	Corrugator supercilii, Depressor supercilii	AU5		Upper Lid Raiser	Levator palpebrae superioris
AU6		Cheek Raiser	Orbicularis oculi, pars orbitalis	AU7		Lid Tightener	Orbicularis oculi, pars palpebralis
AU9		Nose Wrinkler	Levator labii superioris alaquae nasi	AU12		Lip Corner Puller	Zygomaticus Major
AU14		Dimpler	Buccinator	AU15		Lip Corner Depressor	Depressor anguli oris
AU17		Chin Raiser	Mentalis	AU20		Lip Stretcher	Risorius platysma
AU23		Lip Tightener	Orbicularis oris	AU24		Lip Pressor	Orbicularis oris
AU25		Lip Part	Depressor labii inferioris or relaxation of Mentalis, or Orbicularis oris	AU27		Mouth Stretch	Pterygoids, Digastric

Fig. 10. Several AUs with their corresponding face regions, interpretations, and facial muscles (Li et al., 2016b; Ekman et al., 2002) (In http://www.cs.cmu.edu/face/facs.htm the facial muscles, pictures, and descriptions of all AUs are represented.).

Therefore, the appearance-based techniques have been also suggested. The appearance-based facial expression recognition techniques model the appearance changes like wrinkles and furrows. The appearance-based features are extracted by applying holistic analysis such as Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), and Independent Component Analysis (ICA) to whole or parts of face image. In (Long et al., 2012), the ICA transformation is used for production of spatiotemporal filters from videos. Then, by applying these filters, the feature vector is extracted from the input video. The face appearance changes modeled by Local Binary Patterns (LBP), Histograms of Oriented Gradients (HOG) (Li et al., 2009), Gabor wavelets and discriminant sparse local spline approaches (Lei et al., 2015) represent the appearance features.

Many recent researches are based on local texture extraction from face images. It is shown that the local descriptors are more robust against misalignment, occlusions, and pose changes compared to holistic ones. A spatial representation of LBP, called pyramid of local binary pattern (PLBP), is proposed in (Khan et al., 2013). It is a facial expression recognition method for low resolution images that uses a pyramid of LBP for creation of features only from salient regions of face. PLBP reveals stimuli and the spatial layout by local textures where the face image is tilled into regions at multiple resolutions to obtain its spatial layout. The features extracted from different regions are fused to form the final feature vector. In other words, the pyramidal approach is combined with LBP descriptor to analysis the face images. Similar to other facial emotion recognition methods, the framework in (Khan et al., 2013) consists of three stages: face detection, feature extraction and classification. The face and salient facial regions are detected using Viola-Jones object detection algorithm, which is the most cited, fastest and the most accurate method for face detection (Viola and Jones, 2001). Feature extraction is done by PLBP algorithm where the optimal features should simultaneously maximize between-class variations of expressions and minimize within-class variations. Two different approaches are prevalent for facial expression recognition in literature: recognition of expressions through FACS or direct recognition of prototypic expressions. Five different classifiers such as SVM are used for expression classification. The proposed framework is reliable on low resolution images, illumination invariant, and works properly for both spontaneous and posed expressions.

The facial emotion recognition (FER) system in (Loa et al., 2013) applies the Viola-Jones algorithm to detect the face in an image. For reduction of varying illumination effects, the pixel values are normalized. The LBP algorithm with a 3×3 sliding window is used for local texture features extraction from mouth region. The resulting matrix is computed using the histogram of the LBP values/The dimensionality of achieved feature vector is reduced by PCA and finally SVM is used for classification. In (Ouyang et al., 2015), the LBP and HOG are used for feature extraction, and then, the sparse representation based classifier robust to face occlusions is used for facial expression recognition. The extension of LBP, called Adaptively Weighted Extended Local Binary Pattern Pyramid (AWELBPP), is introduced in (Goa et al., 2013). In this method, firstly, the pyramid transform is applied to represent the input image into different multi-resolution images. Then, sub-extended local binary pattern (ELBP) is applied to sub-images and the important information is measured within each sub-image by using entropy. Finally, different ELBP sub regions are fused to form the final feature vector. The results indicated the superiority of the AWELBPP method compared to conventional LBP. Another extension of LBP, called expression specific local binary (ES-LBP), is proposed in (Chao et al., 2015). In ES-LBP, the class independency is maximized using the class regularized locality preserving projection.

LBP from three orthogonal planes (LBP- TOP) (Zhao and Pietikäinen, 2007) is an operator based on co-occurrences of LBPs on three orthogonal planes, i.e., XY, XT and YT. In other words, LBP-TOP is an extension of LBP by considering the temporal information of image. So, it is an appropriate tool for motion analysis. Some extended versions of LBP-TOP are the spatiotemporal local monogenic binary pattern (STLMBP) (Huang et al., 2012) and the local ordinal contrast pattern from three orthogonal planes (LOCP-TOP) (Chan et al., 2012).

Weber local descriptor (WLD) is also a powerful local technique which utilizes the ratio of changes in gray levels of pixels in an image (Chen et al., 2010). It is formed by using differential excitation and also gradient orientation for description of direction of edges. Both of WLD and LBP features are used through feature level fusion in (Muhammad et al., 2012).

Some facial expression recognition techniques use the both of geometric and appearance features. In (Zhang et al., 2014a), at first ASM is used for determination of 68 fiducial facial points. Then, 53 interior points are selected among them. The Scale invariant feature transform (SIFT) is applied to each of 53 interior points to obtain a feature vector of 6784 elements from each of them. Next, the facial animation parameter (FAP) (Pandzic and Forchheimer, 2002) is used for calculation of 43 distances among 53 points for geometric features extraction. Both texture and geometric features are normalized between 0 and 1 and then concatenate to form the final feature vector.

The static frames are used for detection of action units but the

temporal dynamics are characterized by using the frame sequences. The performance of action unit detection is improved by fusion of appearance-based features and geometric-based features and also with taking into account the temporal dynamics. The geometric-based features are used to define state (present, closed, open, etc.) of facial components such as cheek, brow, eyes, and mouth. For example, the geometric features are used for action unit detection in static frames in (Tian et al., 2001). The geometric features in frame sequences are used in (Valstar and Pantic, 2006). The appearance features, namely LBP and LBP-TOP are used for units detection in both static and dynamic manners in (Jiang et al., 2011). The combination of geometric features with Gabor wavelets is proposed in (Tian et al., 2002). The combination of Gabor wavelets and LBP-TOP features in a dynamic manner is proposed in (Almaev and Valstar, 2013).

A facial expression recognition system consists of three steps: feature extraction, feature dimensionality reduction, and expression recognition is proposed in (Chin Neoh et al., 2015). At first, the LGBP operator is used for textual feature extraction. Then, two evolutionary algorithms are used for feature selection. Finally, an adaptive ensemble classifier with weighted majority vote and a neural network are used for recognition of seven expressions. The appearance-based facial expression recognition methods have various challenges. For an instance, although textural features reveal significant features such as wrinkles, due to variability and subtleness of facial expressions among different individuals, the representation of face using textural information could be a challenging issue. There is not accurate accordance between the pixel locations in face images of different individuals. For example, the pixel indicating the eye corner of one individual may correspond to the pixel indicating the eyebrow of other individual. So, instead of any particular pixel of face image (Chin Neoh et al., 2015), considers groups of pixels for feature selection and representing the discriminative emotional facial features. Images with the size of 100×100 pixels extracted from the LGBP operator are converted into 25×25 parts where each 4×4 pixel out of the original image represents one facial feature. In other words, 0.16% of the overall textural information presents in each facial feature.

The proposed facial emotion recognition method in (Sanchez-Mendoza et al., 2015) has two contributions. The first contribution is that it fuses both the facial dynamics features of face, i.e., geometric and appearance obtained from the video sequences. The geometric descriptor encodes the evolution in time of the facial structure and the appearance descriptor is based on LBP-TOP. A mid-level feature vector is obtained by combination of these two descriptors. SVM classifier is used for classification. The second contribution is facial emotion recognition in two different scenarios: six basic emotions of Ekman emotional theory and subtle emotions such as the positive or negative contents.

As said before, the facial emotion recognition methods are divided into two general groups: static and dynamic feature based. The static feature based systems observe the facial geometric features or appearance ones. The static feature based systems have two main drawbacks: 1- they ignore the dynamic information of face, 2- they may have a lot of variations (such as width of mouth and shapes of eyes) between different subjects. So, the static feature based systems may have not adequate generalization ability and are not efficient enough for real facial emotion recognition applications. To deal with these problems, the dynamic feature based systems have been proposed that use the temporal variations of face movements (Kotsia et al., 2008; Srivastava and Roy, 2009; Tsalakanidou and Malassiotis, 2010a). The MPEG-4 face animation framework (Pandzic and Forchheimer, 2002), as a part of MPEG-4 FBA [ISO14496] International Standard, defines 84 facial points to best reflect the movement mechanics and facial anatomy learned from facial actions related to muscle actions (see Fig. 11). Based on facial points defined in MPEG-4, the 2D or 3D distance features between key facial points can be computed, and then dynamic changes of these distances can be used for facial emotion recognition.

Many facial emotion recognition methods either only considered static features of face images or limited to 2D facial models. These methods lose possibly valuable dynamic facial information and have not robust enough against illumination changes and subject variation. Moreover, detection of novel emotion classes, which do not belong to one of six basic emotion classes or compound emotions, is in neglected by many popular systems. A real-life 3D AU intensity estimation and facial expression detection system is proposed in (Zhang et al., 2015a).

Expression dynamics play an important role for recognition of facial emotional behavior. They can be used to distinguish between posed and spontaneous emotional expressions; and to recognize complex and subtle states such as mood, pain, embarrassment, and shame. So, expression dynamics contain useful information for robust and accurate expression recognition. The 2D facial images and video have high sensitivity to illumination conditions, facial pose, and facial appearance changes such as sunglasses. To deal with this difficulty of 2D face images, different capture modalities such as 3D-imaging have widely been used. The 3D facial features contain out-of-plane movement that is not present in 2D features. Moreover, 3D imaging removes the problems of pose and illumination inherent to 2D data. Many studies have been tried to use the 2D

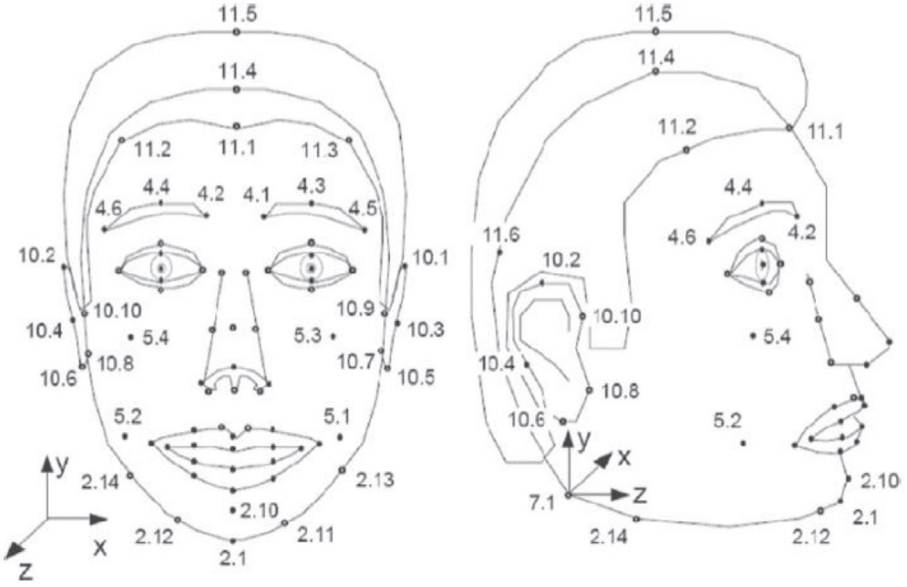


Fig. 11. Defined facial feature points in MPEG-4 (Zhang et al., 2015a), (Pandzic and Forchheimer, 2002).

images for construction of 3D models (Cohen et al., 2003; Sebe et al., 2007). But, these methods are sensitive to the problems of pose and illumination inherent to all 2D methods. So, many recent researches use 3D facial geometry for facial emotion recognition to analysis static models (Mpiperis et al., 2009), model the temporal information (Tsalakanidou and Malassiotis, 2010b), and encode the temporal dynamics of 3D facial expressions models (Sun and Yin, 2008).

The combination of 3D facial geometry and expression dynamics contain a wealth of information leads to improvement of facial expressions analysis. In (Sandbach et al., 2012), firstly Free-Form Deformations (Rueckert et al., 2003) is used to capture 3D motion of the face between frames in an image sequence. Then, a quad-tree decomposition of the motion fields is applied for feature extraction. Next, the extracted features are fed to the GentleBoost classifiers, which are an extension to the traditional AdaBoost classifiers, for simultaneously feature selection and performing the training used for classification. Moreover, HMM is used to model full temporal dynamics of the facial expression. The facial expression recognition method in (Tsalakanidou and Malassiotis, 2010c) uses a combination of 2D and 3D image streams.

One of the main challenges of facial emotion recognition methods is high dimensionality of data. The identification of the features with maximum discrimination ability, which best represent between-class and within-class scatters for emotional facial expression, is an important issue. In (Zhang et al., 2016) an evolutionary algorithm for feature selection of facial features is suggested. The evolutionary algorithm uses the search behavior of the moths and fireflies and deals with difficulties of levy-flight firefly algorithm (Yang, 2010) and the moth-flame optimization algorithm (Mirjalili, 2015).

Because of rapid emergence of 3D movies, a huge volume of demands for face recognition and facial expression recognition have been presented (Ming and Ruan, 2013). In the practical applications, the discrimination and descriptiveness are independent tasks. However, simultaneously performing face recognition and facial expression analysis is preferred. Several researches have been investigated 3D face processing for independent analyzing of 3D face recognition and 3D facial expression categorization (Bowyer et al., 2006; Gao et al., 2011). The studies show that the regional descriptors can better analyze simultaneously both issues of face discrimination and expression classification. Division of a face into a group of regions (Faltemier et al., 2008) and the use of 3D shape description for statistical feature extraction (Alyuz et al., 2010), the use of facial symmetry to cope with large pose variations (Passalis et al., 2011), selection of optimal local features and finding the most discriminative features for 3D face recognition (Wang and Tang, 2010), learning the azimuth angle of subspace normal suited for 3D facial recognition (Marras and Zafeiriou, 2012) are some of algorithms that focus on 3D face recognition and cannot perform facial expression recognition. A robust regional bounding spherical descriptor to balance two aspects of 3D face recognition and expression analysis is introduced in (Yue). This method uses the spherical bands on the face and shape index to segment a group of regions on each 3D facial point cloud (see Fig. 12). To obtain the regional descriptor, the facial areas are projected to the regional bounding spheres. Then, to boost the classification accuracy, the regional and global regression mapping technique is applied to the weighted regional descriptor.

Expression invariant face recognition methods are generally divided into two groups: spatial domain techniques and transform domain techniques. The spatial domain ones are further classified into two groups: model based techniques and based techniques. The model based techniques use the appearance based models and utilize the shape and features of face images. The active appearance model (AAM) based frameworks are examples of model based techniques (Riaz et al., 2009). Although these frameworks provide desired results, but fitting AAM has heavy computations. Subspace based techniques analyze different attributes by using traditional statistical procedures. The use of PCA to construct subspace is an example of subspace based techniques (Li et al., 2006). Transform domain based techniques transform images into another domain for extraction of eminent features. The wavelet transform, Radon transform, and Gabor wavelet transform are some examples of this group. The combination of the wavelet and Radon transforms is used in (Vankayalapati and Kyamakya, 2009). Radon transform extracts local features and wavelet transform provides the time-frequency localization. So, combination of them provides an appropriate and robust

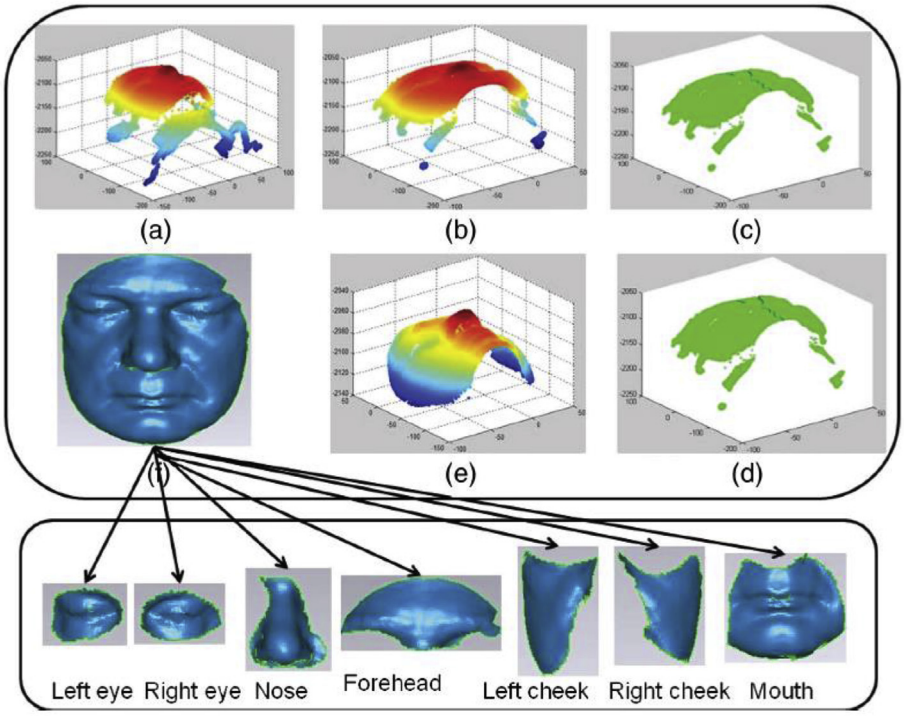


Fig. 12. Facial segmentation in (Yue). (a) Original 3D point cloud. (b) Facial area extraction. (c) Symmetry curve extraction from facial area. (d) Nose tip detection. (e) Facial surface alignment. (f) Aligned facial image.

feature vector. The wavelet transform is also used to deal with expression variation problem. A facial image can be transformed into wavelet domain. By equating the detailed coefficients to zero and taking the inverse transformation, an image representation that is robust for expression variations is obtained (Abbas et al., 2008). The multi-directional transforms such as curvelet transform and contourlet transform can be used for extraction of more robust features. The proposed facial expression recognition method in (Hemprasad et al., 2016) uses both the spatial based techniques (LBP) and transform based techniques (contourlet transform). The Nearest neighbor classifier is used for classification. Because of exhibition of important properties such as anisotropy and directionality, the contourlet transform is an efficient feature extraction. In addition, the coefficients in contourlet subbands are enhanced which results in the robustness of system by enhancing the features of skin region.

The Multiscale morphological is applied to the gray values of pixels in regions of mouth and eye for facial feature extraction in (Sreenivasa Rao et al., 2011). Then, the auto associative neural network is used for classification.

Deep learning has shown its power in different fields. Deep Convolution Neural Network (DCNN) has a wide application for feature extraction from the images. The existence of several layers leads to accurate feature learning. The pre-learned features can be utilized as filters. The input image convolves through filters to produce appropriate features used by subsequent layers of the network (Krizhevsky et al., 2012). These techniques have been also applied for facial expression recognition (Mayya et al., 2016).

Eyebrows are one of the main parts of face which have important role in recognition of emotions. According to this idea (Chen et al., 2015), just uses this part of face for facial emotion recognition. The analysis emotional cues in different parts of the face results in emotion recognition. The top-half or bottom-half of face images readily display some human emotions. For example, sadness, fear, and anger are readily identified from the top-half of face. In contrast, disgust and happiness are readily identified from the bottom-half of face (Calvo and Nummenmaa, 2011). The result is that facial emotion recognition depends on specific regions of faces and this dependency varies by emotion type. Some researchers have investigated the importance of features located in the upper and lower parts of the face individually. For example, the most part of face for recognition of fear and sadness is eye region. The lower part of the face is important for happiness and disgust emotions recognition where the mouth has a unique role for the rapid detection of happiness. However, the real smiles versus posed ones can be discriminated by only looking at the eyes. So, the importance of the mouth is not absolute in happiness recognition. Different cultures also attend to different parts of the face during emotional recognition.

A clustering based facial emotion recognition system is proposed in (Zhang et al., 2015b). At first a facial point detector is proposed to derive 54 facial points. Then, the intensities of 18 selected facial AUs closely associated with the emotional expressions are estimated using SVR and neural networks. Based on the obtained intensities of the 18 AUs, fuzzy c-means (FCM) clustering algorithm is used for recognition of seven basic emotions fear, anger, disgust, happiness, sadness, surprise, and contempt. The used FCM clustering method also deals with classification of compound and newly unseen emotions. In contrast to traditional rigid clustering algorithms, which locate an object only in one cluster, FCM clustering allows an object to belong to multiple clusters. Then, FCM clustering can recognize a compound emotion which may belong to multiple clusters rather than only one emotion cluster.

Most of AU recognition methods work based on this assumption that the complete labeled training images are available while in practice labeling AUs is a time consuming and expensive process. Moreover, because of subjective differences of AUs and their ambiguity, the confident and reliable labeling of some AUs is also difficult. A multi-label learning with missing labels framework is proposed for AU recognition in (Li et al., 2016b). The proposed method contrary to similar previous works, which usually utilize the same features for all emotional classes, uses the most related features to discriminate each AU. The use of the same features causes involving noise from occurrences of other AUs, and so, degrades the performance of model. The proposed method in (Li et al., 2016b) deals with this difficulty, and so, improves the recognition performance.

Usually expression recognition and emotion recognition are used with the same meaning in literature. But, in some cases different meanings are considered for them (Wang et al., 2014a). The facial appearance of the subject's emotion is represented by expression while the latent feeling is represented by emotion. Study in (Lorenzino and Caudek, 2015) showed that the use of facial emotions can facilitate the face discrimination learning. A multi-task facial inference model is proposed in (Zheng et al., 2016) for simultaneously face identification and facial expression recognition. Both of these two independent tasks are done by extracting and using desired shared information across them.

