

Chapter 10

Your Smartphone Knows you Better than you May Think: Emotional Assessment ‘on the Go’ Via TapSense



Surjya Ghosh, Johanna Löchner, Bivas Mitra, and Pradipta De

Introduction

Emotions have an enormous impact on our momentary performance, health, and way of relating to others, hence on the quality of a persons’ life. In particular, the experience of unpleasant (or pleasant) emotions is directly related to an individual’s well-being. Emotions are influenced by subjective experiences and memories and the context the individual is in, and it seems almost impossible to measure this phenomenon objectively, reliably, and validly. Indeed, capturing human emotional states has been a challenging task for researchers for decades, leading to numerous theories about emotions, moods, and feelings. Specifically, psychometrics focuses on the theory and techniques of psychological measurements, including the QoL measurements. The emerging field of affective computing promises to overcome some methodological difficulties that lead to limitations in traditional methods of psychometrics. Affective computing is the study of technologies that can quantitatively measure human emotion from different clues. It is based on the hypothesis that an individual’s digital footprint is highly correlated with their perceptions, feelings, and resulting behaviors and that extracting and analyzing this data collected

S. Ghosh (✉)
BITS Pilani, Goa, India
e-mail: surjyag@goa.bits-pilani.ac.in

J. Löchner
German Youth Institute, Munich, Germany
e-mail: loechner@dji.de

B. Mitra
Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, India
e-mail: bivas@cse.iitkgp.ac.in

P. De
Microsoft Corporation, Atlanta, USA
e-mail: prade@microsoft.co

over time can prove that “your smartphone knows you better than you may think.” In addition to the fact that people use their digital devices extensively, being the first and last thing used during a normal day [1]—this statement becomes even more valid.

In this chapter, we discuss the current development within the affective computing area while focusing on assessing emotions via personal technologies. Firstly, we define emotions as a complex interplay of different components (sensory, cognitive, physiological, expressive, motivational) over time. Several emotion theories have been developed in the past years, along with emotions being classified into three dimensions: valence, arousal, and dominance [2]. Among these emotion theories, the Component Process Model [3] stands out, revealing the emotional process that leads to an individual’s perception and processing of negative and positive life experiences. As an extension of this model, the Emotional Competence Model [4] hypothesizes that mental well-being and adverse psychopathology (e.g., anxiety, depression) greatly depends on a well-functioning emotional process. This depends on the individual’s experienced emotional response, perception of the situation, adequate appraisal, and emotional regulation. Hence, emotional competence plays a key role in maintaining a person’s quality of life. Furthermore, the knowledge and perception of emotions are basic abilities that may elicit more adaptive emotion regulation strategies (like an acceptance of an uncontrollable stressor).

To capture individuals’ emotions (e.g., for clinical or research reasons), mostly self-reports for negative and positive emotions are conducted using psychometrically validated questionnaires of concepts as, e.g., stress, depression, and well-being with a recall period of days to months. These assessment instruments face the problem that (i) self-reports are rather subjective and often biased by, e.g., time and motivation to fill out often many questions and (ii) they are influenced to a great deal by the current psychological state interfering with the recall of someone’s mood days to weeks ago. To overcome these issues, new assessment methods arose in the past years using personal digital technologies. For example, the Experience Sampling Method (ESM), also known as Ecological Momentary Assessment (EMA) [5–7], is increasingly used in psychology to trigger self-reports for emotions and behaviors momentarily, i.e., as closely as possible to the subject’s daily life experiences, periodically (randomly or at fixed intervals) or in an event-driven fashion [6, 8].

Traditionally, different modalities such as facial expression [9–13], speech prosody [14–20], physiological signals like ECG, EEG, HR, GSR blood, brain, posture, [21–27] are explored for emotion assessment. Additionally, other sources can be used to extract emotions using smartphones and internet usage, which is discussed in this chapter. To determine the emotion states, often these affect-aware systems deploy a machine learning model. Therefore, conventional machine learning models like Support Vector Machine (SVM), Random Forest, Bayesian approaches were used in affective computing [28, 29]. In these approaches, first, a set of features (which can distinguish one emotion from another) were extracted manually from, e.g., the physiological signals. These features were correlated with the emotion ground truth labels (self-reports) to construct the emotion inference model. With the latest advances in the Deep Neural Network models, the conventional approaches

were replaced by state-of-the-art AI models such as Convolutional Neural Network (CNN), Long Short Term Memory (LSTM), Recurrent Neural Network (RNN) [28, 29]. The advances in this field have helped to overcome the manual feature engineering effort and helped obtain very high classification performance in the affect classification. The advances in affective computing have led to many affect-aware applications such as emotion-aware music player, affective tutor, mood monitor, which influence the quality of life [30–34]. The key working principle of such emotion-aware applications is to collect physiological and behavioral data from different modalities and to train a machine learning model for emotion inference.

Given the current accelerating scale of developments in personal technologies, new assessment techniques for emotions emerged. We present the design and development of a smartphone keyboard interaction-based emotion assessment application. Specifically, among different types of interactions performed in smartphones, keyboard interactions are highly interesting. They represent the input/output interaction between the user and the phone for, e.g., information, communication, or entertainment [35]. Additionally to the interaction itself, the interactive content may be of high interest. The research shows that not only individuals often express momentary emotions on social networks' platforms [36, 37], but also a person's language was found to reveal his/her momentarily psychological state [38].

In our research, we focus on the smartphone interaction itself. Hence, we have designed and implemented an Android application TapSense, which can unobtrusively log users' typing patterns (without actual content) and trigger self-reports for four types of emotion (happy, sad, stressed, relaxed) leveraging the ESM method. Different typing features like typing speed and error rate are extracted from the typing data and correlated with the emotion self-reports to develop a personalized emotion assessment model. However, as the conventional ESM-driven self-report collection for model construction is labor-intensive and fatigue-inducing, we also investigate how the self-report collection approach can be further optimized for suitable probing moments and reduced probing rate. So, we also have developed an adaptive 2-phase ESM schedule (integrated into TapSense), which balances the probing rate and self-report collection time and probes the individual at the opportune moments. The first phase balances between probing rate and self-report collection time and trains an 'inopportune moment assessment' model. The second phase operationalizes such a model so that no triggering is done at an inopportune moment. We investigate the implications of this ESM design on the emotion classification performance.

We evaluate the proposed approach in a 3-week 'in-the-wild study involving 22 participants. Our first important result demonstrates that using smartphone keyboard interaction; we can determine the emotion states (happy, sad, stressed, relaxed) with an average AUCROC of 78%. The next major result shows the performance of the proposed ESM approach. It demonstrates that the proposed 2-phase ESM schedule (a) assesses inopportune moments with an average accuracy (AUCROC) of 89% (b) reduces the probing frequency by 64% (c) while enabling the collection of the self-reports in a more timely manner, with a reduction of 9% on average in elapsed time between self-report sampling and event occurrence. The

proposed design also helps to improve self-report response quality by (a) improving valid response rate to 96% and (b) yielding a maximum improvement of 24% in emotion classification accuracy (AUCROC) over the traditional ESM schedules.

The chapter is organized as follows. Firstly, in Sect. 2, we discuss the definitions and the nature of emotions and their importance for the daily life experience, which influences an individual's quality of life. In Sect. 3, we present traditional psychometric assessment instruments for emotions and different assessment methods leveraging affective computing developments and beyond. The background, study design, and empirical evaluation of TapSense—a smartphone-based approach for assessing emotions are presented in Sect. 4. We discuss and conclude the chapter findings in Sect. 5.

Background and Related Work

This section presents the definition and models covering the concept of emotions (2.1), their importance for the quality of life (2.2). It attempts to capture emotions via traditional (2.3) and more novel, data-driven approaches (2.3).

Definition of Relevant Domain Concepts

Emotions, moods, and psychological states play a key role in our personal, professional, and social life. Also, interpersonal relationships, professional success, and mental well-being depend greatly on how we cope with stressful events and navigate adverse emotional experiences. Often the terms: emotions, effects, feelings, and moods are confused in language usage. *Feelings* most commonly involve invariably a direct response of the autonomic nervous system (ANS) involving organ functions (e.g., change in respiration pattern, adrenaline rush). At the same time, an umbrella term refers to all basic senses of feelings, ranging from unpleasant to pleasant (*valence*) and from excited to calm (*arousal*). *Moods* differ from feelings, emotions, and affects in that they are experienced as extended in time (c.f., mood stability by Peters, 1984) but are also subject to certain situational fluctuations [39]. Very similar is the psychological concept of *state*, referring to a person's mental state at a certain point in time, introduced by Cattell and Scheier [40] as a counterpart to the concept of timely persisting (personality, motivational, cognitive) *traits*. In contrast, *emotions* are a much more complex mental construct consisting of several components as the physiological response and lasting from minutes to hours.

Research on emotions is usually based on the central evolutionary importance of emotions for human survival. It defines emotions as “*a genetic and acquired motivational predisposition to respond experientially, physically and behaviorally to certain internal and external variables*” [41]. In the context of survival, emotions imply complex communication patterns and information [42–44], as the feedback

of an individual's inner state on different levels enables a biological adaptation to the physical and psychosocial environment. Therefore, emotions are further viewed as complex, genetically anchored behavioral chains that contribute to an individual's homeostasis through various feedback loops [45]. Since the research of Ekman [46], it has become known that elementary emotions such as fear, joy, or sadness show themselves independently of the respective culture. These basic emotions are closely coupled to simultaneously occurring neuronal processes. However, how people communicate and express visible parameters, such as facial expression, are influenced by the values, roles, and socialization practices that vary across cultures [47, 48], age, and gender [49].

Emotions can be divided categorically into primary, secondary, and combined forms: primary emotions are fundamental, while secondary emotions are emotions about emotions (such as guilt over gloating). Ekman distinguishes six basic emotions: happiness, sadness, anger, fear, surprise, and disgust [46]. In contrast, Izard [50] speaks of ten fundamental emotions (1) interest/excitement, (2) pleasure/joy, (3) surprise/fright, (4) sorrow/pain, (5) anger/rage, (6) disgust/repugnance, (7) disdain/contempt, (8) fear/contempt, (9) shame/shyness/humiliation, and (10) guilt/repentance. Another way to categorize emotions relates to their highly variable multidimensional nature: emotions can be categorized into the dimensions as positive or negative (polarity/valence), strong or weak (intensity/arousal), easy or hard to arouse (reactivity), and based on the situation they occur (idiosyncratic vs. universal situation) [51]. Following partly those dimensions, Russel's Circumplex model is the most commonly used emotion model to capture emotion on a continuous scale [52]. It represents every emotion as a tuple of valence and arousal. There exist also a valence, arousal, and dominance model (more commonly known as the VAD model), which captures every emotion as a set of triplets (valence, arousal, dominance) in a continuous scale [2] (see Fig. 10.1).

Following the earlier mentioned, there exists an empirically validated theoretical process model of emotion—the Component Process Model (CPM) [54] and the Emotional Competence Process Model (ECP), further implying the adaptive and maladaptive emotional functioning [4]. In the CPM, emotions are identified within the overall process in which low-level cognitive appraisals, particularly relevance processing, trigger physical reactions, behaviors, and subjective feelings (Fig. 10.2). As a foundation, CPM provides a differentiated theory of the various dimensions of emotions. Emotions are interpreted here as synchronizing several components that interact over time during a defined process regarding the emergence, appraisal, awareness, regulation, and knowledge of emotions. For example, an emotional situation emerges due to a specific trigger (e.g., a job interview), evoking in the appraisal (“I hope I perform well—the interviewer does not look friendly at all”) which results in a certain emotional reaction (e.g., adrenaline release, sweating, nervousness, tension). At the moment, a person is aware of this condition and the fact that they can regulate their condition (e.g., by taking a deep breath). They can categorize their emotional reaction (and those of others in return).

The emotional response components cover sensory, cognitive, physiological, motivational, and expressive components [55]. Initially, the **sensory component**

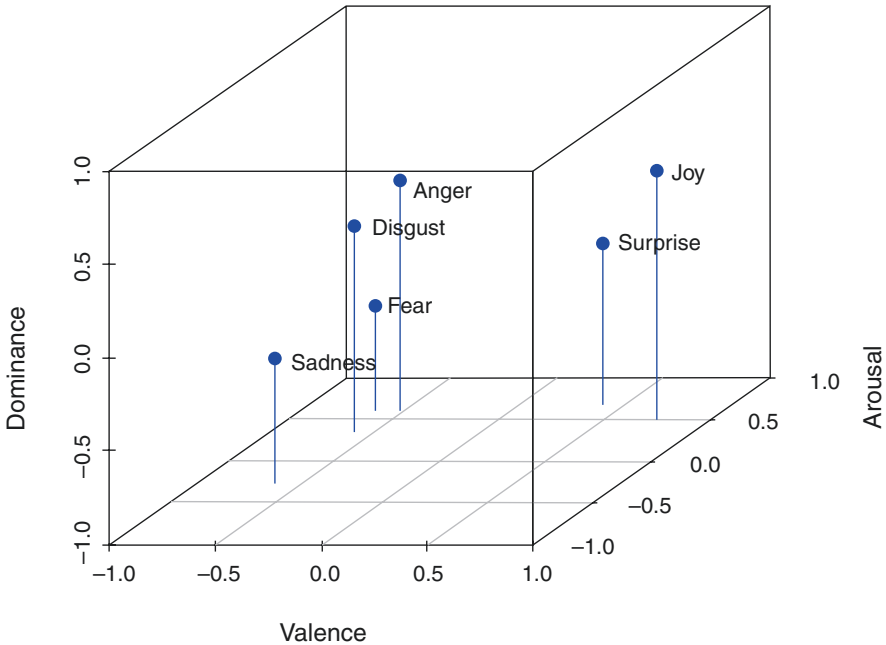


Fig. 10.1 The VAD (Valence-Arousal-Dominance) model spanned across the six basic emotions [53]

enables a subject to recognize an emotional event through the senses (e.g., see, feel). Through the **cognitive component**, the individual can identify possible relationships between itself and the event based on its subjective experiences. The individual then makes a subjective evaluation of the perception of the event (appraisal). This goes along with the two-factor theory of emotions. As early as the 1960s, Schachter and Singer pointed to the cognitive evaluation of a physical response as key to the subsequent emotional sensation [56]. A subject can react to the same event with a different evaluation—depending on his personal world view, value system, and current physiological state. It is resulting in different physical responses to feelings. Depending on the subjective evaluation outcome, the individual reacts by releasing certain neurotransmitters and hormones, thus changing its physiological state (**physical component**). This altered state corresponds to the experience of an emotion. According to Lazarus' appraisal theory [57], the emotional experience first arises through cognitive evaluation and interpreting an emotional stimulus as manageable or not. The **motivational component** follows the event's evaluation and is modulated by the current physiological (or emotional) state. The motivation to a certain action of a person is oriented to an actual-target comparison and the prediction of the effect of conceivable actions. For example, the emotion of anger can result in both the motivation for an attack action (e.g., in the case of a supposedly inferior opponent) and the motivation for a flight action (e.g., in the case of a

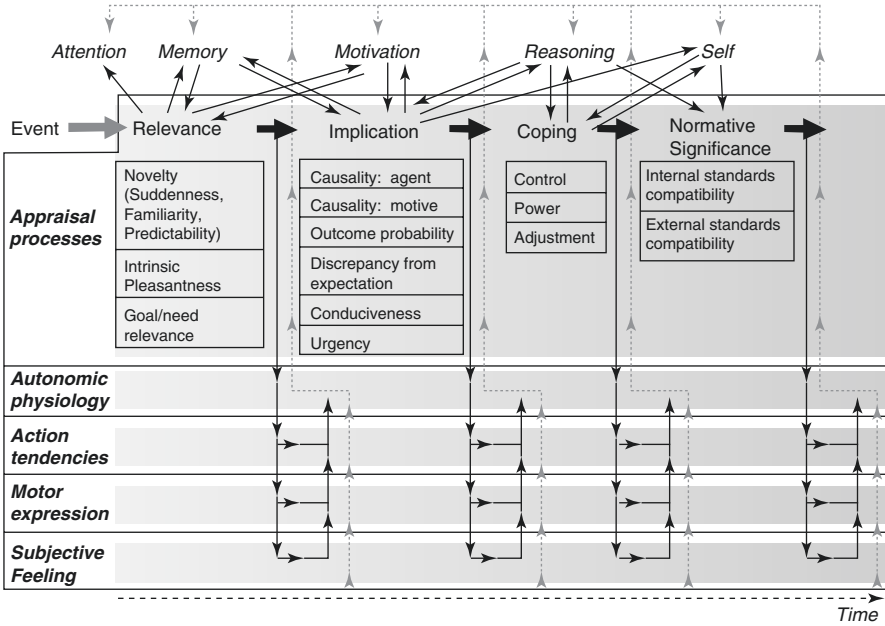


Fig. 10.2 Comprehensive illustration of the Component Process Model of Emotion (adapted from Scherer 2001) [58]

supposedly superior opponent). The **expressive component** refers to the way emotion is expressed. This primarily concerns nonverbal behavior, such as facial expressions, speech, and gestures.

Importance of Feelings Moods, Psychological States

From an evolutionary perspective, emotions play a very important role in motivation, behavior, and attention: they make us act, direct our attention to certain stimuli that might have pleasant or unpleasant consequences for us, and give us signals so that we adjust our behavior to obtain or avoid those consequences. In a modern context, it is not only about our survival and related incentives but also about secondary incentives, such as money, status, entertainment, or others. Moreover, emotions regulate the intensity and duration of different behaviors and cause the learning of those behaviors that were successful under certain conditions (e.g., joy has a pleasant effect on us and motivates us to repeat the behavior) and mark in memory (e.g., via disgust, anger) those that led to failure. In the same way, emotions function in regulating our social interactions, being reinforced by a pleasant (or non-pleasant) effect. This leads to the formation of bonds or rivalries, which provide us with orientation in the social structure. Some distinct emotions may have even a more

Table 10.1 Example Functions of Emotions

Emotion	Function
Fear	Protective function: Avoidance of danger, injury, creating distance to source of danger
Anger	Destruction function: Overcoming something that stands in the way of need satisfaction (e.g. nutrition)
Pleasure	Intake function: Receiving positive stimulus from the external world that supports or nourishes the individual
Disgust	Rejection function: Excrete harmful substance
Surprise	Orientation function: Increased attention to take in information about a new stimulus

specific function, as shown in Table 10.1 [59, 60]. It is part of our daily lives to be confronted with the experience of such emotions and to respond to them. How people deal with different emotional events varies widely. The distinction between “positive” and “negative” emotions is questionable due to emotions’ functionality.

As depicted in the ECP, there is a great body of literature showing that both positive and negative emotions greatly impact our well-being and mental health (see meta-analysis [61]). It has been demonstrated, for example, that people suffering from depression have difficulties in identifying (Rude & McCarthy, 2003), bearing and accepting their emotions [62, 63]. In numerous disorders, the presence of undesirable affective states (such as anxiety or depressed mood) in an inappropriate intensity or duration is among a diagnostic criterion of the disorder (e.g., in anxiety disorders or depression). Also, a whole range of cognitive and behavioral symptoms of mental disorders can be understood as dysfunctional attempts to avoid or terminate such undesirable states. Examples include alcohol or substance abuse, self-injurious behavior, or eating attacks.

Given the nature of variety in the emotional response and experience regarding intensity, duration, personality, and situational aspects, it is poorly defined when a certain emotional phenomenon can be considered inappropriate, abnormal, or even psychopathological [64]. This poses a challenge to psychometricians, researchers, and clinical diagnosticians to make the most valid (clinical) judgment or classification and eventually initiate appropriate and effective treatment.

