



Using Wearable Devices for Emotion Recognition in Mobile Human- Computer Interaction: A Review

Tao Wang  and Hanling Zhang  

School of Design, Hunan University, Changsha 410082, Hunan, China
jt_hl zhang@hnu.edu.cn

Abstract. With the popularity of wearable devices, human-computer interaction (HCI) has begun to expand to mobile environments. Since emotion is an essential part of the interaction experience, capturing users' emotional experience in mobile HCI using wearable devices has become a promising research area. We examined research papers in the above-mentioned fields in the past five years (2017–2021) and finally included 29 papers in the review results. We divide these articles by application area, analyze them in terms of several key elements of the experimental process, and give a clear view of the data. This review unravels the current practical applications and research methods of using wearable devices to capture physiological signals for emotion recognition, aiming to provide readers with guidelines for research methods and inspiration for future research directions.

Keywords: Mobile human-computer interaction · Physiological signals · Wearable devices · Emotion recognition

1 Introduction

The proliferation of wearable devices [1] and the development of mobile technologies have enabled the spread of HCI to all areas of daily life. Accordingly, mobile HCI and the corresponding user experience are receiving increasing attention [2].

Emotions, as an important part of the user experience, play a key role in positive human-computer interaction. Many psychologists believe that most human behavior and thinking are driven by emotions [3]. Several studies have also shown that emotions affect human productivity [4, 5], physical and mental health [1, 6], which has a huge impact on human-computer interaction.

Current emotion recognition methods are difficult to meet the requirements of real mobile HCI. Most research stays in the laboratory testing stage, and emotion recognition tests are rarely conducted in outdoor environments. Emotion recognition methods can generally be divided into those based on behavioral performance (expressions, speech, body movements, etc.) and those based on physiological signals. To date, there has been more research on affective computing around the former than the latter [7].

Although behavioral performance-based emotion recognition methods, such as representative facial expression recognition, have the advantage of being intuitive. However, in some cases, expressions are not obvious or can be disguised, which reduces the reliability and accuracy of recognition results [8]. Compared to methods based on behavioral performance, methods based on physiological signals are more advantageous in mobile human-computer interaction. Physiological signals are influenced by the human endocrine system and autonomic nervous system and are less influenced by the subjective human consciousness. Therefore, physiological signals are almost spontaneous and uncontrollable, which makes them a more objective indicator of true emotional response [8, 9].

Cameras in portable devices often capture only part of the facial expressions of users as they move [10, 11] and complex environments lead to unstable lighting conditions, which reduce the accuracy and feasibility of emotion recognition. In addition, behavioral performance-based methods may have problems in terms of device portability, user privacy, and data computation volume [9]. In contrast, wearable devices that measure changes in physiological signals are often user-friendly and consumer-grade, becoming better tools for measuring emotions outside the laboratory [12].

Under the theme of “Emotion recognition through physiological signal acquisition by wearable devices”, this paper focuses on practical applications and reports the process and results of the literature review. Specifically, this work focuses on the following research questions:

- RQ1: What wearable devices are used in these emotion recognition methods? What physiological signals are these methods based on?
- RQ2: What are the practical applications of this type of emotion recognition technology in HCI? What are the characteristics of each application?
- RQ3: What are the important links in the specific process of emotion recognition experiments? What are the similarities, differences and innovations?

These research questions address the key elements of emotion recognition in mobile HCI: wearable devices, physiological signals, and application domains, focusing on the testing process of emotion recognition in different application scenarios, including experimental environments, emotion-eliciting materials, emotion classification, and benchmarks for evaluating emotion recognition results. In this review, we visualized the results and related information in graphical form and answered each of the research questions in detail. The main contributions of the review are as follows:

- It provides a comprehensive review and summary of the physiological signals and wearable devices employed for emotion recognition in mobile HCI over the past five years, and can serve as a guide for relevant researchers in their selection and use.
- It categorizes the practical applications of emotion recognition in HCI and sorts out the specific process methods used in testing experiments.
- It summarizes the widespread problems and foreseeable future trends when performing emotion recognition in mobile HCI and provides inspiration for future research.

The paper is organized as follows: Sect. 2 presents the sentiment classification model and related work, and Sect. 3 describes the methodology and research methods used in the review. The review results are presented in Sect. 4 and discussed in Sect. 5. As a final remark, Sect. 6 concludes the paper.

2 Background

2.1 Emotion Models

Since human emotions are complex and variable, psychology has many emotion models to quantify feelings. Currently, two main emotion theories are widely used to classify and represent emotions, discrete emotion theory and dimensional emotion theory [13]. The first category is the basic emotion theory, which classifies emotions into discrete categories. The best known of these is the classification model [14], which divides emotions into six basic emotions: anger, fear, sadness, happiness, surprise, and disgust. These basic emotions combine to form other, more complex, non-basic emotions. Another is the dimensional emotion theory, where emotions are composed of different dimensions and discrete emotional states correspond to regions in a multidimensional space. The most classic is the circumplex model [15], which assesses emotions using two dimensions: arousal and valence. Arousal measures the intensity of the emotion, from low to high, while valence measures the pleasantness of the emotion, from negative to positive.

2.2 Related Work

In the past, several studies have reviewed physiological signal-based methods for emotion recognition. Some surveys related to the topic of our review are shown below, and their limitations are analyzed.

Two reviews [16, 17] have reviewed physiological signal-based emotion recognition methods. Similarly, they both stated that the use of wearable devices to detect heart-related physiological signals is a good way to accurately and inconspicuously measure emotions in real life. The limitation is that their testing environments are restricted to controlled laboratories. One difference is that one study [16] added facial expression recognition to the physiological signal.

Close to our research questions, some reviews [18–20] reviewed studies using wearable devices to measure physiological signals to recognize emotion recognition. The difference is that the wearable devices they summarize cannot all meet requirements such as lightweight and wireless, and the tests mostly stay in controlled laboratories. It is worth noting that they started to focus on emotion recognition in outdoor environments.

A more novel review [7] address non-invasive mobile sensing methods for emotion recognition in smartphone devices. Although cell phones are common mobile devices, cell phone cameras have many limitations in real mobile human-computer interaction.

On the topic of “emotion recognition around physiological signal detection by wearable devices,” previous reviews rarely focus on mobile HCI or practical applications. Therefore, we hope to fill the gap in this field with this paper.

3 Review Methodology

This study was conducted according to the systematic literature review guidelines [21, 22]. Following these two guidelines, our literature review was divided into three main steps: (1) information sources and search strategy, (2) study selection and quality assessment, (3) data extraction and synthesis.

3.1 Data Sources and Search Strategy

We searched the Web of Science database in December 2021 using the keywords “emotion recognition AND wearable device”. For publication time, we focused on the latest advances and trends in the field, so we only selected research articles from the last five years (2017 to 2021). For document types, we chose “Articles” and “Meeting” to exclude invalid data. For language, we set it to English.