Japanese and Caucasian facial expressions of emotion (JACFEE) has developed as a picture pool which displays emotional expressions including fear, anger, disgust, happiness, sadness, surprise, and contempt (Matsumoto and Ekman, 1988). The posers in the pictures are Japanese and American. JACFEE has good cross-cultural validity, i.e., high consistency of the emotional labeling in different cultures that is evidenced in many researchers (Shioiri et al., 1999). However, the boundaries between some emotional expressions are not very clear. For example, disgust is sometimes mistaken with anger; and fear is often confused with surprise. In some investigations, PCA is applied on the perceived image intensities to efficiently delineate facial expressions. The researchers show that PCA is an efficient psychological metaphor to encode the structure and represent facial emotions. Further, PCA can represent the personality structures and its universality. Because of similarities between the facial emotions and personality, PCA may help to determine the facial emotions.

To develop an efficient facial recognition system, a huge amount of facial images are needed where gathering of them is a time consuming and difficult task. Since advancement of web technologies, collection of such data is now possible with low human efforts. The different real world facial expressions can be collected from web using search-based approaches. But, sometimes, the returned images may be not matched with the keyword. To deal with this limitation, a machine learning method can be used to separate relevant images from noisy images. The obtained database has more sample images per expression than other standard databases like CK and JAFFE. So, it is more diverse. Moreover, it has more similarities with real world face images (Khan et al., 2016).

Many FER systems such as above instances provide reasonable performance for images taken under controlled conditions and contain non-occluded face. Designing a robust FER system for applying on partially occluded facial images is a challenging issue where face occlusions can be divided into temporary and systematic types. Hair, head, or other movements that partially occlude the face are temporary occlusions. Wearing scarves, sunglasses, and so on result in systematic occlusions (Buciu and Pitas, 2008). However, in both types of occlusions, the system should be trained using non-occluded face images while it should be tested with both occluded and non-occluded images. The FER system proposed in (Zhang et al., 2014b) detects face using Viola-Jones algorithm. Then, it simulates a set of occluded face images by adding masks to some extracted regions of face. Then, Gabor filters are applied to these images. The obtained templates are used as local features to replace the occluded templates in the testing stage. The FER system proposed in (Hernandez-Matamoros et al., 2016) segments the facial image in two regions of interest: the mouth and the forehead/eyes containing some non-overlapping blocks. Then, the Gabor filters are applied to image blocks, and next, PCA is used for feature reduction. The extracted features are fed to a classifier based on fuzzy logic and clustering. This classifier showed high recognition rates with reasonable computational complexity even for occluded facial images.

There are several main challenges for applying the most of introduced

automatic FER methods in real life applications. Most of these methods are designed for more or less standardized environments conditions such as constant/acceptable illumination and frontal faces. But, in the real applications images taken by cameras in intelligence systems have not satisfied these conditions. Moreover, most of facial emotion recognition methods are designed to determine six emotion categories. Although considering six discrete emotions simplifies the emotion recognition, but we know that the six given categories of emotions are not applicable for the majority of everyday applications. As another challenge, many researchers reported that although facial emotion classification reasonably work for posed expressions, such as posed smiles, they have no good efficiency on spontaneous expressions elicited during natural daily interactions (Zeng et al., 2009; Sariyanidi et al., 2015). Obtaining ground truth labels for spontaneous emotions is one of the biggest problems in these scenarios (Gunes and Hung, 2016). The most of the automatic emotion recognition systems have been designed to use the static face images under appropriate illumination conditions, and so, they cannot obtain accurate inferences on observed emotions that are expressed during naturalistic settings. Most of face recognition systems are unable to perform a context-dependent interpretation of the facial expressions. In other words, interpretation of the context in which the facial behavior occurs is difficult.

Beside plenty of studies on emotions recognition from full-face images, some researchers believe that emotion recognition using half-face images is simpler than it using full-face images. In (Wu and Huang, 1990), some studies are done on half-face images. An outline curve is determined on the half-face image and 9 critical points on this curve are recognized (see Fig. 13). Then these nine points are used to generate six feature characteristics. The use of half-face images may lead to more precise results than full-face images, but full-face images are more useable in the context of e-learning. Most of computers have webcams that is located in front of the face. Thus, capturing full-face images seems to be simpler. In comparison to other emotion recognition methods, facial image processing is the most precise. The weakness of this approach for using in e-learning environments is that the learner's face is relatively static during the learning process, because the learning process cannot produce huge emotional excitements.

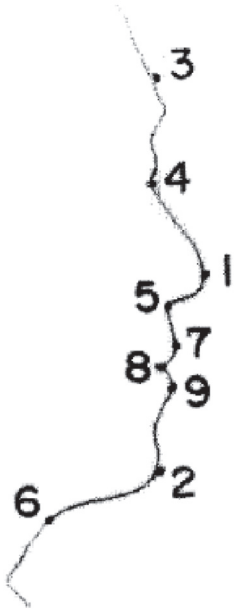


Fig. 13. Half-face image processing (Wu and Huang, 1990).

9. Vital signals

The human body is magnetic, mechanical, electrical, thermal and chemical in nature (Allanson and Fairclough, 2004). Different types of data such as sight (electromagnetic), touch (mechanical, thermal), sound (mechanical), smell and taste (chemical) are received by individual human senses. For detection of information such as brain signals and heart rate that reflect emotional states of human, physiological sensors can be used. The physiological and physical variables can have a main role in generation of an accurate and robust emotion recognition system. Table 6 lists several emotion recognition methods by analyzing the vital signals. According to neurophysiological idea, feeling an emotion causes changes in the autonomous nervous system (ANS), depending on the intensity and nature of emotion (Nasoz et al., 2003). Therefore, physiological reactions observed on the person can be associated with a special emotion. For instance, the emotion of fear leads the ANS to provoke a stress response where activation of special types of hormones into the bloodstream causes an increase in skin sweat, heart rate, and breathing

Table 6
Emotion recognition methods by using vital signals.

Method			References
Vital signal	Features	Classifier	
EEG	kernel density estimation (KDE) and MFCC	multi-layer perceptron (MLP)	Othman et al. (2013)
EEG	PSD, differential asymmetry of PSD and correlation with frequency bands	Correlation based classifier	Kumar and kumar (2016)
EEG	PSD	SVM	Jatupaiboon et al. (2013)
EEG (ERP)	Echo State Networks (ESN)	FCM, LDA, KNN, SVM, Naïve Bayes and decision tree	Bozhkov et al. (2016)
EEG	Time, frequency and wavelet	MLP	Mehmood Bhatti et al. (2016)
EEG	Differential entropy + Subspace alignment auto-encoder (SAAE)	Linear regression classifier	Chai et al. (2016)
EEG	Bispectrum	SVM, ANN	Kumar and KhaundShyamanta (2016)
EEG	ICA, KDE	ANN	Lahane and Kumar Sangaiah (2015)
EEG	mRMR	SVM	Atkinson and Campos (2016)
ECG	trend-based and parameter-based of HRV+(KBCS, LDA, PCA)	KNN	Wang et al. (2013)
EOG, EMG, EEG, GSR, skin temperature, respiration pattern, and blood volume pressure	Discrete Wavelet Transform	SVM, KNN, MLP, Meta-multiclass (MMC) classifiers	Verma and Tiwary (2014a)
fEEG, fEMG, fEOG, SC, BVP, IBI	Mean of amplitude, standard deviation, energy of signal, skewness, kurtosis, frequency energy, sub band frequency power	SVM, KNN	Khezri and FiroozabadiAhmad (2015)
ECG, electrodermal activity, facial EMGs and respiration	Manual feature selection	A classifier based on SVM and adaptive neuro-fuzzy inference	Katsis et al. (2008)

frequency and the mouth becomes dry. To quantify emotional states, different types of physiological signals are used. Blood pressure, breathing, electro-dermal activity, electroencephalograms, electro-cardiograms, and electromyograms are some of frequently used sensors to measure physiological signals (Koelstra et al., 2012). Among them, the use of lightweight wearable sensors is pleasant because it decrease the intrusiveness.

The Electroencephalography (EEG) signals have been used as quantified variables for emotion recognition. An emotion recognition method by using EEG signal is proposed in (OthmanAbdul Wahab et al., 2013), which uses two 2D emotional models (rSASM and 12-PAC).

The kernel density estimation (KDE) and MFCC are used as feature extractor and the multi-layer perceptron (MLP) is used as classifier. With no assumption about the distributions of data, KDE estimates the probability distribution function of random variables. The mel-frequency properties of MFCC allow a linear/logarithmic spacing below/above 1 kHz. The results show that the highest accuracy in EEG emotion recognition is obtained by 12-PAC model for both feature extraction methods. Moreover, the results represent accuracy improvement of the EEG emotion recognition method by increasing the precision of the dimensional models.

An emotion recognition method using EEG signal is proposed in (Kumar and kumar, 2016) that is implemented as follows. Because of existence of an extensive number of features in frequency domain compared to time domain, the Fast Fourier Transform (FFT) is applied to the EEG signal and, the Butterworth notch filter is applied to remove the electrical noises. Interpolation of electrodes is done for binding the data. To identify the local sources and find the independent components, ICA is performed. The resultant spectra is isolated into five frequency bands: delta (δ: 1–3 Hz), theta (θ: 4–7 Hz), alpha (α: 8–13 Hz), beta (β: 14–30 Hz), and gamma (γ: 31–50 Hz). The characteristics of these bands lead to different types of features related to different emotions. Three features types are extracted: power spectral density (PSD), differential asymmetry of power spectral density and correlation with frequency bands. Emotion recognition is done using a correlation based classifier. The emotion recognition method using EEG signals in (Lahane and Kumar Sangaiah, 2015) uses ICA, KDE and ANN for signal enhancement, feature extraction and classification, respectively. The ICA transformation is able to split up a mixed signal into its sources. So, ICA is used for signal enhancement through artifact removing. KDE is a feature extraction method that uses the kernel-smoothing technique to estimate the kernel density. In (Atkinson and Campos, 2016) the mRMR method and kernel-based classifiers such as SVM are used for feature selection and classification of EEG signals, respectively, for emotion recognition based on a dimensional model of emotions (Valence and Arousal).

Selection of an appropriate feature space from EEG signal to achieve robust prediction of emotional states across subjects is an important problem. Because of the high variability between different individuals, the construction of inter-subject models is a harder task than intra-subject ones. For example, for classification of happy/unhappy emotions in (Jatupaiboon et al., 2013), 75.62% classification accuracy is obtained for the intra-subject models while 62.12% accuracy is obtained for the inter-subject models. The emotions are elicited by pictures and classical music. Emotional features are extracted from PSD of the EEG signal, and SVM is used as classifier.

The researchers in (Bozhkov et al., 2016) propose a feature selection method from Event Related Potential (ERP) signals where ERPs, which are transient components in EEG, show the brain activity. They proposed that ESN can be served as an effective feature selection that improves the discrimination between human emotions from ERP signals. ESN is a class of RNN where RNN is randomly generated and also remains unchanged during network training. FCM is used as an unsupervised method and LDA, KNN, SVM, Naïve Bayes and decision tree are used as supervised classification methods. In (Mehmood Bhatti et al., 2016), the music is used for eliciting emotions in an experiment. The EEG signals are observed and analyzed for emotion recognition. Three types of features: time, frequency and wavelet are extracted from EEG signals and MLP is known as the best classifier for recognition of human emotion elicited from audio music.

Many EEG signal classification methods have been proposed such as SVM, KNN, and neural networks which assume that the testing samples and training ones share similar or the same distributions. The mismatch between the testing and training samples may be due to sampling of them from different subjects or experimental sessions because of varying impedances, different electrode placements, user fatigue, etc. In an EEG classification problem, the conventional classifiers involve testing and training samples for the same subject and sometimes for the same experimental session. Typically, training samples and testing samples are called as source domains and target domains, respectively where they have different distributions. Recently, some machine learning techniques using domain adaptation have been investigated (Pan and Yang, 2010). The aim of domain adaptation is to utilize the information of both the source and target domains in the learning phase to build classifiers robust to mismatched distributions with this assumption that the target and source domains have the same aim. Some of domain adaptation techniques employ a linear/nonlinear transformation without considering a consistency constraint. So, they are not so appropriate to achieve a common distribution from highly non-stationary EEG signals. The subspace alignment auto-encoder (SAAE) has been proposed in (Chai et al., 2016) to deal with this problem. The SAAE method combines a subspace alignment solution and an auto-encoder network to take the advantages of both consistency constraint and nonlinear transformation. An auto-encoder is constructed with two hidden layers where the optimal number of neurons is obtained by searching among $\{50; 100; 200; 400; 600; 1000\}$. Moreover, an efficient and simple feature called differential entropy is used as an input of the domain adaptation model and a linear regression classifier is used for supervised classification.

Bispectral analysis is proposed for emotion recognition from EEG signals in (Kumar and KhaundShyamanta, 2016). Bispectral analysis detects the phase relationships between frequency components and characterizes the non-Gaussian information of EEG signals. Emotions represented in Valence-Arousal emotion model are recognized by features derived from bispectrum.

Different features extracted from EEG signals have been used for emotion recognition. The spectral power in various frequency bands of EEG signal is a discriminable indicator of the emotional state. The valence states and also some discrete emotions such as fear, sadness, and happiness change the alpha power (Verma and Tiwary, 2014a). Specifically, the correlation between the frontal asymmetry of the alpha power and valence states has been frequently reported (Lee et al., 2014a). As other distinguishable features for emotion recognition, the interactive characteristics observed between a pair of EEG oscillations such as coherence and phase synchronization have been used (Miskovic and Schmidt, 2010). Because of complexity of brain system, the nonlinear features should be considered to model and analysis of it. Many nonlinear methods such as fractal dimension (Liu et al., 2010), entropy, and Hurst exponent (Wang et al., 2014b) have been used to discriminate between different emotions. Different types of machine learning methods such as SVM, MLP, Bayesian network, KNN and LDA have been also used for classification. There is no single best classifier. The choice of an appropriate classifier greatly depends on the properties of the used datasets.

The physiological-based emotion recognition methods may have inaccurate results if the existent differences in individuals' response patterns are ignored. To solve this problem, a group-based individual response specificity model is proposed for emotion recognition in user-independent scenario by taking user's individual response specificity into account (Li et al., 2016c).

The emotion recognition method proposed in (Verma and Tiwary, 2014b) fuses some different physiological signals: electrooculogram (EOG), electromyogram (EMG), Electroencephalogram (EEG), Galvanic skin response (GSR), skin temperature, respiration pattern, and blood

volume pressure. It suggested a three continuous dimensional representation emotional model. The discrete Wavelet Transform is used for feature extraction and SVM, KNN, MLP, and Meta-multiclass (MMC) classifiers are used for classification. The Wavelet divides the signal into different frequency components, and then analyzes them individually with a different resolution associated to its scale. A multimodal emotion recognition method by fusion of some different types of biosignals has been also proposed in (Khezri et al., 2015). The physiological signals are simultaneously collected. The forehead electroencephalogram (fEEG), forehead electromyogram (fEMG), forehead electrooculogram (fEOG), skin conductance (SC), blood volume pressure (BVP) and interbeat intervals (IBI) signals are measured. Fig. 14 shows some examples of these measurements. The different features such as mean of amplitude, standard deviation, energy of signal, skewness, kurtosis, frequency energy, sub band frequency power are extracted from these signals and then, the sequential forward floating selection algorithm (Breazeal, 2003) is used for feature selection among the extracted features. The SVM and KNN classifiers are used for classification. The recorded signals are classified by using several classification units independently. Then, the achieved results are fused through an adaptive weighted linear model to acquire the final result. The fusion methods are divided into three main levels: input signals, extracted features and decision fusion (classification units). Fusion at the decision level provides some effective characteristics and advantages. It breaks the features into several smaller sets. Therefore, it may increase the classification performance. The use of more effective features of the input data is possible independently. The correct classifiers can compensate the error of incorrect classifiers. So, the classification accuracy would be increased.

In (Katsis et al., 2008), the emotional modalities such as electrocardiogram (ECG), electrodermal activity, facial EMGs and respiration are fused in feature-level where a central classifier based on SVMs and adaptive neuro-fuzzy inference is used for classification.

Two feature generation methods (trend-based and parameter-based methods) have been used for generation of significant features from ECG signals in (Wang et al., 2013). The trend-based features, which are statistical features from long-term heart rate variability (HRV) variations, reveal the long-term variations of a certain HRV parameter. The parameter-based features are features from 5-min HRV analysis. The kernel-based class separability (KBCS) (Flavell, 1999) is used for feature selection and then, PCA and LDA are utilized for dimensionality reduction. The experiments of this work show that the combination of KBCS, LDA, and PCA provides satisfactory recognition accuracy. The KNN classifier is used for classification. This method effectively recognizes the driving stress conditions using only ECG signals.

According to studies in (Castro et al., 2009), the low-frequency band power and the low-frequency/high-frequency (LF/HF) ratio increase when an individual is under stress. The measured ECG signals include three different driving stress conditions: rest before driving and rest after driving (low stress) and during driving (medium/high stress). Therefore, the physiological condition can effectively detected by observing the variation on trend or a physiological parameter.

Some difficulties of emotion recognition using measuring vital signals are represented in the following. This approach requires the use of often expensive physical sensors, limits the participants' mobility, and distracts emotional reactions of person. Moreover, unanticipated changes in physiological characteristics such as surgery may introduce noise in the measurement of the vital signals (Chandra and Calderon, 2005). Providing a proper mapping between vital signals and specific emotions is sometimes impossible. In addition, each user should get a special expertise to use the special sensors and equipment. Similar to emotions detection by voice, it must be considered that we cannot separate emotions completely by vital signals. They separate arousal emotions from passive emotions but they don't separate emotions with the same arousal. However these parameters are various and more researches may show new relationships between them and emotional states. The use of this method for e-learning context seems to be not applicable. Because it needs to different sensors and connecting them into the learner's body. It may make the learning process difficult and annoying. Furthermore, these sensors may not be available for all learners.

10. Gesture recognition

Because of development of new sensors for human action and gesture recognition, the field of human activity and gesture recognition has recently rapid growth. These novel sensors are categorized according to three types: physiological (heart rate, temperature or electromyography), vision (depth, color or heat), and position (motion capture, global positioning system or inertial motion units). With advances of technology, cheaper, smaller, and more efficient sensors have been produced which can be embedded in wearable devices such as watches. The recognition of the gesture performed by a human in the air is a sub domain of activity recognition. The use of multi-purposes corpora, which is used for different types of activity recognition, the lack of standards and increasing complexity and cost may confuse the researchers.

The first issue is concerning about mixing different types of activity recognition fields. The exploration of activity recognition consists of three paths: human activity and action recognition, human surveillance, and human gesture recognition. High level activity recognition such as recognizing walking and eating is action recognition. The recognition result is used for monitoring or generation of statistics about users for

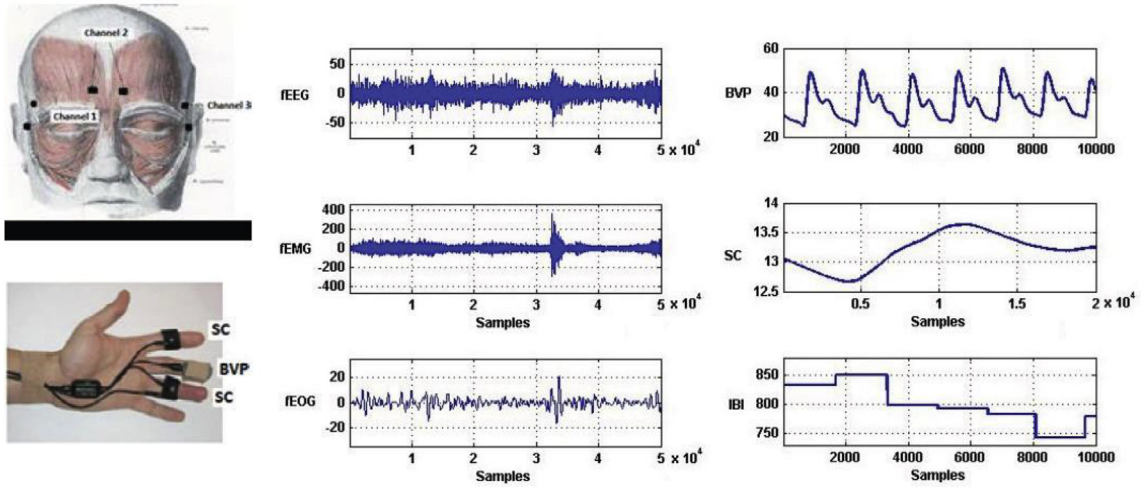


Fig. 14. Some examples of sampling from different types of physiological signals (Khezri et al., 2015).

application such as automation of the environment. The aim of human surveillance is activity recognition for detection of undesired behavior such as shoplifting or facilitating daily life such as ambient-assisted living. The purpose of gesture.

Recognition is different from two latter. The main aim of gesture recognition is recognizing the human gestures in order to interact with machines. The notion of intention has importance in gesture recognition. Table 7 shows some gesture recognition methods.

Even in the absence of vocal and facial cues, some human emotions can be revealed by movement of the body or its parts. We can refer to works such as static body postures (Coulson, 2004), whole-body movement (de Meijer, 1989), and arm movement (Pollick et al., 2001). The information relevant to body actions and gestures are divided into three main classes: 1- form or structural information, 2- kinematics such as displacement, velocity, and acceleration, 3- dynamics, i.e., motion due to force and mass. Kinematics cues have attracted considerable attention for action perception. Typically, motion information (dynamics ad kinematics) and motion-mediated structural information are employed for displays of biological motion while static form information is absent or minimal. Point-light displays of body movements can provide sufficient information for making accurate judgments about the individual making the movement. For example, the sex of person can be identified from gait (Barclay et al., 1978). In some cases displays provide equal performance with full-light displays where whole body is visible. Generally, motion cues are rather more important than static form cues (Atkinson et al., 2007).

In the sophisticated applications such as emotion recognition ones, processing and assessment of changes in the form of body over the time is important (Casile and Giese, 2005). The evidences show the important role of kinematics of body and also body-part movements in perception of emotional states. For example, according to (Pollick et al., 2001), by considering knocking arm movements as stimuli, jerky and fast movements represent anger and happiness while smooth and slow movements reveals sadness expression. For another example, the arm movements due to sadness, anger or joy expressions are varied in terms of their displacement, velocity, and acceleration (Sawada et al., 2003). Nonetheless, the form-related or structural cues in body movements contribute to emotion perception in addition to kinematics. The full-light facial movements provide more emotion classification accuracy than point-light ones except for happy expressions in [430]. Similarly, full-light gestures provide more emotion recognition accuracy compared to point-light and patch-light body gestures displays. The emotion classification accuracy is impaired by inverting the gesture movies. The motion reversal and stimulus inversion have a greater effect on recognizing of threat-related emotions such as disgust, anger, and particularly emotion of fear than on the other emotions (Atkinson et al., 2007).