Assessment of Emotions

Traditional assessment of emotions is based on self-reports via questionnaire and interview (3.1).¹ For many years now, other approaches aiming to capture a person’s emotions more objectively and independently than subjective self-reports were developed. In the following, assessment methods deriving from physiology-based

¹Of note, not all psychometric approaches for assessing the whole emotional process (or parts of it) can be displayed here. Further research attempts to improve questionnaires’ reliability, like eye-tracking experiments (while filling in a questionnaire), appraisal biases, and knowledge of emotions.

emotion assessment such as blood and brain (3.2), expression-based emotion assessment such as facial expression, speech, and posture (3.3) are discussed. Moreover, digital footprints of emotions can be approached via social network platforms and smartphone data collection (3.4), including momentary ecological assessment. In the final part of this section, we compare the advantages and disadvantages of the presented assessment methods (3.5).

Self-Report-Based Methods for Emotion Assessment

In the traditional psychometric approaches, emotions are usually measured with self-report questionnaires, leveraging time- and cost-efficient instruments and enabling access to the cognitive component of someone's emotions. It is noteworthy that those instruments aim to assess single emotions and approach "concepts" as stress, depression, or well-being, indicating well-functioning (or dysfunctional) emotional processes over time. Questionnaires differ on how they are evaluated regarding their standardized psychometric properties as test objectivity, reliability, and validity [65]. Usually, factor analyses are conducted to evaluate the underlying constructs' factor structure in the designed questionnaire. Standardization is usually carried out using large and representative samples, which allow the classification of individual test results compared to the norm sample. Ideally, t-values or percentiles are defined for this purpose based on extensive psychometric studies. For the distinction between pathological and healthy reactions and clinical diagnosis, clinical samples are also collected.

To capture the emotion self-reports in the general population, often different scales are used, guided by the above-mentioned emotion models. In Table 10.2, an overview of some well-established questionnaires, including the item number and recall period, is given. For example, the Self-Assessment Manikin (SAM) is a non-verbal pictorial assessment technique that directly measures the pleasure, arousal, and dominance associated with a person's affective reaction [77]. Similarly, there is an Affect Balance Scale (ABS), based on a model that posits the existence of two independent conceptual dimensions—positive affect (PAS) and negative affect (NAS)—each related to overall psychological well-being by an independent set of variables [68]. The more widely accepted PANAS scale is a 10-item mood scale that comprises the Positive and Negative Affect Schedule (PANAS) [67]. Additionally, a flourishing scale determines psychological flourishing and feelings [92] in relevant areas such as purpose in life, relationships, self-esteem, feelings of competence, and optimism.

To make a clinical diagnosis of an affective disorder, clinical questionnaires are based on the clinical diagnostic criteria: symptom description, duration, intensity, distress, and psychosocial consequences based on the classification criteria of DSM-V [93] and ICD-10 [94]. However, for a valid diagnosis, *a structured clinical interview* is the gold standard (e.g., SKID [95]). Nevertheless, clinical questionnaires are frequently used to validate the diagnosis and evaluation of the treatment process

Table 10.2 Overview of example, well-established questionnaires assessing emotions

Construct	Scale Name	Number of Items	Recall Period	Reference
Affect	The affect balance scale (ABS)	10	Past few weeks	Bradburn (1969) [66]
Affect	The positive and negative affect scale (PANAS)	20	n/a	Watson et al. (1988) [67]
Affect	Affect balance scale	10	Past few weeks	Moriwaki (1974) [68]
Anxiety	Manifest anxiety scale	50/28	n/a	Taylor (1953) [69]
Anxiety	Self-rating anxiety scale	20	Overall	Zung (1965) [70]
Anxiety	Beck anxiety inventory	21	n/a	Beck et al., (1988) [71]
Anxiety, depression	The depression anxiety stress scales	21	Past week	Antony, Bieling, Cox, Enns, & Swinson (1998) [72]
Anxiety	State-trait anxiety inventory	40	Current	Spielberger (1989) [73]
Anxiety	Generalized anxiety Disorder-7 (GAD-7)	7	Past 2 weeks	Spitzer et al., (2006) [74]
Anxiety	The Hamilton anxiety rating scale	14	n/a	Hamilton, (1959) [75]
Coherence	Sense of coherence scale	29	Overall	Antonosky, (1993) [76]
Emotions	Self-assessment manikin	3	Current	Bradley & Lang (1994) [77]
Depression	Patient health questionnaire (PHQ-9)	9	Past 2 weeks	Kroenke, Spitzer, & Williams, 2001 [78]
Depression	The hospital anxiety and depression scale	14	Past week	Zigmond & Snaith, (1983) [79]
Depression	Beck depression inventory	21	Past 2 weeks	Beck, Steer, & Brown (1996) [80]
Depression	Zung self-rating depression scale	20	Current	Zung, (1965) [70]
Depression	Center for Epidemiologic Studies Depression Scale (CES-D)	20	Past week	Lewinsohn et al., (1997) [81]
Depression	Carroll rating scale (CRS) for depression	52	n/a	Carroll, Feinberg, Smouse, Rawson, & Greden, (1981) [82]
Emotional competence	Geneva emotional competence test (GECe)	110	Current	Schlegel & Mortillaro (2019) [83]
Emotion recognition	Geneva emotion recognition test (GERT/-S)	83/42	Current	Schlegel, Grandjean, & Scherer (2014) [84], [85]
Psychopathology	Symptom-checklist SCL-20-R	90	Past week	Hamilton, (1959) [75]
Rumination	Ruminative response scale	22	n/a	Treynor, Gonzalez, & Nolen-Hoeksema, (2003) [86]

Table 10.2 (continued)

Construct	Scale Name	Number of Items	Recall Period	Reference
Stress	Perceived stress scale perceived stress test (PSS)	14	Past month	Cohen, Kamarck, Mermelstein, (1983) [87]
Well-being	General health questionnaire	60	Past few weeks	Goldberg & Hillier (1979) [88]
Well-being	Warwick-Edinburgh mental Well-being scale—WEMWBS	14	Past 2 weeks	Stewart-Brown, et al. (2011) [89]
Well-being	EuroQol instrument EQ-5D-3L	15	Current	Herdman et al.,(2011) [90]
Worrying	Penn State worry questionnaire—(PSWQ)	16	n/a	Hopko et al., (2003) [91]

and follow-up. For example, the Beck’s Depression Inventory (BDI-II, [80]) is commonly used in research to diagnose depression. The BDI-II contains 30 items referring to depressive symptoms someone may have experienced the past two weeks with different intensity. Assessing depressive symptoms in nine items, the PHQ-9 corresponds to the depression module of the Patient Health Questionnaire (PHQ, [78]). Unlike many other depression questionnaires, the PHQ-9 captures one of the nine DSM-IV criteria for diagnosing “major depression” with each question. Again, due to depression classification criteria, the recall period is two weeks. The Depression Anxiety Stress Scale (DASS, [72]) covers 42 Items on those three related emotional states in the last week by not focusing on one specific affective disorder. In contrast to assessing mental illness, there are fewer instruments for assessing mental well-being. For example, the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) aims to assess positive feelings an individual may have experienced to a certain extent in the past two weeks.

Overall, questionnaires to assess emotions (for clinical use) have the advantage of not requiring experienced experts, lead to scalable, comparable results, and are time and cost-efficient than clinical interviews. Besides, the object is an individual’s emotional experience, and therefore the subjectivity of self-reports makes sense to capture someone’s psychological strain. However, self-report questionnaires face big shortcomings in the assessment of emotions over time. Most instruments (covering the classification criteria of DSM-V and ICD-10) refer to a recall period of up to two weeks (see Table 10.2 for examples). Therefore, they are greatly biased by memory effects [96] and the motivation of an individual [97]. Studies are showing the questionnaire results are more negative when filled on Mondays while Saturday mornings are “the happiest moment during the week” and consequently provoke more positive test results [98, 99]. Other bias factors are fatigue and other non-assessed personal conditions (e.g., bad news occurring just that specific assessment day), as well as misunderstanding of the items and social desirability (aiming “to look good,” even in anonymous surveys) [100, 101]. This specific subjectivity is

wanted to a somewhat extent since it also informs the examiner about a person's perception and interpretation. Nevertheless, the objectivity of assessment is always limited greatly by these interpretation biases.

Physiology-Based Emotion Assessment

As mentioned above, emotions are represented not only by the cognitive but also physical and expressive components. While the cognitive aspect of emotion cannot be observed directly, the physical and expressive aspects are often manifested in different bodily signals. An emotional state influences underlying human biology and psychophysiology, as well as the resulting behaviors. For example, stress can be manifested in hormonal changes; anxiety can be manifested in terms of a high pulse rate, while happiness can be expressed via laughter. As we point out in the following subsections, some of these manifestations are captured and modeled to determine the emotion in affective computing.

Emotional Assessment from Blood

The emotional state of the human may be assessed via blood-based analytics, as the emotional state influences the individual's hormonal status. Especially in the field of biological psychiatry research, plasmatic biomarkers have been leading endeavors. The five most named plasmatic biomarkers (BDNF, TNF- alpha, IL-6, C-reactive protein, and cortisol) are classically used to predict psychiatric disorders like schizophrenia, major depressive disorder, or bipolar disorder [21]. In a meta-analysis, patterns of variation of those features were identified between those most important psychiatric disorders. The results indicated robust variations across studies but also showed similarities among disorders. The authors conclude that the implemented biomarkers may be interpreted as transdiagnostic systemic consequences of psychiatric illness rather than diagnostic markers. This is in line with another review of Funalla et al. showing evidence for diagnostic biomarkers associated with obsessive-compulsive disorder (OCD) but did not show diagnostic specificity [102]. A commonly used indicator for stress is the cortisol concentration found in human blood or saliva [103]. Overall, in this chapter, we do not focus on plasmatic biomarkers; we mention these for completeness.

Emotional Assessment from Brain

Other biological features to determine emotions with a rather high accuracy can be extracted from brain electroencephalography (EEG) [22]. With the help of EEG, the assessment of the brain's summed electrical activity is enabled by recording the voltage fluctuations on the surface of the head. Emotion extraction is

consequently based on arousal and categorized into valence and excitation. EEG evaluation is traditionally performed by pattern recognition by the trained evaluator or by an automatic evaluation. For emotions, the patterns of alpha and beta waves are key classifiers. The alpha wave is associated with mild relaxation or relaxed alertness, with eyes closed (frequency range between 8 and 13 Hz). A beta wave has different causes and meanings and may occur during constant tensing of a muscle or during active concentration (frequency range between 13 and 30 Hz). According to Choppin (2000), high valence is associated with high beta power in the right parietal lobe and high alpha power in the brain's frontal lobe [104]. High beta power in the parietal lobes is associated with higher arousal in emotions, while the alpha activity is lower but also located in the parietal lobes. More specifically, negative emotions are represented by activity in the right frontal lobe, whereas positive emotions result in high power in the brain's left frontal part. EEG was found to achieve 88.86% accuracy for four emotions: sad, scared, happy, and calm [105]. After assessing the EEG waves and extracting the particular emotional features, classifiers are trained for emotion identification. Popular is the Canonical Correlation Analysis (CCA) [106], Artificial Neural Network (ANN) [107], Fisher linear discriminant projection [24], and Adaptive Neuro-Fuzzy Interference System (ANFIS) [108]. Using K-Nearest Neighbor (KNN) [109], and Support Vector Machine (SVM) [105, 110] Mehmood and Lee (2016) used five frequency bands (beside alpha and beta, delta, theta, gamma waves) and identified the four emotions *sad*, *scared*, *happy* and *calm* with an accuracy rate of 55% (KNN) and 58% (SVM).

In a more clinical setting, Khodayari-Rostamabd [111] used EEG data to predict the pharmaceutical treatment response of schizophrenic patients. A set of features was classified using the kernel partial least squares regression method to perform response prediction on the positive and negative syndrome scale (PANSS) with 85% accuracy. In another sample aiming to predict psychopharmacological treatment response (SSRI), the same research group extracted candidate features from the subject's pre-treatment EEG using a mixture of factor analysis (MFA) model in a sample with patients suffering from depression [112]. The proposed method's specificity is 80.9%, while sensitivity is 94.9%, for an overall prediction accuracy of 87.9%. Besides EEG, also functional Magnetic Resonance Imaging (fMRI) is used to assess emotions, especially by exploring the amygdala activity [113, 114]. Zhang et al. [114] analyzed connectivity change patterns in an fMRI data-driven approach in 334 healthy participants before and after inducing stress. Besides, the participants' cortisol level was taken to classify pre- and post-stress states. The machine learning model revealed that the discrimination relied on activation in the dorsal anterior cingulate cortex, amygdala, posterior cingulate cortex, and precuneus and with a 75% accuracy rate. The advantage of using EEG for assessing emotions is that the data extraction is independent of facial or verbal expression that could be impaired due to, e.g., paraplegia, facial paralysis, but the necessity of a lab and the complex, costly installation and maintenance of equipment is a big disadvantage not only for the practical field but also research projects [22].

Emotional Assessment from Physiological Signal Collection

A lot has been written about assessment of the emotional state of the individual from the physiological state—via, e.g., EEG (mentioned above), Electrocardiogram (ECG), Electromyography (EMG), Electrooculography (EOG), Galvanic Skin Response (GSR), Heart Rate (HR), Body Temperature (T), Blood Oxygen Saturation (OXY), Respiration Rate (RR), or Blood Pressure analytics (BP) [115–118]. From the technical perspective, the existing physiological signal-driven emotion assessment methods can also be divided into three categories (a) traditional machine learning methods, (b) deep neural network-based method, and (c) sequence-based models.

Conventional Machine Learning Approach In the case of traditional machine learning-based approaches, first, a set of features are extracted from the captured data, and then different algorithms are used for model construction. Apart from the time domain characteristics, to leverage the spectral-domain characteristics such as power spectral density (PSD), spectral entropy (SE) is computed using Fast Fourier Transform (FFT) or Short-term Fourier Transform (STFT).

SVM is probably the most widely used in physiological signal base emotion recognition among various machine learning algorithms. Das et al. extracted Welch's PSD of ECG and Galvanic Skin Response (GSR) signals for emotion recognition [119]. Liu et al. extracted a set of features from EEG and eye signals and used a linear SVM to determine three emotion states [120]. However, as regular SVM does not work in the imbalanced dataset, Liu et al. constructed an imbalanced support vector machine to solve the imbalanced dataset problem, which increased the punishment weight to the minority class and decreased the punishment weight majority class [121]. A few authors also used KNN ($K = 4$) to classify four emotions with the four features extracted from ECG, EMG, GSR, and RR [121]. In [115], the authors collected 14 features of 34 participants as they watch three sets of 10-min film clips eliciting fear, sadness, and neutrality, respectively. Analyses used sequential backward selection and sequential forward selection to choose different feature sets for 5 classifiers (QDA, MLP, RBNF, KNN, and LDA). Wen et al. used RF to classify five emotional states with features extracted from OXY, GSR, and HR [122].

Deep Learning-Based Approach Among different deep learning-based approaches, CNN is one of the most widely used. Martinez et al. trained an efficient deep convolution neural network (CNN) to classify four cognitive states (relaxation, anxiety, excitement, and fun) using skin conductance and blood volume pulse signals [123]. Giao et al. used the Convolutional Neural Network (CNN) for feature abstraction from EEG signal [124]. In [125], several statistical features were extracted and sent to the CNN and DNN. Song et al. used dynamical graph convolutional neural networks (DGCNN), which could dynamically learn the intrinsic relationship between different EEG channels represented by an adjacency matrix to facilitate feature extraction [126]. DBN is also widely used for emotion recognition. It learns a deep input feature through pre-training. Zheng et al. introduced a recent

advanced deep belief network (DBN) with differential entropy features to classify two emotional categories (positive and negative) from EEG data, where a Hidden Markov Model (HMM) was integrated to accurately capture a more reliable emotional stage switching [127]. Huang et al. extracted a set of features and applied DBN in mapping the extracted feature to the higher-level characteristics space [128]. In the work of [129], instead of the manual feature extraction, the raw EEG, EMG, EOG, and GSR signals directly inputted to the DBN, where the high-level features according to the data distribution could be extracted.

Sequence-Based Models To capture the temporal aspects of the physiological signals, often sequence-based models are used. For example, Li et al. applied CNN first to extract features from EEG and then applied LSTM to train the classifier, where the classifier performance was relevant to the output of LSTM in each time step [130]. In the work of [131], an end-to-end structure was proposed, in which the raw EEG signals in 5 s-long segments were sent to the LSTM networks, in which autonomously learned features. Liu et al. proposed a model with two attention mechanisms based on multi-layer LSTM for the video and EEG signals, which combined temporal and band attentions [132].

Overall, as the captured signals are noisy, several pre-processing techniques are often used to eliminate the noise introduced from different sources such as crosstalk, measurement error, and instrument interference. The commonly used pre-processing techniques include filtering [133], Discrete Wavelet Transform (DWT) [134], Independent Component Analysis (ICA) [135], Empirical mode decomposition (EMD) [136].

The overall psychophysiological approach for emotional assessment is cumbersome and not straightforward to be deployed in real-time for real-time accurate emotion recognition. Besides, the required complex laboratory setup (e.g., EEG) is time and cost-intensive.

Expression-Based Emotion Assessment

Following the cognitive and physical components of emotions, we focus on expressing emotions in this section. We describe facial and verbal emotion recognition, as well as how posture may reflect human emotional states.

Facial Emotion Recognition (FER)

We broadly divide facial emotion recognition-related works into the following two groups (a) conventional FER approach and (b) deep learning-based FER approach.

Conventional FER Approach For automatic FER systems, various types of conventional approaches have been studied. First, all these approaches assess the face

region and then extract a set of geometric features, appearance features, or a hybrid of geometric and appearance features on the target face. For the geometric features, the relationship between facial components is used to construct a feature vector for training [137, 138]. For example, Ghimire and Lee [138] used two types of geometric features based on 52 facial landmark points' position and angle. First, the angle and Euclidean distance between each pair of landmarks within a frame are calculated. Second, the distance and angles are subtracted from the corresponding distance and angles in the video sequence's first frame. For the classifier, two approaches are used, using multi-class AdaBoost with dynamic time warping, and SVM on the boosted feature vectors.

The appearance features are usually extracted from the global face region [139] or different face regions containing different types of information [140–143]. An example of using global features includes the exploration by Happy et al. [139]. The authors utilized a local binary pattern (LBP) histogram of different block sizes from a global face region as the feature vectors. They classified various facial expressions using a principal component analysis (PCA). This method's classification performance is poor as it cannot reflect local variations of the facial components to the feature vector. A few explorations also used features from different face regions as they may have different levels of importance, unlike a global-feature-based approach. For example, the eyes and mouth contain more information than the forehead and cheek. Ghimire et al. [144] extracted region-specific appearance features by dividing the entire face region into domain-specific local regions. An incremental search approach is used to identify important local regions, reducing feature vector size, and improving classification performance.

For hybrid features, some approaches [144, 145] have combined geometric and appearance features to complement the two approaches' weaknesses and provide even better results in certain cases.

Deep Learning-Based FER Approach The most adopted one in deep neural network-based FER is CNN. The main advantage is to completely remove or highly reduce the dependence on physics-based models and/or other pre-processing techniques by enabling “end-to-end” learning directly from input images [146]. Breuer and Kimmel [147] investigated the suitability of CNNs on different FER datasets and showed the capability of networks trained on emotion assessment and FER-related tasks. Jung et al. [148] used two different types of CNN: the first extracts temporal appearance features from the image sequences. The second CNN extracts the temporal geometry features from temporal facial landmark points. These two models are combined using a new integration method to boost the performance of facial expression recognition.

