INVITED PAPER

Approaches, Applications, and Challenges in Physiological Emotion Recognition—A Tutorial Overview

By Yekta Said Can⬤, Bhargavi Mahesh⬤, and Elisabeth André⬤, *Senior Member IEEE*

ABSTRACT | An automatic emotion recognition system can serve as a fundamental framework for various applications in daily life from monitoring emotional well-being to improving the quality of life through better emotion regulation. Understanding the process of emotion manifestation becomes crucial for building emotion recognition systems. An emotional experience results in changes not only in interpersonal behavior but also in physiological responses. Physiological signals are one of the most reliable means for recognizing emotions since individuals cannot consciously manipulate them for a long duration. These signals can be captured by medical-grade wearable devices, as well as commercial smart watches and smart bands. With the shift in research direction from laboratory to unrestricted daily life, commercial devices have been employed ubiquitously. However, this shift has introduced several challenges, such as low data quality, dependency on subjective self-reports, unlimited movement-related changes, and artifacts in physiological signals. This tutorial provides an overview of practical aspects of emotion recognition, such as experiment design, properties of different physiological modalities, existing datasets, suitable machine learning algorithms for physiological data, and several applications. It aims to provide the necessary psychological and physiological backgrounds through various emotion theories and the physiological manifestation of emotions, thereby laying a foundation for emotion recognition. Finally, the tutorial discusses open research directions and possible solutions.

I. INTRODUCTION

Emotions serve a significant role in human lives as they assist in decision-making and forging social relationships. The short-lasting emotional responses distinguish themselves from affective states, such as mood or stress. However, enduring negative emotions may have severe effects if they are not managed well early. They may inhibit learning among students [1], lead to burnout among workers [2], and eventually lead to mental health disorders, such as anxiety- and mood-related disorders, schizophrenia, and substance abuse [3].

Automatic recognition of emotions (specifically negative emotions, such as sadness, anxiety, fatigue, and anger) can contribute significantly to a prescreening tool to prevent adverse health consequences. Suppose that one is driving a car for a long distance under time pressure and

cannot afford to rest sufficiently. This condition reduces one's attention on the road and makes one vulnerable to mistakes and accidents. The U.S. Department of Transportation claims that driving-related errors cause around 95% of fatal road accidents [4]. A huge proportion of these driving errors are caused by drowsiness or fatigue. However, an intelligent emotion detection system in a car, which can continuously monitor and detect fatigue or drowsiness using our physiological signals, could save lives by preventing accidents. Sending personalized alerts to the driver ahead of time to pause for a coffee break and change the music tempo or ambient temperature could ensure a safer and more comfortable driving experience.

The advancement in sensing technologies has enabled computer scientists to develop automatic emotion recognition tools. Facial expressions [5] and speech [6] are adopted for emotion recognition due to the ease of associating typical facial expressions and speech with emotions. Physiology-based solutions have emerged as another alternative for emotion recognition research due to their suitability for continuous monitoring in everyday life and relatively fewer privacy issues. Wearable devices have emerged as pervasive instruments for passive quantitative data collection. More than 330 million smart watches, fitness trackers, and similar wearable devices have been sold, and the market has been growing each year [7]. Most wearable devices can capture physiological, environmental, and activity-related information without interfering with the user's activities, making them a promising candidate for emotion recognition, especially in daily life.

Numerous emotion recognition studies in laboratory environments have been conducted over the past decade [8], [9], and several public datasets are created in these settings. However, the focus of research has recently shifted from laboratory to daily life [10] since emotion recognition in the laboratory differs significantly from daily life in terms of emotional stimulus characteristics, responses, and labeling [11], [12]. People can differentiate between the artificial stimuli induced in the laboratory and daily emotional stimuli that matter to them and react accordingly [11]. Researchers proposed several emotion recognition techniques and tested them in the wild [13], [14], [15]. Shu et al. [16] describe a framework for emotion recognition using physiological signals and emphasize that emotion recognition in the wild faces several challenges apart from the stimuli itself, such as emotion labeling and intersubject variability. Saganowski et al. [12] systematically reviewed the literature on wearable devices for emotion recognition in daily life and noted that most studies involved laboratory data. The approaches developed in a laboratory setting do not have sufficient robustness to be employed in a real-time monitoring system. Precise and robust emotion recognition in daily life is crucial for developing emotion-aware systems (i.e., personal agents or robots) that employ the user's emotions as feedback to adapt its behavior. It can be used to find personalized emotion regulation strategies,

teach emotional responses to people with certain conditions, such as autism, monitor enduring negative emotions, report emotions to physicians and psychologists through a prescreening tool, track workers in dangerous lines of work, and notify authorities in the case of accumulated fatigue, anxiety, or stress, thereby decreasing work accidents.