Based on view of some researchers, gestures instead of determination of emotion types indicate the intensity of it. Some other researchers provide evidence that certain body movements are associated with specific emotions. The body movements, specifically hand gestures, have been interested in the human–computer interaction community. Two of the most recently works in hand gesture recognition are the glove-based approaches (El Hayek et al., 2014; Kim and Chien, 2001; Dipietro et al., 2008; Lamberti and Camastra, 2012) and the vision-based ones. The glove-based approaches need a 3-D spatial description of hands and use the model-based techniques. In the vision-based approaches the hands must be appeared in images where appearance-based techniques are used for gesture analysis (Lopatovska and Arapakis, 2011a).

In many cases, body gestures are used to complement spoken language. The voluntary or non-voluntary movements of hands contain additional information, which may not present in speech, regarding the emotional states of the speaker. In (Malatesta et al., 2016), gestures are isolated from speech and focus is on emotional content of hand movement instead of the function. Many researchers have coded human movement in binary classes such as pleasant/unpleasant, big/small, strong/weak, wide/restricted, and fast/slow. The most complete approach for modeling of expressivity is the expressivity dimensions in (Caridakis et al., 2013), which covers the entire parameters of expressivity related to affect and emotional states. The neuro-fuzzy network is used for affective states recognition using hand gestures.

Generally, the glove-based methods extract hand posture and in some cases are able to recover hand orientation and position. But, the use of them is annoying. So, the vision-based hand gestures recognition methods are more popular (Christianet al., 1997). However, the vision-based recognition methods have also several challenges such as extensive diversity in hand gestures, space-time varieties, and multi-meanings. The vision-based hand gesture recognition methods are divided into two general categorizes: 3D hand/arm models (Mikio, 1996; Mu-Chun, 2003) and appearance models (Hou et al., 2001; Huang and Huang, 1998). Although 3D hand/arm models are suitable for recognition of different kinds of hand gestures, they have highly computational complexity and also are fallibility due to many approximation considered for estimation of model parameters. So, they are often used in animation. On the other hands, the appearance based models are suitable and efficient for communicative hand gestures. They have low computational complexity and can be done real-time. So, the appearance-based approaches are adopted by many researchers. 15 different hand gestures are recognized by 3D neural network in (Huang and Huang, 1998). HMM has been successfully applied to the field of gesture recognition [450]. A non-HMM-based system by using slow segmentation scheme is developed in (Cui and Weng, 1996) for recognition of 28 different hand gestures. The optical flow is used to estimate the changes of gesture motion direction for human gestures recognition in (Ohknishi and Nishikawa, 1997). PCA is proposed for representation and recognition of the shapes of animated objects (Zhu and Yuille, 1995). A hand gesture recognition method is proposed in (Gupta et al., 2012), which uses the Gabor filters for feature extraction, PCA followed by LDA for feature reduction and SVM for classification. By applying a Gabor filter bank with 3 scales and 5 orientations on a 64×64 image, the dimensionality of output pattern vector will be $15 \times 64 \times 64$. So, feature reduction of extracted feature vector can improve the classification accuracy in small sample size situations.

A vision-based method for hand gesture recognition is proposed in (Teng et al., 2005) which uses 2D video sequences as input. It consists of four sections: detection of hand gesture region, preprocessing, locally linear embedding (LLE) (Roweis and Saul, 2000), and hand gesture recognition. The application of this method is real-time recognition of Chinese sign language alphabet. The first step is determination of location of hands in image frames or hand detection. The skin color is used to discriminate skin regions from other regions in the image. Two largest skin regions are considered as hand and face. To discriminate between hands and face, different constraints such as the mass center, compactness, perimeter length and area of two skin regions can be used. The

Table 7
Gesture recognition methods.

Parts of body	References
Static body postures	Coulson (2004)
Whole-body movement	[441], (Darrell et al., 1998)
Arm movement	(Pollick et al., 2001; Sawada et al., 2003), (Shimizu et al., 2007)
Form of body over the time	Casile and Giese (2005)
Hand gestures (vision-based) (3D body models)	Caridakis et al. (2013)
Hand gesture (vision-based) (3D hand/arm models)	(Mikio, 1996; Mu-Chun, 2003)
Hand gesture (vision-based) (appearance models)	(Hou et al., 2001; Zhu and Yuille, 1995), (Gupta et al., 2012; Teng et al., 2005)
Hand gesture (glove-based)	(El Hayek et al., 2014; Kim and Chien, 2001; Dipietro et al., 2008; Lamberti and Camastra, 2012)
Head and hands	(Wren et al., 1997; Azarbayejani and Pentland, 1996), (Nickel and Stiefelhagen, 2007)

morphological filters are implemented to the images and then, normal interpolation operation is applied. This process causes that method becomes invariant respect to the scale changes. Then, the LLE transformation is applied for non-linear dimensionality reduction with preserving the neighboring relationships.

Although the use of skin model, detection of skin regions within images, is a simple and efficient method for hand postures detection, in complex backgrounds including various lighting conditions and skin-like background regions, the use of just skin information is not sufficient for accurate hand posture detection. To deal with this difficulty, the researchers in (Chuang et al., 2014) integrate image saliency with skin information to improve hand postures recognition in the complex backgrounds. A visual cortex-based method and SVM are used for feature extraction and classification, respectively.

In traditional motion detection systems for gesture recognition, the infrared markers or colored balls are attached to body, which is hard, costly and takes extra time. So, the most recent researches focus on markerless gesture recognition. These studies usually utilize the skin colors for position detection and fittings located on small regions of body on joints of the skeleton model. A gesture recognition system using model fitting and stereo vision is proposed in (Shimizu et al., 2007). To implement this gesture recognition process, at first, the skin color is used for detection of regions and edges. Then, face is detected, and from the face position, the shoulder position is also estimated. Next, the arm model is fitted with the real arm through little by little rotation of the joints of the arm model. Then, stereo disparity is used to check whether the arm model overlap and match with the center region of the human arm. Where the overlapped regions are maximized, the position of arm model (its joint angles) is fixed. The use of traditional image processing approaches such as edge detection is difficult for detection of such overlaps between the arm model and human arm. So, to deal with this problem, three kinds of fins according to Fig. 15 are considered and used. The fins are initially set on the neighboring regions of the arm and on the wrist end of arm. The fines are projected on both images with stereo disparity. The projected fins on one image are compared with other image. This comparison is done in different possible positions in the range of the arm movement. If the fins projected onto both images have the maximum overlap with the real human arm, then the arm model is correctly fitted. The advantage of this approach is that it is not necessary to cut out the human arm region from the image frame for searching the corresponding points for the stereo vision.

Many approaches for body feature extraction have been proposed using one or more cameras. Different approaches concentrate on different parts of the body such as head and hands (Ekman, 1992). The Pfinder system with a statistical model is proposed in (Wren et al., 1997), which uses the color and shape information to provide a 2D representation of head and hands. A stereo camera setup is used to add 3D-coordinates to a Pfinder for head and hands (Azarbayejani and Pentland, 1996). A dense stereo processing using disparity images provides additional information for body tracking. Face detection, color cues, disparity images and integrated framework are used for person tracking in (Darrell et al., 1998).

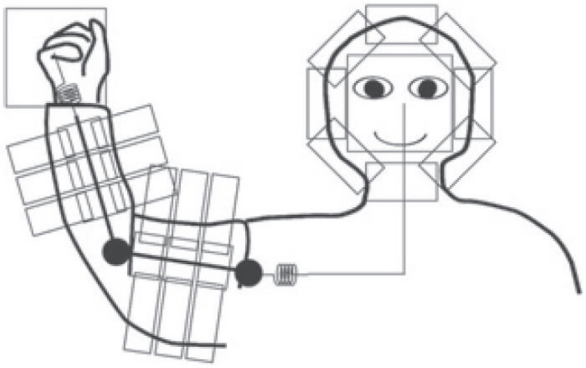


Fig. 15. Three kinds of fins for model fitting used for proposed hand gesture recognition (Shimizu et al., 2007).

The color and disparity information provided from images acquired by a calibrated stereo camera are combined through a multi-hypothesis tracking framework for finding 3D-positions of body parts in (Nickel and Stiefelhagen, 2007). An appearance-based neural-network method estimates the orientation of the head. For detection of pointing gestures, a HMM-classifier based on hands' movements is used. The gesture recognition performance is significantly improved by using additional information of head orientation.

The elastic graph matching (EGM) is an appropriate technique for object recognition (Wiskott et al., 1997). EGM has different applications in face verification, face recognition (Shin et al., 2007) and gesture recognition (Li and Wachs, 2013). EGM has robustness to lighting variation and expression changes. Another variant of EGM is morphological elastic graph matching (MEGM) (Kotropoulos et al., 2000). One of the main disadvantages of this method is its high computational complexity.

(Kleinsmith and Bianchi-Berthouze, 2013) provides a survey on body expression recognition. There are many challenges and questions in emotion recognition using body expressions and gestures: how to collect data, label, model, and set benchmarks to compare automatic emotion recognition systems. Moreover, the major difficulties in emotion recognition using hand gestures are as follows: The hands may not be tracked by variations of lighting conditions, a temporal matching method may be needed to simultaneously track both hands, an appropriate method may be needed to deal with occlusion of one hand by the other hand, and finally, the hands may be covered by some objects or clothing. The lack of standards and common datasets and also the availability of corpora are other concerns related to activity recognition systems.

Some researchers have studied the relationship between emotional state and body gesture (Caridakis et al., 2008a; Castellano et al., 2007). These relationships are shown in Table 8. Recognizing emotions from

Table 8
Relationships between emotions and body gestures.

Emotion	Gesture	References
Happiness	Hand clapping—high frequency	Caridakis et al. (2008a)
	Circular Italianate movement	Castellano et al. (2007)
	body extended	Gunes et al. (2005)
	hands kept high	
	hands made into fists and kept high	
sadness	Hands over the head—posture	Caridakis et al. (2008a)
	Smooth falling hands	Castellano et al. (2007)
Anger	Lift of the hand—high speed	Caridakis et al. (2008a)
	Italianate gestures	
	Violent descend of hands	Castellano et al. (2007)
	body extended	Gunes et al. (2005)
	hands on the waist	
	hands made into fists and kept low, close to the table surface	
fear	Hands over the head—gesture	Caridakis et al. (2008a)
	Italianate gestures	
	body contracted	Gunes et al. (2005)
	body backing	
	hands high up, trying to cover bodily parts	
Disgust	Lift of the hand—low speed	Caridakis et al. (2008a)
	Hand clapping—low frequency	
	body backing	Gunes et al. (2005)
	left/right hand touching the neck	
Surprise	Hands over the head—gesture	Caridakis et al. (2008a)
Despair	Leave me alone	Castellano et al. (2007)
Interest	Raise hands	Castellano et al. (2007)
Pleasure	Open hands	Castellano et al. (2007)
Irritation	Smooth go away	Castellano et al. (2007)
Pride	Close hands towards chest	Castellano et al. (2007)
anxiety	hands close to the table surface	Gunes et al. (2005)
	fingers moving	
	fingers tapping on the table	
uncertainty	shoulder shrug	Gunes et al. (2005)
	palms up	

body gesture can be used in affective tutoring systems, but there are some difficulties that must be considered: most of body gestures such as hand clapping are not feasible in e-learning or are feasible with a low probability. Furthermore, body gesture is highly depended to culture and location. Similar to verbal words, body signs have different meanings in different countries or cultures. So, these differences must be considered in the emotions recognition system.

11. Hybrid methods

Human individuals commonly use the multiple modalities for emotion recognition in human–human interaction. So, it is desirable that human–computer interactions also use the multiple modalities instead of one modality such as face or gestures and body movements. Table 9 represents some different hybrid methods used for emotion recognition. Generally, vision systems obtain higher recognition rates than speech systems (Anagnostopoulos et al., 2015). However, a higher recognition rate can be obtained by a multimodal system rather than the individual speech or vision systems (Song et al., 2008). Recognition of some emotions can be better done based on speech features rather than visual features and vice versa. For example in (Perez-Gaspar et al., 2016), it is observed that emotions of sadness, neutral, and anger are better recognized by performing of speech system while the emotion of happiness is better identified by vision system. This is reason why speech and vision emotion recognition systems are integrated to develop a multimodal emotion recognition system. This integration is done to consider the strengths and advantages of each emotion recognition system to give a global answer in respond to the recognition task. While most of the facial emotion recognition systems consider six emotions, most of the speech emotion recognition systems consider four and five emotions.

Table 9
Emotion recognition methods by using hybrid methods.

Multimodal system	Used techniques	References
Facial and voice features	Tripled HMMs (THMMs)	Song et al. (2008)
Facial and voice features	HMMs (used in the speech system) and PCA + ANN (used in the vision system)	(Perez-Gaspar et al., 2016)
Facial expression and audio information	Probabilistic Combination of Multiple Modalities	Kapoor et al. (2004)
Face (geometric features), behavior (persons' movement), physiological signals	SVM classifier, multimodal fusion using ANN	Fernández-Caballero et al. (2016)
Visual and EEG signals	3D fuzzy GIST, adaptive neuro-fuzzy inference (ANFIS) classifier	Lee et al. (2014b)
Facial expressions and hand gestures	Facial analysis (morphological operations + gradient filters over eyes mouth), hand-tracking (moving skin masks + HMM)	Balomenos et al. (2004)
Facial expressions and body gestures	Decision level fusion through an extension of a neural network	Caridakis et al. (2008b)
Face expression (geometrical features, gray-level information and edge maps) and upper body gestures (color based and silhouette based)	Feature-level fusion and decision-level fusion	Gunes and Piccardi (2007)
Body motion and face expression	Convolutional Neural Networks (CNN)	(Barros et al., 2015)
Facial features (dynamic and static points) and physiological information (EMG of eyebrows, skin temperature of the finger, skin conductance (SC) of the finger, respiration and BVP)	Feature-level (mutual information approach and the PCA) and decision-level (voting process and dynamic Bayesian Networks)	Maaoui et al. (2014)
Facial expression (LBP) and EEG (spectral power features)	KNN and SVM classifiers, feature-level and decision-level fusion	Huang et al. (2016)
Facial expressions (motions of eyebrows, eyes, etc.), shoulder movements and audio cues (MFCC and prosody features)	Output-Associative Relevance Vector Machines (OA-RVM)	Nicolaou et al. (2012)
Video, voice and thermal images	Feature level fusion	Yoshitomi et al. (2000)
Face and speech	HMM-based and Adaboost algorithms, both feature-level and decision-level fusion	Datcu and Rothkrantz (2009)
EEG and eye gaze data	Feature-level and decision-level fusion	Soleymani et al. (2012)
EEG (pre-frontal asymmetry features and band-power features) and acoustic features (spectral, cepstral, perceptual, temporal, and beat features)	A regression model	(Daly et al., 2015)
Acoustic (MFCC and temporal derivatives), linguistic (non-linguistic vocalizations such as sighing, breathing, and), and visual (facial movement features)	Long Short-Term Memory (LSTM) networks	Wöllmer et al. (2013)
EEG signals and self-report	FFT, ANN (for feature extraction and classification of EEG), qualitative analysis of questionnaire	Meza-Kubo et al. (2016)

HMMs and ANNs are popular techniques which have been extensively used in recent researches to address the emotion recognition problems. The Tripled HMMs (THMMs) is used in (Song et al., 2008) to provide a multimodal system. The used THMMs synchronize facial and voice features in the time domain for recognition of emotions of happiness, sadness, anger, surprise, fear, and neutral. The distances between eye-nose, mouth-nose, eyes, and width of the mouth are used as facial features. 48 prosodic and 16 formant features are used as speech features. Speech and vision systems individually obtained 81.45% and 87.40% recognition rates, respectively while the integration of both systems provides an improved recognition rate of 93.31%. Similarly (Busso et al., 2004a), and (Haq et al., 2008) are also reported improvement in recognition rates.

A multimodal emotion recognition system has been proposed in (Perez-Gaspar et al., 2016) which uses the evolutionary optimization of HMMs, ANNs, and PCA. It uses the genetic algorithms to obtain the appropriate values of parameters defined in HMMs and ANNs. The HMMs are used in the speech system while PCA + ANN are used in the vision system.

In (Fernández-Caballero et al., 2016), the emotion detection and regulation architecture is proposed for using in smart health environments. This architecture detects facial emotion, behavior, and valence/arousal. Then, the detected outputs are fused to generate a unified representation from the partial emotional states. The needed information for emotion detection is acquired and processed from different sensors in the environment. The facial emotions are taken by cameras, the behavior is also tacked by camera, and the physiological signals are acquired by

the body sensors.

The output of facial emotion detection is an emotion interpretation in terms of negative/positive (unhealthy/healthy) mood where one of the following basic emotions is obtained as output: fear, anger, disgust, happiness, sadness, and surprise. The output label of behavior detection is active or inactive. The output of valence/arousal detection is one of four possible orthogonal dimensions of affect, i.e., bored, excited, relaxed, and nervous. In the facial emotion detection stage, the geometric features of face are extracted from the shape or salient locations of facial components such as eyes and mouth. Then, the SVM classifier is used for classification. The amount of persons' movement is considered as the level of activation of him/her where the output can be the values of active or inactive. An ANN is used for multimodal fusion (see Fig. 16). The basic emotions obtained from facial emotion detection stage, the activation level obtained from the behavior detection stage, and the detected emotions obtained from the valence/arousal detection stage are used as inputs of ANN. For each of emotions in the input layer, the probability of the dominant emotion is calculated. How much probable that patient's emotion is located in one of the four possible combinations of pleasantness/activation is offered by the output layer.

A hybrid emotion recognition system using both the visual and EEG signals is proposed in (Lee et al., 2014b). Because of permeation of uncertainty within emotion recognition, possibility theory and fuzzy sets have been used to model and process such uncertainty (Russell, 1994). The conceptual and perceptual levels and their uncertainties of visual information have been modeled by Fuzzy GIST (Russell, 1980). An emotion recognition system to analyze the emotional states of human individuals while watching a video clip is proposed in (Lee et al., 2014b), which uses a 3D fuzzy GIST to extract dynamic emotional features from low-level visual features and uses a 3D fuzzy tensor to extract semantic-level features from EEG signals. The 3D fuzzy GIST containing dynamic visual features such as Lightness, Chroma, Hu, and orientation information of a video clip. To extract features from EEG signals, ICA and Short Time Fourier Transform (STFT) are applied for elimination of artifacts and extraction of reliable features, respectively. The 3D tensor data is constructed for the brain signals. The FCM clustering is applied to visual and EEG signals to obtain the 3D fuzzy GIST and 3Dfuzzy tensor, respectively. The obtained 3D fuzzy GIST and 3Dfuzzy tensor are fed to an adaptive neuro-fuzzy inference (ANFIS) classifier for emotion recognition.

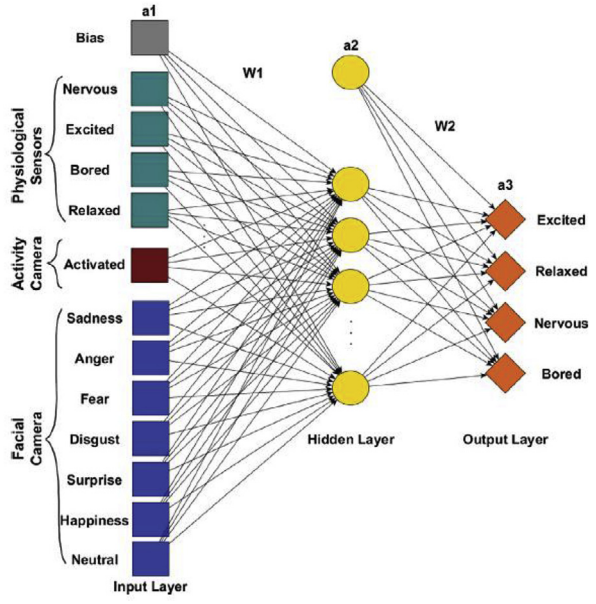


Fig. 16. The used ANN for multimodal fusion in (Fernández-Caballero et al., 2016).

The facial expression is combined with audio information for detection of emotional states of low-interest, high-interest, and refreshing in a child when is solving a puzzle (Kapoor et al., 2004). The combination of facial expressions and hand gestures is reported in (Balomenos et al., 2004). According to (Ambady and Rosenthal, 1992), the most significant cues to judgment of behaviors is visual channel through facial expressions and body gestures. Humans often combine the visual channels of face and body more than other channels when judge about other human behavior.

The combination of face expression and upper body gestures is used in (Gunes and Piccardi, 2007). One example of the recognized states of face and body when a person feels anxiety is as follows: face (stretching of the mouth, lip wipe, lip bite, and eyes turn up/down/left/right), and body (hands near to the surface of table, fingers tapping on the table, fingers moving). Two different cameras are simultaneously used. One camera captured the head only and the second camera captured upper-body movement. The face and body features are extracted individually and then, the face and body states are fused in two different levels: feature-level fusion and decision-level fusion (see Fig. 17). Examples of images recorded by two cameras are shown in Fig. 18.

For information fusion in feature-level, the features extracted from each modality are concatenated into one larger vector, and the obtained feature vector is fed to a single classifier to assign the appropriate label to testing samples. For information fusion in decision-level, at first, each modality is pre-classified independently. Then, the outputs of the different modalities are fused to obtain the final classification. Various approaches such as majority vote, sum rule, product rule, minimum/maximum/median rule, using weights, etc. have been proposed for decision fusion (Kuncheva, 2002). The sum, product and weight criteria are used for decision fusion of face and body gesture information in (Gunes and Piccardi, 2007). The experimental results indicate that the emotion recognition accuracy obtained by two combined modalities is better than classification results obtained by using only the face modality or only the body modality. Moreover, the results show that feature-level fusion provides more classification accuracy than decision-level fusion. Furthermore, among three decision fusion rule (sum, product and weight criteria), the sum rule provides the best results to fuse two modalities.