However, as CNN-based methods are not suitable for capturing the temporal sequence, a hybrid approach combining both CNN (for spatial features) and LSTM (for temporal sequence) was developed. LSTM is a special type of RNN capable of learning long-term dependencies. Kahou et al. [149] proposed a hybrid RNN-CNN framework for propagating information over a sequence using a

continuously valued hidden-layer representation. In this work, the authors presented a complete system for the 2015 Emotion Recognition in the Wild (EmotiW) Challenge [150]. They proved that a hybrid CNN-RNN architecture for a facial expression analysis could outperform a previously applied CNN approach using temporal averaging for aggregation. Kim et al. [151] utilized representative expression-states (e.g., the onset, apex, and offset of expressions), specified in facial sequences regardless of the expression intensity. Hasani and Mahoor [152] proposed the 3D Inception-ResNet architecture followed by an LSTM unit that together extracts the spatial relations and temporal relations within the facial images between different frames in a video sequence. Graves et al. [153] used a recurrent network to consider the temporal dependencies present in the image sequences during classification. This study compared the performance of two types of LSTM (bidirectional LSTM and unidirectional LSTM) and proved that a bidirectional network provides significantly better performance than a unidirectional LSTM.

In summary, hybrid CNN-LSTM (RNN) based FER approaches combine an LSTM with a deep hierarchical visual feature extractor such as a CNN model. Therefore, such a hybrid model can learn to recognize and synthesize temporal dynamics for tasks involving sequential images. Each visual feature determined through a CNN is passed to the corresponding LSTM, and it produces a fixed or variable-length vector representation. The outputs are then passed into a recurrent sequence-learning module. Finally, the predicted distribution is computed by applying softmax [154, 155]. A limitation of this approach is the challenge of capturing the facial expression in real-time in daily life as a natural reaction to an emotional experience. Additionally, privacy issues can be a problem if a person does not want to be visually recorded during such intimate moments.

Speech Based Emotion Recognition (SER)

The existing literature on speech emotion recognition (SER) is also broadly divided into the following two categories (a) Conventional SER approach and (b) deep learning-based SER approach.

Conventional SER Approach In traditional SER systems, there are mainly three steps—(a) signal pre-processing, (b) feature extraction, and (c) classification. At first, acoustic pre-processing such as denoising, segmentation is carried out to determine relevant units of the speech signal [156–158]. Once the pre-processing is done, several short-term characteristics of the signal such as energy, formants, and pitch are extracted, and short-term classification of the speech segment is done [159]. On the contrary, for long-term classification, mean, standard deviation is used [160]. Among prosodic features, the intensity, pitch, rate of spoken words, and variance play an important role in identifying various types of emotions from the input speech signal [161]. The relationship between different vocal parameters and their relation to emotion is often explored in SER. Parameters such as intensity,

pitch, and rate of spoken words, and quality of voice are frequently considered [162]. The intensity and pitch are often correlated to activation so that the value of intensity increases along with high pitch and vice versa [163, 164]. Factors that affect the mapping from acoustic variables to emotion include whether the speaker is acting, there are high speaker variations, and the individual's mood or personality [165, 166].

In the existing SER literature, there are two types of classifiers—linear and non-linear. Linear classifiers usually perform classification based on object features with a linear arrangement of various objects [166]. In contrast, non-linear classifiers are utilized for object characterization in developing the non-linear weighted combination of such objects [167–170]. The GMMs are utilized for the representation of the acoustic features of sound units. The HMMs, on the other hand, are utilized for dealing with temporal variations in speech signals [171].

Deep Learning-Based SER Approach In SER approaches, different types of deep neural networks are used. Senigallia et al. used a 2D CNN with Phoneme data as input data to determine 7 emotion states [172]. Zhao et al. combined Deep 1D and 2D CNN for high-level learning features from input audio and log-mel spectrograms for emotion classification [173]. Convolutional Neural Network (CNN) also uses the layer-wise structure and can categorize the seven universal emotions from the defined speech spectrograms [174]. In [175], an SER technique based on spectrograms and deep CNN is presented. The model consists of three fully connected convolutional layers for extracting emotional features from the speech signal's spectrogram images. Pablo et al. obtained emotional expressions that are spontaneous and can easily be classified into positive or negative [176]. Mao et al. trained the CNN to learn affect salient features and achieved robust emotion recognition with the variational speaker, language, and environment [177].

The hybrid networks consisting of CNN and RNN are also used in SER [178–180]. This enables the model to obtain both frequency and temporal dependency in a given speech signal. Sometimes, a reconstruction-error-based RNN for continuous speech emotion recognition is also used [181]. SER algorithms based on CNNs and RNNs have been investigated in [180]. The deep hierarchical CNNs architecture for feature extraction has also been combined with LSTM network layers. It was found that CNN's have a time-based distributed network that provides results with greater accuracy. Zhao et al. used a hybrid RCNN model to determine basic emotion [182]. Wootae et al. used a deep hierarchical feature extraction architecture of CNNs combined with LSTM network layers for better emotion recognition [180]. Like FER, capturing the speech in real-time as a reaction to the emotional experience is a challenging issue. In addition to interfering with privacy, someone may express emotions without using expressed language because they are alone or do not want to speak, especially in emotional moments. Although individuals tend to adapt quickly to being observed, the awareness of being recorded might interfere with someone's speech and expression of emotions.

Emotional Assessment from Posture

In contrast to research on automatic emotion recognition focusing on facial expressions or physiological signals, little research has been done on exploiting body postures. However, they can be useful for emotion recognition and even more accurate than facial features [23]. Bodily postures refer to the physical expression component of emotions and an important channel of communication. Since it is challenging to categorize the expressed posture to discrete emotions due to the variety of validated emotion poses, researchers in this field focus first on defining features, aiming for understanding the cohesion and dimensional ratings with high validity [183]. Therefore, studies recorded actors displaying concrete, discrete emotions, and the recordings were then rated by the study participants categorizing these emotions. For example, Lopez et al. asked study participants to categorize emotion postures depicting five emotions (joy, sadness, fear, anger, and disgust) and to rate valence and arousal for each emotion pose. Besides a successful categorization of all emotion categories, participants accurately identified multiple distinct poses within each emotion category. The dimensional rating of arousal and valence showed interesting overlaps and distinctions, increasing further granularity of distinct emotions. Similarly, the Emotion Recognition Test (GERT) [85] was developed to test the emotional recognition ability using video clips with sound simultaneously presenting facial, vocal, and emotional expression based on the posture by using various data and video material and relatively large samples to validate. Individuals using the GERT test material are asked to watch video clips and rate the displayed emotions to assess their emotion recognition ability.

Postures are also captured using movement sensor data from smartwatches or mobile phones [184]. Quiroz et al. observed the movement sensor logged by the smartwatches of 50 participants differentiated between the emotions *happy*, *sad*, and *neutral* as a response to an emotional stimulus in an experimental setting. Furthermore, the response was validated by additional data from the Positive Affect and Negative Affect Schedule Questionnaire (PANAS). Emotional states could be assessed well by self-report and data obtained from the smartwatch with high accuracy across all users for classification of happy versus sad emotional states. Although the categories here depict only two emotions and other categorization difficulties need to be evaluated, movement sensors' usage still appears promising for emotion recognition purposes. Another smartphone-based approach also uses information about postures for recognizing emotions but uses self-reported body postures [185]. A mobile application was developed that classifies (based on the nearest neighbor algorithm) inserted poses into Ekman's six basic emotion categories and a neutral state. Emotion recognition accuracy was evaluated using poses reported by a sample of users.

Although digital devices may be included to capture postures in real-time, this data collection is challenging to conduct in a person's daily life and close to their natural expression of emotions. Due to the necessity of leveraging, e.g., the use of cameras, this method is obtrusive and implies the same privacy issues we discussed in the context of FER and SER.

Internet-Use Based Emotion Assessment

This subsection presents the emotion assessments based on the internet usage of an individual. Firstly, analysis on social networks (3.4.1) is discussed, followed by smartphone-based emotion assessment (3.4.2) and smartphone-based experience sample methods (3.4.3).

Social Network Analysis

Digital records of an individual's behavior are extracted, including linguistic style, sentiment, online social networks, and other activity traces, which can be used to infer the individual's psychological state. In particular, social networking platforms are becoming increasingly popular. They have recently been used more extensively to study emotions, as they are easily accessible to users, and researchers can collect the necessary information with the users' consent. Based on this approach, Chen et al. aimed to identify users with depression or at risk of depression by assessing the individual's expressed emotions from Twitter posts over time [36]. In another study, voluntarily shared Facebook Likes for $N = 58,000$ users were used to predict several highly sensitive personality attributes [37]: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use addictive substances, parental separation, age, and gender. All attributes were predicted with high accuracy, especially the ethnic origin and gender. Other emerging approaches focus on Spotify music or Instagram picture extraction as a feature to predict personality or mood outcomes [186].

Furthermore, studies focus on language content in social media networks or messaging systems to discover depressive symptoms, the so-called Natural Language Processing (NLP) [187]. The depressive language was characterized by more negative and extreme words such as "always, everybody, never" [188]. The challenge with this data is that communication on Twitter, Facebook may be heavily distorted by aspects of social desirability and specific motivations that drive someone to express themselves on the Internet. Furthermore, data collection from someone's account may face privacy issues.

Smartphone-Based Emotion Assessment

Smartphones are personal devices that individuals carry around with them almost all the time [189]. They include a plethora of onboard sensors (e.g., accelerometer, gyroscope, GPS) and can sense different user activities passively (e.g., mobility, app usage history) [190]. In this subsection, we review the smartphone-based methods for emotion assessment in its user's natural daily environments. We consider the usage-based assessment methods, as well as touch-based ones.

Usage-Based Emotion Recognition Methods The smartphone provides numerous data sources for collecting real-world data about emotions. For example, defined FER and SER assessments can also be extracted from the smartphone's camera and microphone. Other smartphone-based sensing sources are connectivity (WIFI on/off), smartphone status (screen, battery, power-saving mode), calls (type, duration), text messages (type, length), notifications (apps, category), calendar (initial query, logging of new entries), technical data (anonymized user ID, IP address, mobile phone type) [191]. Harari et al. [192] sensed conversations, phone calls, text messages, and messaging and social media applications for individual trait assessment. Namely, they collected sensing data in five semantic categories (communication & social behavior, music listening behavior, app usage behavior, mobility, and general day- & night-time activity) and used a machine learning approach (random forest, elastic net) to predict personality traits.² MoodScope proposed to infer mood exploiting multiple information channels, such as SMS, email, phone call patterns, application usage, web browsing, and location [143]. In EmotionSense, Rachuri et al. used multiple Emotional Prosody Speech and Transcripts library features to train the emotion classifier [195]. In the same vein, researchers also demonstrated that aggregated features obtained from smartphone usage data could indicate the Big-Five personality traits [196]. We also find that there are multiple works, which use different information sources to infer the presence of a particular emotional state. For example, Pielot et al. tried to infer boredom from smartphone usage patterns like call details, sensor details, and others [197]. In their work on assessing stress, Lu et al. built a stress classification model using several acoustic features [198]. Similarly, Bogomolov et al. showed that daily happiness [199] and daily stress [200] could be inferred from mobile phone usage, personality traits, and weather data.

Touch-Based Emotion Recognition Methods Widespread availability of touch-based devices and a steady increase [35] in the usage of instant messaging apps open a new possibility of inferring emotion from touch interactions. Therefore, research groups started to focus on typing patterns (Shapsough et al., 2016) using a built-in sensor (a smart keyboard) and using machine learning techniques to assess emotions based on different aspects of typing. For example, Lee et al. designed a Twitter client app and collected data from various onboard sensors, including typing (e.g., speed), to predict one user's emotion in the pilot study [201]. Similarly, Gao et al. used multiple finger-stroke-related features to identify different emotional states during touch-based gameplay [202]. Ciman et al. assessed stress conditions by analyzing multiple features of smartphone interaction, including swipe, scroll, and text input interactions [203]. Kim et al. [204] proposed an emotion recognition framework analyzing touch behavior during app usage, using 12 attributes from 3

²Personality traits are generally described as relatively stable patterns of thought, feelings, and behaviors and are therefore well related to emotional theories [193, 194].

onboard smartphone sensors. Although focused on narrow application scenarios, all of these works point to the value of touch patterns in emotion assessment.

Smartphone-Based Experience Sampling Method Design

One of the key requirements to develop a smartphone-based emotion assessment system is to collect emotion ground truth labels, which are typically collected as emotion self-reports by deploying an Experience Sampling Method (ESM), also known as an Ecological Momentary Assessment (EMA). The Experience Sampling Method (ESM) is a widely used tool in psychology and behavioral research for in-situ sampling of human behavior, thoughts, and feelings [5, 205]. The ubiquitous use of smartphones and wearable devices helps in the more flexible design of ESM, aptly termed as mobile ESM (mESM) [206–208]. It allows the collection of rich contextual information (e.g., sensor information, application usage data) along with behavioral data at an unprecedented scale and granularity. Frameworks like Device Analyzer, UbiqLog, AWARE, ACE, MobileMiner, or MQoL-Lab [190] have been designed to infer user's context based on sensor data, application usage details of the smartphones [142, 209–213]. While these frameworks help in the automatic logging of sensor data, self-reports related to various aspects of human life (like emotion) still require direct input from the user.

Balancing Probing Rate and Self-Report Timeliness In the ESM studies, the participant burden mainly arises from repeatedly answering the same survey questions. Time-based, event-based schedules are the most commonly used ESM schedules [214]. Time-based approaches aim to reduce probing rate (at the cost of fine granularity), while event-driven ones try to collect self-report timely (at the cost of a high probing rate). Recently, hybrid ESM schedules are designed combining time-based and event-based ones to trade-off between probing rate and self-report timeliness [215]. With the proliferation of smartphones, and other wearable devices, more intelligent and less intrusive survey schedules, including these limiting the maximum number of triggers, increasing the gap between two consecutive probes, have been designed. Several open-source software platforms, like ESP [216], my experience [217], psychology [218], Personal Analytics Companion [219], are available on different mobile computing platforms to cater to ESM experiments.

Maintaining Response Quality Via Interruptibility-Aware Designs Recent advancements in interruptibility-aware notification management recommend several strategies to probe at opportune moments leveraging contextual information (e.g., placing between two activities like sitting and walking, after completing one task such as messaging or reading a text on mobile) [220–222]. In [223], the authors showed that features like the last survey response, phone's ringer mode, and user's proximity to the screen could predict whether the recipient will see a notification within a few minutes. Leveraging these findings, intelligent notification strategies were developed, which resulted in a higher compliance rate and improved response quality [224, 225]. However, one of the major challenges of using such details in mobile-based ESM design is resource overhead and privacy. Designing ESM sched-

Table 10.3 Summarization of the existing ESM design approaches

ESM Schedule	Probing rate	Timeliness	Opportune probing
Time-based (TB) (e.g. [31], [144])	✓	✗	✗
Event-based (EB) (e.g. [194], [226])	✗	✓	✗
Hybrid [214]	✓	✓	✗
Interruptibility-aware (e.g. [8], [220], [223], [227])	✗	✗	✗
Ideal ESM schedule	✓	✓	✓

ules based on the underlying study can overcome such limitations [226]. Ideally, an ESM schedule shall optimize the probing rate (like time-based schedules), reduce latency (like event-based schedules), and probing moments (like interruptibility-aware schedules) (Table 10.3).

Pros and Cons in Different Emotion Assessment Approaches

In the previous sections, we have presented different approaches to assess emotions. Table 10.4 summarizes the pros and cons of the different self-report and sensing methods of data collection. We pointed out that self-report questionnaires have the advantage of being rather a time and cost-efficient (for assessors) and enable to reveal cognitions that are otherwise hard to capture. Furthermore, the subjectivity of an individual’s view on their emotions expressed via self-reporting might be wanted in some contexts (e.g., for clinical diagnostics). However, self-reports are challenged by numerous confounding variables as fatigue, interpretation and memory biases, non-assessed personal conditions, misunderstanding of the items, and social desirability [100, 101]. Some physiological assessment methods might be more objective (like EEG or blood) but require a laboratory and complex setup and controlled environment. Due to this limitation, real-time assessment of emotions close to an individual’s everyday life experience is not possible.

Additionally, some research is based on induced emotional states. Emotional reactions can be induced in experimental settings. However, the transfer and generalizability of such results into an individual’s real-life is doubtful. Besides, the period of data collection is often limited, and collecting a high volume of data from a large number of participants is difficult. Finally, the participants need to participate actively and contribute to the data collection effort (via self-reports).

In most cases, the data collection cannot be done passively and, consequently, lacks unobtrusiveness. Moreover, most of the discussed methods focus only on one data collection source (e.g., speech, EEG alpha waves, or social network analysis). They are, therefore, very limited regarding the complex emotional process described in the CPM [4].

In the context of emotional assessment, novel, personal sensing methods embedded in daily life via wearables and smartphones promise to overcome some of those issues by providing real-time observations [143, 197, 207, 209–211]. These devices can capture data passively from different modalities without user intervention, log app usage behavior, and leverage different computational models on the device for

Table 10.4 Pro's and Con's of Emotion Assessment Methods

Method	Pro	Con
<i>Questionnaire-based Emotion Assessment</i>		
Self-report questionnaires	<ul style="list-style-type: none"> – Easy to conduct – Access to cognitions – Time-efficient for examiners – Established and accepted in practical fields – Privacy is given – Standardization for the classification of individual results available 	<ul style="list-style-type: none"> – Several biases (memory, interpretation, social desirability) – Time consuming for participants (not examiners)
<i>Physiology-based emotion assessment</i>		
Blood	<ul style="list-style-type: none"> – Objective assessment of autonomous nervous system (ANS) reaction – Standardization for the classification of individual results available 	<ul style="list-style-type: none"> – Intrusive – Not applicable outside a laboratory – Focus on a single factor
Brain	<ul style="list-style-type: none"> – Objective assessment of brain activity as emotional response – Standardization for the classification of individual results available 	<ul style="list-style-type: none"> – Not applicable outside a laboratory – High costs – High setup up requirements – Focus on a single factor
Posture	<ul style="list-style-type: none"> – Naturalistic expression – Setup needed (cameras, posture logging devices) – Important channel of communication 	<ul style="list-style-type: none"> – Categories vary greatly, no standardization – Interpretation of postures is limited – Focus on one class of emotion expression only
<i>Expression-based emotion assessment</i>		
Facial emotion recognition	<ul style="list-style-type: none"> – Objective assessment of expression of emotions – No language needed – Passive data collection possible 	<ul style="list-style-type: none"> – Recording in real-time in daily life may be difficult – Privacy issues – Camera needs to capture facial expression – Awareness of being recorded might interfere natural reaction
Speech emotion recognition	<ul style="list-style-type: none"> – No laboratory setup needed, objective – No body language and physical parameters needed – Passive data collection possible 	<ul style="list-style-type: none"> – Recording in real-time in daily life may be difficult – Privacy issues – Awareness of being recorded might interfere natural reaction – No standardization
<i>Smartphone—And internet-based emotion assessment</i>		

Table 10.4 (continued)

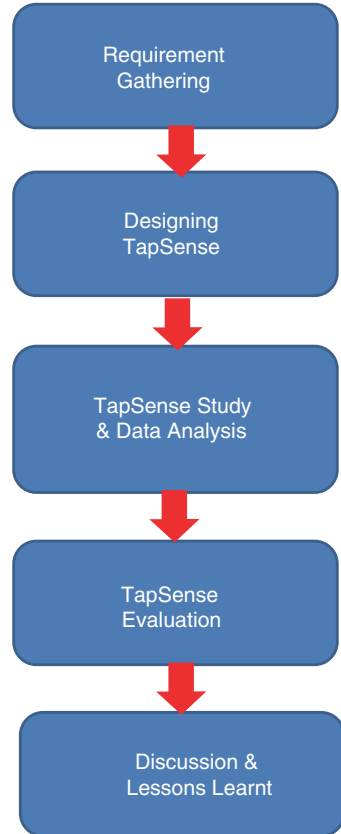
Method	Pro	Con
<i>Questionnaire-based Emotion Assessment</i>		
Social network	<ul style="list-style-type: none"> – Time and cost efficient – No special laboratory or technical equipment needed – Real-time behaviour accessible – Passive data collection possible 	<ul style="list-style-type: none"> – Only specific part of daily expressed emotions visible as embedded in a social network use context – Focusing on just a single class of behaviour – Biased by social desirability and other motivations to communicate on social platforms – Objectivity and/or standard values not given
Ecological momentary assessment	<ul style="list-style-type: none"> – No specific setup needed to conduct – Unobtrusive – Real time assessment in daily life – Access to cognitions 	<ul style="list-style-type: none"> – Self-report with potential biases (social desirability, interpretation bias) – Data protection issues – No standardization
Smartphone-based sensing	<ul style="list-style-type: none"> – No specific setup needed to conduct – Unobtrusive – Real time assessment in daily life – Passive data collection possible – Great variety of data sources possible – Execute different models for emotion inference 	<ul style="list-style-type: none"> – Data protection issues – Perceived as “scary” by participants – No standardization

emotion inference. As a result, these devices are very promising to determine emotion-based behavior based on different usage patterns. However, the approach is still novel, and some participants are concerned about their privacy [227]. These concerns must be taken into account seriously but contrasted by the fact that personal information is shared openly on the Internet nowadays. This ambiguity is labeled as *Privacy Paradox* and known since 2006 [228].