3.2 Study Selection and Quality Assessment

After the initial review, a total of 205 articles potentially relevant to the current research question were shortlisted. Before the formal screening, 7 duplicate articles were excluded, yielding 198 articles. Then 108 articles were excluded based on the exclusion and inclusion criteria review, leaving 90 articles. The following are the inclusion and exclusion criteria we specified:

- Wearable devices. Do they meet the requirements of inconspicuousness and user-friendliness in mobile HCI? Are they consumer-grade products?
- Signal source. Does it mainly rely on capturing physiological signals to identify emotions? Some wearable devices rely on built-in miniature cameras for local facial expression recognition, with physiological signals as a supplement; they are excluded from the scope of this paper.
- Experimental testing. Whether real-life experimental tests have been conducted to test the emotion recognition method. Some articles only present a feasible framework for emotion recognition without actual experimental testing on subjects, which are excluded from this paper.

In addition, to ensure that each article has sufficient and appropriate information to demonstrate the methodology of their studies, we set additional quality criteria:

- Research focus. Does the research focus on practical applications and experimental methods?
- Application potential or value. Does the research have close to real-life use cases or tests in mobile HCI?
- Experiment Introduction. Are the experiments or cases described in detail and clearly? Does it include key elements such as test environment, stimulus materials, equipment, and data processing?

The quality criteria review excluded the remaining articles and the final 29 remaining articles were included in this review. The detailed process is illustrated in the PRISMA flowchart in Fig. 1.

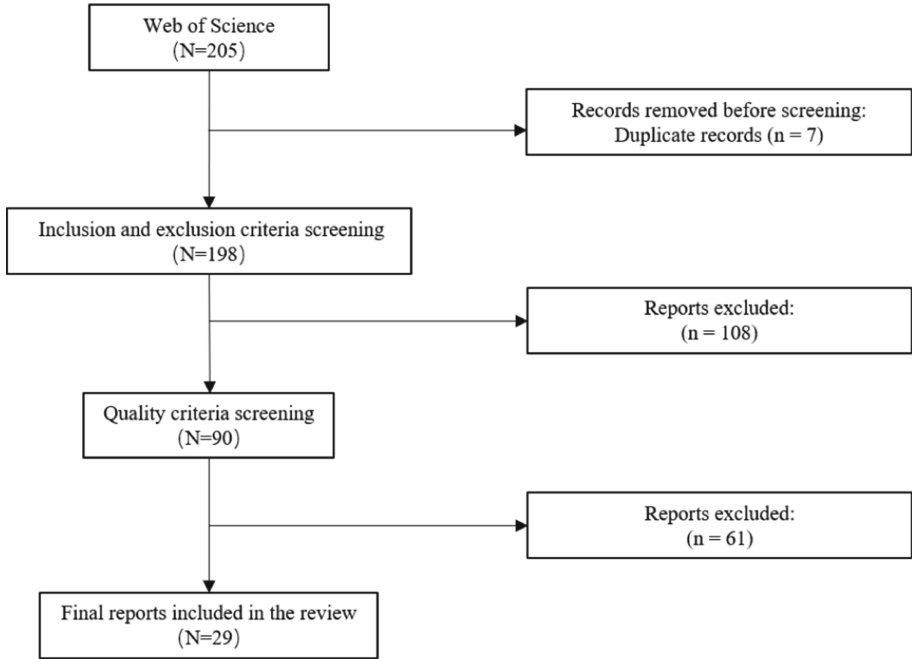


Fig. 1. PRISMA flowchart.

3.3 Data Extraction and Data Synthesis

Each article reviewed was read in more detail to extract the following key data.

- Practical applications of the study
- The wearable device, wireless transmission method, type of device, sensors carried
- Experimental environment
- Emotional stimulus materials
- Emotion classification
- Evaluation benchmark

First, we summarized the realistic application scenarios of emotion recognition in the field of human-computer interaction by classifying different studies according to the attributes of the target population, the experimental environment, and the type of emotion for their practical applications. Then, the experimental methods are organized in several dimensions as a guide for future research methods.

4 Results

4.1 Wearables and Physiological Signals

This section focuses on answering RQ1 using data charts and detailed information. First, we counted the number of times each type of device was used and its corresponding research literature, as shown in Table 1 below.

Table 1. Wearable devices and related research.

Device	Number of uses	Related Studies
Empatica E4	10	[9, 24, 28, 30, 31, 36, 39, 40, 45, 46]
Microsoft Band 2	5	[1, 25, 32, 39, 40]
Shimmer3 GSR + Unit	4	[4, 12, 26, 27]
Emotiv EPOC X	3	[6, 33, 44]
Shimmer3 ECG Unit	3	[24, 26, 38]
Interaxon MUSE	2	[12, 41]
TDK Silmme W20	2	[5, 34]
Emotiv Insight	1	[43]
Algoband F8	1	[8]
Analog ADI-VSM	1	[37]
Mio® alpha	1	[42]
NeuroSky	1	[32]
B-Alert x10	1	[23]

In addition, to provide a reference guide for future researchers in selecting wearable devices and physiological signals, we analyzed which devices researchers would choose more often when using specific signals, as shown in Fig. 2 below.

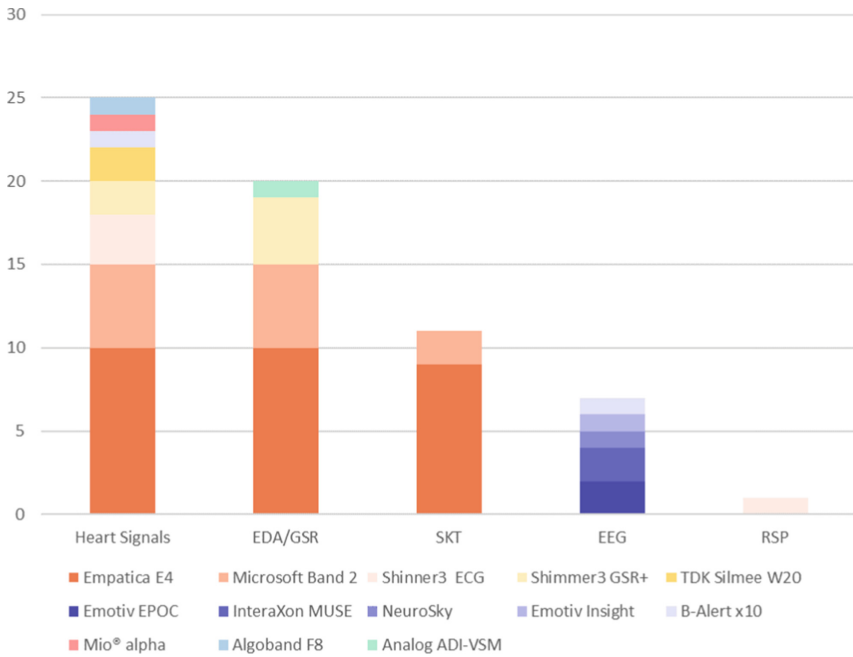


Fig. 2. Equipment usage in each signal.