In feature-level fusion, the features extracted from each modality are merged into a single cumulative structure and fed to a single classifier such as multiple HMM or ANN. In this approach, the correlation between different modalities can be taken into account across training of the classifier. Feature fusion is appropriate for synchronized and closely coupled modalities such as lip movements and speech. But, if the temporal characteristics of features of be substantially different from each other, feature fusion cannot be generalized very well for, an example when facial expression, speech or gesture are used as input modalities. Furthermore, because of the high dimensionality of the obtained feature vectors, lots of training samples must be collected. On the other hand, decision-level fusion is appropriate for integrating asynchronous but temporally correlated modalities. In this framework, at first, each modality is independently classified. Then, the outputs obtained from different modalities are fused to achieve the final classification. Decision-level fusion cannot model the interplay between different modalities. Then, it can fortify the obtained results of each individual modality or deal with uncertainty in cases that one or more modalities are unreliable (Caridakis et al., 2008b). An emotion recognition system which fuses the facial expressions and body gestures in decision-level is proposed in (Caridakis et al., 2008b).

To fuse face and speech using HMM-based and Adaboost algorithms, both feature-level and decision-level fusion methods are utilized in (Datcu and Rothkrantz, 2009). It concludes that the fusion at the semantic level achieves better efficiency for the multimodal emotion analysis. EEG and eye gaze data are also fused in both feature-level and decision-level in (Soleymani et al., 2012). The better performance of decision-level fusion compared to feature-level fusion can be seen. But in (Busso et al., 2004b) similar results are obtained for feature-level and

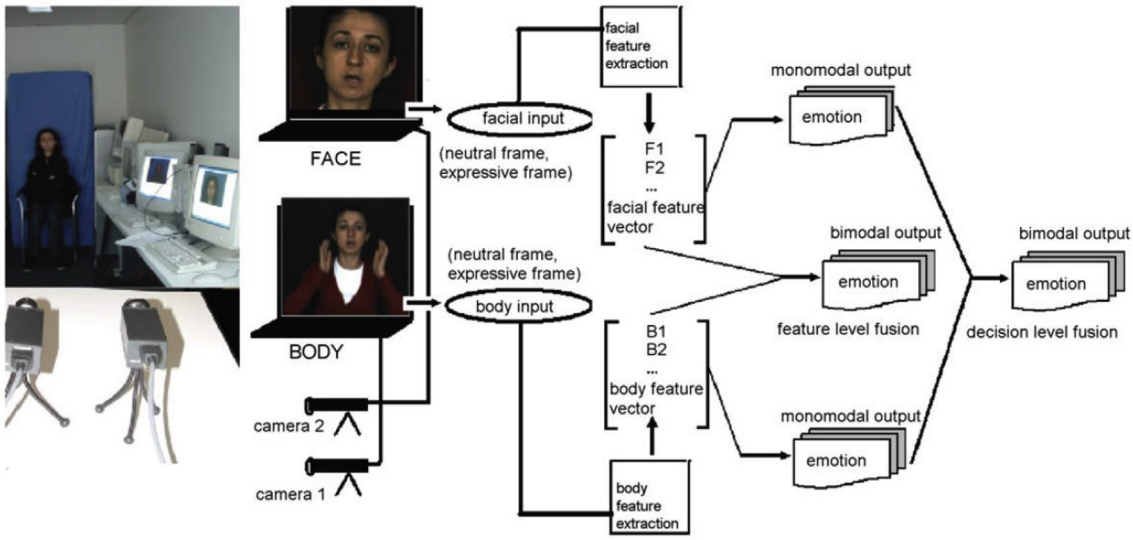


Fig. 17. Fusin of affective face and body gesture for emotion classification (Gunes and Piccardi, 2007).

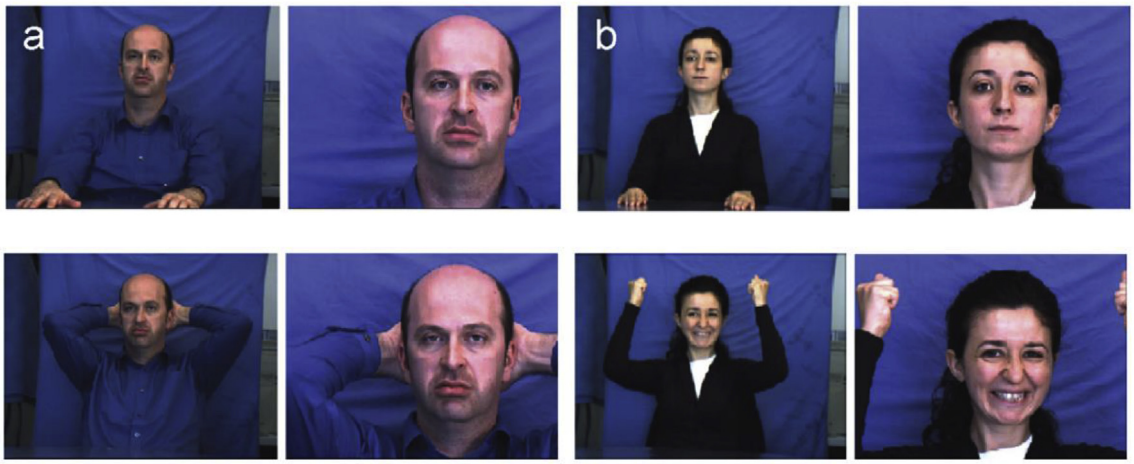


Fig. 18. Examples of images recorded by camera a (for body) and camera b (for face) (Gunes and Piccardi, 2007).

decision-level fusion of face and speech data. According to this result, the best method for modalities fusion depends on the application.

Human brain recognizes emotions aroused by different stimuli correlating varied information from different areas (Adolphs, 2002). The human brain correlates motion information and face expressions in the visual stimuli, past experiences, and other available cues and stimuli to integrate this multimodal information for emotion recognition. Neural models in computer systems, particularly hierarchical ones such as convolutional Neural Networks (CNN), can simulate this process of brain. CNNs were formally introduced by (Lecun et al., 1998). The human brain contains simple and complex cells which have a hierarchical process for extraction and learning different information from varied stimuli such as visual inputs. CNNs are inspired by this hierarchical process of cells of the human brain. Each layer of a CNN reacts to a different type of information. When, layers are stacked together, a complex representation of the visual input is created. The simplest information of the input stimuli is extracted in the first layers, which acted as edge-like detectors. More complex representations such as orientation, position, shape and image transformation can be extracted in deeper layers.

A CNN is applied for face expression recognition in (Fasel, 2002). A CNN-based emotion recognition model is proposed in (Barros et al., 2015) which extends the hierarchical visual representation to recognize the multimodal emotional states using body motion and face expression. In the first layers, an edge-like representation is generated for each stimulus. Subsequently, in the deeper ones, a complex representation of feelings is created. In manual fusion techniques, the constraints of each individual technique are accumulated and results in reduction in the generalization capability of models. But, the fusion learnt in CNN, does not sum the restrictions of the individual modalities.

The fusion of facial and physiological information is done in feature-level and decision-level in (Maaoui et al., 2014). The mutual information approach and the PCA transformation are used for feature selection and dimensionality reduction, respectively in the feature-level fusion. Two different approaches, the voting process and dynamic Bayesian networks are used for decision-level fusion.

The pyramidal Lucas-Kanade algorithm (Bouguet, 2000), with this assumption that the brightness of facial points does not change over time, is used to track the facial points detected in the first frame of the image sequence. Further, an Euclidean distance is used for representation of each facial muscle. Also, five physiological signals are used: EMG of eyebrows, skin temperature of the finger, skin conductance (SC) of the finger, respiration and BVP. 30 features are extracted for emotion recognition. The features are computed according to six parameters: the mean and standard deviation of the raw signals, the mean of the absolute values of the first/second differences of the raw signals, the mean of the absolute values of the first/second differences of the normalized signals.

In each modality, SVM is used as classifier.

A multi-modal emotion recognition framework induced by video based on facial expression (as an external channel) and EEG signals (as an internal channel) is developed in (Huang et al., 2016). The LBP and its various versions are used for feature extraction from face. The spectral power features are also extracted from EEG signals. A sequential forward floating search approach is used for dimensionality reduction. The KNN and SVM are also used as classifiers. Finally, for multi-modal emotion recognition, these two internal and external channels are fused in feature-level and decision-level to analyze emotion in the valence and arousal dimensions.

Kernel methods such as SVM and Relevance Vector Machines (RVM) are the most dominant methods used in machine learning. The output-associative RVM (OA-RVM) regression, which is an extension of traditional RVM regression capable of learning the temporal output correlations, is proposed in (Nicolaou et al., 2012) for continuous emotion prediction in a 2D arousal-valence emotional space. It uses multiple emotional cues such as facial expressions, shoulder movements and audio cues.

The automatic emotion recognition (AER) systems deal with different challenges. For instance, an emotion recognition system designed in a laboratory environment has may not by efficient enough to reflect real-life scenarios (Schuller et al., 2008). In a real environment, the emotion recognition system has to observe and listen time-continuously. Such challenges have been developed the second generation of AER systems with focusing on realistic data. These systems are able to account continuity, subtlety, complexity, and dynamics of human emotions (Gunes et al., 2011). Many researches are shifting from prototypical emotions such as happiness or anger to a continuous way by using emotional dimensions, for example, including arousal and valence. The discretized emotional dimensions such as negative vs. positive or low vs. high arousal can easily detect a defined set of user states. Another challenging problem is how to model the contextual information. Judgment an individual's emotional state just from a short isolated utterance is difficult even for humans. So, the modern AER systems consider the importance of contextual information role in perceiving and expressing emotions. Because of slowly evolving emotions over time, HMM can be used to model the feature-level contextual information within a video segment or spoken utterance. Some classification frameworks such as Long Short-Term Memory (LSTM) networks are suited to model long-range context in emotion recognition problems (Hochreiter and Schmidhuber, 1997). LSTM incorporates a self-learned and arbitrary amount of context into the decoding process. But, applying the LSTM architecture is a relatively hard for audiovisual emotion recognition methods such as method introduced in (Schuller et al., 2011). A LSTM framework has been proposed for emotion classification through exploiting acoustic, linguistic, and visual information in (Wöllmer et al., 2013). The block diagram of the proposed LSTM-based audio-visual emotion recognition method of (Wöllmer et al., 2013) is shown in Fig. 19. The acoustic features, linguistic feature, and facial features are fused for prediction of current emotion state in the LSTM network.

The precision of emotion recognition is increased when in addition to individuals' expressions, the context information is also considered. A study on children from the age of 5–15 years indicated that integration of visual context information with facial expression improves facial emotion recognition (Theurel et al., 2016). This study emphasize that this ability of integration exists in human from the age of 5 years. By inspiration of integration ability of human, we can design machines that automatically detect emotions expressions, extract the context information, and then integrate them to do emotion recognition task with high precision.

The emotion recognition method proposed in (Meza-Kubo et al., 2016) utilizes 3 complementing methods to increase the recognition accuracy. The FFT extracts features of EEG signals. The signal powers are classified into brain waves, namely, Alpha, Beta and Theta. Then, a neural network with hidden layer containing 151 neurons is used to detect pleasant emotions from unpleasant ones. To validate the results obtained by the neural network, additional self-report and qualitative analysis is also done. To do the self-report evaluation, the individuals are asked to answer a questionnaire after concluding the cognitive stimulation activity. To do the qualitative analysis evaluation, the cognitive stimulation activities of participants are recorded and analyzed by an expert observer. Moreover, to reduce the studies costs, the researchers suggest to do just two emotion recognition methods with the least cost demanding, i.e., self-report and neurophysiological, and conduct the most cost demanding method, i.e., qualitative analysis by expert observer only on the cases that two former evaluations do not provide the consistent results. Facial expression recognition, questionnaires, Heart rate and blood pressure are also used together for emotion recognition during cocaine intoxication in (Kuypers et al., 2015).

12. Comparison among different emotion recognition methods for e-learning goals

All the reviewed emotion recognition methods have their advantages and disadvantages. Which method is appropriate is application-dependent. For example, when the users' perceptions and explanations of the felt emotions are important in a study, the self-report methods should be selected. However, it is better that a researcher considers several methods to get a comprehensive representation of the users' emotional states, ensures the consistency and accuracy of the collected data, and increases the reliability of inferences. In some cases that there is Poke face and the lack of facial expression, other representatives of emotional expressions are used.

A comparison between different discussed emotion recognition methods is given in Table 10. This comparison is done by considering their suitability for e-learning purposes. There are some works in the case of emotion recognition in e-learning systems. The Affective Tutoring System (ATS) firstly has been used by Alexander and Vicente (Alexander et al., 2003; de Vicente, 2003). Although some researchers have worked on emotion recognition in ATSs, but they have utilized one of the emotion recognition methods with no changes. Except a few researches, e-learning context is not considered enough in selecting or designing an emotion recognition method (Kardan and Einavypour, 2008b). None of

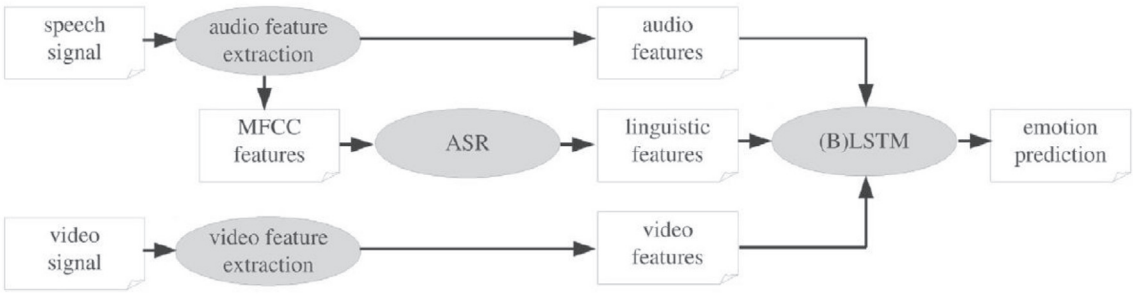


Fig. 19. The proposed emotion recognition method in (Wöllmer et al., 2013).

Table 10
Comparison of emotion recognition methods for e-learning environments.

Method	Advantages	Disadvantages
Asking from user	1- Simplicity 2- No need to additional instrument	1- Probably is annoying for learner 2- Emotions are temporary and are changed during the session 3- Learner may be not honest
Tracking implicit parameters	1- No need to additional instrument 2- Using some information of learner model for deduction	1- Sophistication 2- Low preciseness 3- Need to save a lot of information
Voice recognition	1- Suitability for learning languages 2- Distinguishing between high arousal and low arousal emotions	1- Not useable for all learning contents 2- Not able to distinguish emotions with the same arousal
Facial recognition	1- Ability to determine 7 basic emotional states 2- Availability of webcams	1- Time consuming processes 2- Learners may have no enough changes in their faces
Vital signals	1- Distinguishing between high arousal and low arousal emotions 2- High preciseness if many parameters being used 3- Its psychological basic	1- It needs some specific instruments and sensors which are not available for all learners 2- Connection of sensors may annoy the learner
Body gesture	1- Relatively is precise for some emotions	1- Time consuming processes 2- Culture dependency 3- Many body signs are not accrue in learning process

them are applicable in the actual e-learning systems. E-learning systems need a particular method that should be applicable for real learners that are utilizing the system in their home. Therefore, this comparison could be used as a guide for designing an efficient emotion recognition method for e-learning environments. In the following, more comparisons among emotion recognition methods are represented.

Based on the type and amount of user's involvement in emotion recognition process, the learners' emotion detection approaches are divided into three main groups: manual, semi-automatic and automatic (AliAkber Dewan, 2019). In the manual approaches, learner has direct involvement in the emotion detection process. Self-reporting and observational checklist are two popular manual methods. In self-reporting approach, the learners represent their own level of excitement, distraction, attention or boredom.

Self-reporting is simply implemented and provides useful information about learners' emotions. But, the validity of it is dependent to the learners' honesty and their willingness to represent their feelings and also the true perception of learners about their emotions. In the observational checklist, the questionnaire is completed by an external observer such as teacher. The external observer can complete the questionnaire by observing the online or pre-recorded videos of the learning activities. As a limitation of this approach, the assessment metrics of observer may be not related to the learner's emotion. For example, measures such as good behavior and sitting quietly is more related to compliance of regulations rather than emotional state. The manual methods such as self-reporting and observational checklist are time spending and require effort from the learners or the observers for emotion detection.

Some emotion detection methods using implicit parameters are located in the semi-automatic approaches. In these cases, the learners indirectly involve in the emotion detection process. A popular method in this category is engagement tracking where the accuracy and timing of learner for responding to the test questions and practice problems are evaluated. For example, given random answers to the test questions without any effort for thinking or simply giving short responses on easy questions indicate that the learner is not engaged, he/she is not interest in testing contents or is bored.

Many used emotion recognition methods in e-learning environments are located in the automatic category. Emotion recognition using facial expressions, voice (speech), vital (physiological) signals and gestures recognition or some techniques of tracking implicit parameters such as learners' messages analysis are done automatically. In the automatic emotion recognition methods, the data is acquired from the log files or sensors and then, some computer vision or machine learning tools are used for emotion recognition through three main steps of pre-processing, feature extraction and classification.

To have a comprehensive comparison among various emotion recognition methods, the advantages and disadvantages of each of them are represented in the following.

- *Asking from user*

The online self-reports that acquire the learners current experience are more valid than self-reports concerning past or future experiences of emotions. Although the self-report methods are implemented simply, they can be intrusive for the learner since asking irrelevant questions has negative effect on the ongoing process. Due to the temporary nature of emotions, it is essential to interrupt several times during the learning process that is annoying for the learner. In addition, the learners may lie about their emotions to relate their mistakes to inappropriate emotional states no their weakness in understanding the educational contents. So, the self-report methods are not lonely used in the e-learning systems. They should be used in parallel with other methods or for validating them. That is the emotions are detected by other methods and the detection accuracy can be investigated through asking the learner.

- *Tracking implicit parameters*

Tracking the implicit parameters of learner such as the number of mistakes, written sentences or analyzing the mouse movements is applicable for e-learning purposes because it does not need any additional devices. But, this method has two main disadvantages: complexity and low precision. Finding a reasonable relationship among the implicit parameters and specific emotional states is a hard task. This method should be used for detection of emotional states polarity not for determination of specific emotional states.

A basic tool for information exchanging in an e-learning environment is message where the learner's emotions are sensed from the sentences and even individual words. In addition to words, the learners use a large number of pictures and signs which are a direct reference of emotions. Sending positive pictures and words show interest to lesson while sending negative ones indicate that the learners are not interested in the learning context.

- *Voice recognition*

The speed and level of learners sounds can be reflect their emotions. For example, when learners accelerate their speech speed and raise their tone, they are interested in what they talk about it; otherwise, they slowly speak in a serious tone. Speech contains features such as pitch, timing and articulation that have strong relation to the emotional state of human. Pitch is considered as an index of arousal. While some emotions such as happiness, anger and fear lead to fast and loud speech with high frequency energy, higher average pitch and wider pitch range, passive emotions such as sadness result in slow speech with low pitch and low frequency energy.

Among various emotions, sadness, anger and fear are best recognized through the voice and disgust is the worst. Distinguishing between active emotions such as happiness and anger or between passive emotions such as sadness and boredom are also challenging problems. The features of voice can be extracted locally from each frame or globally from the whole voice. Although global feature extraction is faster (due to less number of global features with respect to the local ones), but, it is efficient only in distinguishing between high-arousal emotions such as anger, fear and joy

versus low-arousal ones such as sadness. The global features fail to distinguish between emotions with similar arousal, for example, in discrimination between anger and joy. In addition, the temporal information in speech signals is completely lost in the global features. Also, due to limited number of training global feature vectors, the use of global features in complex classifiers such as SVM or HMM is not suggested.

The use of voice for emotion recognition in e-learning systems has two main difficulties. First, in many cases such as mathematics, there is no need for the learner to speak, and therefore, there is no voice for emotion recognition. Second, discrimination between some emotions such as happiness and anger, boredom or sadness is a difficult task. So, an additional emotion recognition should be applied with voice recognition method.

- *Facial expression recognition*

Human face is an important indicator of emotions and non-verbal means of communication. The facial expression recognition methods have been divided into two main groups: 1) judgment based methods where a message is communicated by facial expression; for example, a smile indicates happiness, and 2) sign based methods where occurring frequency and types of movements describe a behavior. Factors such as how long a face movement lasts or how many times face moves leads to an emotion detection. For instance, movement of lip corners towards up during a specific time composes a smile. The sign based approaches function as coders such as FACS systems. According to the FACS, the face movements described by a set of action muscular units compose various facial expressions. However, many defined facial action units have not meaningful facial expressions and do not show any specific emotional states. In addition, the facial emotion recognition methods are based on appearance without considering any reliable information about cognitive and learning process.

For facial expression recognition in a learning environment, a real facial expression database containing facial expressions of learners is required. But, in most studies, databases containing only six basic emotions have been used. Usually, facial emotion recognition methods can detect surprise, neutral and happy with the highest accuracy. In contrast, disgust and anger have many similar facial expressions that makes confusion in recognition of them.

Some factors can disturb the facial expression recognition: partial covering of face with a hand, looking around, uneven illumination and location of camera. When there are two cameras in the e-learning environment, one in up and one in low, some inconsistency can be observed between the results obtained by two cameras. For example, the upper camera tends to overestimate anger while the lower one overestimates surprise (Landowska et al., 2017).

Facial emotion recognition is suitable for using in e-learning context. Most computers have webcams located in front of the face. In addition, facial image processing can be done with a high precise. The weakness of this approach is the relatively static state of the learner's face during the learning process because the learning process cannot make the emotional excitements. It means that a learner usually does not change his/her face significantly or does not laugh during learning.

Investigations of user's movement magnitude can be also used for estimation of emotions (Coombes et al., 2006). The most robust behavioral component is the eye blink. That index startles magnitude of human. Although this method is suitable in computer games, but, it is not an appropriate approach in e-learning environment because a learner has not significant movement during the learning process.

- *Vital signals*

Based on physiological studies, emotions have significant effects on vital signals of human body. While the sympathetic nervous system is aroused by emotions of anger, fear and joy, the parasympathetic nervous system is aroused by sadness. The result is change in blood pressure, heart rate, depth of respiratory movements, occasional muscle tremor and so on.