TapSense: Smartphone Typing Based Emotion Assessment

This chapter specifically focuses on assessing the individual’s emotional state from the smartphone usage patterns via the authors’ TapSense study. Therefore, in this section, we first describe the overall research approach of the keyboard interaction study. We focus on a typing-based emotion assessment scenario, which helps to identify the key requirements to design the emotion assessment model and the self-report collection approach using an Experience Sampling Method (ESM) (Sect. 5.1.). In Sect. 5.2, the TapSense field study and data analysis are presented, and in Sect. 5.3. evaluated. In Sect. 5.4, the study is discussed.

Fig. 10.3 Overall approach of the TapSense requirements gathering, study design, data analysis, performance evaluation, and discussion



Background

The overall approach for the TapSense study is shown in Fig. 10.3. First, we gathered a set of requirements to design the keyboard interaction-based emotion assessment tool, followed by the actual design and implementation of the TapSense application. We discuss the study in detail and analyze the collected data. Finally, we evaluate the performance of the TapSense application and discuss the lessons learned from this study.

Requirements

TapSense study relies on the users' smartphone usage patterns. We explain the scenario of typing-based emotion assessment in Fig. 10.4. As the user performs typing activity, we extract his/her typing sessions and the amount of time he/she stays in a single mobile application without. For example, when a user uses

WhatsApp without switching to other applications from t_1 till t_2 , then we define elapsed time between t_1 and t_2 as a Typing Session. Once the user completes the session, he/she is probed via an ESM, i.e., emotion self-report, which is considered the emotion ground truth. Later, several features are extracted from the typing sessions and correlated with the emotion self-report to develop an emotion assessment model. This scenario suggests consideration of the following requirements,

- *Trace keyboard interaction for emotion assessment:* The key requirements while determining emotions from the typing sessions is to make sure that (a) the typing details are captured correctly so that the relevant features can be extracted (b) the emotion ground truths are collected (ESM) and (b) an accurate emotion assessment model is constructed. We discuss these aspects further in this section.
- *ESM design for self-report collection:* Probing a user after every session may induce fatigue due to many probes. So, the probing moments should be chosen in such a manner that it captures the user's response accurately (i.e., before it fades away from the user's memory) and at the same time, the probing rate is not too high. We discuss in detail the ESM design further in this section.

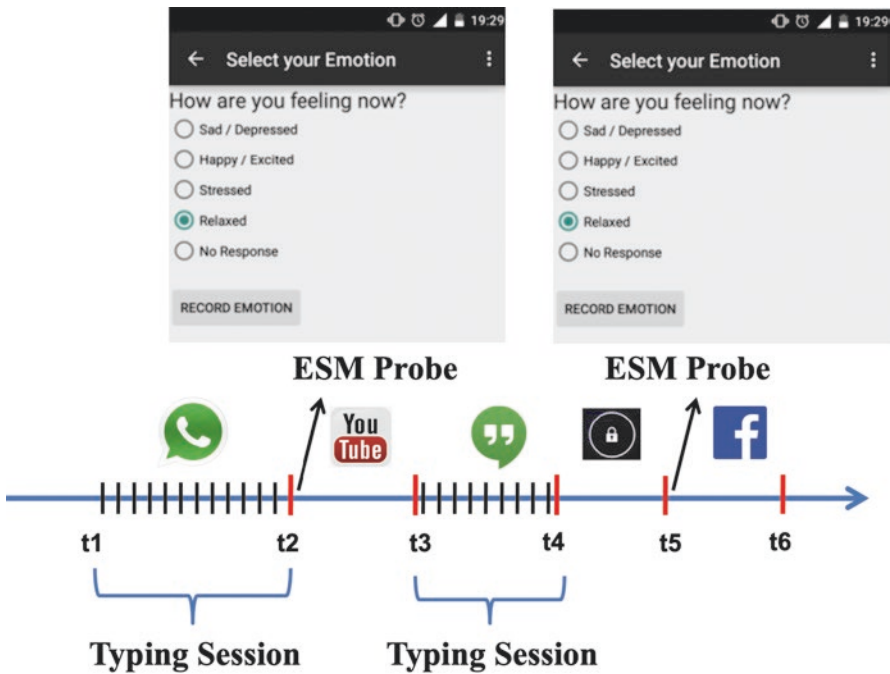


Fig. 10.4 TaspSense Approach: Typing Session-based ESM-triggering scenario

Design and Implementation of TapSense

TapSense consists of the following key components as in Fig. 10.5. TapLogger records the user's typing activity. It implements a virtual keyboard for tracing keyboard interactions. ProbeEngine runs on the phone to generate the user's ESM notifications and collects the ESM responses. The typing details and the associated emotion self-reports are made available at the server via the Uploader module that synchronizes with the server occasionally, or, if the user is offline, once the user connects to the Internet. The emotion assessment model is constructed on the server-side to determine the different emotional states from the typing details and the emotion self-reports. In parallel, a set of typing features is also extracted to construct the inopportune moment assessment model, which feeds back the ProbeEngine to optimize the probe generation. Next, we discuss the two key components of TapSense (a) emotion assessment from keyboard interaction, (b) ESM design for the emotion self-report collection.

TapLogger: Keyboard Interaction Collection

The TapLogger module of TapSense implements an Input Method Editor (IME) [229] provided by Android OS, and we refer to it as the TapSense keyboard (Fig. 10.6). It is the same as any QWERT keyboard; it provides similar functionalities as any Google keyboard. We have selected a standard keyboard because we aimed to provide similar functionalities. The user's keyboard interaction experience does not deviate much from what he/she is used to. It differs from others, as it has the additional capability of logging user's typing interactions, which, for security reasons, is not available in Google keyboard. To ensure user privacy, we do not store or record the characters typed. The logged information is the timestamp of each tap event, i.e., when a character is entered and the key input's categorical type, such as an alphanumeric key or delete key.

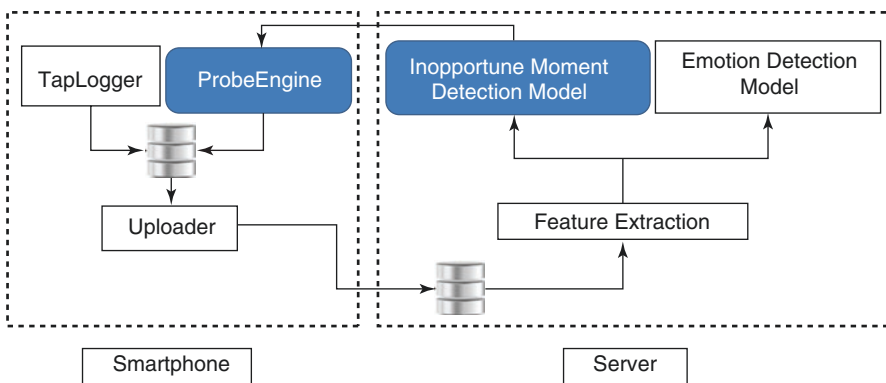


Fig. 10.5 TapSense High-level System Architecture

Fig. 10.6 TapSense Keyboard

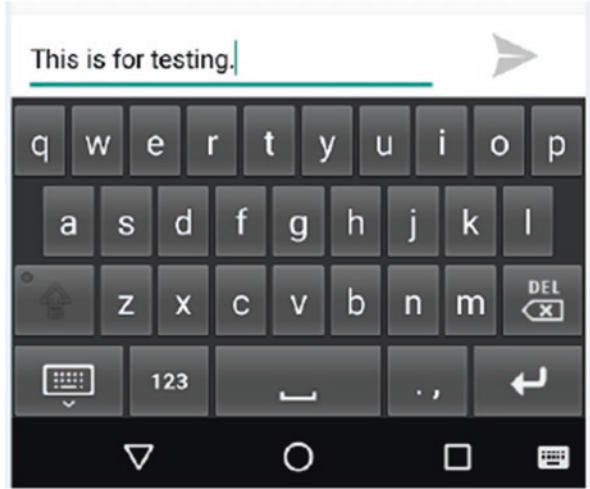
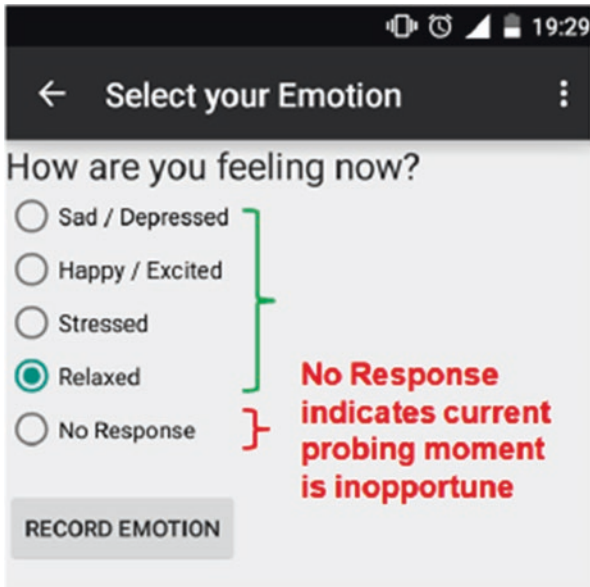


Fig. 10.7 Emotion Self-report UI



ProbeEngine: Emotion Self-Report Collection (ESM)

The ProbeEngine module of TapSense issues the ESM self-report probes by delivering a self-report questionnaire (Fig. 10.7). This survey questionnaire provides the option (happy, sad, stressed, relaxed) to record ground truth about the user's emotion while typing. This captures four largely represented emotions from four

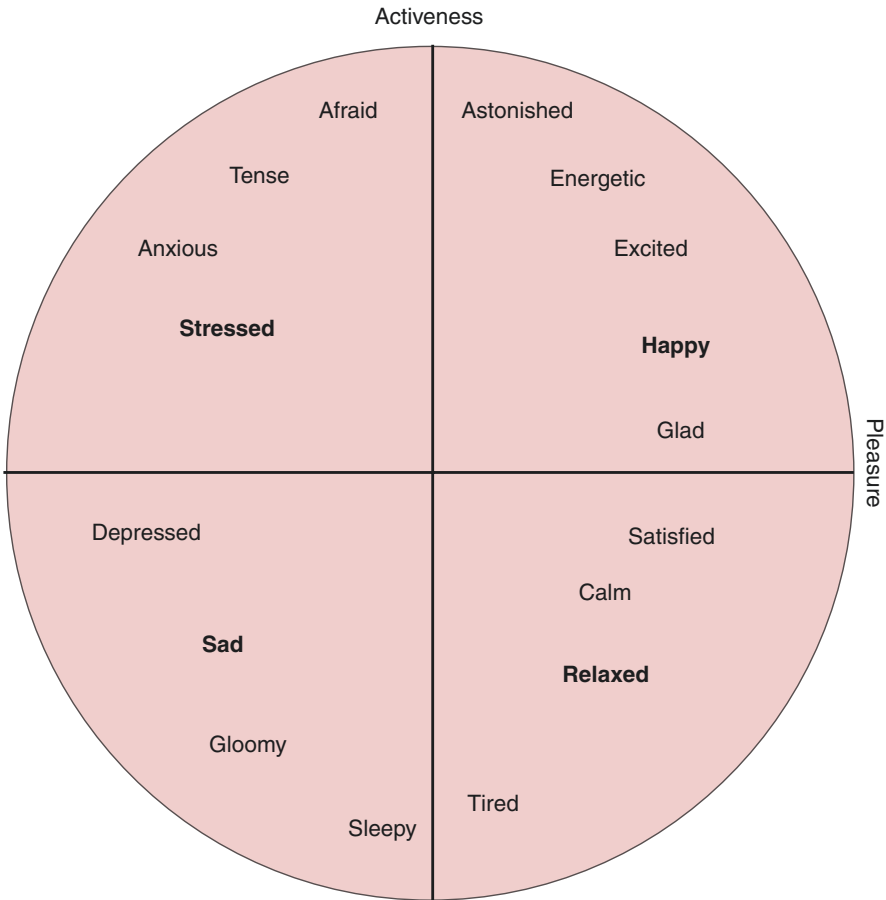


Fig. 10.8 Circumplex Model

different quadrants of the Circumplex model [52], as shown in Fig. 10.8. We select these discrete emotions as their valence-arousal representation is unambiguous on the Circumplex plane. Any discrete emotion and its unambiguous representation on the valence-arousal plane are equivalents [230]. We also include the “No Response” option to select this option to indicate the current probing moment is inopportune.

Emotion Assessment Model Construction

The emotion assessment model in TapSense is responsible for determining the four emotion states based on the keyboard interaction pattern. This is implemented on the server-side once the typing interaction details and the emotion self-reports details are available.

Table 10.5 Features used to Construct TapSense Emotion Assessment Model

Feature Name	Feature Description
Session typing speed (MSI)	Average of all ITDs present in the typing session
Session length	Number of characters typed in the typing session
Session duration	Time duration of the typing session
Backspace percentage	Fraction of backspace and delete keys typed in the typing session
Special character percentage	Fraction of special characters (non-alphanumeric) typed in the typing session
Last ESM trigger response	Emotion label as provided by the user

ITD elapsed time between two consecutive keypress events [ms]

Emotion Assessment Features From raw data collected within every Typing Session, we extract a set of typing features as defined in Table 10.5. For every session, we compute the ITDs, i.e., the elapsed time between two consecutive keypress events for all the presses. We derive the mean of all ITDs in the session and use it as typing speed. We define it as the Mean Session ITD (MSI). We compute the backspace and delete keys present in a session and use it as a feature. This is used as the representation of typing mistakes made in a session.

Similarly, we use the fraction of special characters in a session, session duration, and typed text length in a session as features. Any non-alphanumeric character is considered a special character. We use the last emotion self-report as a label for the model [215, 231]. However, at the later stage, when the TapSense model is operational, we use the predicted emotion for the last session as the feature value for the current session.

Emotion Assessment Model Trees-based machine learning approaches have been accurate in the context of emotion assessment in the past [201, 232]. We design a Random Forest (RF) based personalized multi-state emotion assessment model using the features described in Table 10.5 to assess the emotions. As typing patterns vary across individuals, derived, features will vary. Hence we construct a personalized model. We implement these models in Weka [233], building 100 Random Forest decision trees with a maximum depth of the tree set as ‘unlimited’ (i.e., the tree is constructed without pruning). We then derive the RF models’ performance by deriving the mean and variability of the accuracy for the 100 RF-based models.

Experience Sampling Method Design

The ESM used in TapSense is optimized in two phases. Phase 1 balances probing rate and timeliness of self-report collection, and Phase 2 tries to probe at the opportune moments when the user’s attention is available. We achieve this by designing a two-phase ESM [234]. We summarize it in Fig. 10.9. In Phase 1, we combine policy-based schedules to balance probing rate and timeliness and learn the

inopportune moment assessment model. In Phase 2, we make the inopportune moment assessment model operational. We discuss both phases in detail now.

Phase 1: Balancing ESM Probing Frequency and Timeliness The collection of ESM emotion self-reports at the end of every typing session would help collect the labels close to the event, but it would lead to the generation of too many probes and user burden. To trade off these two conflicting requirements, we first assess the quality of the session itself, i.e., we make sure that there is a sufficient amount of typing done in a typing session for it to be considered. We issue the ESM probe only (a) if the user has performed a sufficient amount of typing, i.e., a minimum $L = 80$ characters in a typing session, and (b) a minimum time interval, i.e., $W = 30$ minutes has elapsed since the last ESM probe. To ensure the labels are collected close to the typing session, we use the polling interval parameter ($T = 15$ seconds) to check if the user has performed a sufficient amount of typing within a session. We describe the selection of threshold values based on initial field trials in Appendix 1. We name this ESM schedule the *Low Interference High Fidelity (LIHF)* ESM schedule (Fig. 10.9 (Phase 1)).

Phase 2: Inopportune Moment Assessment Model As we collect self-reports, we obtain both “No Responses” and valid emotion responses. We leverage these labels to build the inopportune moment assessment model (Table 10.6).

We use typing session duration and the typing length in a session as features since lengthy and longer typing sessions may indicate high user engagement and not be the ideal moment for triggering a probe. Besides, there may be some types of

Fig. 10.9 Balancing ESM Probing Frequency and Timeliness: ESM Triggering Steps

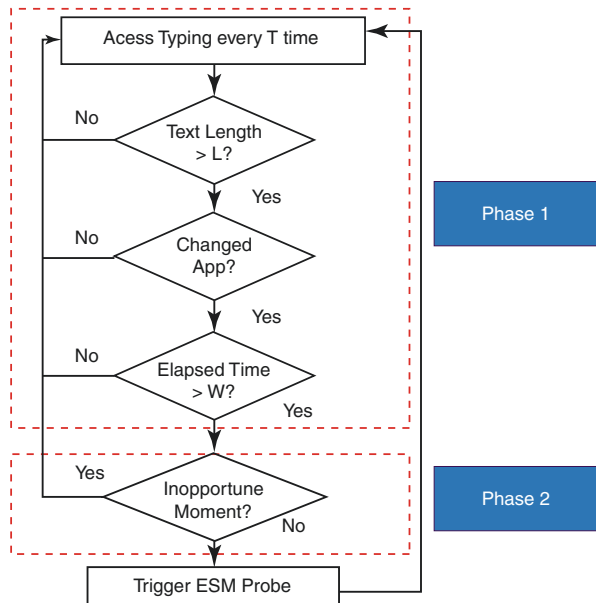


Table 10.6 Features Used To Detect Inopportune Moments

Feature Name	Feature Description
Session duration	Duration of the typing session
Session length	Length of the text in the typing session
App category	Category of the application
Last ESM trigger response	Last ESM trigger response

applications like media, games when the users may not be interrupted for probing. So, we include the application type also as a feature. We categorize the applications into one of the 7 categories: Browsing, Email, Media, Instant Messaging (IM), Online Social Network (OSN), SMS, and ‘Misc,’ following the application’s description in the Google Play Store. Moreover, we use the label of the last ESM probe response as a feature. We use it to determine whether the user continues to remain occupied in the current session and if he/she marked the previous session with “No Response.” However, once the model is operational and deployed, we use the predicted value of the inopportune moment for the last session as the current session’s feature value. Table 6 summarizes the features used to implement the model. We construct a Random Forest-based prediction model to assess the inopportune moments for all the users. The model is augmented with the LIHF schedule to assess and eliminate inopportune probes (Fig. 10.9 [Phase 2]).

TapSense: Field Study and Data Analysis

In this section, we discuss the TapSense field study and the dataset collected from the study.

Study Participants

We recruited 28 university students (22 males, 6 females, aged 24–35 years) to evaluate TapSense. We installed the application on their smartphones and instructed them to use it for 3 weeks. Three participants left the study in between, and the other three participants have recorded less than 40 labels. We have discarded these 6 users and collected data from the remaining 22 participants (18 males, 4 females). The ethics committee approved the study under the approval order IIT/SRIC/SAO/2017.