According to the position of wearing, wearable devices are mainly divided into two types: wristband and headband. A special one, Shimmer3 ECG was worn on the chest. Compared to the headband devices, the wristband devices represented by Empatica E4 are more frequently used.

Physiological signals can be broadly classified into five categories: cardiac-related signals, electrical skin signals, skin temperature, electroencephalographic signals, and respiration.

Cardiac-related biosignals are used the most (up to 25 times), which include photoplethysmography (PPG) and electrocardiography (ECG). The PPG is more commonly used, with 21 times. Its measurement electrodes are usually integrated into smartwatches to make it easier for users to wear and measure while on the move. The ECG method is used less frequently, with only four times. Measurement devices are also only provided by Alert x10 [23] and Shimmer3 ECG units [24–26].

The EDA/GSR was used the second most often, reaching 20 times. Emotional stimuli cause an autonomous activation of the sweat glands of the skin, which in turn causes changes in the electrical activity of the skin [27]. The measurement sensors are usually installed on the finger (Shimmer3 GSR + Unit) or wrist (Empatica E4, Microsoft Band 2, Analog Devices ADI-VSM). Studies [4, 27, 28] have mentioned that GSR signals are more sensitive to the arousal of emotions compared to valence.

The skin temperature (SKT) was used the third most often, amounting to 11 times. The SKT is measured with the Empatica E4 and Microsoft Band 2, which have the infrared (IR) thermometer. The stimulation causes changes in sympathetically controlled smooth muscle and further causes changes in blood vessels [28]. Its more specific phenomenon is that in a relaxed state a person's blood vessels dilate and skin temperature rises, while in stress or anxiety the blood vessels contract and the skin becomes cold. Thus, measuring SKT can identify the degree of relaxation or stress in a person.

The EEG signal was used 8 times, which is relatively few. EEG signals are usually measured from the head, and the fixed acquisition position limits the measurement of other signals. Exceptionally, the B-Alert x10 measures both EEG signals and ECG signals. EEG can help researchers accurately detect changes in brain activity and perform unconscious and second-by-second assessments. However, EEG wearable devices face challenges of time-consuming setup and noise sensitivity in mobile use environments.

Respiration (RSP) was mentioned only once [26], collected by the Shimmer3 ECG Unit. The breathing patterns (speed and depth) contain rich information about the emotional state [29].

In addition, there are several noteworthy points about these devices:

- The Empatica E4 device has an event marker button for easy self-annotation by the user.
- The Microsoft Band 2 was discontinued in 2016, and the companion app was discontinued in 2019. This means that users can continue to use their devices (track heart rate, record exercise, track sleep, etc.), but the features provided by the cloud or mobile apps will no longer work.
- The Emotiv EPOC model mentioned in the studies is the Emotiv EPOC+, which has been discontinued and replaced by the upgraded EPOC X. The main features of both are similar.
- The Silmme W20 can measure pulse rate while moving, but in some cases (strenuous exercise, cold environments, poor blood flow, etc.) the signal may be lost.
- Some studies used other signal characteristics in addition to physiological signals in identifying emotions, including eye-tracking data [9] and acceleration data representing body movement [30].

4.2 Specific Applications

This section answers RQ2, focusing on different application scenarios' concerns and experimental procedures. Twenty-two studies account for specific application scenarios of emotion recognition, while the other seven articles only account for emotion recognition methods and performance without practical applications. To understand the research approaches of researchers in different fields, we divided the 22 articles by application areas into negative emotion detection, context-aware systems, analytical assessment, worker emotion analysis, and communication assistance. We summarized the characteristics of different application scenarios (application meaning, experimental setup, emotional stimulus material, and emotional classification), as shown in Table 2.

Table 2. Experimental procedures for different applications.

Application	References	Environment	Pre-experiment	Stimulation	Emotion Model	Baseline
Negative Emotion Detection	[6]	Labs	Assessment	Psychology Approach	Negative emotion	Self-report
	[36]	Labs	Assessment	Psychology Approach	Negative emotion	Self-report
		Dynamic	n/a	Real Scenes		Self-report
	[24]	Labs	Adjustment	Psychology Approach, Video Clips	Negative emotion	Self-report
	[25]	Dynamic	n/a	Real Scenes	Negative emotion	Self-report
	[26]	Labs	Adjustment	Video Clips	Negative emotion	n/a
	[1]	Labs	Adjustment	Emotional Pictures	Negative emotion	n/a
	[37]	Static	n/a	Real Scenes	Negative emotion	Self-report, Biomarkers
	[38]	Dynamic	n/a	Real Scenes	Negative emotion	n/a
[27]	Labs	Assessment	Audio Clips	Negative emotion	Self-report	
Unclassified	[8]	Labs	n/a	Video Clips	Valence	Emotional labels
	[12]	Labs	Adjustment	Emotional Pictures	Dimensional	Self-report
	[32]	Labs	Adjustment	Video Clips	Valence	Self-report
	[33]	Labs	n/a	Games	Valence	Self-report
	[31]	Labs	Adjustment	Video Clips	Dimensional, Discrete	Self-report
	[34]	Dynamic	n/a	Real Scenes	Dimensional, Discrete	Manual labeling
	[28]	Static	n/a	Audio Clips	Valence	Self-report
Context-aware Systems	[40]	Labs	n/a	Emotional Pictures	Dimensional	Self-report
		Labs	n/a	Games		n/a
	[39]	Labs	n/a	VR	Dimensional	Self-report
		Labs	n/a	Emotional Pictures		Emotional labels
	[41]	Labs	n/a	Video Clips	Dimensional, Discrete	Manual labeling
		Dynamic	n/a	Real Scenes		n/a

(continued)

Table 2. (continued)

Application	References	Environment	Pre-experiment	Stimulation	Emotion Model	Baseline
	[42]	Labs	n/a	n/a	Dimensional, Discrete	Self-report
	[43]	Static	n/a	Video Clips	Discrete	n/a
	[30]	Labs	Adjustment	Video Clips	Laughter	Manual labeling
Analysis and Evaluation	[9]	Dynamic	n/a	Video Clips	Dimensional	Manual labeling
	[23]	Labs	Adjustment	Emotional Pictures	Dimensional	Self-report
		Labs	n/a	VR		Self-report
		Dynamic	n/a	Real Scenes		Self-report
Worker Emotion Detection	[5]	Dynamic	n/a	Real Scenes	Dimensional, Discrete	n/a
	[4]	Labs	n/a	Emotional Pictures	Dimensional	Self-report
	[44]	Dynamic	n/a	Real Scenes	Dimensional	Biomarkers
Communication Assistance	[45]	Dynamic	n/a	Real Scenes	Dimensional	Manual labeling
	[46]	Labs	Measurement	Video Clips	Discrete	Self-report