An ideal emotion recognition cycle in the wild comprises emotion recognition and regulation components (see Fig. 1 for details). This tutorial provides the necessary background and guidelines for developing such an emotion recognition system. Section II briefly describes the evolution of theories on how emotions are caused, represented, and regulated. Section III describes the physiological correlates of emotions and presents empirical evidence for emotion manifestation through physiological changes in the body. Section IV describes the physiological signals, devices to obtain them, and discriminatory features. Section V presents the guidelines for designing and implementing scientific experiments for data collection, along with prominent public datasets. Section VI describes state-of-the-art machine learning and deep learning techniques appropriate for physiological time-series data. This article concludes with open research issues, insights, and recommendations for recognizing emotions in the wild.

II. BACKGROUND

Several psychophysiologists proposed emotion theories to model the elicitation of an emotional experience. Although most of the emotion elicitation theories bear similarities in the psychological and physiological elements that constitute an emotional experience, they differ in the occurrence order of these elements or the depth of description of the underlying process. Nevertheless, these theories laid the foundation for emotion representation and regulation. Emotion representation frameworks consist of single or multidimensional spaces where several emotions are arranged. Such frameworks for emotion representation promote emotion modeling—detection and recognition. This section concludes with the emotion regulation theory that describes the potential strategies at various stages of an emotional experience where individuals can regulate their emotions.

A. Emotion Elicitation

Emotion refers to a change in the mental state arising from a complex interaction between a stimulus in the external environment and the internal state of an individual. Although details of the emotion definition have been controversial, the theories around emotion elicitation have converged on the crucial components of emotion, which define the characteristics of an emotional response. These theories, however, differ based on the process details of an emotional experience. James [18], [19] proposed that the physiological response precedes an emotional experience. According to this theory, an emotional stimulus
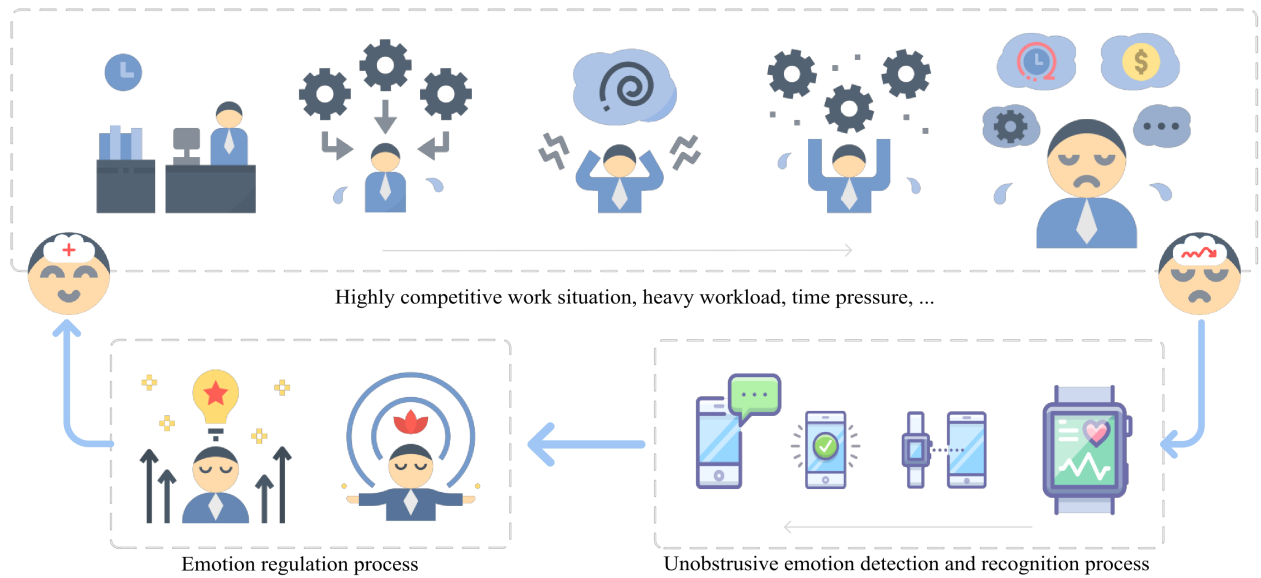
Fig. 1. *Ideal emotion recognition system for daily life is shown. It should continuously monitor the signals, and if it detects negative emotions, it should suggest appropriate relaxation methods (emotion regulation support scheme) to return individuals to their baseline state [17].*