Instruction and Study Procedure

During the field study, we executed only Phase 1, where we implement the LIHF schedule for self-report collection. We instructed participants to select the TapSense keyboard as the default keyboard. We informed the participants that

when they switch from an application after completing typing activity, they may receive a survey questionnaire as a pop-up to record their emotions. We also advised the participants not to dismiss the pop-up if they are occupied; instead, they were asked to record “No Response” if they do not want to record emotion at that moment.

Collected Dataset

We have collected 4609 typing sessions during this study period, which constitute close to 200 hours of typing labeled with an emotional state of all the participants ($N = 22$). Out of these sessions, we record 642 “No Response” sessions, which is nearly 14% of all recorded sessions. Notably, the actual number of ESM triggers is less than the number of typing sessions because, as per the LIHF policy, if two sessions are close (as defined by W in Fig. 10.8), only one ESM will be triggered to cater to both the sessions. We summarize the final dataset in Table 10.7.

EMA Self-Report Analysis

The users have reported two types of responses (a) One of the four valid emotions or (b) “No Response.” While the valid emotion labels are used to construct the emotion assessment model, the “No Response” labels are important to design the inopportune moment assessment model for the ESM.

Emotion Labels Analysis We show the distribution of different emotion states for every user in Fig. 10.10. We have observed that ‘relaxed’ is the most dominant emotional state for most of the users. Overall, we have acquired 14%, 9%, 30%, 47% sessions tagged with happy, sad, stressed, and relaxed emotion states.

Table 10.7 Collected Data Summary

Number of participants	$N = 22$ (18 m, 4 f)
Total typing events	942,827
Total typing sessions	4609
Total typing duration (in Hr.)	199.1
Mean typing sessions (per user)	209 (std. dev 167.2)
Minimum number of typing sessions for a user	46
Maximum number of typing sessions for a user	549
Total ESM triggers	2554
Mean ESM trigger (per user)	116.1 (std. dev 71.9)

No Response Analysis We show the user-wise distribution of “No Response” sessions in Fig. 10.11a. Although for most users, the fraction of “No Response” labels is relatively low, for a few users, it is more than 40%. We observe the application-wise distribution of “No Response” sessions in Fig. 10.11b; the majority of the “No Response” labels are associated with Instant Messaging (IM) applications like WhatsApp. We also compare the distribution of total “No Response” and total valid emotion labels at weekday, weekend, working hour (9 am-9 pm), and non-working hour in Fig. 10.11c. We infer the working hour based on the timestamp of the ESM response. We compute the percentage of total “No Response,” and the percentage of total other sessions is recorded at these times. However, in our dataset, we do not observe any major differences among these distributions. We also explore the time-wise distribution of No Response sessions in Fig. 10.11d, which indicates that a small number of No Response sessions were recorded during the late-night from 3 am onwards. This can be attributed to overall less engagement during late night.

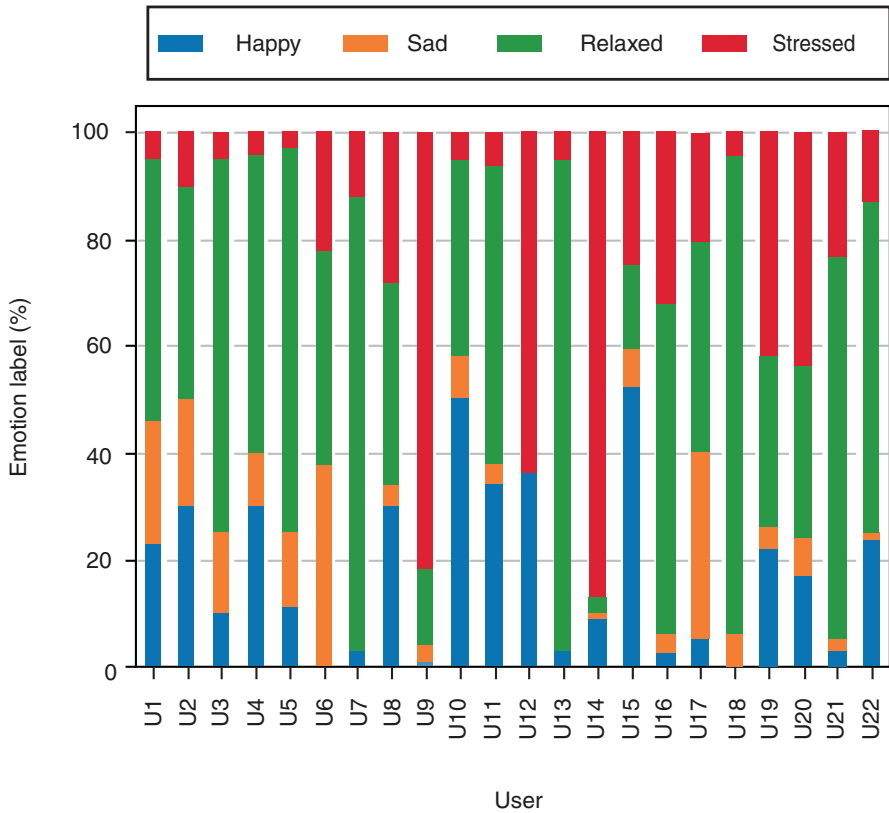


Fig. 10.10 Distribution of emotion labels for every user. All but 5 users (6,7,12, 13, 18) recorded all four emotions

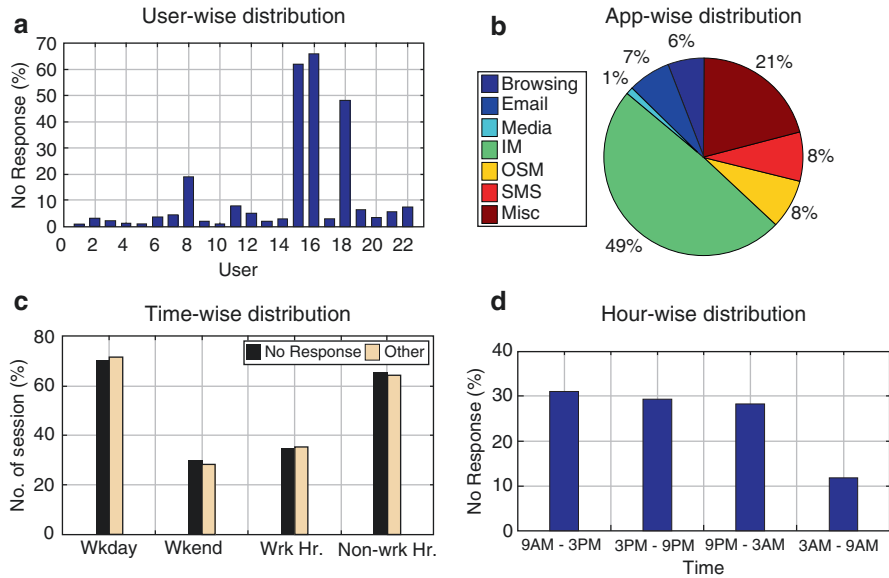


Fig. 10.11 Distribution of No Response sessions (a) User-wise (b) App-wise (c) Time-wise (d) Hour-wise

TapSense Evaluation

In this section, first, we discuss the experiment setup. Then we evaluate the emotion classification performance and the ESM performance. Finally, we discuss the limitations of the study.

Experimental Setup

During the field study, we used the LIHF ESM schedule for collecting self-reports. However, to perform a comparative study across different policies, we require data from time-based and event-based ESM schedules under identical experimental conditions from every participant. In the actual deployment, identical conditions are impossible to repeat over different time frames. Hence, we generate traces for the other policy-based schedules from the data collected using LIHF ESM. We outline the generation steps for these traces in Appendix 2. We show the distribution of emotion labels obtained from different schedules after trace generation in Fig. 10.12.

Baseline ESM Schedules

Different ESM schedules, listed in Table 10.8, used for comparison are described.

- *Policy-based ESM*: We focus on the Phase 1 approach (i.e., without optimizing the triggering) and use three policy-based ESM schedules—Time Based (TB),

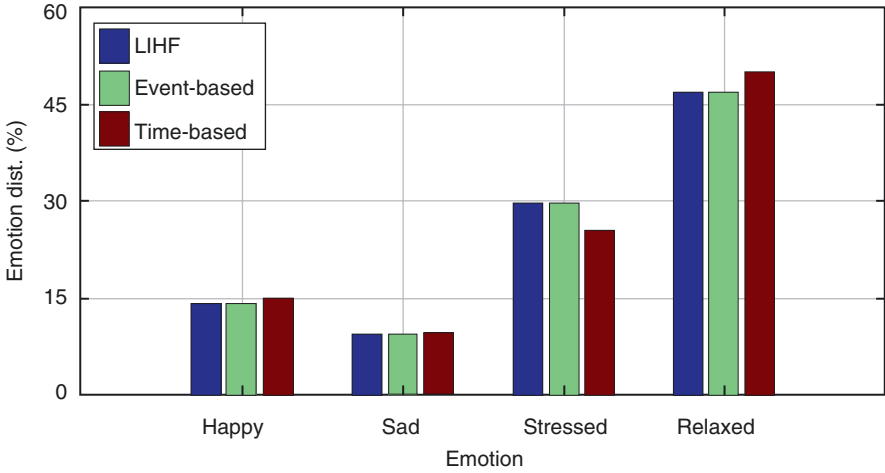


Fig. 10.12 Frequency distribution of different emotions across various ESM schedules

Table 10.8 Different ESM Schedules based on the Policy used in Phase 1 and Usage of the Model in Phase 2

ESM Schedule	Phase 1	Phase 2
TB	Time-based	No model is used.
EB	Event-based	No model is used.
LIHF	LIHF	No model is used.
TB-M	Time-based	Inopportune moment detection model
EB-M	Event-based	Inopportune moment detection model
LIHF-M	LIHF	Inopportune moment detection model

Event-Based (EB), and LIHF. In the case of TB, probes are issued at a fixed interval (3 hours). In EB’s case, after every typing session, a probe is issued while LIHF implements the LIHF policy. These approaches do not use an inopportune moment assessment model. Comparing these schedules helps to understand their effectiveness in reducing the probing rate and collecting self-reports timely.

- *Model-based ESM:* We use the following model-based ESM schedules—TB-M, EB-M, and LIHF-M. These ESM schedules implement TB, EB, and LIHF schedules in Phase 1, respectively, followed by the inopportune moment assessment model operational in Phase 2. In all these schedules, the model is constructed using the same set of features (Table 10.6) extracted from relevant trace (i.e., for TB-M, the model is constructed from the trace of TB and similarly). Comparison of these model-driven schedules helps to understand the efficacy of the model in assessing the inopportune moments and whether applying the model with any off-the-shelf ESM is good enough to improve survey response quality.

Overall Performance Metrics

We use the classification accuracy to measure the emotion classification performance, and as for the ESM performance, we assess it along with the probing rate index and its reduction wrt. The classical self-report approach, timely self-report collection, inopportune moment identification, and valid response rate collection.

Emotion Assessment: Classification Accuracy (Weighted AUCROC) The performance of supervised learning algorithms highly depends on the quality of labels [235]. The label quality can adversely impact classification accuracy [236, 237]. In our research, we use Typing Session emotion classification accuracy. We measure it in terms of the weighted average of AUCROC (auc_{wt}) using AUCROC from four different emotional states. Let f_i , auc_i indicate the fraction of samples and AUCROC for emotion state i respectively, then $auc_{wt} = \sum_{\forall i \in \{\text{happy, sad, stressed, relaxed}\}} f_i * auc_i$.

ESM Performance Metrics

Probe Frequency Index (PFI) We compare the probing frequencies of different ESM schedules using PFI, defined as follows. Let there be different ESM schedules ($e \in E$) and N_i^e denotes the number of probes issued for the user i for an ESM schedule e , then PFI for user i for ESM schedule e is

$$\text{expressed as, } PFI_i^e = \frac{N_i^e}{\forall e, \max(N_i^e)}.$$

The Recency of Label (RoL) The timeliness of self-report response collection is measured using RoL defined as follows. Let there be different ESM schedules ($e \in E$), and d_i^e denotes average elapsed time between typing and probing for user i for an ESM schedule e , then RoL for user i for ESM schedule e is expressed as, $RoL_i^e = \frac{1}{\forall e, \max(d_i^e)}$.

Inopportune Moment Identification We measure Precision, Recall and F-score for inopportune moment assessment. We also compute the weighted AUCROC (auc_{wt}) for the inopportune and opportune moments. Let f_i , auc_i indicate the fraction of samples and AUCROC for class i respectively, then $auc_{wt} = \sum_{\forall i \in \{\text{inopportune, opportune}\}} f_i * auc_i$.

Valid Response Rate (VRR) We also compare the percentage of valid emotion labels for different ESM schedules. Let there be different ESM schedules ($e \in E$), and nr_e denotes the fraction of No Response sessions recorded for ESM e , then Valid Response Rate for ESM e is expressed as $VRR_e = (1 - nr_e) * 100$.

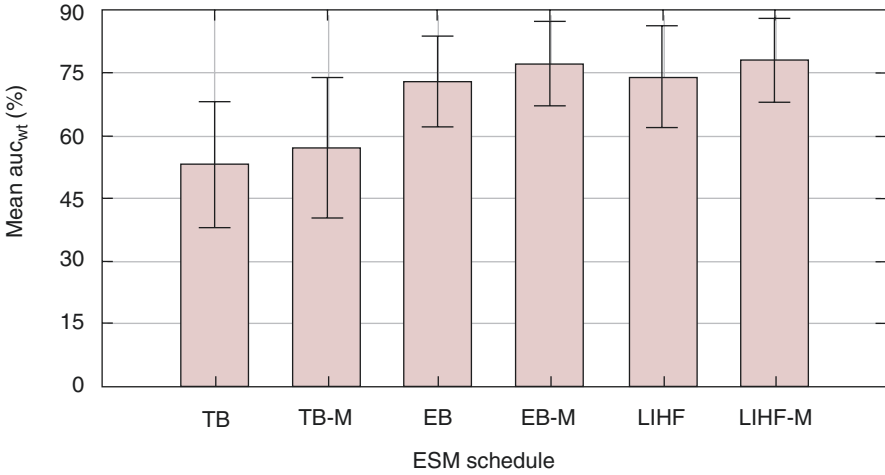


Fig. 10.13 Emotion classification accuracy (AUCROC) using different ESM schedules

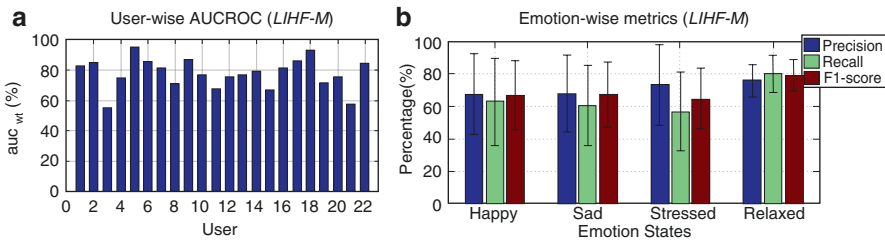


Fig. 10.14 Emotion classification (a) AUC_{wt} for LIHF-M per user (b) Emotion-wise accuracy LIHF-M

Emotion Assessment: Classification Performance

The emotion classification accuracy for different ESMs is shown in Fig. 10.13. We observe that the LIHF-M outperforms other schedules with a mean AUCROC of 78%. It returns a maximum improvement of 24% with respect to TB and an improvement of 5% with respect to EB. We also observe that after applying the inopportune moment assessment model, the mean AUCROC (auc_{wt}) improves (by 4%) for each corresponding schedule (TB, EB, LIHF).

We also show the user-wise emotion assessment AUCROC (auc_{wt}) corresponding to the LIHF-M schedule in Fig. 10.14a. The quality of the prediction for each emotion category is presented in Fig. 10.14b. The emotion states are identified with an average f-score between 54% and 74%. We observe that the relaxed state is identified with the highest f-score, followed by sad, stressed, and happy states, respectively. As data volume increases, as in the case of the relaxed state, the performance metrics improve.

Table 10.9 Ranking of Features Used to Construct TapSense Emotion Assessment Model

Feature name	Rank	Average IG.
Last ESM trigger response	1	0.468
Session typing speed	2	0.376
Backspace percentage	3	0.270
Session length	4	0.231
Special character percentage	5	0.203
Session duration	6	0.181

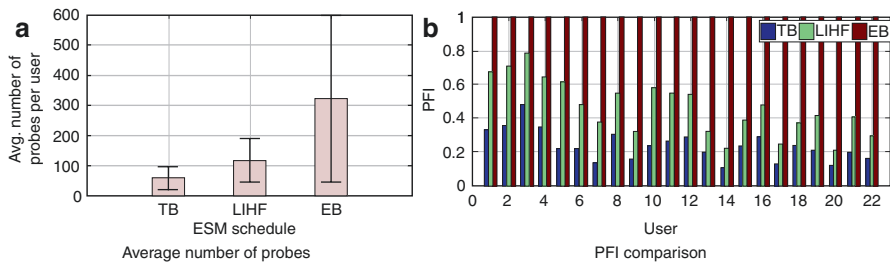


Fig. 10.15 Probing rate across ESM schedules: Average number of probes and the Probe Frequency Index. (a) Average number of probes (b) PFI comparison

Influence of Emotion Assessment Features

We find the importance of the input features used for emotion assessment using the ‘InfoGainAttributeEval’ method from Weka. We compute the average Information Gain (IG) of every feature and rank them in Table 10.9. We observe that the last ESM response is the most discriminating feature, followed by features like typing speed and backspace percentage. All the features are found to have an input into the model for the emotion assessment.

ESM Performance

In this section, we evaluate the ESM’s performance in terms of the three parameters (ESM probing rate, self-report timeliness, and opportune probing moments).

Probing Rate Reduction

We compare the average number of probes issued by each ESM schedule in Fig. 10.15a. We observe that time-based ESM (TB) issues the minimum number of probes, event-based ESM (EB) issues the maximum number of probes, while LIHF ESM lies in between. It is observed that the average number of probes is reduced by 64% for LIHF ESM policy.

We also perform the user-wise comparison using the Probe Frequency Index (PFI) metric in Fig. 10.15b. For all users, PFI for LIHF ESM is lower than that of event-based ESM. Across all users, there is an average improvement of 54% in PFI. Time-based ESM is the best in PFI but does not capture self-reports timely, as

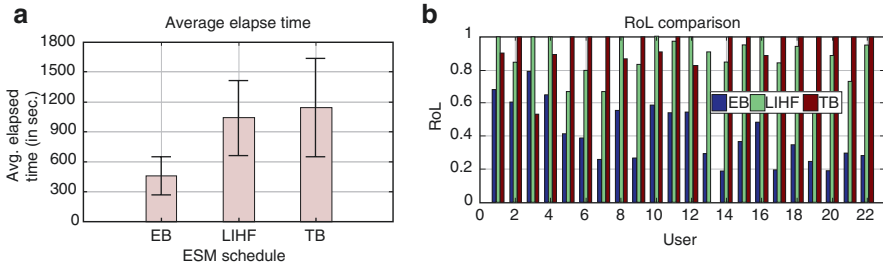


Fig. 10.16 Timeliness of the ESM collection across schedules: Average elapsed time and the Recency of Label. (a) Average elapsed time (b) RoL comparison

shown later. LIHF ESM schedule enforces a minimum elapsed time between two successive probes; it generates fewer probes and reduces probing rate compared to event-based ESM.

Timely Self-Report Collection

We measure how close to the event (i.e., typing session completion) the ESM schedule collects the self-report. We compare the average elapsed time between typing completion and self-report collection for different ESM schedules in Fig. 10.16a. The average elapsed time is the least for event-based ESM, highest for time-based ESM, while for the LIHF, it lies in between. The average elapsed time for label collection is reduced by 9% for LIHF.