- Uncategorized.** The seven studies in this category used emotion recognition methods that did not apply to specific scenarios but had potential and value for practical application. Most of these studies were tested in a controlled experimental setting, as no practical application scenarios were specified. Four articles set up a relaxation period before the experiment to allow subjects to maintain a neutral mood. Emotional stimuli relied primarily on visual and auditory stimuli, including still pictures [12], audio clips [28], and video clips [8, 31, 32], and video games [33]. Studies outside the laboratory setting [34] targeted five subjects and collected their physiological signals during 10 weeks of work. All seven studies had a similar feature in terms of mood categorization, all categorizing on the valence dimension (positive and negative). In addition, some of them [8, 35] adding neutral.
- Negative emotion detection.** Nine articles addressed negative emotion detection: six examined psychological assessment in everyday life [6, 24–27, 36] and three focused on negative emotion detection in specific scenarios, such as work [1], pre-surgery [37] and driving [38]. Studies in this category used a variety of emotionally stimulating materials. Four studies used realistic scenarios for their experiments: vehicle driving [38], pre-surgery [37], classroom lectures [36] and daily life [24, 36]. Six used laboratory simulations, divided into psychological tests [6, 26, 36] and audiovisual materials [1, 25–27]. These studies specifically address the emotion of stress, some of which are also elaborated as anxiety. Anxiety can essentially be explained as chronic

stress [6]. One study [26] classifies stress more carefully into three levels: relaxed, mildly stressful, and moderately stressful. In particular, a study on negative emotions in driving [38] considered both anxiety and anger.

- **Context-aware system (CAS).** Context-aware systems (CAS), including Cognitive assistants [39] and personal assistants [40], provide users with information prompts and decision support services by collecting data from users and the environment. The addition of emotion recognition allows the CAS to understand users' feelings and needs better and provide more humanized services. This type of research has affluent subdivisions: mobile scenarios [39], emotional games [40], life and health logs [41], and music recommendations [42]. Two studies are more specific, Cai et al. [43] studying the relationship between human personality, behavior and emotion; and Di Lascio et al. [30] using physiological signals and body movement data to detect laughter. The emotionally stimulating materials used in these studies were essentially visual and or auditory.
- **Analysis and Evaluation.** Emotions play an essential role in user experience in consumption, travel, and entertainment. Two articles use emotion recognition for targeted assessment tests. One study [9] had subjects walk or stand in a real outdoor scenario while watching video clips. The other [23] was more complex, with a controlled laboratory and an actual pavilion. First, subjects viewed emotional pictures of the IAPS or experienced a VR virtual exhibit in the controlled laboratory. In addition, subjects assessed the show in an actual pavilion. Similarly, both studies used the dimensional emotions model.
- **Worker state detection.** The emotional state of workers affects key production factors such as safety, efficiency, and employee turnover. Therefore, understanding the emotional state of employees at work can help workers do their jobs better, and companies improve their efficiency [4, 5, 44]. Three articles investigated the relationship between mood and working conditions [44] and between mood and productivity [4, 5], respectively. Two studies [5, 44] used different work scenarios in actual workplaces to elicit emotions. Another study in a laboratory setting [4] used photographs of workers' daily lives to evoke emotions. Similarly, all three articles categorized emotions according to an arousal-valence dimensional model.
- **Communication Assistance.** Individuals with autism or developmental delays have difficulty expressing their feelings or seeking help. Unobtrusive wearable devices can identify emotions to address such communication impairments. The study on children with developmental abnormalities or delays [45] conducted experiments in real everyday life; another study on individuals with autism [46] used video clips to elicit emotions under laboratory conditions.

4.3 Experimental Environment

Unlike the previous section, the sections from 4.3 to 4.7 are not limited to individual application areas but focus more on the experimental process in overall mobile HCI to answer RQ3.

We divided the different experimental settings into the controlled laboratory (22 papers) and realistic environment (13 papers). Since multiple experiments in different environments may be conducted in a single study, the cumulative number of experiments

exceeds the finalized 29 papers. Based on the motion state of the wearable device, the realistic environments were further divided into static and dynamic categories. Static means that the subjects need to follow regulations to maintain a specific state, similar to a controlled laboratory, while dynamic represents the freedom of movement that the subjects can perform without restrictions. There were 3 static experiments and 10 dynamic experiments. The dynamic experiments in the natural environment are very close to the real mobile human-computer interaction and have guiding and reference values, so we analyzed the wearable devices applied in the naturally dynamic environment, as shown in Fig. 3.

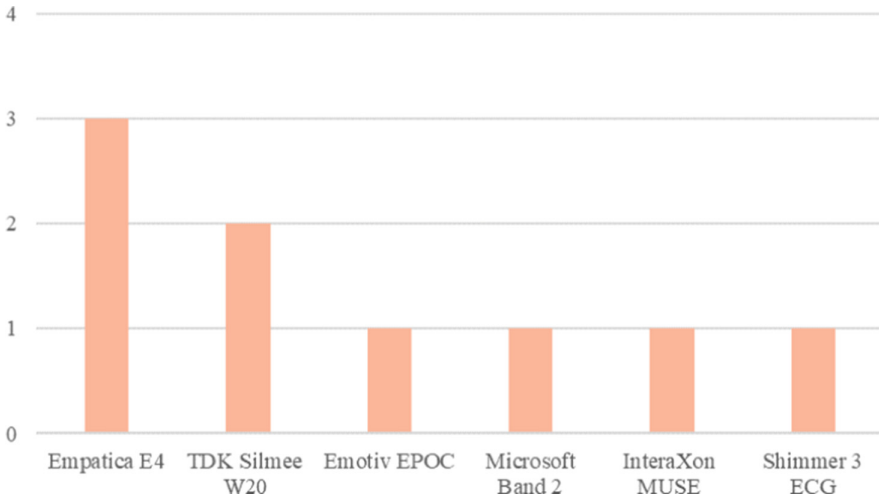


Fig. 3. Wearable device usage in dynamic real-world environments.

Some patterns can be found: these wearable devices in the figure all have the potential for practical application in mobile HCI. Among them, the more used ones are the wristband wearables like Empatica E4 [9, 36, 45], Silmee W20 [5, 34], and Microsoft Band2 [24]. Compared to EEG headbands such as the Emotiv Epoc [44] and InteraXon MUSE [41], wristband wearables are lighter, less conspicuous, measure a wider variety of signals, are more preferred by researchers.

4.4 Pre-experimental Sessions

Many studies will set up a pre-experimental session to measure, assess, or stabilize subjects' emotional baseline (often referred to as a neutral mood or relaxed state). We counted this literature, and the results are shown in Table 3.

The obvious point is that only studies in which the testing environment was in a controlled laboratory (12 studies) arranged some pre-experimental sessions. This was also done for all studies where the application was a negative emotion test and the experimental setting was a controlled laboratory.

Table 3. Specific details of the pre-experimental session.