Physiological measures and vital signals are more difficult to manipulate or conceal than vocal utterances, facial expression, gestures and implicit parameters such as user texts and messages. So, they are more reliable measures for inner feelings and emotion recognition in e-learning environments. But, as a difficulty in discrimination among various emotions, the same emotion could elicit different physiological patterns where different appeared features lead to wrong detected emotions. In addition, comparing to speech recognition and computer vision, collecting reliable vital signals as bio data is a challenging and hard task that can be annoying for the learner. While the face images are simply taken by cameras and webcams and speeches are easily acquired by microphones, but there are some factors that affect the reliability of vital signals acquired by bio-sensors. For instance, the value of gel used under electrode, the air humidity and how tight the electrodes are placed can impact the readings. In addition, the learning lab that the learner is sat should be comfortable enough. The inconvenience feeling of measuring tools and sensors should not effect on real emotions of the learner. The appropriate number and type of the used sensors also must be chosen accurately. Skin conductance, blood volume pressure and EEG are the most popular signals sensed for measuring of physiological patterns of the learners. Moreover, because the physiological features do not significantly change during learning, emotion detection is hard in the learning process.

Although by using the vital signals, arousal emotions can be separated from the passive ones, but, complete discrimination among emotions with the same arousal is not possible. In addition, emotion recognition using vital signals is not suitable especially for e-learning environment because the use of bio-sensors is intrusive for the learner and also the bio-sensors may be not available for all learners.

- *Gesture recognition*

Body gesture and posture can indicate the people emotions. Posture and body movements have been considered as indicates of emotional states in many works. While some methods use 2D photos with low level visual features, other complicated ones capture the position of body parts using 3D information. Computing distances and angles between different parts of body can be used as appropriate features for recognition of body posture. In some works, video is used for automatic tracking of body movements. However, body movement coding is a hard task, and up to now, no specific theoretical framework has introduced to define basic units of movements for emotion detection through gesture and body movements. In other words, no standard body movement coding method has been introduced for gesture expression. Other difficulty is difference of body gestures in various cultures and ecological validity. An observed behavior from a learner can be different in the laboratory environment with that in a normal environment with natural situation. However, the use of body movements beside the face images can be more effective than just using static images of body gestures or facial expression.

The information offered by posture is sometimes unavailable from the conventional nonverbal measures such as face over long distances. The main advantage of gesture based emotion detection is that the body motions are usually unconscious and unintentional. Emotion recognition using body gesture is relatively a precise technique especially in combination with facial expression. But, the use of body gesture for emotion recognition in e-learning systems have some deficiencies. Most body gestures such as hand clapping are rarely happen during the learning process. In addition, body gesture is highly dependent to location and culture where the body signs have different meanings in different counties or cultures.

13. Conclusions and future works

In recent years, emotions recognition has become an active research

area. It is also considered in e-learning systems. There are varied emotion recognition approaches but all of them are not suitable for e-learning environments. This study has reviewed different methods for emotion recognition such as asking from the user, tracking implicit parameters, voice recognition, facial expression recognition, vital signals and gesture recognition. The advantages and disadvantages of each category of these methods are discussed. Especially, the methods' suitability are compared for using in e-learning systems. It is hoped that this work can be utilized as a guide for designing emotion recognition component of the affective tutoring systems.

Three main aspects have been seen in recently advanced works of emotion recognition in electronic environments and continued in future works:

1) Internet of things (IOT)
2) Information fusion
3) Deep learning

Due to rapid development of IOT and availability of mobile internet, smart phones and great achievements in wearable technology such as body and brain wearable equipment, communication between human and machine becomes more and more (Santos et al., 2020). Almost all persons have smartphones that can acquire, process or send information of users. Non-intrusive, convenient and comfortable body wearable equipment have been designed to collect user's EEG, blood pressure and other physiological data. The voice related and camera modules have been used to collect the user's voice and facial expression. In addition, thanks to development of intelligent terminals and mobile networks, a large number of users announce their ideas and emotions in social networks. So, a huge source of information about users emotions have been collected in real time and transfer to the data centers through the wireless communication systems for data processing and decision making about human-machine interaction. Due to availability of various data sources and due to lack of single modal emotion recognition methods, the trend of researchers is toward multi-modal recognition techniques through information fusion. In other words, the new emotion recognition methods often use two or more types of emotion data such as facial expression, vital signals, voice and text (Jiang et al., 2020).

Many recent works have been focused on artificial intelligent based methods especially deep learning framework (Qin et al., 2020). A multi-layer deep structure can automatically do feature extraction and classification with a high accuracy. A deep learning model can be implemented in both supervised ad unsupervised manner where a large number of training samples is required for supervised learning. Thanks to IOT and advanced sensors, a big data is available from various sources of human (Muzammal et al., 2020).

The most studies about emotion recognition consider just six basic emotions of angry, disgust, sad, scared, happy and surprised. Each group of these emotions can indicate special states of the learners. For example, angry or disgusted indicates weary; scared or sad shows quiet while surprised or happy indicates concentrated. When the learners are in a state of excitement, they are concentrated and encouraged to be kept in this state for a long time, when the learners are in an unhappy or uninterested state are quiet and it is time to conduct them for going on studying. The teacher agent should wake up the learner's excitement. When the learners are in sleep or painful state, they are weary and it is time to cheer them up. Emotions, cognitive states and learning rate of learners are radical interactive (Ma, 2016).

Among various emotion recognition techniques, the methods based on voice recognition or vital signals can just appropriately separate active emotions with high arousal from the passive ones with low arousal. But, precise separation of emotions with the same arousal using voice or vital signals is impossible. In addition, the learner may not speak during the learning process or the bio-sensors are not available for all the learners. Emotion recognition by tracking implicit parameters such as mouse movements or time duration for responding to test questions is also a hard task because finding an exact relationship between the implicit parameters and human emotions is unknown. Self-reporting from the learners is a simple but sometimes intrusive tool for emotion recognition of learners. Due to availability of cameras on the computers and wide advances in the machine learning domain for image and video processing, the facial emotion recognition beside the body gesture analysis can provide an accurate emotion recognition in the e-learning environment. Face and body gesture analysis indicate different emotions of learners more obvious than other methods. In addition, the learners communicate through the virtual learning environment by sending messages containing texts, pictures and signs that analysis of them can be useful for increase of emotion detection accuracy. Occasionally self-reports of learner about his/her emotions beside the other methods help to decrease the error rate of emotion detection. As a conclusion, a hybrid emotion recognition framework combining face, gesture, text and self-reporting analysis can be a precise and applicable system for e-learning environments.

Until now, a comprehensive method for emotion recognition in e-learning systems is not presented. E-learning is a specific activity and it has some particular characteristics such as learner model. Considering these characteristics, a more tailored way for emotion recognition can be opened. Thus, there are two main research areas that should be more considered in the future:

1 Designing a particular method (probably a combined method) for learner's emotions recognition in the affective tutoring systems.
2 Using information about learner's emotional states for adjusting teaching method or educational content.

To implement automatic emotion recognition methods using voice signals, gestures and facial expressions and also to have a fair comparison among different methods, presence of public, comprehensive, large enough and various datasets is necessary. It should be noted that the learner's emotion detection is biased toward the learner's age, demographic variables, geographic location and culture. So, the provided dataset should have demographic variability. In addition, the validity and reliability of the training label have to be ensured by human experts. Specially, due to fast growth of deep learning methods such as CNNs and the superior advances of them in recognition and classification problems such as emotion recognition, providing more volumes of data is an essential requirement. Future works can be concentrated on designing appropriate deep learning and creating real databases for training the deep networks.

References

Abbas, A., Khalil, M., Abdel-Hay, S., Fahmy, H., 2008. Expression and illumination invariant preprocessing technique for face recognition. In: International Conferenceon Computer Engineering & Systems. IEEE, Cairo, pp. 59–64.
Aceto, Giuseppe, Persico, Valerio, Pescapé, Antonio, 2018. The role of Information and Communication Technologies in healthcare: taxonomies, perspectives, and challenges. J. Netw. Comput. Appl. 107, 125–154.
Adolphs, R., 2002. Neural systems for recognizing emotion. Curr. Opin. Neurobiol. 12 (2), 169–177.
Ahmad, Baylari, Montazer, Gholam Ali, 2009. Design a personalized e-learning system based on item response theory and artificial neural network approach. Expert Syst. Appl. 36, 8013–8021.
al-shalchi, Olla najah, March 2009. The effectiveness and development of on-line discussions. MERLOT journal of on-line learning and teaching 5 (1).
Alexander, S.T.V., Sarrafzadeh, A., Fan, C., 2003. Pay attention! the computer is watching: affective tutoring systems. In: Proceedings of e-learn, Phoenix, Arizona.
Alexander Lerch, July 2012. An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics. Wiley- IEEE Press, p. 272.
Ali, Rahimi, Askari Bigdeli, Rouhollah, 2014. The broaden-and-build theory of positive emotions in second language learning. Procedia - Social and Behavioral Sciences 159, 795–801.
Ali Akber Dewan, M., 2019. Mahbub Murshed and Fuhua Lin, Engagement detection in online learning: a review, Dewan et al. Smart Learning Environments 6 (1), 1–20.
Allanson, Jennifer, Fairclough, Stephen H., October 2004. A research agenda for physiological computing. Interact. Comput. 16 (5), 857–878.
Almaev, T.R., Valstar, M.F., 2013. Local gabor binary patterns from three orthogonal planes for automatic facial expression recognition. In: Affective Computing and

Intelligent Interaction (ACII), 2013 Humaine Association Conference on, IEEE, pp. 356–361.
Alyuz, N., Gokberk, B., Akarun, L., 2010. Regional registration for expression resistant 3D face recognition. IEEE Trans. Inf. Forensics Secur. (3), 425–440.
Ambady, N., Rosenthal, R., 1992. Thin slices of expressive behavior as predictors of interpersonal consequences: a metaanalysis. Psychol. Bull. 111 (2), 256–274.
an, Treur, Umair, Muhammad, October 2015. Emotions as a vehicle for rationality: rational decision making models based on emotion-related valuing and Hebbian learning. Biologically Inspired Cognitive Architectures 14, 40–56.
Anagnostopoulos, C.N., Iliou, T., Giannoukos, I., 2015. Features and classifiers for emotion recognition from speech: a survey from 20 0 0 to 2011. Artif. Intell. Rev. 43, 155–177.
Arapakis, I., Moshfeghi, Y., Joho, H., Ren, R., Hannah, D., Jose, J.M., 2009. Enriching user profiling with affective features for the improvement of a multimodal recommender system. In: Proceeding of the ACM International Conference on Image and Video Retrieval. ACM, New York, NY, USA, pp. 1–8.
Atkinson, John, Campos, Daniel, 2016. Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. Expert Syst. Appl. 47 (1 April), 35–41.
Atkinson, Anthony P., Tunstall, Mary L., Dittrich, Winand H., July 2007. Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. Cognition 104 (1), 59–72.
El Ayadi, Moataz, Kamel, Mohamed S., Karray, Fakhri, 2011. Survey on speech emotion recognition: features, classification schemes, and databases. Pattern Recognit. 44 (3), 572–587.
Azarbayejani, A., Pentland, A., 1996. Real-time self-calibrating stereo person tracking using 3-D shape estimation from blob features. In: Proceedings of 13th ICPR.
Ball, G., Breese, J., 1998. Emotion and personality in a conversational character. In: Workshop on Embodied Conversational Characters, pp. 189–219. Published in book: Embodied conversational agents.
Balomenos, T., Raouzaiou, A., Ioannou, S., Drosopoulos, A., Karpouzis, K., Kollias, S.D., 2004. Emotion analysis in man–machine interaction systems. In: Lecture Notes in Computer Science, vol. 3361. Springer, Berlin, pp. 318–328.
Bao, S., Xu, S., Zhang, L., Yan, R., Su, Z., Han, D., Yu, Y., 2012. Mining social emotions from affective text. IEEE Trans. Knowl. Data Eng. 24, 1658–1670.
Barclay, C.D., Cutting, J.E., Kozlowski, L.T., 1978. Temporal and spatial factors in gait perception that inXuence gender recognition. Percept. Psychophys. 23, 145–152.
Barros, Pablo, Jirak, Doreen, Weber, Cornelius, Wermter, Stefan, 2015. Multimodal emotional state recognition using sequence-dependent deep hierarchical features. Neural Netw. 72, 140–151.
Bilal, D., Bachir, I., 2007. Children's interaction with cross-cultural and multilingual digital libraries ii: information seeking, success, and affective experience. Inf. Process. Manag.: Int. J. 43 (1), 65–80.
Bilal, D., Kirby, J., 2002. Differences and similarities in information seeking: children and adults as web users. Inf. Process. Manag.: Int. J. 38 (5), 649–670.
Bloom, B., Engelhart, M., Murst, E., Hill, W., Drathwohl, D., 1956. Taxonomy of Educational Objectives: Handbook I, Cognitive Domain. Longman.
Bo, Cheng, et al., 2013. Silentsense: silent user identification via touch and movement behavioral biometrics. In: Proceedings of the 19th Annual International Conference on Mobile Computing & Networking. ACM.
Bouguet, J., 2000. Pyramidal Implementation of the Lucas Kanade Feature Tracker. Intel Corporation Microprocessor Research Labs.
Bowyer, K.W., Chang, K., Flynn, P., 2006. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. Comput. Vis. Image Understand. (1), 1–15.
Bozhkov, Lachezar, Koprinkova-Hristova, Petia, Georgieva, Petia, June 2016. Learning to decode human emotions with echo state networks. Neural Netw. 78, 112–119.
Breazeal, C., 2003. Emotion and sociable humanoid robots. Int. J. Hum. Comput. Stud. 59, 119–155.
Buciu, K., Pitas, I., 2008. An analysis of facial expression recognition under partial face image occlusion, Image and Vision. Computing 26 (7), 1052–1067.
Busso, C., Narayanan, S.S., 2008. The expression and perception of emotions: comparing assessments of self versus others. In: Proc. Interspeech 2008, pp. 257–260.
Busso, C., Lee, C.M., Yildirim, S., Bulut, M., Kazemzadeh, A., Deng, Z., Lee, S., Neumann, U., Narayanan, S., 2004. Analysis of emotion recogition using facial expressions, speech and multimodal information. In: Proceedings of the Interna-Tional Conference on Multimodal Interfaces. ICMI '04, pp. 205–211.
Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C.M., Kazemzadeh, A., Lee, S., Neumann, U., Narayanan, S., 2004. Analysis of emotion recognition using facial expressions, speech and multimodal information. In: 6th International Conference on Multimodal Interfaces. ICMI '04, pp. 205–211.
Busso, C., Lee, S., Narayanan, S., 2009. Analysis of emotionally salient aspects of fundamental frequency for emotion detection. IEEE Trans. Audio Speech Lang. Process. 17 (4), 582–596.
Busso, C., Metallinou, A., Narayanan, S.S., 2011. Iterative feature normalization for emotional speech detection. In: ICASSP, pp. 5692–5695.
Calvo, M.G., Nummenmaa, L., 2011. Time course of discrimination between emotional facial expressions: the role of visual saliency. Vis. Res. 51 (15), 1751–1759.
Cao, Houwei, Verma, Ragini, Nenkova, Ani, January 2015. Speaker-sensitive emotion recognition via ranking: studies on acted and spontaneous speech. Comput. Speech Lang 29 (1), 186–202.
Caputi, Valentina, Garrido, Antonio, 2015. Student-oriented planning of e-learning contents for Moodle. J. Netw. Comput. Appl. 53, 115–127.
Caridakis, G., Karpouzis, K., Kollias, S., 2008. User and context adaptive neural networks for emotion recognition. Neurocomputing 71, 2553–2562.
Caridakis, G., Karpouzis, K., Kollias, S., 2008. User and context adaptive neural networks for emotion recognition. Neurocomputing 71, 2553–2562.
Caridakis, G., Moutselos, K., Maglogiannis, I., 2013. Natural Interaction expressivity modeling and analysis. In: Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments, ACM, May, p. 40.
Casile, A., Giese, M.A., 2005. Critical features for the recognition of biological motion. J. Vis. 5, 348–360.
Castellano, G., Kessous, L., Caridakis, G., 2007. Multimodal Emotion Recognition from Expressive Faces, Body Gestures and Speech.
Castro, M.N., Vigob, D.E., Chu, E.M., Fahrer, R.D., Achá val, D., Costanzo, E.Y., Leiguarda, R.C., Nogué sa, M., Cardinali, D.P., Guinjoan, S.M., 2009. Heart rate variability response to mental arithmetic stress is abnormal in first-degree relatives of individuals with schizophrenia. Schizophr. Res. 109, 134–140.
Chaffar, S., Frasson, C., 2004. Using an emotional intelligent agent to improve the learner's performance. In: Social and Emotional Intelligence in Learning Environments Workshop, 7th International Conference on Intelligent Tutoring System, Brazil, pp. 37–43.
Chai, Xin, Wang, Qisong, Zhao, Yongping, Liu, Xin, Ou, Bai, Li, Yongqiang, 1 December 2016. Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition. Comput. Biol. Med. 79, 205–214.
Chakraborty, Rupayan, Pandharipande, Meghna, Kumar Kopparapu, Sunil, 2016. Knowledge-based framework for intelligent emotion recognition in spontaneous speech. Procedia Computer Science 96, 587–596.
Chambers, S., Breazel, C., Atkins, A., Revis, M., Asher, J., Craft, A., Westelman, R., Kotelly, B., Smith, L., 2015. Meet Jibo, the world's first family robot. http://www.jibo.com/. (Accessed 30 June 2015).
Chan, C.H., Jones, G.J.F., 2005. Affect-based indexing and retrieval of films. In: Proceedings of the 13th Annual ACM International Conference on Multimedia. ACM, New York, NY, USA, pp. 427–430.
Chan, C., Goswami, B., Kittler, J., Christmas, W., 2012. Local ordinal contrast pattern histograms for spatiotemporal, lip-based speaker authentication. IEEE Trans. Inf. Forensics Secur. 7 (2), 602–612.
Chandra, A., Calderon, T., 2005. Challenges and constraints to the diffusion of biometrics in information systems. Commun. ACM 48 (12), 101–106.
Chao, W.L., Ding, J.J., Liu, J.Z., 2015. Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection. J. Signal Process. 2, 552–561.
Chapman, Alan. Body language, how to read body language signs and gestures - non-verbal communications - male and female, for work, social, dating, and mating relationships. Available at. http://www.businessballs.com/body-language.htm. download on 2013.
Chen, Chih-Ming, Wang, Hui-Ping, July 2011. Using emotion recognition technology to assess the effects of different multimedia materials on learning emotion and performance. Libr. Inf. Sci. Res. 33 (3), 244–255.
Chen, J., He, C., Zhao, G., Pietikainen, M., Chen, X., Gao, W., 2010. WLD: a robust local image descriptor. IEEE Trans. Pattern Anal. Mach. Intell. 32 (9), 1705–1720.
Chen, Yumiao, Yang, Zhongliang, Wang, Jiangping, 30 November 2015. Eyebrow emotional expression recognition using surface EMG signals. Neurocomputing 168, 871–879.
Chenchah, Farah, Lachiri, Zied, 2014. Speech emotion recognition in acted and spontaneous context. Procedia Computer Science 39, 139–145.
Chenchah, Farah, Lachiri, Zied, 1 January 2017. A bio-inspired emotion recognition system under real-life conditions. Appl. Acoust. 115, 6–14.
Chen, Lijiang, Xia, Mao, Xue, Yuli, Lee, Lung Cheng, 2012. Speech emotion recognition: features and classification models. Digit. Signal Process. 22 (6), 1154–1160.
Chien, Shing Ooi, Kah, Phooi Seng, Ang, Li-Minn, Chew, Li Wern, 1 October 2014. A new approach of audio emotion recognition. Expert Syst. Appl. 41 (13), 5858–5869.
On, Chin Kim, Pandiyan, Paulraj M., Yaacob, Sazali, Saudi, Azali, 2006. Mel-frequency cepstral coefficient analysis in speech recognition. In: Proceedings of International Conference on Computing & Informatics, pp. 1–5.
Chin Neoh, Siew, Zhang, Li, Mistry, Kamlesh, Alamgir Hossain, Mohammed, Lim, Chee Peng, Aslam, Nauman, Kinghorn, Philip, 2015. Intelligent facial emotion recognition using a layered encoding cascade optimization model. Appl. Soft Comput. 34 (September), 72–93.
Choi, D., Park, J., Oh, Y., 2015. Unsupervised rapid speaker adaptation based on selective eigen voice merging for user-specific voice interaction. Eng. Appl. Artif. Intell. 40, 95–102.
Christian, K., et al., 1997. Human–machine interface for a VR-based medical imaging environment. In: Proceedings of SPIE—The International Societyfor Optical Engineering, vol. 3031, pp. 527–534.
Chuang, Yuelong, Chen, Ling, Chen, Gencai, 10 June 2014. Saliency-guided improvement for hand posture detection and recognition. Neurocomputing 133, 404–415.
Chunling, M., Prendinger, H., Ishizuka, M., 2005. Emotion estimation and reasoning based on affective textual interaction. Affective Comput. Intell. Interact. 3784, 622–628.
Clavel, C., Vasilescu, I., Devillers, L., Richard, G., Ehrette, T., June 2008. Fear-type emotion recognition for future audio-based surveillance systems. Speech Commun. 50 (6), 487–503.
Cohen, I., Sebe, N., Garg, A., Chen, L., Huang, T., 2003. Facial expression recognition from video sequences: temporal and staticmodeling. Comput. Vis. Image Understand. 91 (1–2), 160–187.
Consulting, e-Learning. What is e-Learning. Available at. http://www.e-learningconsulting.com/consulting/what/elearning.html. download on March 2013.
Coombes, Stephen A., Cauraugh, James H., Janelle, Christopher M., 2006. Emotion and movement: activation of defensive circuitry alters the magnitude of a sustained muscle contraction. Neurosci. Lett. 396 (3), 192–196.