We also compare the recency of labels using RoL in Fig. 10.16b. We observe that for every user, RoL is minimum for EB, and for most of the users, RoL is maximum in the case of TB, while for LIHF, the RoL lies in between. In the case of EB, we issue the probe as soon as the typing event is completed; it can collect self-reports very close to the event, resulting in the lowest RoL. On the contrary, in TB, we perform probing at an interval of 3 hours. As a result, there is often a large gap between typing completion and self-report collection, resulting in high RoL. However, in the case of LIHF, we keep accumulating events and separate two consecutive probes by at least half an hour; we compromise to some extent in the label recency, yet less than in the case of TB.

Inopportune Moment Assessment

We compare the inopportune moment classification performance of three model-based approaches in Fig. 10.17a. We observe that the LIHF-M attains an accuracy (AUCROC) of 89%, closely followed by EB-M (88%), while TB-M (75%) performs poorly. We also note the precision, recall, and F-score values of identifying inopportune moments in Fig. 10.17b using the LIHF-M schedule. We also report the recall rate of inopportune moments for every user in Fig. 10.17c. We observe that for 14% of the users, the recall rate is greater than 75%, and for 60% of the users, the recall rate is greater than 50%. It is observed that users with many “No Response” (Fig. 10.8a) get more benefit using the inopportune moment assessment model. In summary, the proposed model combined with LIHF ESM performs best, while other ESM schedules also assess the inopportune moments accurately with this model.

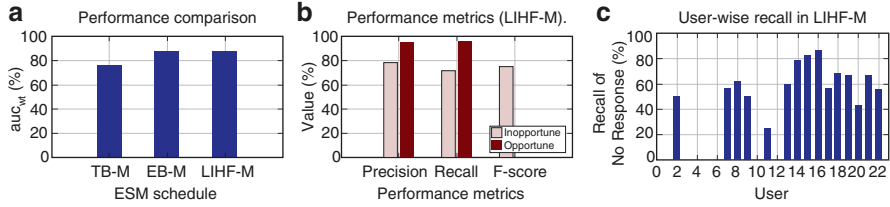


Fig. 10.17 Inopportune moment assessment performance (a) different ESM schedules (b) LIHF-M schedule (c) User-wise LIHF-M schedule

Table 10.10 Ranking of Features Used to Construct Inopportune Moment Model

Feature Name	Rank	Average IG
Last ESM trigger response	1	0.669
App category	2	0.053
Session length	3	0.019
Session duration	4	0.012

Influence of Inopportune Moment Assessment Features

We find the importance of every feature by ranking them based on the information gain (IG) achieved by adding it for predicting the inopportune moment. We use the InfoGainAttributeEval method from Weka [233] to obtain the information gain of each feature. Our results (in Table 10.10) show that the last ESM probe response is the most important feature, followed by the application category.

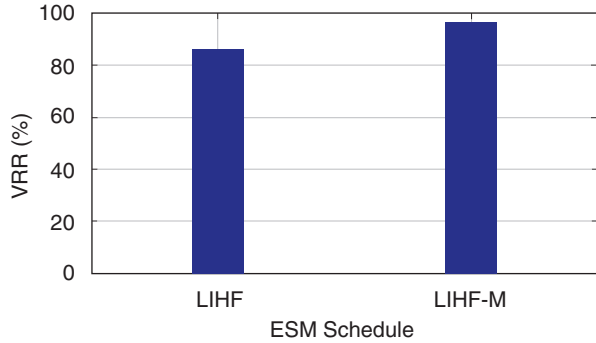
Valid Response Collection

We compare the valid response rate (VRR) for LIHF, LIHF-M schedules in Fig. 10.18. We do not consider other schedules as those labels were generated synthetically. The VRR for LIHF is 86%, and the same for LIHF-M is 96%. This further proves the effectiveness of the inopportune moment assessment model. As the model is in place for LIHF-M, it assesses and skips probing at the inopportune moments, thereby improving the number of valid emotion responses..

TapSense Study Discussion and Lessons Learnt

In our research, we leverage the user’s smartphone for accurate and timely emotion assessment. We designed, developed, and evaluated TapSense, which passively logs typing behavior and develops a personalized machine learning model for multi-state emotion detection. We log the keyboard interactions (typing patterns and not the actual content) of the user and infer four types of emotions (*happy, sad, stressed, relaxed*). We also proposed an intelligent ESM-based self-report collection method

Fig. 10.18 Comparing the number of valid responses for LIHF, LIHF-M schedules



and integrated the same with TapSense, optimizing the manual self-report collection. We evaluate the emotion classification performance and the ESM performance of TapSense in a 3-week in-the-wild study involving 22 participants. The empirical analysis reveals that the TapSense can infer emotions with an average accuracy of 78%. It also demonstrates the efficacy of the proposed ESM in terms of probing rate reduction (on avg. 24%), self-report timeliness (on avg. 9%), and probing at opportune moments (on avg. 89%); all of which improve the emotion classification performance.

However, a few factors need to be considered before deploying sensing technologies as TapSense as an emotion assessment tool. First, it is crucial to consider keyboard interaction experience should not be impacted while using the TapSense keyboard as most of the participants are conversant with the Google keyboard. However, we do not observe a significant effect in the app usage due to this, as we record 86% valid emotion labels and, on average, 209 typing sessions per user. Second, the model-driven probing strategy at opportune moments may not perform well for some users if the number of “No Response” labels are very few (less than 4% of all sessions). If the number of no response labels is very less, then the model may not perform well and may not detect all the inopportune moments.

Another factor to consider is which ESM strategy is to be adopted during self-report collection. We recommend using the LIHF strategy to reduce survey fatigue compared to fixed event-based (EB) schedules, but it may suffer from latency in the self-report collection. Time-driven schedules may not be used if long-time-interval separates two probes. This may miss capturing the fine-grain event details, which are likely to carry emotional signatures. Finally, during self-reporting, if the participants have skipped the pop-up instead of selecting “No Response,” we could not capture those moments in our study. However, this can be easily incorporated by logging the skipping events.

The study we have carried out involving TapSense has several limitations that may limit the results’ generalization. First of all, the study has been of small size, as only 22 users have been engaged in the study for only 3 weeks. On the other hand, given the number of self-reports and user typing sessions, we may assume we have captured a representative sample of the general population, with 25 minutes a day

of interaction being logged and labeled for an emotional expression. As we have shown, the representation of emotional states was diverse. The additional limitation may stem from the Android OS only study; the iOS participants may have different traits and emotional states. The emotional modeling may have led to different results. Nevertheless, we present the results as indicative and will follow future research in this context, given a larger population sample and longer study duration.

Conclusive Remarks

In this chapter, we focused on emotions as indicators of quality of life due to the association of positive/negative life experiences and positive/negative emotional states and other aspects of quality of life, such as health, safety, economic and mental well-being. We highlighted functions and the importance of emotions in a persons' daily life and the challenge of assessing emotions in traditional vs. novel methods of assessing emotions. The advantages and disadvantages of the diverse methods are especially rooted in objectivity, required assessment setup, self-report bias, privacy, real-time measurement, obtrusiveness, and inclusion of a wide range of emotion components.

In more depth, in this chapter, we have leveraged smartphone interactions to assess a user's mental state. As smartphones have become a true companion of our daily life, they can passively sense the usage behavior and mental state. With numerous typing-based communication applications on a smartphone, typing characteristics provide a rich source to model user emotion. In the specific case, the users' keyboard interactions predicted four different emotion categories with an average accuracy of 78%, confirming and outperforming other approaches of smartphone-based emotion-sensing technologies. Hence, emotional assessment is feasible to conduct via different technologies.

TapSense study shows promising results for minimally obtrusive, smartphone-based emotional state assessment, which may be leveraged for further studies and, if largely improved, in clinical practice for a 'companion assessment' of individuals' mental health, accompanying the current gold standard assessment methods and approaches. It is in line with the recent research results showing a potential for co-calibration of the self-reported, gold-standard approaches with the technology-reported ones [238, 239]. The aspects of minimal obtrusiveness may be of interest, especially for the leaders in the self-assessment space—the co-called QuantifiedSelfers [190, 240] who leverage diverse self-assessment technologies for better self-knowledge and optimization of daily life activities for better well-being, health, and other outcomes in the long term. Overall, the TapSense may be seen as an example of the emerging Quality of Life Technologies [241, 242] to assess the individual's behavioral patterns for better life quality. Once proven accurate and timely, and highly reliable in the context of daily life assessment, TapSense, and similar technologies may pave the way for technologies for a better

understanding of the physical, mental, and emotional well-being of populations at large.

Appendices

Appendix 1: Parameter Threshold Value

We use three parameters L , W , T as defined in details in sect. 5.1 (Phase 1: Balancing ESM Probing Frequency and Timeliness) to balance between probing frequency and timeliness in label collection. L is defined as the minimum amount of typing performed in a typing session. W is defined as the minimum time elapsed since last ESM trigger, and T is defined as the polling interval (i.e. how frequently the typing session will be checked for sufficient amount of typing). Based on our initial dataset, we observe the CDF of session length (L) in Fig. 10.19a, which reveals that frequency distribution of session length is highly skewed. So, we select 66th percentile value as the threshold so that two-third values are less than this value. We observe similar CDF and frequency distribution (Fig. 10.19b) for inter-session gap (W) and use the 66th percentile value as the threshold.

However, polling interval (T) is to be chosen in such a way that for most of the sessions, the event of interest is captured within this interval. In this case, the event is change of application after typing in a session. For this purpose, we measure the elapsed time between two successive key pressing events (ITD) in a session. We note the CDF of all ITD values from all sessions in Fig. 10.20. We observe that 99% of the inter-tap duration (ITDs) are less than 15 seconds i.e. for most of the sessions the application change happens after 15 seconds. So, we decide to use 15 seconds as the threshold for T .

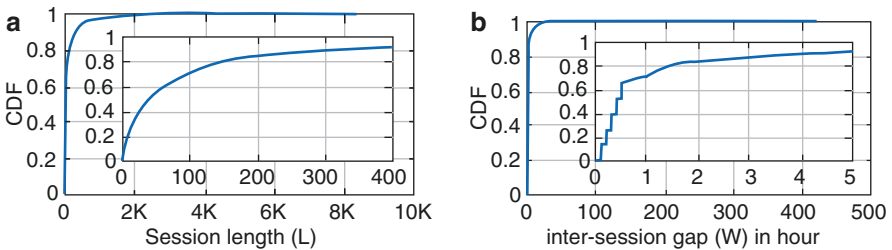
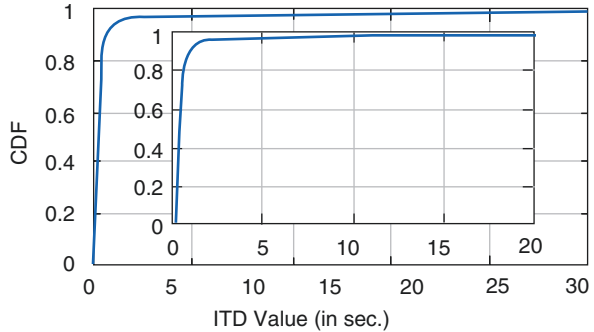


Fig. 10.19 CDFs for session length (L) and inter-session gap (W) reveal that the distribution is skewed. 66th percentile value is chosen as threshold so that two-third values are less than threshold. (a) CDF for session length (L). (b) CDF for inter-session gap (W)

Fig. 10.20 CDF of polling interval (T). It reveals the distribution is skewed and 99% ITD values are close to 15 seconds



Appendix 2: The ESM Trace Generation

In this section, we discuss in detail the steps followed to generate trace for Time-based (TB) ESM and Event-based (EB) ESM schedule from the data collected using LIHF ESM schedule. In Fig. 10.21, a schematic is given to depict the same. E_i denotes the application switching event after sufficient typing. In case of LIHF ESM, there are 6 such events, however only 5 probes were issued (Fig. 10.21a). No probe is issued after E_3 because it occurs within time-window ($W = 30$ minutes) since last probe (Probe 2). In order to generate the corresponding Time-based trace, probes are considered at 3 hour interval. As a result, there will be only one probe Probe 1 and all events E_1 to E_6 will be labeled with the single emotion response collected via it (Fig. 10.21b). But in case of conversion to Event-based ESM, all events are treated separately, as a result there will be in total 6 probes and the emotion labels will be assigned accordingly to the respective events (Fig. 10.21c). Next, we define the formal procedure for trace generation.

Generation of Time-based Trace

We take the trace collected from LIHF schedule »C.lihf p₅ as p₅ matrix where p denotes the total number of key press events. We generate the respective Time based trace »C.time p₅ following the Algorithm 1. We consider the sampling interval of Time-based ESM as 3 hours. We parse through (line 5–12) the LIHF trace »C. lihf p₅ and all key press events. As in case of LIHF, two responses may be recorded less than 3 hour interval, we may need to down-sample, which is performed in following way. If two emotion responses for key press events are collected within 3 hours, both are considered as a part of single session and the later is labeled with the previous emotion. Otherwise, they belong to different session and the new emotion response is considered (line 7–9).

Generation of Event-based Trace

We design Algorithm 2 to generate the corresponding event-based trace »C.event p₅ from the collected LIHF trace »C. lihf p₅. We consider changing application after typing as an event. We parse through (line 5–12) the trace obtained from LIHF schedule and all key press events. If two consecutive key press events are associated with different application, they belong to separate session (line 6–7). Otherwise,

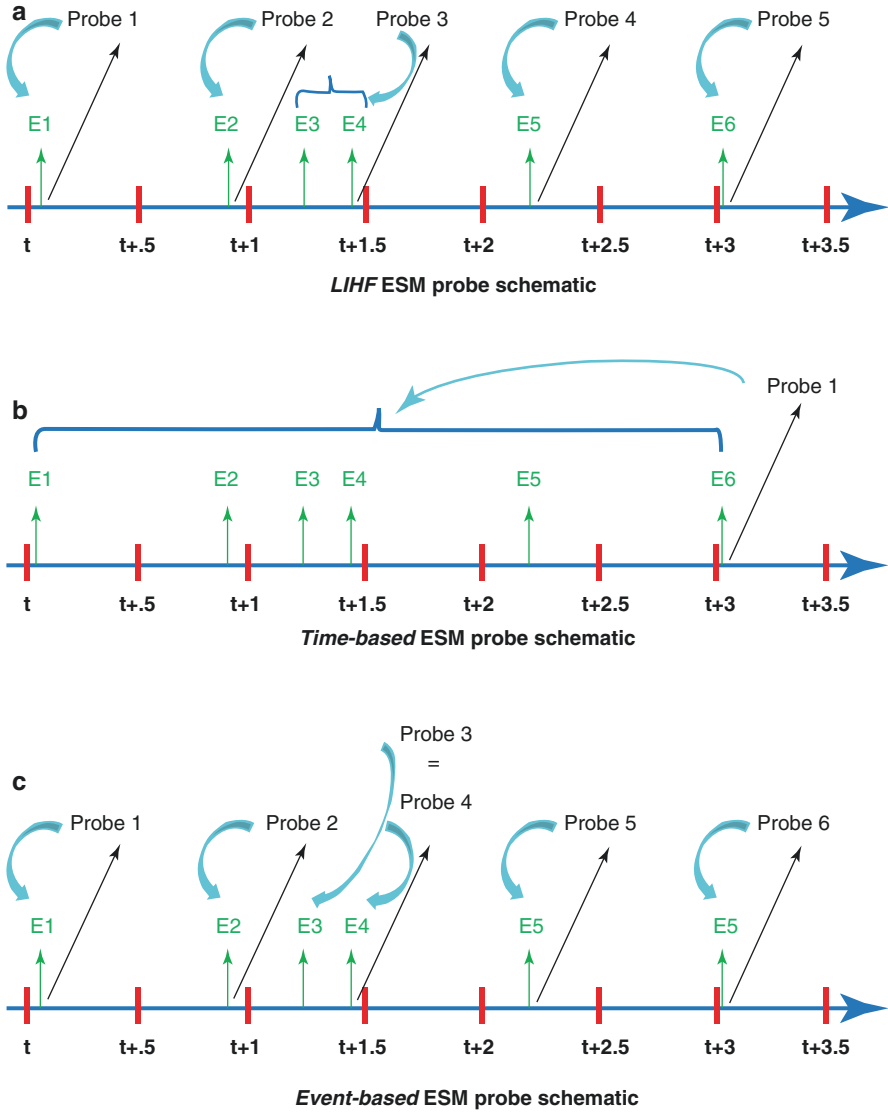


Fig. 10.21 Schematic to display how Event-based and Time-based traces are generated from LIHF ESM

they are considered as part of the same session. In both these cases, no emotion response is dropped (unlike time-based), they are associated with different sessions. In case of LIHF, multiple sessions are grouped and tagged with single emotion, but in case of event-based schedule, this grouping is not done and every session is labeled with the same response. This is how the over-sampling is done in case of event-based schedule.

Algorithm 1: Time-based trace generation

Input: $[C]_{p \times 5}^{lihf}$, key pressing details of a user as obtained from LIHF schedule; $[C]_{:,1}^{lihf}$ contain session number, $[C]_{:,2}^{lihf}$ contain the application names, $[C]_{:,3}^{lihf}$ contain the timestamp of the key press, $[C]_{:,4}^{lihf}$ contain emotion recording timestamp for the session, $[C]_{:,5}^{lihf}$ contain the associated emotion.; p denotes the total number of key press performed by the user.

Output: $[C]_{p \times 5}^{time}$; Corresponding time-based trace

/* Copy the first row from LIHF schedule (session number, application name, emotion and other) */

```

1 sessiontime ← [C]lihf(1, 1)
2 apptime ← [C]lihf(1, 2)
3 emotiontime ← [C]lihf(1, 5)
4 [C]time(1, :) ← [C]lihf(1, :)
5 for i ← 2 to p do
    /* Find elapsed time between two consecutive emotion recording timestamp */
    6 δ ← time_difference([C]lihf(i, 5), [C]lihf(i - 1, 5))
    7 if δ ≥ 3 hours then
    8     sessiontime ← sessiontime + 1
    9     emotiontime ← [C]lihf(i, 5)
10    [C]time(i, 1) ← sessiontime
11    [C]time(i, 2 : 4) ← [C]lihf(i, 2 : 4)
12    [C]time(i, 5) ← emotiontime
13 return [C]time

```

Algorithm 2: Event-based trace generation

Input: $[C]_{p \times 5}^{lihf}$, key pressing details of a user as obtained from LIHF schedule; $[C]_{:,1}^{lihf}$ contain session number, $[C]_{:,2}^{lihf}$ contain the application names, $[C]_{:,3}^{lihf}$ contain the timestamp of the key press, $[C]_{:,4}^{lihf}$ contain emotion recording timestamp for the session, $[C]_{:,5}^{lihf}$ contain the associated emotion.; p denotes the total number of key press performed by the user.