Application	References	Pre-experiment	Specific operation
Negative emotion detection	[6]	Assessment	The Hamilton Anxiety Rating Scale (HAM-A)
	[36]	Assessment	The Perceived Stress Scale (PSS)
	[24]	Adjustment	Audio for meditation
	[26]	Adjustment	Relaxing video clips
	[1]	Adjustment	IAPS (Non-stress videos)
	[27]	Assessment	The Visual Analogue Scale for Anxiety (VAS-A)
Unclassified	[12]	Adjustment	2 min of relaxation time
	[32]	Adjustment	Rest time (10 s open eyes and 10 s closed eyes)
	[31]	Adjustment	60 s of relaxation time
Context-aware systems	[30]	Adjustment	60 s of relaxation time
Analysis and evaluation	[23]	Adjustment	Watch the white screen for a 4-min break
Communication Assistance	[46]	Measurement	Calculate the average value of ten-second signal

Assessment procedures typically use psychological questionnaires or physiological signal measures to assess the subjects' emotional state to determine whether they are in the desired neutral state. Stabilization procedures often use blank or neutral audiovisual material to relax subjects for 10 s to 4 min to ensure that the emotional state is neutral (relaxed) before the formal experiment.

4.5 Emotionally Stimulating Material

The previous Sect. 4.2 analyzed the stimulus materials used in different application scenarios, while this section will sort, analyze and summarize the emotional stimulus materials as a whole, as shown in Fig. 4.

The video clips were used 10 times and were commonly used in all areas. Six of these were video clips selected by the researchers themselves, and these were more flexible and adapted to the needs of the study. The next most popular video clip was the Database for Emotion Analysis (DEAP) [25, 41], which was used twice.

Real-life scenarios occurred 10 times, and the emotional stimuli faced by subjects in real-life and work scenarios were more complex and could not be specifically counted. Specific to the study of anger anxiety while driving [38], it developed an Android app to capture photos in front of the car. These photos were then manually analyzed to extract potential stimuli that caused changes in driver mood, including traffic density, road complexity, and any obstacles that could cause stress.

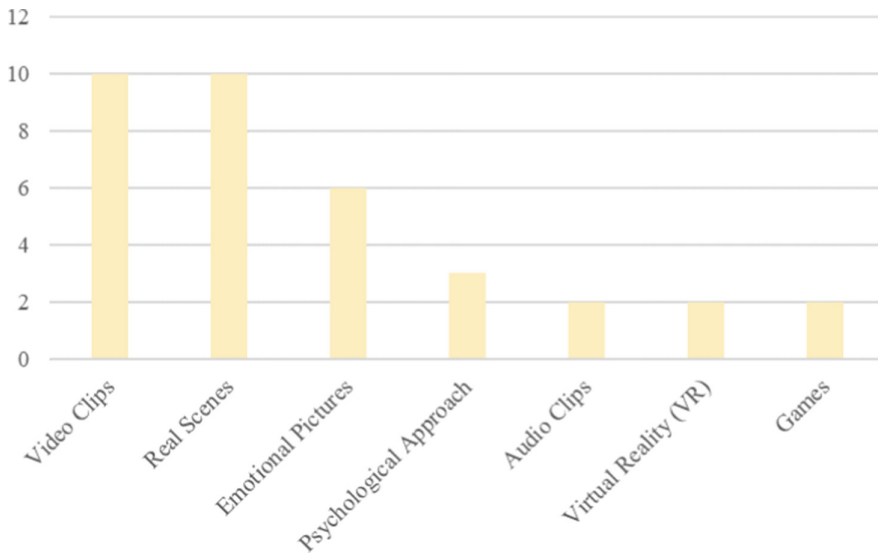


Fig. 4. The use of different emotionally stimulating materials.

Emotional pictures appeared 6 times, commonly using large databases of validated emotional pictures such as the International Affective Picture System (IAPS) and the Nencki Affective Picture System (NAPS). However, different individuals may have biased feelings and assessments of these pictures, so Fortune et al. (2020) chose personal photographs as emotional stimulus material in the hope that they would elicit more realistic and easily measurable emotional responses.

The psychological method was used 3 times, all of them appearing in studies in the category of negative emotion (stress and anxiety) detection. Of these, the Trier Social Test (TSST) [26, 36] was used more often.

The audio material was used 2 times. One was the canonical database the International Affective Digitized Sounds (IADS-2) [28] and the other was a researcher-selected audio clip [27].

Both VR and games were used 2 times, but they do not have canonical databases and were mostly screened by the researchers themselves.

4.6 Emotion Classification

According to Table 2, we count the number of times different emotion classification methods were used. As shown in the Fig. 5, there are broadly four types of emotion classification: dimensional emotions, negative emotions, discrete emotions, and positive-negative emotions. The arousal-potency dimensional model was the most widely used, with 13 studies basing their mood classification on this. Studies in the negative emotion detection category all focused on detecting stress, anxiety, or anger emotions.

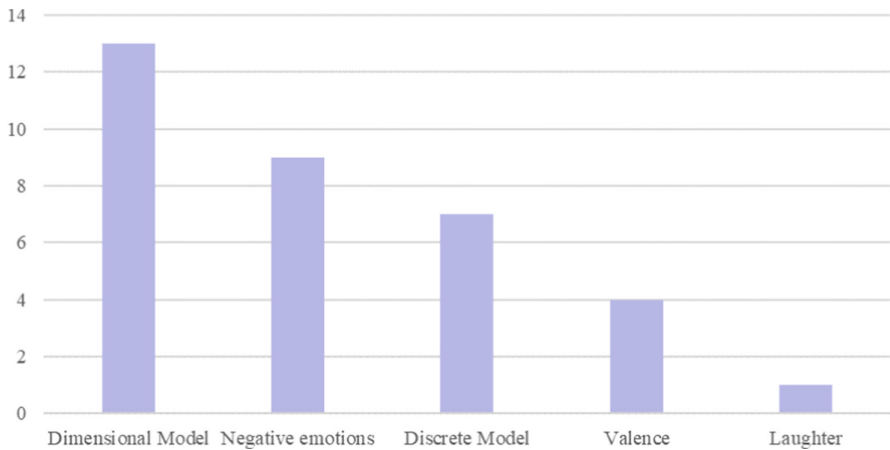


Fig. 5. The use of different emotion models.

The discrete emotion model is not usually used alone. Discrete emotions corresponded to different regions on the two-dimensional plane of arousal-valence, and five of the seven studies used the discrete emotion model together with the arousal-valence dimension model. Only two used the discrete emotion model alone in the classification of emotions.

The categorization of emotions according to valence (negative-positive) is somewhat ambiguous in its definition of different emotions and appears only in studies that do not specify an application scenario, which may indicate its difficulty in adequately describing the user's emotional experience in real emotion recognition applications.

There is only one study on laughter detection [30].