Corneanu, C.A., Simón, M.O., Cohn, J.F., Guerrero, S.E., Aug. 1 2016. Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: history, trends, and affect-related applications. IEEE Trans. Pattern Anal. Mach. Intell. 38 (8), 1548–1568.
Coulson, M., 2004. Attributing emotion to static body postures: recognition accuracy, confusions, and viewpoint dependence. J. Nonverbal Behav. 28, 117–139.
Critcher, C.R., Ferguson, M.J., 2011. Affect in the abstract: abstract mindsets promote sensitivity to affect. J. Exp. Soc. Psychol. 47 (6), 1185–1191.
Cui, Y., Weng, J.J., 1996. Hand sign recognition from intensityimage sequences with complex backgrounds. In: Proceedings of the IEEE Second International Conference on Automatic Face and Gesture Recognition.
Dai, Hua, Luo, Xin (Robert), Liao, Qinyu, Cao, Mukun, 2015b. Explaining consumer satisfaction of services: the role of innovativeness and emotion in an electronic mediated environment. Decis. Support Syst. 70, 97–106.
Dai, Weihui, Han, Dongmei, Dai, Yonghui, Xu, Dongrong, 2015a. Emotion recognition and affective computing on vocal social media. Inf. Manag. 52 (7), 777–788.
Daly, Ian, Williams, Duncan, Hallowell, James, Hwang, Faustina, Kirke, Alexis, Malik, Asad, Weaver, James, Miranda, Eduardo, Nasuto, Slawomir J., 2015. Music-induced emotions can be predicted from a combination of brain activity and acoustic features. Brain Cogn. 101, 1–11.
Daniels, Lia M., Stupnisky, Robert H., June 2012. Not that different in theory: discussing the control-value theory of emotions in online learning environments. Internet High Educ. 15 (3), 222–226.
Daniels, L.M., Stupnisky, R.H., Pekrun, R., Haynes, T.L., Perry, R.P., Newall, N.E., 2009. Affective antecedents, mastery and performance goals, emotion outcomes, and academic attainment: testing a longitudinal model. J. Educ. Psychol. 101, 948–963. https://doi.org/10.1037/a0016096.
Darban, Mehdi, Polites, Greta L., 2016. Do emotions matter in technology training? Exploring their effects on individual perceptions and willingness to learn. Comput. Hum. Behav. 62, 644–657.
Darrell, T., Gordon, G., Harville, M., Woodfill, J., 1998. Integrated person tracking using stereo, color, and pattern detection. In: IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara, CA.
Darwin, C., 1872. The expression of the emotions in man and animals. London, England: John Murray.Thirteenth Thousand. Post 8vo. 7s. 6d. MURRAY. https://doi.org/10.1037/10001-000.
Darwin, C., 2005. The Expression of the Emotions in Man and Animals. Kessinger Publishing.
Dash, Manoranjan, Liu, Huan, 1997. Feature selection for classification. Intell. Data Anal. 1 (3), 131–156.
Datcu, D., Rothkrantz, L., 2009. Multimodal recognition of emotions in car environments. In: Driver Car Interaction & Interface Conference (DCII'09), pp. 98–106.
Devlin, A.M., Canavan, B., Magill, J., Lally, V., 2012. Evaluation of the Novel Inter-lifevirtual World as an Innovative Technology to Support Transition to University,European Conference on Educational Research (ECER). University of Cadiz.
Dipietro, L., Sabatini, A., Dario, P., 2008. A survey of glove-based systems and their applications. IEEE Transactions on Systems, Man and Cybernetics 38 (4), 461–482.
Doulik, Pavel, Skoda, Jiri, Simonova, Ivana, 2017. Learning Styles in the e-Learning Environment: the Approaches and Research on Longitudinal Changes. IJDET 15.2, 45–61. https://doi.org/10.4018/IJDET.2017040104. Web. 11 Jul. 2019.
Dweck, C.S., 1999. Self-theories: Their Role in Motivation, Personality and Development. Taylor and Francis/Psychology Press, Philadelphia.
Dweck, C.S., Molden, D.C., 2005. Self-theories: their impact on competence motivation and acquisition. In: Elliot, A., Dweck, C.S. (Eds.), The Handbook of Competence and Motivation. Guilford, New York.
Dweck, C.S., 2011. Self-theories. In: Van Lange, P., Kruglanski, A., Higgins, E.T. (Eds.), Handbook of Theories in Social Psychology.
Egges, A., Kshirsagar, S., Magnenat-Thalmann, N., 2003. A model for personality and emotion simulation. J Knowl Based Intell Inf Eng Syst 2773, 453–461.
Ekman, P., 1984. Expression and the Nature of Emotion. Approaches to Emotion. Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 319–344.
Ekman, P., 1992. An argument for basic emotions. Cognit. Emot. 6, 169–200.
Ekman, P., Friesen, W.V., 1971. Constant across cultures in face and emotions. J. Personal. Soc. Psychol. 17 (2), 124–129.
Ekman, P., Friesen, W.V., 1978. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, palo alto. CA. Ellsworth, PC, & Smith, CA (1988). From appraisal to emotion: Differences among unpleasant feelings. Motivation and Emotion 12, 271–302.
Ekman, P., Friesen, W.V., Hager, J.C., 2002. Facial Action Coding System. A Human Face, Salt Lake City, UT.
Eyben, F., Wöllmer, M., Schuller, B., 2010. Opensmile: the munich versatile and fast open-source audio feature extractor. In: ACM Multimedia, pp. 1459–1462.
Faltemier, T., Bowyer, K., Flynn, P., 2008. A region ensemble for 3D face recognition. IEEE Trans. Inf. Forensics Secur. (1), 62–73.
Faria, R., Almeida, A., Martins, C., Gonçalves, R., 2015. Emotional Adaptive Platform for Learning. Methodologies and Intelligent Systems for Technology Enhanced Learning. Springer, pp. 9–16.
Faria, A.R., Almeida, A., Martins, C., Gonçalves, R., Martins, J., Branco, F., 2017. A global perspective on an emotional learning model proposal. Telematics Inf. 34 (6), 824–837.
Fasel, B., 2002. Head-pose invariant facial expression recognition using convolutional neural networks. In: Proceedings of the Fourth IEEE International Conference on Multimodal Interfaces, 2002, pp. 529–534.
Fatahi, S., Ghasem-Aghaee, N., 2010. An effective intelligent educational model using agent with personality and emotional filters. In: Ao, S.I., Gelman, L., Hukins, D.W.L., Hunter, A., Korsunsky, A.M. (Eds.), Lecture Notes in Engineering and Computer Science, Proceedings of the World Congress on Engineering, June 30-July 2, 2010, London, U.K., vol. 1. International Association of Engineers, Hong Kong, pp. 142–147.
Fatahi, S., Moradi, H., Kashani-Vahid, L., 2016. Artif. Intell. Rev. 46, 413. https://doi.org/10.1007/s10462-016-9469-7.
Feng, Tao, et al., 2012. Continuous mobile authentication using touchscreen gestures. In: Homeland Security (HST), 2012 IEEE Conference on Technologies for. IEEE.
Fernandez, R., Picard, R., 2003. Modeling drivers' speech under stress. Speech Commun. 145–159.
Fernández-Caballero, Antonio, Martínez-Rodrigo, Arturo, Pastor, José Manuel, Castillo, José Carlos, Lozano-Monasor, Elena, López, María T., Zangróniz, Roberto, Latorre, José Miguel, Fernández-Sotos, Alicia, December 2016. Smart environment architecture for emotion detection and regulation. J. Biomed. Inform. 64, 55–73.
Flavell, J., 1999. Cognitive development: children's knowledge about other minds. Annu. Rev. Psychol. 50, 21–45.
Fox, S., Karnawat, K., Mydland, M., Dumais, S., White, T., 2005. Evaluating implicit measures to improve web search. ACM Trans. Inf. Syst. 23 (2), 147–168.
Fredrickson, B.L., 1998. What good are positive emotions? Rev. Gen. Psychol. 2, 300–319.
Fredrickson, B.L., 2001. The role of positive emotions in positive psychology: the broaden-and-build theory of positive emotion. Am. Psychol. 56, 218–226.
Frijda, N.H., 1994. Varieties of Affect: Emotions and Episodes, Moods, and Sentiments. The Nature of Emotion. Oxford University Press, New York, pp. 59–67.
Gao, Y., Wang, M., Zha, Z., Tian, Q., Dai, Q., Zhang, N., 2011. Less is more: efficient 3D object retrieval with query view selection. IEEE Trans. Multimed. (5), 1007–1018.
Ghimire, G., Lee, J., 2013. Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines. Journal of sensors 13, 7714–7734.
Ghimire, D., Lee, J.W., Li, Z.N., Jeong, S.W., Park, S.H., Choi, H.S., 2015. Recognition of facial expressions based on tracking and selection of discriminative geometric features. International Journal of Multimedia and Ubiquitous Engineering 10 (3), 35–44.
Goa, T., Fenga, X.L., Lub, H., Zhaia, J.H., 2013. A novel face feature descriptor using adaptively weighted extended LBP pyramid. Journal of Optik 124, 6286–6291.
Gobl, C., Chasaide, A.N., 2003. The role of voice quality in communicating emotion, mood and attitude. Speech Commun. 40 (1–2), 189–212.
Grassi, M., Cambria, E., Hussain, A., Piazza, F., Senticweb, 2011. A new paradigm for managing social media affective information. Cognit. Comput. 3, 480–489.
Gratch, J., Marsella, S., 2004. A domain-independent framework for modeling emotion. Cogn. Syst. Res. 5 (4), 269e306.
Grimm, M., Kroschel, K., Narayanan, S.S., 2007. Support vector regression for automatic recognition of spontaneous emotions in speech. In: ICASSP, pp. 1085–1088.
Gunes, Hatice, Piccardi, Massimo, November 2007. Bi-modal emotion recognition from expressive face and body gestures. J. Netw. Comput. Appl. 30 (4), 1334–1345.
Gunes, H., Piccardi, M., 2005. In: Tao, J., Tan, T., Picard, R.W. (Eds.), Fusing Face and Body Display for Bi-modal Emotion Recognition: Single Frame Analysis and Multi-Frame Post Integration, pp. 102–111. ACII 2005, LNCS 3784.
Gunes, H., Schuller, B., Pantic, M., Cowie, R., 2011. Emotion representation, analysis and synthesis in continuous space: a survey. In: Proc. Of IEEE Conference on Face and Gesture Recognition, Santa Barbara, CA, USA, pp. 827–834.
Gunes, Hatice, Hung, Hayley, November 2016. Is automatic facial expression recognition of emotions coming to a dead end? The rise of the new kids on the block. Image Vis Comput. 55 (1), 6–8.
Gupta, Shikha, Jaafar, Jafreezal, Wan Ahmad, Wan Fatimah, 2012. Static hand gesture recognition using local gabor filter. Procedia Engineering 41, 827–832.
Gwizdka, J., Lopatovska, I., 2009. The role of subjective factors in the information search process. J. Am. Soc. Inf. Sci. Technol. 60 (12), 2452–2464.
Hancock, J.T., Landrigan, C., Silver, C., 2007. Expressing emotion in text-based communication. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 929–932.
Haq, S., Jackson, P.J.B., Edge, J., 2008. Audio-visual feauture selection and Re- duction for emotion classification. In: Proceedings of Auditory-Visual Speech Pro- Cessing. AVSP '08, pp. 185–190.
Harris, M., 2000. Correlates and characteristics of boredom proneness and boredom. J. Appl. Soc. Psychol. 30, 576–598. https://doi.org/10.1111/j.1559-1816.2000.tb02497.x.
Hartmann, P., 2006. The five-factor model: psychometric, biological and practical perspectives. Nord. Psychol. 58, 150–170.
El Hayek, H., Nacouzi, J., Kassem, A., Hamad, M., 2014. Sign to letter translator system using a hand gloves. In: International Conference on e-Technologies and Networks for Development.
Heidig, Steffi, Julia Müller, Reichelt, Maria, March 2015. Emotional design in multimedia learning: differentiation on relevant design features and their effects on emotions and learning. Comput. Hum. Behav. 44, 81–95.
Hemprasad, Y. Patil, Ashwin, G. Kothari, Bhurchandi, Kishor M., March 2016. Expression invariant face recognition using local binary patterns and contourlet transform. Optik - International Journal for Light and Electron Optics 127 (5), 2670–2678.
Hermansky, H., 1990. Perceptual linear predictive (PLP) analysis of speech. Acoustical Soc Am J 87, 1738–1752.
Hernandez, Javier, et al., 2014. Under pressure: sensing stress of computer users. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM.
Hernandez-Matamoros, Andres, Bonarini, Andrea, Escamilla-Hernandez, Enrique, Nakano-Miyatake, Mariko, Perez-Meana, Hector, 15 October 2016. Facial expression recognition with automatic segmentation of face regions using a fuzzy based classification approach. Knowl. Based Syst. 110, 1–14.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Hou, X., Li, S.Z., Zhang, H., Cheng, Q., 2001. Direct appearance models. IEEE Conference on Computer Vision and Pattern Recognition 1, 828–833.
Hrastinski, Stefan, 2008. A study of asynchronous and synchronous e-Learning methods discovered that each supports different purposes. Educ. Q. 4.
Huang, C.L., Huang, W.Y., 1998. Sign language recognition using model-based tracking and a 3D Hopfield neural network. Mach. Vis. Appl. 10, 292–307.
Huang, X., Zhao, G., Zheng, W., Pietikäinen, M., 2012. Spatiotemporal local monogenic binary patterns for facial expression recognition. IEEE Signal Process. Lett. 19 (5), 243–246.
Huang, Xiaohua, Kortelainen, Jukka, Zhao, Guoying, Li, Xiaobai, Moilanen, Antti, Seppänen, Tapio, Pietikäinen, Matti, June 2016. Multi-modal emotion analysis from facial expressions and electroencephalogram. Comput. Vis. Image Understand. 147, 114–124.
Jaimes, A., Sebe, N., 2007. Multimodal human–computer interaction: a survey. Comput. Vis. Image Understand. 108 (1–2), 116–134.
Jan, Treur, Arlette van Wissen, 2013. Conceptual and computational analysis of the role of emotions and social influence in learning. Procedia - Social and Behavioral Sciences 93, 449–467.
Jatupaiboon, N., Panngum, S., Israsena, P., 2013. Real-time EEG-based happiness detection system. The ScientificWorld Journal, 618649.
Jiang, B., Valstar, M.F., Pantic, M., 2011. Action unit detection using sparse appearance descriptors in space-time video volumes. In: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, IEEE, pp. 314–321.
Jiang, Yingying, Li, Wei, Shamim Hossain, M., Chen, Min, Alelaiwi, Abdulhameed, Al-Hammadi, Muneer, 2020. A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition. Inf. Fusion 53, 209–221.
Joachims, T., 2002. Optimizing search engines using clickthrough data. In: Proceedings of the ACM Conference on Knowledge Discovery and DataMining (KDD), pp. 133–142.
Johnson, E.A., 1965. Touch display—a novel input/output device for computers. Electron. Lett. 1 (8), 219–220.
Kapoor, A., Picard, R.W., Ivanov, Y., 2004. Probabilistic combination of multiple modalities to detect interest. In: Proceedings of IEEE International Conference on Pattern Recognition, pp. 969–972.
Kapoor, A., Burleson, W., Picard, R.W., 2007. Automatic prediction of frustration. Int. J. Hum. Comput. Stud. 65 (8), 724–736.
Kardan, A., Einavypour, Y., October, 2008. Multi-criteria learners' classification for selecting an appropriate teaching method. In: Proceeding of International Conference on Education and Information Technology (ICEIT'08) San Francisco, USA, pp. 22–24.
Kardan, A., Einavypour, Y., 2008. Involving learner's emotional behaviors in learning process as a temporary learner model. In: Proceeding of International Conference on Virtual Learning (ICVL), Bucurest, Romania.
Katsis, C.D., Katertsidis, N., Ganiatsas, G., Fotiadis, D.I., 2008. Toward emotion recognition in car-racing drivers: a biosignal processing approach. IEEE Trans. Syst. Man Cybern. A Syst. Hum. 38 (3), 502–512.
Katz, P., Singleton, M., Wicentowski, R., 2007. Swat-mp: the semeval-2007 systems for task 5 and task 14. In: Proc. 4th International Workshop on SemanticEvaluations (ACL), pp. 308–313.
Khan, Rizwan Ahmed, Meyer, Alexandre, Hubert, Konik, Bouakaz, Saïda, 15 July 2013. Framework for reliable, real-time facial expression recognition for low resolution images. Pattern Recognit. Lett. 34 (10), 1159–1168.
Khan, Sajid Ali, Hussain, Ayyaz, Usman, Muhammad, August 2016. Facial expression recognition on real world face images using intelligent techniques: a survey. Optik - International Journal for Light and Electron Optics 127 (15), 6195–6203.
Khezri, Mahdi, Firoozabadi, Mohammad, Reza Sharafat, Ahmad, November 2015. Reliable emotion recognition system based on dynamic adaptive fusion of forehead biopotentials and physiological signals. Comput. Methods Progr. Biomed. 122 (2), 149–164.
Kim, J., Andre, E., 2008. Emotion recognition based on physiological changes in music listening. IEEE Trans. Pattern Anal. Mach. Intell. 30, 2067–2083.
Kim, In-Cheol, Chien, Sung-Il, 2001. Analysis of 3D hand trajectory gestures using stroke-based composite hidden Markov models. Appl. Intell. 15 (2), 131–143.
Kim, Jae-Bok, Park, Jeong-Sik, June 2016. Multistage data selection-based unsupervised speaker adaptation for personalized speech emotion recognition. Eng. Appl. Artif. Intell. 52, 126–134.
Kim, J., Park, J., Oh, Y., 2009. Feature vector classification based speech emotion recognition for service robots. IEEE Trans. Consum. Electron. 55 (3), 1590–1596.
Kim, C., Stern, R.M., 2016. Power-normalized cepstral coefficients (PNCC) for robust speech recognition. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4101–4104.
King, Ronnel B., McInerney, Dennis M., Watkins, David A., 2012. How you think about your intelligence determines how you feel in school: the role of theories of intelligence on academic emotions. Learn. Individ. Differ. 22 (6), 814–819.
Klein, J., Moon, Y., Picard, R.W., 1999. This computer responds to user frustration. In: CHI '99 Extended Abstracts on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 242–243.
Kleinginna, P., Kleinginna, A., 2005. A categorized list of motivation definitions, with a suggestion for a consensual definition. Motiv. Emot. 5 (3), 263–291.
Kleinsmith, A., Bianchi-Berthouze, N., 2013. Affective body expression perception and recognition: a survey. IEEE Trans. Affect. Comput. 4 (1), 15–33.
Koelstra, S., Mühl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I., 2012. Deap: a database for emotion analysis; using physiological signals. IEEE Trans. Affect. Comput. 3 (1), 18–31.
Kotropoulos, C., Tefas, A., Pitas, I., Apr 2000. Frontal face authentication using morphological elastic graph matching. IEEE Trans. Image Process. 9 (4), 555–560.
Kotsia, I., Zafeiriou, S., Pitas, I., 2008. Texture and shape information fusion for facial expression and facial action unit recognition. Pattern Recognit. 41, 822–851.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Pereira, F., Burges, C., Bottou, L., Weinberger, K. (Eds.), Advances in Neural Information Processing Systems 25. Curran Associates, Inc., pp. 1097–1105
Krouska, A., Troussas, C., Virvou, M., 2017. Comparative evaluation of algorithms for sentiment analysis over social networking services. J. Univers. Comput. Sci. 23 (8), 755–768.
Kuhlthau, C.C., 1991. Inside the search process: information seeking from the user's perspective. J. Am. Soc. Inf. Sci. 42 (5), 361–371.
Kuhnert, Rebecca-Lee, Sander, Begeer, Fink, Elian, de Rosnay, Marc, February 2017. Gender-differentiated effects of theory of mind, emotion understanding, and social preference on prosocial behavior development: a longitudinal study. J. Exp. Child Psychol. 154, 13–27.
Kumar, Nitin, Khaund, Kaushikee, Shyamanta, M., 2016. Hazarika, bispectral analysis of EEG for emotion recognition. Procedia Computer Science 84, 31–35.
Kumar, Jyotish, kumar, Jyoti, 2016. Affective modelling of users in HCI using EEG. Procedia Computer Science 84, 107–114. ISSN 1877-0509.
Kuncheva, L.I., 2002. A theoretical study on six classifier fusion strategies. IEEE Trans. Pattern Anal. Mach. Intell. 24 (2), 281–286.
Kurihara, Kenzo, Sugiyama, Daisuke, Matsumoto, Shigeru, Nishiuchi, Nobuyuki, Masuda, Kazuaki, March 2009. Facial emotion and gesture reproduction method for substitute robot of remote person. Comput. Ind. Eng. 56 (2), 631–647.
Kuypers, K.P.C., Steenbergen, L., Theunissen, E.L., Toennes, S.W., Ramaekers, J.G., November 2015. Emotion recognition during cocaine intoxication. Eur. Neuropsychopharmacol. 25 (11), 1914–1921.
Lahane, Prashant, Kumar Sangaiah, Arun, 2015. An approach to EEG based emotion recognition and classification using kernel density estimation. Procedia Computer Science 48, 574–581.
Lamberti, L., Camastra, F., 2012. Handy: A real-time three color glove-based gesture recognizer with learning vector quantization. Expert Syst. Appl. 39 (12), 10489–10494.
Landowska, Agnieszka, Brodny, Grzegorz, Wróbel, Michał, 2017. Limitations of Emotion Recognition from Facial Expressions in e-Learning Context, pp. 383–389. https://doi.org/10.5220/0006357903830389.
Larsen, R.J., Diener, E., 1992. Promises and problems with the circumplex model of emotion. Rev. Personal. Soc. Psychol. 13, 25–59.
Latham, A., Crockett, K., McLean, D., Edmonds, B., 2012. A conversational intelligent tutoring system to automatically predict learning styles. Comput. Educ. J. 59 (1), 95–109.
Lazarus, R.S., 1984. Thoughts on the Relations between Emotion and Cognition. Approaches to Emotion. Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 247–259.
Lazarus, R., 2001. Relational meaning and discrete emotions. Appraisal Process. Emotion: Theor., Methods, Res. 37–67.
Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 2278–2324.
Lee, Chi-Chun, Mower, Emily, Busso, Carlos, Lee, Sungbok, Narayanan, Shrikanth, November–December 2011. Emotion recognition using a hierarchical binary decision tree approach. Speech Commun. 53 (9–10), 1162–1171.
Lee, G., Kwon, M., Kavuri Sri, S., Lee, M., 2014. Emotion recognition based on 3D fuzzy visual and EEG features in movie clips. Neurocomputing 144, 560–568.
Lee, Giyoung, Kwon, Mingu, Kavuri Sri, Swathi, Lee, Minho, 20 November 2014. Emotion recognition based on 3D fuzzy visual and EEG features in movie clips. Neurocomputing 144, 560–568.
Legree, P.J., Psotka, J., Tremble, T., Bourne, D.R., 2005. Using consensus based measurement to assess emotional intelligence. In: Schulze, R., Roberts, R.D. (Eds.), Emotional Intelligence: an International Handbook. Hogrefe, Cambridge, MA, pp. 155–179.
Lei, Y., Han, H., Hao, X., 2015. Discriminant sparse local spline embedding with application to face recognition. Knowl. Based Syst. 89, 47–55.
Leslie, A.M., Friedman, O., German, T.P., 2004. Core mechanisms in ‘‘theory of mind’’. Trends Cogn. Sci. 8 (12), 528–533.
Li, Yu-Ting, Wachs, Juan P., September 2013. Recognizing hand gestures using the weighted elastic graph matching (WEGM) method,. Image Vis Comput. 31 (9), 649–657.
Li, X., Mori, G., Zhang, H., 2006. Expression-invariant face recognition with expres-sion classification. In: The Third Canadian Conference on Computer and RobotVision, Canada, pp. 77–79.
Li, Y.S., Chen, P.S., Tsai, S.J., 2007. A comparison of the learning styles among different nursing programs in Taiwan: implications for nursing education. J Nurse Educ Today 28 (1), 70–76.
Li, Z., Imai, J.i., Kaneko, M., 2009. Facial-component-based bag of words and phog descriptor for facial expression recognition. In: Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on, IEEE, pp. 1353–1358.
Li, Jun, Rao, Yanghui, Jin, Fengmei, Chen, Huijun, Xiang, Xiyun, 19 October 2016. Multi-label maximum entropy model for social emotion classification over short text. Neurocomputing 210, 247–256.
Li, Yongqiang, Wu, Baoyuan, Ghanem, Bernard, Zhao, Yongping, Yao, Hongxun, Ji, Qiang, December 2016. Facial action unit recognition under incomplete data based on multi-label learning with missing labels. Pattern Recognit. 60, 890–900.
Li, Chao, Xu, Chao, Feng, Zhiyong, 20 February 2016. Analysis of physiological for emotion recognition with the IRS model. Neurocomputing 178, 103–111.