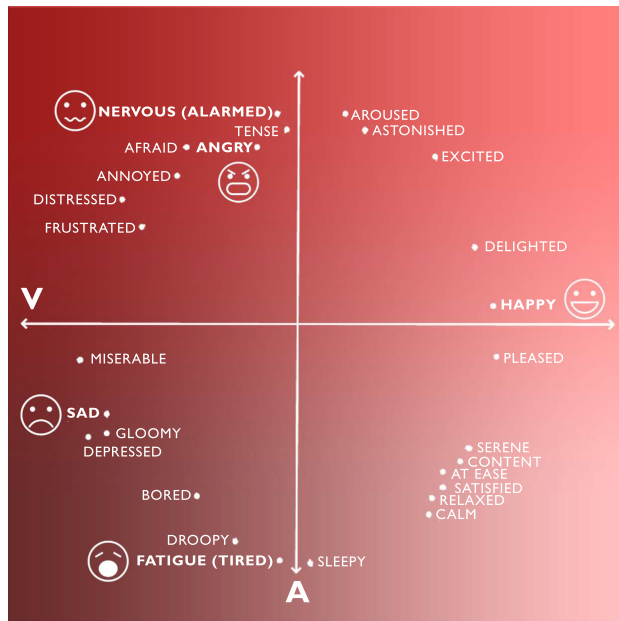activates the sensory cortex, thereby eliciting peripheral responses. The feedback from the peripheral responses then triggered an emotional response. This theory emphasizes that the stimulus elicits a specific response pattern that governs emotion quality. However, the theory does not explain how physiological responses are initiated. Later, Cannon [20] argued against the specificity of physiological responses for a given emotion and claimed that different emotions could elicit an undifferentiated physiological (autonomic) response, such as an increased heart rate (HR). Following Cannon's empirical understanding, Schachter [21] proposed that different stimuli are likely to produce similar physiological arousal, but a specific emotional experience is produced by the cognitive process of consciously attributing the arousal to characteristics of the stimulus. Therefore, according to this theory, attributing arousal to different characteristics of the eliciting stimulus produces different emotions. Several researchers, including Arnold [22], Scherer [23], and Lazarus [24], argued against the conscious attribution of the physiological arousal and instead claimed that the cognitive appraisal of stimulus or the situation with respect to the individual's goals is likely to occur unconsciously, and it precedes the physiological arousal. This concept of appraisal gave rise to appraisal theories of emotion. While the first level of appraisal focused on the situation itself, Lazarus furthered his theory by adding the concept of coping or secondary appraisal of a potentially dangerous situation by individuals based on their capabilities. Roseman et al. [25] suggested that the subjective evaluation of the situation concerning an individual's goals and accountability also influences emotions, thereby making an appraisal individualistic. The component process model proposed by Scherer [23] suggested that a cognitive appraisal

involves a sequence of appraisal checks, the response to which varies over different personalities and cultures. Emotion researchers have proposed several such variables influencing the appraisal. Meanwhile, Ekman et al. [26] challenged the theory of undifferentiated physiological responses using empirical studies and showed that autonomic responses are specific to emotions. Although there have been several propositions regarding emotion elicitation throughout history, the class of appraisal theories is preferred the most. Fig. 2 depicts the main components and the sequence of occurrence of these components based on the appraisal theory of emotions. The research concludes that there is an endless and inconsistent list of components leading to an emotional experience in an individual. However, it is worthwhile to consider various factors that influence the subjectivity of the cognitive appraisal.
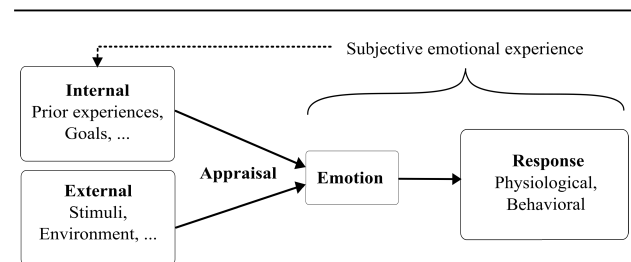
Fig. 2. *Sequence of emotion elicitation based on the appraisal. Cognitive appraisal of the external situation and the internal state of the individual result in emotions that further trigger various physiological and behavioral responses. The changes in the mental and physiological states of an individual constitute an emotional experience.*