4.7 Baseline for Assessment

The vast majority of experiments established baselines for judging the accuracy of emotion recognition results, including self-reports completed by subjects, emotion labels that came with the emotion stimulus material, manual observation and labeling, and biological benchmarks. Table 4 shows the baseline used for all experiments to determine the accuracy of the emotion recognition results; Fig. 6 reveals the usage of different baselines.

Self-reports were used 20 times, mostly on scales or questionnaires that are widely used in psychology. Of these, the Self-Assessment Manikin (SAM) was used the most (seven times); the Likert scale was used four times; the State-Trait Anxiety Inventory (STAI) was used twice. The Perceived Stress Scale (PSS) was used three times, all in negative emotion detection studies. Nalepa, Kutt, Giżycka, et al. [40] used an off-the-shelf application, the PsychoPy program. Two studies [32, 42] used questionnaires custom-designed by the researchers.

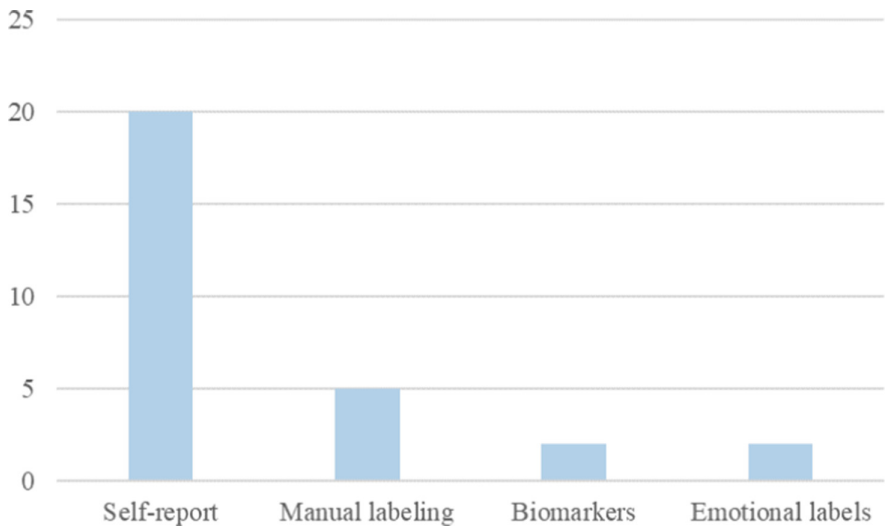
Table 4. The specific content of the baseline for assessment.

Application	References	Baseline	Specific content
Negative Emotion	[6]	Self-report	Psychological scales (SAM)
	[36]	Self-report	Psychological scales (PSS)
		Self-report	Psychological scales (PSS, EMAs)
	[26]	Self-report	Psychological scales (VASS)
	[24]	Self-report	Psychological scales (PSS)
	[37]	Self-report,	Psychological scales (STAI)
		Biomarkers	Cortisol
[27]	Self-report	Psychological scales (a Likert scale)	
Unclassified	[8]	Emotional labels	From video clips
	[12]	Self-report	Psychological scales (SAM)
	[31]	Self-report	Psychological scales (a Likert scale)
	[33]	Self-report	Psychological scales (SAM)
	[32]	Self-report	Psychological scales (Customized questionnaire)
	[34]	Manual labeling	Subjects labeled themselves
	[28]	Self-report	Psychological scales (PANAS, STAI, SSSQ)
Context-aware Systems	[40]	Self-report	Psychological scales (PsychoPy program)
	[39]	Self-report	Psychological scales (a Likert scale)
		Emotional labels	From emotional pictures
	[41]	Manual labeling	By subjects
	[42]	Self-report	Psychological scales (Customized questionnaire)
	[30]	Manual labeling	By professional annotator
Analysis and Evaluation	[9]	Manual labeling	Subjects labeled themselves
	[23]	Self-report	Psychological scales (SAM)
		Self-report	Psychological scales (SAM)
		Self-report	Psychological scales (SAM)
	[4]	Self-report	Psychological scales (a Likert scale)
[44]	Biomarkers	Cortisol	

(continued)

Table 4. (continued)

Application	References	Baseline	Specific content
Communication Assistance	[45]	Manual labeling	Diary of observations by people around subjects
	[46]	Self-report	Psychological scales (SAM)

**Fig. 6.** Usage of different benchmarks.

Manual annotation was usually divided into annotation by the subjects themselves and annotation by professional observers. Subjects' own labeling is usually done in between video clips [9, 41] or when emotions are strong [34]. Professional observer labeling varies depending on the experimental setting. Daily life scenarios relied on people around the subjects, such as in a study of emotional expression in children with autism [45], where family members and teachers observed children's behaviors and activities and labeled them in the form of a diary. In laboratory scenarios, professional observers annotated by watching back video recordings [30].

A novel approach is the detection of cortisol in saliva, a reliable biomarker proven by research to reflect levels of stress, anger and depression [37, 44].

In addition, two studies [8, 39] used mood labels that come with the mood stimulus material as a baseline for testing the identification results. These mood stimulus materials were derived from widely used databases, and using their self-contained mood labels is a more convenient method.

5 Discussion

This section discusses our findings around the review results and research questions, as well as analyzes their similarities and differences with findings in other review articles. Based on the review findings, we discuss the problems and future directions in the reviewed studies.

For the selection of physiological signals, the use of wearable devices to detect cardiac-related physiological signals in real life is a good way to accurately and inconspicuously measure emotions, which is in line with the findings of other reviews [16, 17]. Slightly less than the former, the EDA/GSR signal is also widely used and proves to be a well-suited signal source. In addition, far more studies have used multimodal signals than single-modal signals, proving that the combination of multimodal signals performs better in terms of recognition performance.

In the use of wearable devices, based on the type of signal corresponding to the device and the ease of wearing the device, wristband devices equipped with signal sensors such as heart-related signals and EDA are more widely used, compared to headband devices based on EEG signals.

Among the classifications of practical applications, negative emotion detection and context-aware systems are the most widely used areas of emotion recognition technology, but most of the remaining studies only propose a framework for emotion recognition without practical applications, which can reflect that the exploration of practical applications is still insufficient. This paper's review and classification of practical applications of emotion recognition in mobile HCI is a novel exploration that has not yet been found in other research reviews.

In terms of experimental settings, there are even fewer studies on the application of emotion recognition in real-life HCI. Only 10 dynamic experiments in real-life environments have been conducted, which implies that there are still a large number of issues to be explored and studied in the process of transforming technology into applications.

Before the stimulation begins, studies that place experiments in controlled laboratories usually arrange procedures to keep subjects in a neutral emotional state to ensure the accuracy of the data. Given the complexity of the experimental setting and the time-consuming nature of the pre-experimental phase, real-life experiments usually ignore the pre-experimental phase.

The widely used emotional stimulus materials are video clips, real-life scenes or emotional pictures, which is consistent with the finding of another review [18]. The stimulus materials used in studies on negative emotion detection are noteworthy: some psychological methods are more widely used, such as the Trier Social Test (TSST) [26, 36] which is used to elicit stress or anxiety in subjects.