Output: $[C]_{p \times 5}^{event}$; Corresponding event-based trace

/* Copy the first row from LIHF schedule (session number, application name, emotion and other) */

```

1 sessionevent ← [C]lihf(1, 1)
2 old_appevent ← [C]lihf(1, 2)
3 [C]event(1, :) ← [C]lihf(1, :)
4 for i ← 2 to p do
    /* Find current application name */
    5 curr_appevent ← [C]lihf(i, 2)
    6 if curr_appevent ≠ old_appevent then
    7     sessionevent ← sessionevent + 1
    8     [C]event(i, 1) ← sessionevent
    9     [C]event(i, 2 : 5) ← [C]lihf(i, 2 : 5)
10 return [C]event

```

References

1. Harari GM, Müller SR, Aung MS, Rentfrow PJ. Smartphone sensing methods for studying behavior in everyday life. *Curr Opin Behav Sci.* 2017;18:83–90.
2. Mehrabian A. Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament. *Curr Psychol.* 1996;14(4):261–92.
3. Scherer KR. The dynamic architecture of emotion: evidence for the component process model. *Cogn Emot.* 2009;23(7):1307–51.
4. Mehu M, Scherer KR. Normal and abnormal emotions—the quandary of diagnosing affective disorder. *Emot Rev.* 2015;7(Special issue)
5. Larson R, Csikszentmihalyi M. The experience sampling method. In: *Flow and the foundations of positive psychology.* Springer; 2014. p. 21–34.
6. Conner TS, Tennen H, Fleeson W, Barrett LF. Experience sampling methods: a modern idiographic approach to personality research. *Soc Personal Psychol Compass.* 2009;3(3):292–313.
7. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol.* 2008;4:1–32.
8. Pejovic V, Musolesi M. InterruptMe: designing intelligent prompting mechanisms for pervasive applications. In: *Proceedings of ACM UbiComp; 2014.* p. 897–908.
9. Bartlett MS, Littlewort GC, Sejnowski TJ, Movellan JR. A prototype for automatic recognition of spontaneous facial actions. In: *Advances in neural information processing systems; 2003.* p. 1295–302.
10. Cohn JF, Reed LI, Ambadar Z, Xiao J, Moriyama T. Automatic analysis and recognition of brow actions and head motion in spontaneous facial behavior. In: *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583), vol. 1; 2004.* p. 610–6.
11. Kapoor A, Burleson W, Picard RW. Automatic prediction of frustration. *Int J Hum Comput Stud.* 2007;65(8):724–36.
12. Littlewort GC, Bartlett MS, Lee K. Faces of pain: automated measurement of spontaneous-allfacial expressions of genuine and posed pain. In: *Proceedings of the 9th International Conference on Multimodal interfaces; 2007.* p. 15–21.
13. Bartlett MS, Littlewort G, Frank M, Lainscsek C, Fasel I, Movellan J. Recognizing facial expression: machine learning and application to spontaneous behavior. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 2; 2005.* p. 568–73.
14. Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol.* 1996;70(3):614.
15. Lee CM, Narayanan SS. Toward detecting emotions in spoken dialogs. *IEEE Trans speech audio Process.* 2005;13(2):293–303.
16. Devillers L, Vidrascu L. Real-life emotions detection with lexical and paralinguistic cues on human-human call center dialogs. In: *Ninth International Conference on Spoken Language Processing; 2006.*
17. Schuller B, Stadermann J, Rigoll G. Affect-robust speech recognition by dynamic emotional adaptation. In: *Proc. Speech Prosody 2006, Dresden; 2006.*
18. Litman D, Forbes-Riley K. Predicting student emotions in computer-human tutoring dialogues. In: *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04); 2004.* p. 351–8.
19. Schuller B, Villar RJ, Rigoll G, Lang M. Meta-classifiers in acoustic and linguistic feature fusion-based affect recognition. In: *Proceedings(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005, vol. 1; 2005.* p. 1–325.
20. Fernandez R, Picard RW. Modeling drivers' speech under stress. *Speech Commun.* 2003;40(1–2):145–59.

21. Pinto J, Moulin T, Amaral O. On the transdiagnostic nature of peripheral biomarkers in major psychiatric disorders: a systematic review. In: *Transdiagnostic Nat. Peripher. Biomarkers major Psychiatr. Disord. A Syst. Rev.*, p. 086124; 2016.
22. Ejger M, Ley M, Hanke S. Emotion recognition from physiological signal analysis: a review. *Electron Notes Theor Comput Sci.* 2019;343:35–55.
23. Aviezer H, Trope Y, Todorov A. Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science* (80-). 2012;338(6111):1225–9.
24. Healey J, Picard R. Digital processing of affective signals. In: *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process.–Proc*, vol. 6; 1998. p. 3749–52.
25. AlZoubi O, Calvo RA, Stevens RH. Classification of EEG for affect recognition: an adaptive approach. In: *Australasian Joint Conference on Artificial Intelligence*; 2009. p. 52–61.
26. Bashashati A, Fatourechi M, Ward RK, Birch GE. A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. *J Neural Eng.* 2007;4(2):R32.
27. Lotte F, Congedo M, Lécuyer A, Lamarche F, Arnaldi B. A review of classification algorithms for EEG-based brain–computer interfaces. *J Neural Eng.* 2007;4(2):R1.
28. Shu L, et al. A review of emotion recognition using physiological signals. *Sensors.* 2018;18(7):2074.
29. Ko BC. A brief review of facial emotion recognition based on visual information. *Sensors.* 2018;18(2):401.
30. Lane ND, et al. Bewell: a smartphone application to monitor, model and promote wellbeing. In: *5th International ICST Conference on pervasive computing technologies for healthcare*; 2011. p. 23–6.
31. Wang R, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: *Proceedings of the ACM UbiComp*; 2014. p. 3–14.
32. D’mello S, Graesser A. AutoTutor and affective autotutor: learning by talking with cognitively and emotionally intelligent computers that talk Back. *ACM Trans Interact Intell Syst.* 2013;2(4)
33. Zheng Y, Mobasher B, Burke RD. The role of emotions in context-aware recommendation. *Decis RecSys.* 2013;2013:21–8.
34. Tkalcic M, Odic A, Kosir A, Tasic J. Affective labeling in a content-based recommender system for images. *IEEE Trans. Multimed.* 2013;15(2):391–400.
35. Lee U, et al. Hooked on smartphones: an exploratory study on smartphone overuse among college students. In: *Proceedings of the 32nd annual ACM conference on human factors in computing systems*; 2014. p. 2327–36.
36. Chen X, Sykora MD, Jackson TW, Elayan S. What about mood swings: identifying depression on twitter with temporal measures of emotions. In: *Web Conf. 2018–Companion World Wide Web Conf. WWW 2018*; 2018. p. 1653–60.
37. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. *Proc Natl Acad Sci U S A.* 2013;110(15):5802–5.
38. Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. *Curr Opin Behav Sci.* 2017;18:43–9.
39. Peters UW. *Wörterbuch der Psychiatrie und medizinischen Psychologie*, vol. 4., überar; 1990.
40. Cattell RB, Scheier IH. *The meaning and measurement of neuroticism and anxiety.* Oxford, England: Ronald; 1961.
41. Carlson JG, Hatfield E. *Psychology of emotion.* In: Jovanovich HB, editor. *Psychology of emotion*; 1992.
42. Frijda NH, Mesquita B. The analysis of emotions: dimensions of variation. *What Dev Emot Dev.* 1998:273–95.
43. Reisenzein R. *Emotionen.* In: *Lehrbuch Allgemeine Psychologie.* Bern: Huber; 2005. p. 435–500.
44. Greenberg S, Safran JD. Emotional-change processes in psychotherapy. *Emot Psychopathol Psychother.* 1990:59–85.
45. Plutchik R. The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *Am Sci.* 2001;89(4):344–50.

46. P. Ekman, "Gefühle lesen : wie Sie Emotionen erkennen und richtig interpretieren.," pp. XIX, 389 S.:Ill.; 19 cm, 2010.
47. Matsumoto D, et al. Culture, emotion regulation, and adjustment. *J Pers Soc Psychol.* 2008;94(6):925–37.
48. de Leersnyder J, Mesquita B, Kim HS. Where do my emotions belong? A study of immigrants' emotional acculturation. *Personal Soc Psychol Bull.* 2011;37(4):451–63.
49. Murphy NA, Isaacowitz DM. Age effects and gaze patterns in recognising emotional expressions: an in-depth look at gaze measures and covariates. *Cogn. Emot.* 2010;24(3):436–52.
50. Izard CE. Die Emotionen des Menschen: Eine Einführung in die Grundlagen der Emotionspsychologie; 1981. p. 530.
51. Stanley R, Burrows G. Varieties and functions of human emotion. In: Payne RL, Cooper CC, editors. *Emotions at work. Theory, research and applications for management.* John Wiley & Sons; 2003. p. 3–20.
52. Russell JA. A circumplex model of affect. *J Pers Soc Psychol.* 1980;39(6):1161–78.
53. Bălan O, Moise G, Petrescu L, Moldoveanu A, Leordeanu M, Moldoveanu F. Emotion classification based on biophysical signals and machine learning techniques. *Symmetry (Basel).* 2020;12(1):1–22.
54. Scherer KR. Component models of emotion can inform the quest for emotional competence. *Sci Emot Intell Knowns Unknowns.* 2007:101–26.
55. Mees U. Zum Forschungsstand der Emotionspsychologie—eine Skizze. *Emot und Sozialtheorie Disziplinäre Ansätze.* 2006:104–24.
56. Schachter S, Singer J. Cognitive, social, and physiological determinants of emotional state. *Psychol Rev.* 1962;69(5):379.
57. Lazarus RS. From psychological stress to the emotions.Pdf. *Annu Rev Psychol.* 1993;44:1–21.
58. Lewis MD. Bridging emotion theory and neurobiology through dynamic systems modeling. *Behav Brain Sci.* 2005;28(2):169–94.
59. Petta P, Trappl R. Emotions and agents; 2001. p. 301–16.
60. Averill JR. Anger and aggression: an essay on emotion, vol. 8, no. 2. New York: Springer; 1982.
61. Sánchez-Álvarez N, Extremera N, Fernández-Berrocal P. The relation between emotional intelligence and subjective Well-being: a meta-analytic investigation. *J Posit Psychol.* 2016;11(3):276–85.
62. Campbell-Sills L, Barlow DH, Brown TA, Hofmann SG. Effects of suppression and acceptance on emotional responses of individuals with anxiety and mood disorders. *Behav Res Ther.* 2006;44(9):1251–63.
63. Aldao A, Nolen-Hoeksema S, Schweizer S. Emotion-regulation strategies across psychopathology: a meta-analytic review. *Clin Psychol Rev.* 2010;30(2):217–37.
64. Elizabeth KG, Watson D. Measuring and assessing emotion at work. In: Payne Roy L, Cooper CL, editors. *Emotions at work. Theory, research and applications for management.* John Wiley & Sons; 2003. p. 21–44.
65. Jonkisz E, Moosbrugger H, Brandt H. Planung und Entwicklung von psychologischen Tests und Fragebogen; 2008. p. 27–72.
66. Bradburn NM. The structure of psychological Well-being. Chicago: Aldine; 1969.
67. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol.* 1988;54(6):1063.
68. Moriwaki SY. The affect balance scale: a validity study with aged samples. *J Gerontol.* 1974;29(1):73–8.
69. Taylor JA. A personality scale of manifest anxiety. *J Abnorm Soc Psychol.* 1953;48(2):285–90.
70. W. W. K. Zung, "A self-rating depression scale," *Arch Gen Psychiatry*, vol. 12, no. 1, pp. 63–70, Jan. 1965.
71. "An inventory for measuring clinical anxiety: psychometric properties."
72. Antony MM, Cox BJ, Enns MW, Bieling PJ, Swinson RP. Psychometric properties of the 42-item and 21-item versions of the depression anxiety stress scales in clinical groups and a community sample. *Psychol Assess.* 1998;10(2):176–81.
73. Spielberger CD. *State-trait anxiety inventory: bibliography.* 2nd ed. Palo Alto, CA: Consulting Psychologists Press; 1989.

74. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med.* 2006;166(10):1092–7.
75. Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol.* 1959;32(1):50–5.
76. Antonosky A. The structure and properties of the sense of coherence scale. *Soc Sci Med.* 1993;36:725–33.
77. Bradley MM, Lang PJ. Measuring emotion: the self-assessment manikin and the semantic differential. *J Behav Ther Exp Psychiatry.* 1994;25(1):49–59.
78. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16(9):606–13.
79. Zigmon AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* 1983;67:367–70.
80. A. T. Beck, R. A. Steer, and G. Brown, Beck depression inventory–II, Database. 1996.
81. “Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults.”
82. Carroll BJ, Feinberg M, Smouse PE, Rawson SG, Greden JF. The Carroll rating scale for depression. I. Development, reliability and validation. *Br J Psychiatry.* 1981;138:194–200.
83. Schlegel K, Mortillaro M. The Geneva emotional competence test (GECO): an ability measure of workplace emotional intelligence. *J Appl Psychol.* 2019;104(4):559–80.
84. Schlegel K, Scherer KR. Introducing a short version of the Geneva emotion recognition test (GERT-S): psychometric properties and construct validation. *Behav Res Methods.* 2016;48(4):1383–92.
85. Schlegel K, Grandjean D, Scherer KR. Introducing the Geneva emotion recognition test: an example of Rasch-based test development. *Psychol Assess.* 2014;26(2):666–72.
86. Treynor W, Gonzalez R, Nolen-Hoeksema S. Rumination reconsidered: a psychometric analysis. *Cognit Ther Res.* 2003;27(3):247–59.
87. Cohen S, Mermelstein R, Kamarck T. A global measure of perceived stress. *J Health Soc Behav.* 2016;24(4):385–96.
88. Goldberg DP, Hillier VF. A scaled version of the general health questionnaire. *Psychol Med.* 1979;9:139–45.
89. Stewart-Brown S, et al. The Warwick-Edinburgh mental Well-being scale (WEMWBS): a valid and reliable tool for measuring mental Well-being in diverse populations and projects. *J Epidemiol Community Heal.* 2011;65(Suppl 2):A38–9.
90. Herdman M, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res.* 2011;20(10):1727–36.
91. Hopko DR, et al. Assessing worry in older adults: confirmatory factor analysis of the Penn State worry questionnaire and psychometric properties of an abbreviated model. *Psychol Assess.* 2003;15(2):173–83.
92. Diener E, et al. New Well-being measures: short scales to assess flourishing and positive and negative feelings. *Soc Indic Res.* 2010;97(2):143–56.
93. American Psychiatric Association, Diagnostic and Statistical Manual of. 4th ed. Washington DC; 2013.
94. World Health Organization, ICD-10, the ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. Geneva, 1993.
95. Wittchen H-U, Zaudig M, Fydrich TH. SKID–Strukturiertes Klinisches Interview für DSM-IV. Achse I und II Handanweisungen [Structured Clinical Interview for DSM-IV]. Göttingen: Hogrefe; 1997.
96. Schmier J, Halpern MT. Patient recall and recall bias of health state and health status. *Expert Rev Pharmacoeconomics Outcomes Res.* 2004;4(2):159–63.
97. Walter SD. Recall bias in epidemiologic studies. *J Clin Epidemiol.* 1990;43(12):1431–2.
98. Stone AA, Schneider S, Harter JK. Day-of-week mood patterns in the United States: on the existence of ‘blue Monday’, ‘thank god it’s Friday’ and weekend effects. *J Posit Psychol.* 2012;7(4):306–14.
99. Ryan RM, Bernstein JH, Brown KW. Weekends, work, and Well-being: psychological need satisfactions and day of the week effects on mood, vitality, and physical symptoms. *J Soc Clin Psychol.* 2010;29(1):95–122.

100. Rosenman R, Tennekoon V, Hill LG. Measuring bias in self-reported data. *Int J Behav Healthc Res.* 2011;2(4):320.
101. Berinsky AJ. Can we talk? Self-presentation and the survey response. *Polit Psychol.* 2004;25(4):643–59.
102. Fullana MA, et al. Diagnostic biomarkers for obsessive-compulsive disorder: a reasonable quest or ignis fatuus? *Neurosci Biobehav Rev.* 2020;118:504–13.
103. Galatzer-Levy IR, Ma S, Statnikov A, Yehuda R, Shalev AY. Utilization of machine learning for prediction of post-traumatic stress: a re-examination of cortisol in the prediction and pathways to non-remitting PTSD. *Transl Psychiatry.* 2017;7(3)
104. Choppin A. EEG-based human Interface for disabled individuals : emotion expression with neural networks submitted for the master degree. *Emotion.* 2000;
105. Mehmood RM, Lee HJ. A novel feature extraction method based on late positive potential for emotion recognition in human brain signal patterns. *Comput Electr Eng.* 2016;53:444–57.
106. Li L, Chen JH. Emotion recognition using physiological signals from multiple subjects. In: *Proc.–2006 Int. Conf. Intell. Inf. Hiding multimed. Signal process. IIH-MSP 2006;* 2006. p. 355–8.
107. Uyl DMJ, Kuilenburg VH. The FaceReader: online facial expression recognition TL–30. *FaceReader Online facial Expr. Recognit.* 2005;30 VN-r(September):589–90.
108. Jang JR. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern.* 1993;23(3):665–85.
109. Rani P, Liu C, Sarkar N, Vanman E. An empirical study of machine learning techniques for affect recognition in human-robot interaction. *Pattern Anal Appl.* 2006;9(1):58–69.
110. Najj M, Firoozabadi M, Azadfallah P. Classification of music-induced emotions based on information fusion of forehead biosignals and electrocardiogram. *Cognit Comput.* 2014;6(2):241–52.
111. Khodayari-Rostamabad A, Hasey GM, MacCrimmon DJ, Reilly JP, de Bruin H. A pilot study to determine whether machine learning methodologies using pre-treatment electroencephalography can predict the symptomatic response to clozapine therapy. *Clin Neurophysiol.* 2010;121(12):1998–2006.
112. Khodayari-Rostamabad A, Reilly JP, Hasey GM, de Bruin H, MacCrimmon DJ. A machine learning approach using EEG data to predict response to SSRI treatment for major depressive disorder. *Clin Neurophysiol.* 2013;124(10):1975–85.
113. Robinson S, Hoheisel B, Windischberger C, Habel U, Lanzenberger R, Moser E. FMRI of the emotions: towards an improved understanding of amygdala function. *Curr Med Imaging Rev.* 2005;1(2):115–29.
114. Zhang W, et al. Discriminating stress from rest based on resting-state connectivity of the human brain: a supervised machine learning study. *Hum Brain Mapp.* 2020;41(11):3089–99.
115. Kolodyazhnyi V, Kreibig SD, Gross JJ, Roth WT, Wilhelm FH. An affective computing approach to physiological emotion specificity: toward subject-independent and stimulus-independent classification of film-induced emotions. *Psychophysiology.* 2011;48(7):908–22.
116. Wac K, Tsiourti C. Ambulatory assessment of affect: survey of sensor systems for monitoring of autonomic nervous systems activation in emotion. *IEEE Trans Affect Comput.* 2014;5(3):251–72.
117. Goshvarpour A, Abbasi A, Goshvarpour A. An accurate emotion recognition system using ECG and GSR signals and matching pursuit method. *Biom J.* 2017;40(6):355–68.
118. Hart B, Struiksmma ME, van Boxtel A, van Berkum JJA. Emotion in stories: facial EMG evidence for both mental simulation and moral evaluation. *Front Psychol.* 9(APR):2018.
119. Das P, Khasnobish A, Tibarewala DN. Emotion recognition employing ECG and GSR signals as markers of ANS. In: *2016 Conference on Advances in Signal Processing (CASP);* 2016. p. 37–42.
120. Liu W, Zheng W-L, Lu B-L. Emotion recognition using multimodal deep learning. In: *International conference on neural information processing;* 2016. p. 521–9.