Li, X., Rao, Y., Xie, H., Lau, R.Y.K., Yin, J., Wang, F.L., 1 . Bootstrapping social emotion classification with semantically rich hybrid neural networks. IEEE Transactions on Affective Computing 8 (4), 428–442.
Liang, Weiming, Xie, Haoran, Rao, Yanghui, Lau, Raymond Y.K., Fu, Lee Wang, 2018. Universal affective model for Readers' emotion classification over short texts. Expert Syst. Appl. 114, 322–333.
Lin, K.H.-Y., Yang, C., Chen, H.-H., 2008. Emotion classification of online news articles from the reader's perspective. In: Proc. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI/IAT), pp. 220–226.
Lisetti, C., Nasoz, F., LeRouge, C., Ozyer, O., Alvarez, K., 2003. Developing mul- timodal intelligent affective interfaces for tele-home health care. Int. J. Hum. Comput. Stud. 59, 245–255.
Liu, Y., Sourina, O., Nguyen, M.K., 2010. Real-time EEG-based human emotion recognition and visualization. In: Paper Presented at the Cyberworlds (CW), 2010 International Conference on.
Loa, Y., Wub, C., Zhangb, Y., 2013. Facial expression recognition based on fusion feature of PCA and LBP with SVM. Journal of Optik 124, 2767–2770.
Lockwood, Travis W., April 2015. Redefining the role of emotion in critical language teaching and learning. Linguist. Educ. 29, 90–91.
Loconsole, C., Miranda, C.R., Augusto, G., Frisoli, A., Orvalho, V., 2014. Real-time emotion recognition: a novel method for geometrical facial features extraction. In: Proc International Conf. On Computer Vision Theory and Applications - VISAPP , Lisbon , Portugal, vol. 01, pp. 378–385.
LoganK, Thomas, P., 2002. Learning styles in distance education students learning to program. In: Proceedings of 14th Workshop of the Psychology of Programming Interest Group. Brunel University, pp. 29–44.
Long, F., Wub, T., Movellan, J.R., Bartlett, M.S., Littlewort, G., 2012. Learning spatiotemporal features by using independent component analysis with application to facial expression recognition. Journal of Neurocomputing 93, 126–132.
Lopatovska, I., 2009. Searching for good mood: examining relationships between search task and mood. no. 1. In: Proceedings of the 72th Annual Meeting of the American Society for Information Science and Technology, vol. 46, pp. 1–13, 2009.
Lopatovska, I., 2009a. Emotional Aspects of the Online Information Retrieval Process. Ph.D. Thesis. Rutgers: The State University of, New Jersey.
Lopatovska, Irene, Arapakis, Ioannis, July 2011. Theories, methods and current research on emotions in library and information science, information retrieval and human–computer interaction. Inf. Process. Manag. 47 (4), 575–592.
Lopatovska, rene, Arapakis, Ioannis, July 2011. Theories, methods and current research on emotions in library and information science, information retrieval and human–computer interaction. Inf. Process. Manag. 47 (4), 575–592.
Lopatovska, I., Mokros, H.B., 2008. Willingness to pay and experienced utility as measures of affective value of information objects: users' accounts. Inf. Process. Manag.: Int. J. 44 (1), 92–104.
Lorenzino, Martina, Caudek, Corrado, March 2015. Task-irrelevant emotion facilitates face discrimination learning. Vis. Res. 108, 56–66.
Luengo, I., Navas, E., Hernez, I., Snchez, J., 2005. Automatic Emotion Recognition Using Prosodic Parameters. in: Interspeech, pp. 493–496.
Luo, Wenshu, Tee Ng, Pak, Lee, Kerry, Maung Aye, Khin, August 2016. Self-efficacy, value, and achievement emotions as mediators between parenting practice and homework behavior: a control-value theory perspective. Learn. Individ. Differ. 50, 275–282.
Lyusin, Dmitry, Ovsyannikova, Victoria, December 2016. Measuring two aspects of emotion recognition ability: accuracy vs. sensitivity, Learning and Individual Differences 52, 129–136.
Ma, Chunyan, 2016. Design of an emotional interaction mode in e-learning. World Transactions on Engineering and Technology Education 14 (1), 14–19.
Maaoui, C., Abdat, F., Pruski, A., June 2014. Physio-visual data fusion for emotion recognition. IRBM 35 (3), 109–118.
Majumder, Anima, Behera, Laxmidhar, Subramanian, Venkatesh K., March 2014. Emotion recognition from geometric facial features using self-organizing map. Pattern Recognit. 47 (3), 1282–1293.
Malatesta, Lori, Asteriadis, Stylianos, George, Caridakis, Vasalou, Asimina, Karpouzis, Kostas, May 2016. Associating gesture expressivity with affective representations. Eng. Appl. Artif. Intell. 51, 124–135.
Mannepalli, K., Narahari Sastry, P., Suman, M., 2017. A novel adaptive fractional deep belief networks for speaker emotion recognition. Alexandria Eng. J. 56 (4), 485–497. Available online.
Mano, Leandro Y., Faiçal, Bruno S., Luis, H., Nakamura, V., Gomes, Pedro H., Libralon, Giampaolo L., Meneguete, Rodolfo I., Filho, Geraldo P.R., Giancristofaro, Gabriel T., Pessin, Gustavo, Krishnamachari, Bhaskar, Ueyama, Jó, 1 September 2016. Exploiting IoT technologies for enhancing Health Smart Homes through patient identification and emotion recognition. Comput. Commun. 89–90, 178–190.
Mariooryad, Soroosh, Busso, Carlos, February 2014. Compensating for speaker or lexical variabilities in speech for emotion recognition. Speech Commun. 57, 1–12.
López, Mariza G. Méndez, Cárdenas, Martha Armida Fabela, 2014. Emotions and their effects in a language learning Mexican context. System 42, 298–307.
de Marneffe, M.C., MacCartney, B., Manning, C.D., 2006. Generating typed dependency parses from phrase structure parses. In: Proceedings of LREC, vol. 6, pp. 449–454.
Marras, G.T.I., Zafeiriou, S., 2012. Robust learning from normals for 3D face recognition. In: Computer Vision-European Conference on Computer Vision, pp. 230–239.
Matsui, T., Furui, S., 1998. N-best-basedunsupervisedspeakeradaptationforspeech recognition. Comput.SpeechLanguage 12 (1), 41–50.
Matsumoto, D., Ekman, P., 1988. Japanese and Caucasian Facial Expressions of Emotion (JACFEE) [slides]. San Francisco State University, San Francisco, CA.
Mayer, J.D., Salovey, P., Caruso, D., 2000. Competing models of emotional intelligence. In: Sternberg, R.J. (Ed.), Handbook of Human Intelligence, second ed. Cambridge University Press, New York, pp. 396–420.
Mayya, Veena, Pai, Radhika M., Pai, M.M. Manohara, 2016. Automatic facial expression recognition using DCNN. Procedia Computer Science 93, 453–461.
Mehdi, E., JNico, P.Julie D., Bernard, P., 2004. Modelling character emotion in an interactive virtual environment. In: Proceedings of AISB 2004 Symposium: Motion, Emotion and Cognition.
Mehmood Bhatti, Adnan, Majid, Muhammad, Anwar, Syed Muhammad, Khan, Bilal, 2016. Human emotion recognition and analysis in response to audio music using brain signals. Comput. Hum. Behav. 65 (December), 267–275.
de Meijer, M., 1989. The contribution of general features of body movement to the attribution of emotions. J. Nonverbal Behav. 13, 247–268.
Mencattini, Arianna, Martinelli, Eugenio, Costantini, Giovanni, Todisco, Massimiliano, Basile, Barbara, Bozzali, Marco, Di Natale, Corrado, June 2014. Speech emotion recognition using amplitude modulation parameters and a combined feature selection procedure. Knowl. Based Syst. 63, 68–81.
Mendoza, E., Carballo, G., 1999. Vocal tremor and psychological stress. J. Voice 13 (1), 105–112. http://www.sciencedirect.com/science/article/B7585-4GM9P0N-D/2/01036b570cbfb0ab303f793bc474217a.
Meng, H., Pittermann, J., Pittermann, A., Minker, W., 2007. Combined speech-emotion recognition for spoken human–computer interfaces. In: IEEE International Conference on Signal Processing and Communications, pp. 1179–1182.
Mentis, H.M., 2007. Memory of Frustrating Experiences. Information and Emotion: the Emergent Paradigm in Information Behavior Research and Theory. Information Today, Medford, NJ, pp. 197–210.
Meza-Kubo, Victoria, Morán, Alberto L., Carrillo, Ivan, Galindo, Gilberto, García-Canseco, Eloisa, 2016. Assessing the user experience of older adults using a neural network trained to recognize emotions from brain signals. J. Biomed. Inform. 62 (August), 202–209.
Miguel-Hurtado, Oscar, Stevenage, Sarah V., Bevan, Chris, Guest, Richard, 1 August 2016. Predicting sex as a soft-biometrics from device interaction swipe gestures. Pattern Recognit. Lett. 79, 44–51.
Mikio, Y., 1996. Interface system based on hand gestures and verbal expressions for 3-D shape generation. Terebijon Gakkaishi/Journal of the Institute of Television Engineers of Japan 50 (10), 1482–1488.
Milton, A., Tamil Selvi, S., May 2014. Class-specific multiple classifiers scheme to recognize emotions from speech signals. Comput. Speech Lang 28 (3), 727–742.
Ming, Y., Ruan, Q., 2013. A Mandarin edutainment system integrated virtual learning environments. Speech Commun. (1), 71–83.
Mirjalili, S., 2015. Moth-Flame optimization algorithm: a novel nature-inspired heuristic paradigm. Knowl. Based Syst. 89, 228–249.
Mishne, G., 2005. Experiments with mood classification in blog posts. In: Proceedings of the Style2005: the First Workshop on Stylistic Analysis of Text for Information Access. SIGIR, pp. 15–19.
Miskovic, V., Schmidt, L.A., 2010. Cross-regional cortical synchronization during affective image viewing. Brain Res. 1362, 102–111.
Moataz, E.A., Kamel, M.S., Karray, Fakhri, 2011. Survey on speech emotion recognition: features, classification schemes, and databases. Pattern Recognit. 44 (3), 572–587.
Montazer, Gholam Ali, Sadegh Rezaei, Mohammad, July 2012. A new approach in e-learners grouping using Hybrid Clustering Method, International Conference on Education and e-Learning Innovations (ICEELI).
Morrison, Donn, Liyanage, C., De Silva, 2007. Voting ensembles for spoken affect classification. J. Netw. Comput. Appl. 30 (4), 1356–1365.
Moshkina, L., 2006. An Integrative Framework for Time-Varying Affective Agent Behavior. Georgia Institute of Technology, Atlanta.
Mouratidis, Athanasios, Michou, Aikaterini, Vassiou, Aikaterini, 14 September 2016. Adolescents' Autonomous Functioning and Implicit Theories of Ability as Predictors of Their School Achievement and Week-To-Week Study Regulation and Well-Being, Contemporary Educational Psychology. Available online.
Mpiperis, S., Malassiotis, S., Strintzis, M.G., 2009. Bilinear decomposition of 3D face images: an application to facial expression recognition. In: 10th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2009).
Mu-Chun, S., 2003. A neural-network-based approach to recognizing 3D arm movements. Biomedical Engineering—Applications, Basis and Communications 15 (1), 17–26.
Muhammad, G., Hussain, M., Alenzezy, F., Bebis, G., Mirza, A.M., Aboalsamh, H., 2012. Race classification from face images using local descriptors. Journal of artificial intelligence tools 21 (5), 113–121.
Muzammal, Muhammad, Talat, Romana, Sodhro, Ali Hassan, Pirbhulal, Sandeep, 2020. A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks. Inf. Fusion 53, 155–164.
Nahl, D., Tenopir, C., 1996. Affective and cognitive searching behavior of novice end-users of a full-text database. J. Am. Soc. Inf. Sci. 47 (4), 276–286.
Nanavare, V.V., Jagtap, S.K., 2015. Recognition of human emotions from speech processing. Procedia Computer Science 49, 24–32.
Nasoz, F., Lisetti, C.L., Alvarez, K., Finkelstein, N., 2003. Emotion recognition from physiological signals for user modeling of affect. In: Proceedings of the 3rd Workshop on Affective and Attitude User Modelling (Pittsburgh, PA, USA.
Neiberg, D., Elenius, K., Laskowski, K., 2006. Emotion recognition in spontaneous speech using GMMs. In: Interspeech.
Nicholson, J., Takahashi, K., Nakatsu, R., 2000. Emotion Recognition in Speech Using Neural Networks. Neural Computing and Applications.
Nicholson, J., Takahashi, K., Nakatsu, R., 2000. Emotion Recognition in Speech Using Neural Networks. Neural Computing and Applications.
Nickel, Kai, Stiefelhagen, Rainer, 3 December 2007. Visual recognition of pointing gestures for human–robot interaction. Image Vis Comput. 25 (12), 1875–1884.

Nicolaou, Mihalis A., Gunes, Hatice, Pantic, Maja, March 2012. Output-associative RVM regression for dimensional and continuous emotion prediction. Image Vis Comput. ume 30 (3), 186–196.
Niese, R., Al-Hamadi, A., Farag, A., Neumann, H., Michaelis, B., 2012. Facial expression recognition based on geometric and optical flow features in colour image sequences. IET Comput. Vis. 6 (2), 79–89.
Nkambou, Roger, Gauthier, Gilles, 1996. Integrating WWW resources in an intelligent tutoring system. J. Netw. Comput. Appl. 19 (4), 353–366.
Ogunfunmi, T., Togneri, R., Narasimha, M. (Eds.), 2015. Speech and Audio Processing for Coding, Enhancement and Recognition. Springer Science +Business Media, New York.
Ohknishi, A., Nishikawa, A., 1997. Curvature-based segmentation and recognition of hand gestures. In: Proceedings of the Annual Conference on Robotics Societyof Japan, p. 401.
Olsher, D., 2012. Full Spectrum Opinion Mining: Integrating Domain, Syntactic and Lexical Knowledge. Sentiment Elicitation from Natural Text for Information Retrieval and Extraction, IEEE, pp. 693–700.
Ong, Desmond C., Zaki, Jamil, Goodman, Noah D., 2015. Affective cognition: exploring lay theories of emotion. Cognition 143 (October), 141–162.
Origlia, A., Cutugno, F., Galatà, V., February 2014. Continuous emotion recognition with phonetic syllables. Speech Commun. 57, 155–169.
Ortony, A., Clore, G.L., Collins, A., 1988. The Cognitive Structure of Emotions. Cambridge University Press, New York, NY, USA, pp. 10011–14211.
Othman, Marini, Wahab, Abdul, Karim, Izzah, Adawiah Dzulkifli, Mariam, Taha Alshaikli, Imad Fakhri, 2013. EEG emotion recognition based on the dimensional models of emotions. Procedia - Social and Behavioral Sciences 97, 30–37.
Othman, Marini, Abdul Wahab, Karim, Izzah, Adawiah Dzulkifli, Mariam, Taha Alshaikli, Imad Fakhri, 2013. EEG emotion recognition based on the dimensional models of emotions. Procedia - Social and Behavioral Sciences 97, 30–37.
Ouyang, Y., Sang, N., Huang, R., 2015. Accurate and robust facial expressions recognition by fusing multiple sparse representation based classifiers. Journal of Neurocomputing 149, 71–78.
PAL, Robotics, 2015. Reem - humanoid robot. http://reemc.pal-robotics.com/en/. (Accessed 30 June 2015).
Palaz, Dimitri, Magimai-Doss, Mathew, Collobert, Ronan, 2019. End-to-end acoustic modeling using convolutional neural networks for HMM-based automatic speech recognition. Speech Commun. 108, 15–32.
Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359.
Pandzic, I.S., Forchheimer, R., 2002. MPEG-4 Facial Animation: the Standard, Implementation and Applications. Wiley.
Pantic, M., Rothkrantz, L.J.M., 2000. Automatic analysis of facial expressions: the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. 22 (12), 1424–1445.
Pantic, M., Rothkrantz, L.J., 2003. Toward an affect-sensitive multimodal human–computer interaction. Proc. IEEE 91 (9), 1370–1390.
Partala, T., Surakka, V., 2004. The effects of affective interventions in human–computer interaction. Interact. Comput. 16 (2), 295–309.
Passalis, G., Perakis, P., Theoharis, T., Kakadiaris, I.A., 2011. Using facial symmetry to handle pose variations in real-world 3D face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 10, 1938–1951.
Peeters, G., 2006. Chroma-based estimation of musical key from audio-signal analysis. In: Proceedings of the 7th International Conference on Music Information Retrieval, Victoria (BC), Canada.
Pekrun, R., 2006. The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice. Educ. Psychol. Rev. 18, 315–341.
Pekrun, R., Goetz, T., Titz, W., Perry, R.P., 2002. Academic emotions in students' self-regulated learning and achievement: a program of quantitative and qualitative research. Educ. Psychol. 37, 91–106.
Pekrun, R., Elliot, A.J., Maier, M.A., 2006. Achievement goals and discrete emotions: a theoretical model and prospective test. J. Educ. Psychol. 98, 583–597. https://doi.org/10.1037/0022-0663.98.3.583.
Pekrun, R., Goetz, T., Frenzel, A., Barchfeld, P., Perry, R.P., 2011. Measuring emotions in students' learning and performance: the Achievement Emotions Questionnaire (AEQ). Contemp. Educ. Psychol. 36, 36–48.
Perez-Gaspar, Luis-Alberto, Caballero-Morales, Santiago-Omar, Trujillo-Romero, Felipe, 2016. Multimodal emotion recognition with evolutionary computation for human-robot interaction. Expert Syst. Appl. 66, 42–61.
Perikos, Isidoros, Hatzilygeroudis, Ioannis, May 2016. Recognizing emotions in text using ensemble of classifiers. Eng. Appl. Artif. Intell. 51, 191–201.
Perikos, I., Hatzilygeroudis, I., 2013. Recognizing emotion presence in natural language sentences. In: Iliadis, L., Papadopoulos, H., Jayne, C. (Eds.), Engineering Applications of Neural Networks. Springer, Berlin Heidelberg, pp. 30–39.
Peter, C., Herbon, A., 2006. Emotion representation and physiology assignments in digital systems. Interact. Comput. 18 (2), 139–170.
Picard, R.W., 2001. Building hal: computers that sense, recognize, and respond to human emotion. In: Society of Photo-Optical Instrumentation Engineers. Human Vision and Electronic Imaging VI, vol. 4299, pp. 518–523.
Picard, R.W., Vyzas, E., Healey, J., 2001. Toward machine emotional intelligence: analysis of affective physiological state. IEEE Trans. Pattern Anal. Mach. Intell. 23 (10), 1175–1191.
Plutchik, R., 1980. A General Psychoevolutionary Theory of Emotion. Emotion: Theory, Research and Experience. Theories of Emotion, vol. 1. Academic, New York, pp. 3–33.
Pollick, F.E., Paterson, H.M., Bruderlin, A., Sanford, A.J., 2001. Perceiving aVect from arm movement. Cognition 82, B51–B61.
Putwain, David W., Larkin, Derek, Sander, Paul, October 2013. A reciprocal model of achievement goals and learning related emotions in the first year of undergraduate study. Contemp. Educ. Psychol. 38 (Issue 4), 361–374.
Qin, Zhen, Zhang, Yibo, Meng, Shuyu, Qin, Zhiguang, Raymond Choo, Kim-Kwang, 2020. Imaging and fusing time series for wearable sensor-based human activity recognition. Inf. Fusion 53, 80–87.
Quan, C., Ren, F., 2010. A blog emotion corpus for emotional expression analysis in Chinese, Comput. Speech Lang. 24, 726–749.
Quan, Changqin, Ren, Fuji, 2016. Weighted high-order hidden Markov models for compound emotions recognition in text. Inf. Sci. 329 (1 February), 581–596.
Quraishi, M., Choudhury, J., De, M., Chakraborty, P., February 2012. A framework for the recognition of human emotion using soft computing models. Int. J. Comput. Appl. 40 (17).
Rabiner, L.R., Schafer, R.W., 2004. Digital Processing of Speech Signals. Pearson Education (Singapore) Pte. Ltd. (Indian reprint).
Rajisha, T.M., Sunija, A.P., Riyas, K.S., 2016. Performance analysis of Malayalam language speech emotion recognition system using ANN/SVM. Procedia Technology 24, 1097–1104.
Rao, Yanghui, Li, Qing, Mao, Xudong, Liu, Wenyin, 10 May 2014. Sentiment topic models for social emotion mining. Inf. Sci. 266, 90–100.
Riaz, Z., Mayer, C., Wimmer, M., Beetz, M., Radig, B., 2009. A model based approachfor expressions invariant face recognition. In: Tistarelli, M., Nixon, M. (Eds.), Advances in Biometrics. Springer, Berlin Heidelberg, pp. 289–298.
Rodríguez, Luis-Felipe, Octavio Gutierrez-Garcia, J., Ramos, Félix, July 2016. Modeling the interaction of emotion and cognition in Autonomous Agents. Biologically Inspired Cognitive Architectures 17, 57–70.
Roweis, Sam, Saul, Lawrence, 2000. Nonlinear dimensionalityreduction bylocally linear embedding. Science 290 (5500), 2323–2326.
Rueckert, D., Frangi, A., Schnabel, J., 2003. Automatic construction of 3-Dstatistical deformation models of the brain using nonrigid registration. IEEE Trans. Med. Imaging 22 (8), 1014–1025.
Russell, J.A., 1980. A circumplex model of affect. J. Personal. Soc. Psychol. 39 (6), 1161–1178.
Russell, J.A., 1994. Is there universal recognition of emotion from facial expression? Psychol. Bull. 115, 102–141.
Saberi, Nafiseh, Montazer, Gholam Ali, Feb. 2012. A new approach for learners' modeling in e-learning environment using LMS logs analysis. In: 6th National and 3rd International Conference of E-Learning and E-Teaching, Tehran, Iran, pp. 25–33.
Sadoughi, Najmeh, Busso, Carlos, 2019. Speech-Driven Animation with Meaningful Behaviors. Speech Communication.
Saeed, A., Al-Hamadi, A., Niese, R., Elzobi, M., 2014. Frame-based facial expression recognition using geometrical features. Advances in Human-Computer Interaction 14, 1–14.
Sanchez-Mendoza, David, Masip, David, Lapedriza, Agata, 1 December 2015. Emotion recognition from mid-level features. Pattern Recognit. Lett. 67 (1), 66–74.
Sandbach, Georgia, Zafeiriou, Stefanos, Pantic, Maja, Rueckert, Daniel, October 2012. Recognition of 3D facial expression dynamics. Image Vis Comput. 30 (10), 762–773.
Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., et al., 2005. Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. Neuroimage 28 (4), 848–858 (special Section: Social Cognitive Neuroscience).
Sander, D., Grandjean, D., Scherer, K.R., 2005. A systems approach to appraisal mechanisms in emotion. Neural Netw. 18 (4), 317–352.
Santos, R., Marreiros, G., Ramos, C., Neves, J., Bulas-Cruz, J., 2011. Personality, emotion, andmood in agent-based group decision making. J. Intell. Syst. 26 (6), 58–66.
Santos, Marcus A.G., Munoz, Roberto, Olivares, Rodrigo, Pedro, P., Filho, Rebouças, Javier Del Ser, de Albuquerque, Victor Hugo C., 2020. Online heart monitoring systems on the internet of health things environments: a survey, a reference model and an outlook. Inf. Fusion 53, 222–239.
Sariyanidi, E., Gunes, H., Cavallaro, A., 2015. Automatic analysis of facial affect: a survey of registration, representation and recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37 (6), 1113–1133.
Satpute, A.B., Shu, J., Weber, J., Roy, M., Ochsner, K.N., 2013. The functional neural architecture of self-reports of affective experience. Biol. Psychiatry 73 (7), 631–638.
Sawada, M., Suda, K., Ishii, M., 2003. Expression of emotions in dance: relation between arm movement characteristics and emotion. Percept. Mot. Skills 97, 697–708.
Scheirer, J., Fernandez, R., Klein, J., Picard, R.W., 2002. Frustrating the user on purpose: a step toward building an affective computer. Interact. Comput. 14 (2), 93–118. http://www.sciencedirect.com/science/article/B6V0D-44GF457-1/2/9601b14ed27badfe069784931ea7cc31.
Scherer, K.R., 2002. Emotion, the Psychological Structure of Emotions. International Encyclopedia of the Social & Behavioral Sciences. Harvard Libraries, Oxford.
Scherer, K.R., 2005. What are emotions? And how can they be measured? Soc. Sci. Inf. 44 (4), 695–729.
Schmid, H., 1994. Probabilistic part-of-speech tagging usin decision trees. In: Proceedings of the International Conference on New Methods in Language Processing, pp. 44–49.
Schuller, B., Wimmer, M., Mösenlechner, L., Arsic, D., Rigoll, G., 2008. Brute-forcingHierarchical functionals for paralinguistics: a waste of feature space? Proc. of ICASSP, Las Vegas, NV 4501–4504.
Schuller, B., Valstar, M., Eyben, F., McKeown, G., Cowie, R., Pantic, M., 2011. Avec – the first international audio/visual emotion challenge. In: Proc. Of First International Audio/Visual Emotion Challenge and Workshop (AVEC 2011) Held in Conjunction with ACII, Memphis, Tennessee, USA, pp. 415–424.
Sebe, N., Lew, M., Sun, Y., Cohen, I., Gevers, T., Huang, T., 2007. Authentic facial expression analysis. Image Vis Comput. 25 (12), 1856–1863.