The dimensional model of arousal-potency is widely used by many studies, in contrast to the claim of another survey [7] (that many studies of mobile affective computing aim to identify discrete emotions). The discrete emotion models in the reviewed studies were mostly used in combination with dimensional models.

In contrast to the study by Saganowski et al. [18] that used only self-reports as a basis for emotion assessment, our work incorporates manual annotation, biomarkers, and emotion labels that come with the stimulus material, providing researchers with more flexible options. Of course, self-report is still the primary source of basic facts about

emotions for many researchers, and it includes commonly used questionnaires such as the Likert scale, the Self-Assessment Manikin (SAM), and the State-Trait Anxiety Inventory (STAI).

Despite the progress made, real mobile HCI applications are scarce, with only 1/3 of the studies reviewed attempting emotion recognition applications using wearable devices in real environments. A variety of factors have constrained the progress of research.

First, the user experience of the device. Only a small class of devices such as the Empatica E4 and Microsoft Band 2 have managed to measure inconspicuously, while the rest still suffer from exposed cables [25, 26] and excessive size. EEG headgear such as the Emotiv Epoc and InteraXon MUSE face problems with cumbersome electrode setup processes and interference from motion artifacts, and the headgear limits suitable scenarios. Both Shimmer3 GSR Unit and Shimmer3 ECG have exposed cables, leading to the inevitable interference of human motion in the real environment by them. Second, the problem of equipment setup. Subjects were often asked to undergo a signal stabilization period of 5 to 10 min after putting on the devices [26, 28]. Similarly, subjects were asked to clean the skin on the forehead and behind the ear before wearing the EEG headband to reduce the impedance between the skin and the electrodes [12]. These settings disrupt the coherence and ease of use in real human-computer interaction. In addition, almost all devices inevitably suffer from poor signal quality in mobile conditions.

There are few studies on how to identify the source of stimuli that cause mood changes. T. Zhang et al. [9] attempted to have subjects wear head-mounted eye-tracking devices to collect eye-movement data for video providers to analyze the relationship between video content and users' emotions. Dobbins & Fairclough [38] used a camera to capture the view in front of the driving vehicle to provide a human analysis of the relationship between road conditions and mood changes. In-depth exploration of the user's emotional experience requires clear identification of the source of emotional stimuli in the interaction and then making adjustments, for which we hope and suggest more research be explored in the future.

6 Conclusions

This paper is based on the systematic literature review approach, focusing on the use of wearable devices to measure physiological signals for emotion recognition and highlighting practical applications of emotion recognition in mobile HCI. Closely focused on the above-mentioned topics, we finally obtained 29 articles through a review of the Web of Science database and a rigorous inclusion and exclusion process.

Extracting valid data from these articles and performing statistical analysis, we have the research directions and experimental methods for emotion recognition in mobile HCI, which can be used as a reference guide for future researchers. This review is mainly for researchers who wish to translate emotion recognition technology into practical applications. It is hoped that the results of our review will provide them with references in equipment selection and experimental design, and bring inspiration for their research directions.

The limitations of this review are the number of articles used and the research questions. Due to limited manpower and time, there is a possibility of expanding the scope of the reviewed articles. The research questions focus on research methods and experimental procedures, without sorting out signal processing methods and algorithmic models.

Emotion recognition methods are based on physiological signals only. Some studies were not only based on physiological signals but also incorporated other data sources such as facial expressions, speech, or eye movement signals, for which we can conduct more extensive research in the future.

Acknowledgment. This work was supported by funds for the National Natural Science Foundation of Changsha (kq2202176), Key R&D Program of Hunan(2022SK2104), Leading plan for scientific and technological innovation of high-tech industries of Hunan(2022GK4010), National Key R&D Program of China(2021YFF0900602), the National Natural Science Foundation of China (61672222).

References

1. Khowaja, S.A., Prabono, A.G., Setiawan, F., Yahya, B.N., Lee, S.-L.: Toward soft real-time stress detection using wrist-worn devices for human workspaces. *Soft. Comput.* **25**(4), 2793–2820 (2021)
2. Landowska, A., Miler, J.: Limitations of emotion recognition in software user experience evaluation context. In: 2016 Federated Conference on Computer Science and Information Systems (FedCSIS). pp. 1631–1640 (2016)
3. Agarwal, A., Meyer, A.: Beyond usability: evaluating emotional response as an integral part of the user experience. In: CHI '09 Extended Abstracts on Human Factors in Computing Systems. pp. 2919–2930. Association for Computing Machinery, New York, NY, USA (2009)
4. Fortune, E., Yusuf, Y., Blocker, R.: Measuring arousal and emotion in healthcare employees using novel devices. In: 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE). pp. 835–838 (2020)
5. Kadoya, Y., Khan, M.S.R., Watanapongvanich, S., Binnagan, P.: Emotional status and productivity: Evidence from the special economic zone in Laos. *Sustainability* **12**, 1544 (2020)
6. Baghdadi, A., Aribi, Y., Fourati, R., Halouani, N., Siarry, P., Alimi, A.: Psychological stimulation for anxious states detection based on EEG-related features. *J. Ambient. Intell. Humaniz. Comput.* **12**(8), 8519–8533 (2021)
7. Tzafilkou, K., Economides, A.A., Protogeros, N.: Mobile sensing for emotion recognition in smartphones: a literature review on non-intrusive methodologies. *Int. J. Human-Computer Interaction*. 0, 1–15 (2021). <https://doi.org/10.1080/10447318.2021.1979290>
8. Shu, L., et al.: Wearable emotion recognition using heart rate data from a smart bracelet. *Sensors* **20**, 718 (2020)
9. Zhang, T., El Ali, A., Wang, C., Hanjalic, A., Cesar, P.: Corrnnet: fine-grained emotion recognition for video watching using wearable physiological sensors. *Sensors* **21**, 52 (2021)
10. Kwon, J., Kim, D.-H., Park, W., Kim, L.: A wearable device for emotional recognition using facial expression and physiological response. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 5765–5768 (2016)