121. Wang Y, Mo J. Emotion feature selection from physiological signals using tabu search. In: 2013 25th Chinese Control and Decision Conference (CCDC); 2013. p. 3148–50.
122. Wen W, Liu G, Cheng N, Wei J, Shangguan P, Huang W. Emotion recognition based on multi-variant correlation of physiological signals. *IEEE Trans Affect Comput*. 2014;5(2):126–40.
123. Martinez HP, Bengio Y, Yannakakis GN. Learning deep physiological models of affect. *IEEE Comput Intell Mag*. 2013;8(2):20–33.
124. Qiao R, Qing C, Zhang T, Xing X, Xu X. A novel deep-learning based framework for multi-subject emotion recognition. In: 2017 4th International Conference on Information, Cybernetics and Computational Social Systems (ICSS); 2017. p. 181–5.
125. Salari S, Ansarian A, Atrianfar H. Robust emotion classification using neural network models. In: 2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS); 2018. p. 190–4.
126. Song T, Zheng W, Song P, Cui Z. EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Trans Affect Comput*. 2018;
127. Zheng W-L, Zhu J-Y, Peng Y, Lu B-L. EEG-based emotion classification using deep belief networks. In: 2014 IEEE International Conference on Multimedia and Expo (ICME); 2014. p. 1–6.
128. Huang J, Xu X, Zhang T. Emotion classification using deep neural networks and emotional patches. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2017. p. 958–62.
129. Kawde P, Verma GK. Deep belief network based affect recognition from physiological signals. In: 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON); 2017. p. 587–92.
130. Li X, Song D, Zhang P, Yu G, Hou Y, Hu B. Emotion recognition from multi-channel EEG data through convolutional recurrent neural network. In: 2016 IEEE international conference on bioinformatics and biomedicine (BIBM), vol. 2016. p. 352–9.
131. Alhagry S, Fahmy AA, El-Khoribi RA. Emotion recognition based on EEG using LSTM recurrent neural network. *Emotion*. 2017;8(10):355–8.
132. Liu J, Su Y, Liu Y. Multi-modal emotion recognition with temporal-band attention based on lstm-rnn. In: Pacific Rim Conference on Multimedia; 2017. p. 194–204.
133. Jerritta S, Murugappan M, Wan K, Yaacob S. Emotion recognition from facial EMG signals using higher order statistics and principal component analysis. *J Chinese Inst Eng*. 2014;37(3):385–94.
134. Cheng Y, Liu G-Y, Zhang H. The research of EMG signal in emotion recognition based on TS and SBS algorithm. In: The 3rd International Conference on Information Sciences and Interaction Sciences; 2010. p. 363–6.
135. Valenza G, Lanata A, Scilingo EP. The role of nonlinear dynamics in affective valence and arousal recognition. *IEEE Trans Affect Comput*. 2011;3(2):237–49.
136. Patel R, Janawadkar MP, Sengottuvel S, Gireesan K, Radhakrishnan TS. Suppression of eye-blink associated artifact using single channel EEG data by combining cross-correlation with empirical mode decomposition. *IEEE Sensors J*. 2016;16(18):6947–54.
137. Suk M, Prabhakaran B. Real-time mobile facial expression recognition system-a case study. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2014. p. 132–7.
138. Ghimire D, Lee J. Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. *Sensors*. 2013;13(6):7714–34.
139. Happy SL, George A, Routray A. A real time facial expression classification system using local binary patterns. In: 2012 4th International conference on intelligent human computer interaction (IHCI); 2012. p. 1–5.
140. Siddiqi MH, Ali R, Khan AM, Park Y-T, Lee S. Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. *IEEE Trans Image Process*. 2015;24(4):1386–98.
141. Khan RA, Meyer A, Konik H, Bouakaz S. Framework for reliable, real-time facial expression recognition for low resolution images. *Pattern Recogn Lett*. 2013;34(10):1159–68.

142. Srinivasan V, Moghaddam S, Mukherji A, Rachuri KK, Xu C, Tapia EM. Mobileminer: mining your frequent patterns on your phone. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2014. p. 389–400.
143. LiKamWa R, Liu Y, Lane ND, Zhong L. MoodScope: building a mood sensor from smartphone usage patterns. In: Proceeding of the ACM Mobisys; 2013. p. 389–402.
144. Ghimire D, Jeong S, Lee J, Park SH. Facial expression recognition based on local region specific features and support vector machines. *Multimed Tools Appl.* 2017;76(6):7803–21.
145. Fabian Benitez-Quiroz C, Srinivasan R, Martinez AM. Emotionet: an accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 5562–70.
146. Walecki R, Pavlovic V, Schuller B, Pantic M. Deep structured learning for facial action unit intensity estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 3405–14.
147. R. Breuer and R. Kimmel, “A deep learning perspective on the origin of facial expressions.” *arXiv Prepr. arXiv1705.01842*, 2017.
148. Jung H, Lee S, Yim J, Park S, Kim J. Joint fine-tuning in deep neural networks for facial expression recognition. In: Proceedings of the IEEE international conference on computer vision; 2015. p. 2983–91.
149. Ebrahimi Kahou S, Michalski V, Konda K, Memisevic R, Pal C. Recurrent neural networks for emotion recognition in video. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction; 2015. p. 467–74.
150. Ng H-W, Nguyen VD, Vonikakis V, Winkler S. Deep learning for emotion recognition on small datasets using transfer learning. In: Proceedings of the 2015 ACM on international conference on multimodal interaction; 2015. p. 443–9.
151. Kim DH, Baddar WJ, Jang J, Ro YM. Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition. *IEEE Trans Affect Comput.* 2017;10(2):223–36.
152. Hasani B, Mahoor MH. Facial expression recognition using enhanced deep 3D convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2017. p. 30–40.
153. Graves A, Mayer C, Wimmer M, Schmidhuber J, Radig B. Facial expression recognition with recurrent neural networks. In: Proceedings of the International Workshop on Cognition for Technical Systems; 2008.
154. Donahue J, et al. Long-term recurrent convolutional networks for visual recognition and description. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 2625–34.
155. Chu W-S, la Torre F, Cohn JF. Learning spatial and temporal cues for multi-label facial action unit detection. In: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017); 2017. p. 25–32.
156. Deng J, Frühholz S, Zhang Z, Schuller B. Recognizing emotions from whispered speech based on acoustic feature transfer learning. *IEEE Access.* 2017;5:5235–46.
157. Demircan S, Kahramanli H. Feature extraction from speech data for emotion recognition. *J Adv Comput Networks.* 2014;2(1):28–30.
158. Anagnostopoulos C-N, Iliou T, Giannoukos I. Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011. *Artif Intell Rev.* 2015;43(2):155–77.
159. Dellaert F, Polzin T, Waibel A. Recognizing emotion in speech. In: Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP’96, vol. 3; 1996. p. 1970–3.
160. Zhou Y, Sun Y, Zhang J, Yan Y. Speech emotion recognition using both spectral and prosodic features. In: 2009 International Conference on Information Engineering and Computer Science; 2009. p. 1–4.
161. Haq S, Jackson PJB, Edge J. Audio-visual feature selection and reduction for emotion classification. In: Proceeding International Conference on Auditory-Visual Speech Processing (AVSP’08), Tangalooma, Australia; 2008.

162. Alpert M, Pouget ER, Silva RR. Reflections of depression in acoustic measures of the patient's speech. *J Affect Disord.* 2001;66(1):59–69.
163. Ververidis D, Kotropoulos C. Emotional speech recognition: resources, features, and methods. *Speech Commun.* 2006;48(9):1162–81.
164. Mozziconacci S. Prosody and emotions. In: *Speech prosody 2002, international conference*; 2002.
165. J. B. Hirschberg *et al.*, “Distinguishing deceptive from non-deceptive speech,” 2005.
166. Neiberg D, Elenius K, Laskowski K. Emotion recognition in spontaneous speech using GMMs. In: *Ninth international conference on spoken language processing*; 2006.
167. Dileep AD, Sekhar CC. HMM based intermediate matching kernel for classification of sequential patterns of speech using support vector machines. *IEEE Trans Audio Speech Lang Processing.* 2013;21(12):2570–82.
168. Vyas G, Dutta MK, Riha K, Prinosil J. An automatic emotion recognizer using MFCCs and hidden Markov models. In: *2015 7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*; 2015. p. 320–4.
169. Pan Y, Shen P, Shen L. Speech emotion recognition using support vector machine. *Int J Smart Home.* 2012;6(2):101–8.
170. Schuller B, Rigoll G, Lang M. Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. In: *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1; 2004. p. I–577.
171. Zhao J, Mao X, Chen L. Speech emotion recognition using deep 1D & 2D CNN LSTM networks. *Biomed Signal Process Control.* 2019;47:312–23.
172. Yenigalla P, Kumar A, Tripathi S, Singh C, Kar S, Vepa J. Speech emotion recognition using Spectrogram & Phoneme Embedding. *Interspeech.* 2018:3688–92.
173. Zhao J, Mao X, Chen L. Learning deep features to recognise speech emotion using merged deep CNN. *IET Signal Process.* 2018;12(6):713–21.
174. Zhang Y, Liu Y, Weninger F, Schuller B. Multi-task deep neural network with shared hidden layers: breaking down the wall between emotion representations. In: *2017 IEEE International Conference on acoustics, speech and signal processing (ICASSP)*; 2017. p. 4990–4.
175. Badshah AM, Ahmad J, Rahim N, Baik SW. Speech emotion recognition from spectrograms with deep convolutional neural network. In: *2017 International Conference on platform technology and service (PlatCon)*; 2017. p. 1–5.
176. Barros P, Weber C, Wermter S. Emotional expression recognition with a cross-channel convolutional neural network for human-robot interaction. In: *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*; 2015. p. 582–7.
177. Mao Q, Dong M, Huang Z, Zhan Y. Learning salient features for speech emotion recognition using convolutional neural networks. *IEEE Trans Multimed.* 2014;16(8):2203–13.
178. Lakomkin E, Zamani MA, Weber C, Magg S, Wermter S. On the robustness of speech emotion recognition for human-robot interaction with deep neural networks. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; 2018. p. 854–60.
179. Tzirakis P, Trigeorgis G, Nicolaou MA, Schuller BW, Zafeiriou S. End-to-end multimodal emotion recognition using deep neural networks. *IEEE J Sel Top Signal Process.* 2017;11(8):1301–9.
180. Lim W, Jang D, Lee T. Speech emotion recognition using convolutional and recurrent neural networks. In: *2016 Asia-Pacific signal and information processing association annual summit and conference (APSIPA)*, vol. 2016. p. 1–4.
181. S. Sahu, R. Gupta, G. Sivaraman, W. AbdAlmageed, and C. Espy-Wilson, “Adversarial auto-encoders for speech based emotion recognition,” *arXiv Prepr. arXiv1806.02146*, 2018.
182. Zhao Y, Jin X, Hu X. Recurrent convolutional neural network for speech processing. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2017. p. 5300–4.
183. Lopez LD, Reschke PJ, Knothe JM, Walle EA. Postural communication of emotion: perception of distinct poses of five discrete emotions. *Front Psychol.* 2017;8(MAY)

184. Quiroz JC, Geangu E, Yong MH. Emotion recognition using smart watch sensor data: mixed-design study. *J Med Internet Res*. 2018;20(8)
185. García-Magariño I, Cerezo E, Plaza I, Chittaro L. A mobile application to report and detect 3D body emotional poses. *Expert Syst Appl*. 2019;122:207–16.
186. Stachl C, et al. Predicting personality from patterns of behavior collected with smartphones. *Proc Natl Acad Sci U S A*. 2020;117(30):17680–7.
187. Calvo RA, Milne DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. *Nat Lang Eng*. 2017;23(5):649–85.
188. Eichstaedt JC, et al. Facebook language predicts depression in medical records. *Proc Natl Acad Sci U S A*. 2018;115(44):11203–8.
189. Dey AK, Wac K, Ferreira D, Tassini K, Hong JH, Ramos J. Getting closer: an empirical investigation of the proximity of user to their smart phones. In: *UbiComp'11-Proc. 2011 ACM Conf. Ubiquitous Comput*; 2011. p. 163–72.
190. Berrocal A, Manea V, de Masi A, Wac K. MQOL lab: step-by-step creation of a flexible platform to conduct studies using interactive, mobile, wearable and ubiquitous devices. *Procedia Comput Sci*. 2020;175:221–9.
191. Schoedel R, Oldemeier M. Basic protocol: smartphone sensing panel. Leibniz Inst für Psychol Inf und Dokumentation. 2020;
192. Harari GM, et al. Sensing sociability: individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *J Pers Soc Psychol*. 2019;
193. Ozer DJ, Benet-Martínez V. Personality and the prediction of consequential outcomes. *Annu Rev Psychol*. 2006;57:401–21.
194. Roberts BW, Kuncel NR, Shiner R, Caspi A, Goldberg LR. The power of personality: the comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspect Psychol Sci*. 2007;2(4):313–45.
195. Rachuri KK, Musolesi M, Mascolo C, Rentfrow PJ, Longworth C, Aucinas A. EmotionSense: a Mobile phones based adaptive platform for experimental social psychology research. In: *Proceedings of ACM UbiComp*; 2010.
196. Chittaranjan G, Blom J, Gatica-Perez D. Who's who with big-five: analyzing and classifying personality traits with smartphones. In: *Wearable Computers (ISWC), 2011 15th Annual International Symposium on*; 2011. p. 29–36.
197. Pielot M, Dingler T, Pedro JS, Oliver N. When attention is not scarce-detecting boredom from mobile phone usage. In: *Proceedings of the ACM UbiComp*; 2015. p. 825–36.
198. Lu H, et al. Stresssense: detecting stress in unconstrained acoustic environments using smartphones. In: *Proceedings of ACM UbiComp*; 2012.
199. Bogomolov A, Lepri B, Pianesi F. Happiness recognition from Mobile phone data. In: *Proceedings of the IEEE International Conference on Social Computing (SocialCom)*; 2013.
200. Bogomolov A, Lepri B, Ferron M, Pianesi F, Pentland A. Daily stress recognition from Mobile phone data, weather conditions and individual traits. In: *Proceedings of the 22nd ACM International Conference on Multimedia*; 2014.
201. Lee H, Choi YS, Lee S, Park IP. Towards unobtrusive emotion recognition for affective social communication. In: *IEEE Consumer Communications and Networking Conference (CCNC)*; 2012.
202. Gao Y, Bianchi-Berthouze N, Meng H. What does touch tell us about emotions in touchscreen-based gameplay? *ACM Trans Comput Hum Interact*. 2012;19(4):Dec.
203. Wac K, Ciman M, Gaggi O. iSenseStress: assessing stress through human-smartphone interaction analysis. In: *9th International Conference on Pervasive Computing Technologies for Healthcare-PervasiveHealth*; 2015. p. 8.
204. Kim H-J, Choi YS. Exploring emotional preference for smartphone applications. In: *IEEE Consumer Communications and Networking Conference (CCNC)*; 2012.
205. Hektner JM, Schmidt JA, Csikszentmihalyi M. Experience sampling method: measuring the quality of everyday life. Sage; 2007;
206. Pejovic V, Lathia N, Mascolo C, Musolesi M. Mobile-based experience sampling for behaviour research. In: *Emotions and personality in personalized services*. Springer; 2016. p. 141–61.

207. Van Berkel N, Ferreira D, Kostakos V. The experience sampling method on Mobile devices. *ACM Comput Surv.* 2017;50(6):93.
208. Hernandez J, McDuff D, Infante C, Maes P, Quigley K, Picard R. Wearable ESM: differences in the experience sampling method across wearable devices. In: *Proceedings of ACM MobileHCI*; 2016. p. 195–205.
209. Wagner DT, Rice A, Beresford AR. Device analyzer: understanding smartphone usage. In: *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*; 2013. p. 195–208.
210. Rawassizadeh R, Tomitsch M, Wac K, Tjoa AM. UbiqLog: a generic mobile phone-based life-log framework. *Pers ubiquitous Comput.* 2013;17(4):621–37.
211. Rawassizadeh R, Momeni E, Dobbins C, Gharibshah J, Pazzani M. Scalable daily human behavioral pattern mining from multivariate temporal data. *IEEE Trans Knowl Data Eng.* 2016;28(11):3098–112.
212. Ferreira D, Kostakos V, Dey AK. AWARE: mobile context instrumentation framework. *Front ICT.* 2015;2:6.
213. Nath S. ACE: exploiting correlation for energy-efficient and continuous context sensing. In: *Proceedings of the 10th international conference on Mobile systems, applications, and services*; 2012. p. 29–42.
214. Consolvo S, Walker M. Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Comput.* 2003;2(2):24–31.
215. Ghosh S, Ganguly N, Mitra B, De P. Towards designing an intelligent experience sampling method for emotion detection. In: *Proceedings of the IEEE CCNC*; 2017.
216. Barrett LF, Barrett DJ. An introduction to computerized experience sampling in psychology. *Soc Sci Comput Rev.* 2001;19(2):175–85.
217. Froehlich J, Chen MY, Consolvo S, Harrison B, Landay JA. MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones. In: *Proceedings of the 5th Mobisys*; 2007.
218. Gaggioli A, et al. A mobile data collection platform for mental health research. *Pers Ubiquitous Comput.* 2013;17(2):241–51.
219. “Personal Analytics Companion.”
220. Sahami Shirazi A, Henze N, Dingler T, Pielot M, Weber D, Schmidt A. Large-scale assessment of mobile notifications. In: *Proceedings of the ACM SIGCHI*; 2014. p. 3055–64.
221. Fischer JE, Greenhalgh C, Benford S. Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications. In: *Proceedings of ACM MobileHCI*; 2011. p. 181–90.
222. Ho J, Intille SS. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In: *Proceedings of ACM SIGCHI*; 2005. p. 909–18.
223. Pielot M, de Oliveira R, Kwak H, Oliver N. Didn't you see my message?: predicting attentiveness to mobile instant messages. In: *Proceedings of the ACM SIGCHI*; 2014. p. 3319–28.
224. Kushlev K, Cardoso B, Pielot M. Too tense for candy crush: affect influences user engagement with proactively suggested content. In: *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17)*. New York, NY, USA: ACM; 2017.
225. Weber D, Voit A, Kratzer P, Henze N. In-situ investigation of notifications in multi-device environments. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*; 2016. p. 1259–64.
226. Turner LD, Allen SM, Whitaker RM. Push or delay? Decomposing smartphone notification response behaviour. In: *Human behavior understanding: 6th international workshop, HBU*, vol. 2015; 2015.

227. Gerber N, Gerber P, Volkamer M. Explaining the privacy paradox: a systematic review of literature investigating privacy attitude and behavior. *Comput Secur.* 2018;77:226–61.
228. T. Dienlin, “Das privacy paradox aus psychologischer Perspektive.,” 2019.
229. “No Title.”
230. Mauss IB, Robinson MD. Measures of emotion: a review. *Cogn. Emot.* 2009;23(2):209–37.
231. Verduyn P, Lavrijsen S. Which emotions last longest and why: the role of event importance and rumination. *Motiv Emot.* 2015;39(1):119–27.
232. Cimam M, Wac K. Individuals’ stress assessment using human-smartphone interaction analysis. *IEEE Trans Affect Comput.* 2018;9(1):51–65.
233. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *SIGKDD Explor Newsl.* 2009;11(1):10–8.
234. Ghosh S, Ganguly N, Mitra B, De P. Designing an experience sampling method for smartphone based emotion detection. *IEEE Trans Affect Comput.* 2019:1–1.
235. Tarasov A, Delany SJ, Cullen C. Using crowdsourcing for labelling emotional speech assets. *Proc W3C Work Emot Markup Lang.* 2010;
236. Zhu X, Wu X. Class noise vs. attribute noise: A quantitative study. *Artif Intell Rev.* 2004;22(3):177–210.
237. Frénay B, Verleysen M. Classification in the presence of label noise: a survey. *IEEE Trans neural networks Learn Syst.* 2014;25(5):845–69.
238. Manea V, Wac K. Co-calibrating physical and psychological outcomes and consumer wearable activity outcomes in older adults: an evaluation of the coqol method. *J Pers Med.* 2020;10(4):1–86.
239. Vidal Bustamante CM, Rodman AM, Dennison MJ, Flournoy JC, Mair P, McLaughlin KA. Within-person fluctuations in stressful life events, sleep, and anxiety and depression symptoms during adolescence: a multiwave prospective study. *J Child Psychol Psychiatry Allied Discip.* 2020;61(10):1116–25.
240. Wac K. From quantified self to quality of life; 2018. p. 83–108.
241. Wac K. Quality of life technologies. *Encycl Behav Med.* 2020:1–2.
242. Wac K, Fiordelli M, Gustarini M, Rivas H. Quality of life technologies: experiences from the field and key challenges. *IEEE Internet Comput.* 2015;19(4):28–35.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