Shami, M.T., Kamel, M.S., 2005. Segment-based approach to the recognition of emotions in speech. In: IEEE International Conference on Multimedia and Expo, 2005. ICME 2005, p. 4.
Shanmugarajah, Kumaran, Gaind, Safina, Clarke, Alex, Peter, E., Butler, M., September 2012. The role of disgust emotions in the observer response to facial disfigurement. Body Image 9 (4), 455–461.
Shimizu, Masaki, Yoshizuka, Takeharu, Miyamoto, Hiroyuki, July 2007. A gesture recognition system using stereo vision and arm model fitting. Int. Congr. Ser. 1301, 89–92.
Shin, H.-C., Kim, S.-D., Choi, H.-C., 2007. Generalized elastic graph matching for face recognition. Pattern Recognit. Lett. 28 (9), 1077–1082.
Shioiri, T., Someya, T., Helmeste, D.M., Tang, S.W., 1999. Cultural difference in recognition of facial emotional expression: contrast between Japanese and American raters. Psychiatry Clin. Neurosci. 53, 629–633.
Shivhare, S., Khethawat, S., 2008. Emotion Detection from Text.
Smeaton, A.F., Rothwell, S., 2009. Biometric responses to music-rich segments in films: the cdvplex. In: Seventh International Workshop on Content-Based Multimedia Indexing, pp. 162–168.
Soleymani, M., Chanel, G., Kierkels, J.J., Pun, T., 2008. Affective ranking of movie scenes using physiological signals and content analysis. In: Proceeding of the 2nd ACM Workshop on Multimedia Semantics. ACM, New York, NY, USA, pp. 32–39.
Soleymani, M., Pantic, M., Pun, T., 2012. Multimodal emotion recognition in response to videos. IEEE Transactions on Affective Computing 3 (2), 211–223.
Song, M., You, M., Li, N., Chen, C., 2008. A robust multimodal approach for emo- tion recognition. Neurocomputing 71, 1913–1920.
Sreenivasa Rao, K., Saroj, V.K., Maity, Sudhamay, Koolagudi, Shashidhar G., 15 September 2011. Recognition of emotions from video using neural network models. Expert Syst. Appl. 38 (10), 13181–13185.
Srivastava, R., Roy, S., 2009. 3D facial expression recognition using residues. In: 2009–2009 IEEE Region 10 Conference on TENCON, pp. 1–5.
Stein, M.K., Newell, S., Wagner, E.L., Galliers, R., 2015. Coping with information technology: mixed emotions, vacillation, and nonconforming use patterns. MIS Q. 39 (2), 367e392.
Strapparava, C., Valitutti, A., 2004. WordNet-Affect: an affective extension of WordNet. In: Proceedings of LREC, vol. 4, pp. 1083–1086.
Subramanian, Hariharan, November 2004. Audio Signal Classification, M.Tech. Credit Seminar Report, Electronic Systems Group. EE. Dept, IIT Bombay submitted for publication.
Sun, P.H., Tao, L.M., 2008. Emotion measuring method in PAD emotional space. In: Proceedings of the Fourth Harmonious Man-Machine Environment Joint Academic Conference, pp. 638–645.
Sun, Y., Yin, L., 2008. Facial expression recognition based on 3D dynamic range model sequences. Computer Vision—ECCV 2008, 58–71.
Sun, Yaxin, Wen, Guihua, Wang, Jiabing, April 2015. Weighted spectral features based on local Hu moments for speech emotion recognition. Biomed. Signal Process. Control 18, 80–90.
Sun, Haotong, Lv, Guodong, Mo, Jiaqing, Lv, Xiaoyi, Du, Guoli, Liu, Yajun, 2019. Application of KPCA combined with SVM in Raman spectral discrimination. Optik 184, 214–219.
Suykens, J.A.K., Vandewalle, J., 1999. Least squares support vector machine classifiers. Neural Process. Lett. 9 (3), 293–300.
Syed, Idrus, Syed, Zulkarnain, et al., 2014. Soft biometrics for keystroke dynamics: profiling individuals while typing passwords. Comput. Secur. 45, 147–155.
Teager, H., Teager, S., 1990. Evidence for nonlinear production mechanisms in the vocal tract. In: Speech Production and Speech Modelling, vol. 55. Nato Advanced Institute, pp. 241–261.
Teng, Xiaolong, Wu, Bian, Yu, Weiwei, Liu, Chongqing, October 2005. A hand gesture recognition system based on local linear embedding. J. Vis. Lang. Comput. 16 (5), 442–454.
Tenopir, C., Wang, P., Zhang, Y., Simmons, B., Pollard, R., 2008. Academic users' interactions with sciencedirect in search tasks: affective and cognitive behaviors. Inf. Process. Manag.: Int. J. 44 (1), 105–121.
Theurel, Anne, Witt, Arnaud, Malsert, Jennifer, Lejeune, Fleur, Fiorentini, Chiara, Barisnikov, Koviljka, Gentaz, Edouard, 2016. The integration of visual context information in facial emotion recognition in 5- to 15-year-olds. J. Exp. Child Psychol. 150 (October), 252–271.
Tian, Y.l., Kanade, T., Cohn, J.F., 2001. Recognizing action units for facial expression analysis. Pattern Analysis and Machine Intelligence. IEEE Transactions on 23, 97–115.
Tian, Y.l., Kanade, T., Cohn, J.F., 2002. Evaluation of gaborwavelet- based facial action unit recognition in image sequences of increasing complexity. In: Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on, IEEE, pp. 229–234.
Tomasello, M., Carpenter, M., Call, J., Behne, T., Moll, H., 2005. Understanding and sharing intentions: the origins of cultural cognition. Behav. Brain Sci. 28 (5), 675–691 discussion 691–735.
Tomkins, S.S., 1984. Affect Theory. Approaches to Emotion. Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 163–197.
Trentin, Edmondo, Scherer, Stefan, Schwenker, Friedhelm, 15 November 2015. Emotion recognition from speech signals via a probabilistic echo-state network. Pattern Recognit. Lett. 66, 4–12.
Troussas, C., Krouska, A., Virvou, M., 2019. Trends on sentiment analysis over social networks: pre-processing ramifications, stand-alone classifiers and ensemble averaging. In: Tsihrintzis, G., Sotiropoulos, D., Jain, L. (Eds.), Machine Learning Paradigms. Intelligent Systems Reference Library, vol. 149. Springer, Cham.
Truong, K.P., Neerincx, M.A., Van Leeuwen, D.A., 2008. Assessing agreement of observer- and self-annotations in spontaneous multimodal emotion data. In: Proc. Interspeech 2008, 381–321.
Truong, Khiet P., van Leeuwen, David A., Franciska, M., de Jong, G., November 2012. Speech-based recognition of self-reported and observed emotion in a dimensional space. Speech Commun. 54 (9), 1049–1063.
Tsalakanidou, F., Malassiotis, S., 2010. Real-time 2D+3d facial action and expression recognition. Pattern Recognit. 43 (5), 1763–1775.
Tsalakanidou, F., Malassiotis, S., 2010. Real-time 2D+3D facial action and expression recognition. Pattern Recognit. 43 (5), 1763–1775.
Tsalakanidou, Filareti, Malassiotis, Sotiris, May 2010. Real-time 2D+3D facial action and expression recognition. Pattern Recognit. 43 (5), 1763–1775.
Valstar, M., Pantic, M., 2006. Fully automatic facial action unit detection and temporal analysis. In: Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on, IEEE, pp. 149–149.
Vankayalapati, H.D., Kyamakya, K., 2009. Nonlinear feature extraction approacheswith application to face recognition over large databases. In: Second Inter-national Workshop on Nonlinear Dynamics and Synchronization, Klagenfurt, pp. 44–48.
Vasuki, P., Aravindan, C., 2012. Improving emotion recognition from speech using sensor fusion techniques. In: IEEE Region 10 Conference (TENCON 2012), pp. 1–6.
Verma, G.K., Tiwary, U.S., 2014. Multimodal fusion framework: a multiresolution approach for emotion classification and recognition from physiological signals. Neuroimage 102, 162–172.
Verma, G.K., Tiwary, U.S., 2014. Multimodal fusion framework: a multiresolution approach for emotion classification and recognition from physiological signals. Neuroimage 102, 162–172.
de Vicente, A., 2003. Towards Tutoring Systems that Detect Students' Motivation: an Investigation. Ph.D. Thesis. School of Informatics. University of Edinburgh.
Villiger, Caroline, Hauri, Silke, Tettenborn, Annette, Hartmann, Erich, Näpflin, Catherine, Hugener, Isabelle, Niggli, Alois, 2019. Effectiveness of an extracurricular program for struggling readers: a comparative study with parent tutors and volunteer tutors. Learn. Instr. 60, 54–65.
Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. Computer Vision and Pattern Recognition 511–518.
Vondra, M., Vich, R., 2009. Recognition of emotions in German speech using Gaussian mixture models. In: Multimodal Signals: Cognitive and Algorithmic Issues, pp. 256–263.
Wang, P., Soergel, D., 1998. A cognitive model of document use during a research project. Study i. Document selection. J. Am. Soc. Inf. Sci. 49 (2), 115–133.
Wang, J.L.Y., Tang, X., 2010. Robust 3D face recognition by local shape difference boosting. IEEE Trans. Pattern Anal. Mach. Intell. 10, 1858–1870.
Wang, Xiang-Yang, Niu, Pan-Pan, Qi, Wei, 2008. A new adaptive digital audio watermarking based on support vector machine. J. Netw. Comput. Appl. 31 (4), 735–749.
Wang, Shangfei, Liu, Zhilei, Wang, Jun, Wang, Zhaoyu, Li, Yongqiang, Chen, Xiaoping, Ji, Qiang, October 2014. Exploiting multi-expression dependences for implicit multi-emotion video tagging. Image Vis Comput. 32 (10), 682–691.
Wang, X.-W., Nie, D., Lu, B.-L., 2014. Emotional state classification from EEG data using machine learning approach. Neurocomputing 129, 94–106.
Wang, Yaowei, Rao, Yanghui, Zhan, Xueying, Chen, Huijun, Luo, Maoquan, Yin, Jian, 1 November 2016. Sentiment and emotion classification over noisy labels. Knowl. Based Syst. 111, 207–216.
Wang, Jeen-Shing, Lin, Che-Wei, Yang, Ya-Ting C., 2013. A k-nearest-neighbor classifier with heart rate variability feature-based transformation algorithm for driving stress recognition. Neurocomputing 116, 136–143.
Wilhelm, rank H., Pfaltz, Monique C., Grossman, Paul, March 2006. Continuous electronic data capture of physiology, behavior and experience in real life: towards ecological momentary assessment of emotion. Interact. Comput. 18 (2), 171–186.
Wiskott, L., Fellous, J.-M., Kruger, N., von der Malsburg, C., 1997. Face recognition by elastic bunch graph matching. Int. Conf. Image Process. 1, 129–132.
Wöllmer, Martin, Kaiser, Moritz, Eyben, Florian, Schuller, Björn, Rigoll, Gerhard, February 2013. LSTM-Modeling of continuous emotions in an audiovisual affect recognition framework. Image Vis Comput. 31 (2), 153–163.
Wren, C., Azarbayejani, A., Darrell, T., Pentland, A., 1997. Pfinder: real-time tracking of the human body. IEEE Trans. Pattern Anal. Mach. Intell. 19 (7).
Wu, C.J., Huang, J.S., 1990. Human face profile recognition by computer. Pattern Recognit. 23, 255–260.
Wu, C., Chuang, Z., Lin, Y., Jun. 2006. Emotion recognition from text using semantic labels and separable mixture models. ACM Trans. Asian Lang. Inf. Process 5 (2), 165–183. https://doi.org/10.1145/1165255.1165259.
Wu, S., Falk, T.H., Chan, W.-Y, 2011. Automatic speech emotion recognition using modulation spectral features. Speech Commun. 768–785.
Xia, R., Zong, C., Li, S., 2011. Ensemble of feature sets and classification algorithms for sentiment classification. Inf. Sci. 181, 1138–1152.
Xie, Haoran, Li, Xiaodong, Wang, Tao, Raymond, Y., Lau, K., Wong, Tak-Lam, Chen, Li, Fu, Lee Wang, Li, Qing, 2016. Incorporating sentiment into tag-based user profiles and resource profiles for personalized search in folksonomy. Inf. Process. Manag. 52 (1), 61–72.
Yang, X.S., 2010. Firefly algorithm, levy Flights and global optimization. Res. Develop. Intell. Syst. 26, 209–218.
Yang, Zongkai, Liu, Zhi, Liu, Sanya, Lei, Min, Meng, Wenting, 20 November 2014. Adaptive multi-view selection for semi-supervised emotion recognition of posts in online student community. Neurocomputing 144, 138–150.
Yang, Q., Rao, Y., Xie, H., Wang, J., Wang, F.L., Chan, W.H., 2019. Segment-level joint topic-sentiment model for online review analysis. IEEE Intell. Syst. 34 (1), 43–50, 1 Jan.-Feb.

yasmina, Douiji, Hajar, Mousannif, Hassan, Al Moatassime, 2016. Using YouTube comments for text-based emotion recognition. Procedia Computer Science 83, 292–299.
Yoshitomi, Y., Kim, S., Kawano, T., Kilazoe, T., 2000. Effect of sensor fusion for recognition of emotional states using voice, face image and thermal image of face. In: 9th IEEE International Workshop on Robot and Human Interactive Communication, pp. 178–183.
You, Ji Won, Kang, Myunghee, August 2014. The role of academic emotions in the relationship between perceived academic control and self-regulated learning in online learning. Comput. Educ. 77, 125–133.
Yue, Ming, March 2015. Robust regional bounding spherical descriptor for 3D face recognition and emotion analysis. Image Vis Comput. 35, 14–22.
Zaki, J., Ochsner, K., 2011. Reintegrating the study of accuracy into social cognition research. Psychol. Inq. 22 (3), 159–182.
Zeng, Z., Pantic, M., Roisman, G., Huang, T., 2009. A survey of affect recognition methods: audio, visual, and spontaneous expressions. IEEE Trans. Pattern Anal. Mach. Intell. 31 (1), 39–58.
Zhang, Yan, Hua, Caijian, December 2015. Driver fatigue recognition based on facial expression analysis using local binary patterns. Optik - International Journal for Light and Electron Optics 126 (Issue 23), 4501–4505.
Zhang, Li, Jiang, Ming, Farid, Dewan, Hossain, M.A., 1 October 2013. Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot. Expert Syst. Appl. 40 (13), 5160–5168.
Zhang, Ligang, Tjondronegoro, Dian, Chandran, Vinod, February 2014. Facial expression recognition experiments with data from television broadcasts and the World Wide Web. Image Vis Comput. 32 (2), 107–119.
Zhang, L., Tjondronegoro, D., Chadran, V., 2014. Gabor based templates for facial expression recognition in images with facial occlusion. Neurocomputing 145 (5), 451–464.
Zhang, Yang, Zhang, Li, Hossain, M.A., 15 February 2015. Adaptive 3D facial action intensity estimation and emotion recognition. Expert Syst. Appl. 42 (Issue 3), 1446–1464.
Zhang, Li, Mistry, Kamlesh, Jiang, Ming, Chin Neoh, Siew, Alamgir Hossain, Mohammed, November 2015. Adaptive facial point detection and emotion recognition for a humanoid robot. Comput. Vis. Image Understand. 140, 93–114.
Zhang, Li, Mistry, Kamlesh, Chin Neoh, Siew, Lim, Chee Peng, 1 November 2016. Intelligent facial emotion recognition using moth-firefly optimization. Knowl. Based Syst. 111, 248–267.
Zhao, G., Pietikäinen, M., 2007. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. Pattern Anal. Mach. Intell. 29 (6), 915–928.
Zheng, Hao, Geng, Xin, Tao, Dacheng, Jin, Zhong, January 2016. A multi-task model for simultaneous face identification and facial expression recognition. Neurocomputing 171 (1), 515–523.
Zhou, Mingming, December 2016. The roles of social anxiety, autonomy, and learning orientation in second language learning: a structural equation modeling analysis. System 63, 89–100.
Zhou, G., Hansen, J., Kaiser, J., 2001. Nonlinear feature based classification of speech under stress. IEEE Trans. Speech Audio Process. 9 (3), 201–216.
Zhu, S.C., Yuille, A.L., 1995. FORMS: a flexible object recognition and modeling system. In: Proceedings of the Fifth International Conference on Computer Vision, pp. 465–472.