11. Nie, J., Hu, Y., Wang, Y., Xia, S., Jiang, X.: SPIDERS: Low-Cost Wireless Glasses for Continuous In-Situ Bio-Signal Acquisition and Emotion Recognition. In: 2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI). pp. 27–39 (2020)
12. Laureanti, R., et al.: Emotion assessment using machine learning and low-cost wearable devices. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC). pp. 576–579 (2020)
13. Zhang, J., Yin, Z., Chen, P., Nichele, S.: Emotion recognition using multi-modal data and machine learning techniques: a tutorial and review. *Information Fusion*. **59**, 103–126 (2020). <https://doi.org/10.1016/j.inffus.2020.01.011>
14. Ekman, P.: An Argument for Basic Emotions. *Psychol. Bull.* **6**, 169–200 (1992). <https://doi.org/10.1080/02699939208411068>
15. Russell, J.A.: A circumplex model of affect. *J. Pers. Soc. Psychol.* **39**, 1161–1178 (1980). <https://doi.org/10.1037/h0077714>
16. Egger, M., Ley, M., Hanke, S.: Emotion recognition from physiological signal analysis: a review. *Electronic Notes in Theoretical Computer Science* **343**, 35–55 (2019). <https://doi.org/10.1016/j.entcs.2019.04.009>
17. Shu, L., et al.: A review of emotion recognition using physiological signals. *Sensors* **18**, 2074 (2018). <https://doi.org/10.3390/s18072074>
18. Saganowski, S., et al.: Emotion recognition using wearables: a systematic literature review - work-in-progress. In: 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). pp. 1–6 (2020)
19. Schmidt, P., Reiss, A., Dürichen, R., Laerhoven, K.V.: Wearable-based affect recognition—a review. *Sensors* **19**, 4079 (2019). <https://doi.org/10.3390/s19194079>
20. Wijasena, H.Z., Ferdiana, R., Wibirama, S.: A survey of emotion recognition using physiological signal in wearable devices. In: 2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS). pp. 1–6 (2021)
21. Kitchenham, B.: Procedures for performing systematic reviews. Keele, UK, Keele University. 33, 1–26 (2004)
22. Mohamed Shaffril, H.A., Samsuddin, S.F., Abu Samah, A.: The ABC of systematic literature review: the basic methodological guidance for beginners. *Qual. Quant.* **55**(4), 1319–1346 (2021). <https://doi.org/10.1007/s11135-020-01059-6>
23. Marín-Morales, J., et al.: Real vs. immersive-virtual emotional experience: analysis of psychophysiological patterns in a free exploration of an art museum. *PLOS ONE*. **14**, e0223881 (2019). <https://doi.org/10.1371/journal.pone.0223881>
24. Dobbins, C., Fairclough, S., Lisboa, P., Navarro, F.F.G.: A lifelogging platform towards detecting negative emotions in everyday life using wearable devices. In: 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). pp. 306–311 (2018)
25. Markova, V., Ganchev, T., Kalinkov, K.: Detection of negative emotions and high-arousal negative-valence states on the move. In: 2018 Advances in Wireless and Optical Communications (RTUWO). pp. 61–65 (2018)
26. Montesinos, V., Dell’Agnola, F., Arza, A., Aminifar, A., Atienza, D.: Multi-modal acute stress recognition using off-the-shelf wearable devices. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 2196–2201. IEEE (2019)
27. Gasparini, F., Giltri, M., Bandini, S.: Discriminating affective state intensity using physiological responses. *Multimedia Tools and Applications* **79**(47–48), 35845–35865 (2020)

28. Cosoli, G., Poli, A., Scalise, L., Spinsante, S.: Measurement of multimodal physiological signals for stimulation detection by wearable devices. *Measurement* **184**, 109966 (2021)
29. Zhang, Q., Chen, X., Zhan, Q., Yang, T., Xia, S.: Respiration-based emotion recognition with deep learning. *Comput. Ind.* **92**, 84–90 (2017)
30. Di Lascio, E., Gashi, S., Santini, S.: Laughter recognition using non-invasive wearable devices. In: *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*. pp. 262–271. Association for Computing Machinery, New York, NY, USA (2019)
31. Ahn, I.Y., Sung, N.-M., Lim, J.-H., Seo, J., Yun, I.D.: Development of an oneM2M-compliant IoT platform for wearable data collection. *KSII Trans. Internet Inf. Syst. (TIIS)* **13**, 1–15 (2019)
32. Zhao, B., Wang, Z., Yu, Z., Guo, B.: EmotionSense: emotion recognition based on wearable wristband. In: *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. pp. 346–355. IEEE (2018)
33. Alakuş, T.B., Türkoğlu, İ.: EEG-based emotion estimation with different deep learning models. In: *2019 4th International Conference on Computer Science and Engineering (UBMK)*. pp. 33–37 (2019)
34. Hayano, J., Tanabiki, T., Iwata, S., Abe, K., Yuda, E.: Estimation of emotions by wearable biometric sensors under daily activities. In: *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*. pp. 240–241 (2018)
35. Apicella, A., Arpaia, P., Mastrati, G., Moccaldi, N., Prevete, R.: Preliminary validation of a measurement system for emotion recognition. In: *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. pp. 1–6 (2020)
36. Can, Y.S., Gokay, D., Kılıç, D.R., Ekiz, D., Chalabianloo, N., Ersoy, C.: How laboratory experiments can be exploited for monitoring stress in the wild: a bridge between laboratory and daily life. *Sensors*. **20**, 838 (2020). <https://doi.org/10.3390/s20030838>
37. Anusha, A.S., et al.: Electrodermal activity based pre-surgery stress detection using a wrist wearable. *IEEE J. Biomed. Health Inform.* **24**, 92–100 (2019)
38. Dobbins, C., Fairclough, S.: A mobile lifelogging platform to measure anxiety and anger during real-life driving. In: *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. pp. 327–332 (2017)
39. Nalepa, G.J., Kutt, K., Bobek, S.: Mobile platform for affective context-aware systems. *Futur. Gener. Comput. Syst.* **92**, 490–503 (2019). <https://doi.org/10.1016/j.future.2018.02.033>
40. Nalepa, G.J., Kutt, K., Giżycka, B., Jemiołło, P., Bobek, S.: Analysis and use of the emotional context with wearable devices for games and intelligent assistants. *Sensors*. **19**, 2509 (2019)
41. Jiang, S., Li, Z., Zhou, P., Li, M.: Memento: an emotion-driven lifelogging system with wearables. *ACM Trans. Sen. Netw.* **15**, 8:1–8:23 (2019). <https://doi.org/10.1145/3281630>
42. Chiu, M.-C., Ko, L.-W.: Develop a personalized intelligent music selection system based on heart rate variability and machine learning. *Multimedia Tools Appl.* **76**(14), 15607–15639 (2017)
43. Cai, R., Guo, A., Ma, J., Huang, R., Yu, R., Yang, C.: Correlation analyses between personality traits and personal behaviors under specific emotion states using physiological data from wearable devices. In: *2018 IEEE 16th Intl Conf on Dependable, Autonomous and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*. pp. 46–53 (2018)
44. Hwang, S., Jebelli, H., Choi, B., Choi, M., Lee, S.: Measuring workers' emotional state during construction tasks using wearable EEG. *J. Constr. Eng. Manag.* **144**, 04018050 (2018)

45. Redd, C.B., et al.: Physiological signal monitoring for identification of emotional dysregulation in children. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 4273–4277. IEEE (2020)
46. Pollreisz, D., TaheriNejad, N.: A simple algorithm for emotion recognition, using physiological signals of a smart watch. In: 2017 39th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 2353–2356. IEEE (2017)