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

Fig. 3. *Russell's circumplex model of affect [31] depicting emotions on a 2-D space. V stands for valence, and A stands for arousal. Can and Ersoy [35] selected the five highlighted emotions for their study. The figure becomes lighter when the valence is more positive. When the arousal increases, the red color becomes more evident (similar to an alarm).*

B. Emotion Representation

Emotions have been represented as discrete or categorical emotion states and in continuous or dimensional emotion space. Over the past decades, different sets of primary emotions have been categorized by emotion researchers. James [19] identified emotion categories, such as fear, grief, love, and rage as coarse emotions as they involve strong physiological changes. Ekman [27] proposed a finite set of emotions having distinctive physiological signatures and universal signals, and called them basic emotions, having common features, such as rapid onset, short duration, automatic appraisal, and coherent responses. The basic emotions constitute anger, fear, sadness, enjoyment, disgust, and surprise. However, the universality in the definition of basic emotions limited the representation of the complexity of the emotion generation process among different individuals. Researchers extended the list of basic emotions to 15 [28], 17 [29], and 27 emotions [30]. However, the similarity among the emotions could not be gauged with such emotion categories though they could be broadly classified into positive and negative emotions. Conversely, the dimensional model represents emotions in a continuous multidimensional space that denotes a systematic relationship between different emotions. A prominent example is the Circumplex model proposed by Russell [31], which is defined by two orthogonal dimensions: valence and arousal. The two dimensions depicted the subjective experience and the extent of physiological activity. An example of its application is presented in Fig. 3. Another commonly adopted model is a 3-D space
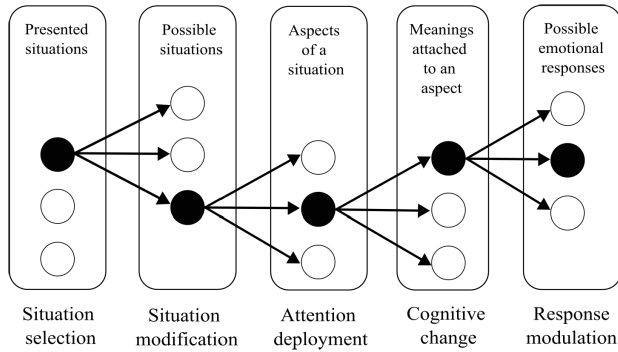
defined by the scales: pleasure, arousal, and dominance (PAD) dimensions [32]. The PAD model was based on the premise that emotions are the foundations for cognitive judgments. Plutchik [33] used a hybrid approach with the emotion wheel made of eight primary emotions represented with different colors and intensities on a polar coordinate system, thereby establishing the spatial relationship between them. The intensity of the emotions is proportional to the color intensity, and opposite emotions are placed diagonally opposite in the wheel. Meanwhile, the mixtures of primary emotions are presented in the spaces in the outermost layer. Prior work [30] has revealed that emotions represented as categories are better at capturing subjective experiences through self-reports than the commonly used dimensions such as valence and arousal. Scherer [23] identified a robust alternative to categorical emotions with eight cognitive dimensions leading to a cognitive appraisal—novelty, pleasantness, fairness of the situation, and the individual's perception of goal, coping ability, accountability, morality, and self-consistency. Each basic emotion was found to exhibit a specific appraisal profile along the eight dimensions, and these appraisal profiles have been used to distinguish different emotion categories qualitatively [34].

C. Emotion Regulation

Emotions are helpful when they enhance our decision-making and motivate socially appropriate behaviors. Nevertheless, they could also be unhelpful when they are inappropriate for a given situation or are of inappropriate intensity, higher frequency, and longer duration. Emotion regulation is required when these unhelpful emotions lead to collateral damage or harm to oneself or others. Emotion recognition systems have the potential to assist in emotion regulation.

One needs to assume a positive goal in order to regulate emotions. Such a goal could be to feel less sad or to lead a healthy lifestyle. Emotion regulation could be intrinsic, where an individual regulates one's own emotions, such as encouraging oneself after a job rejection, or extrinsic, where an individual regulates another person's emotions, such as a parent consoling a child. Individuals have different strategies to regulate emotions, and not all strategies work. Hence, one must find the emotion regulation strategy that works for them. Gross [36] proposed the process model of emotion regulation, which is a framework for identifying emotion regulation strategies at several steps involved in emotion generation. The steps involved in emotion generation and the regulation strategies are depicted on a time axis in Fig. 4. Each step presents a potential opportunity for regulation. The first strategy is situation selection, where an individual can choose the situation that will have the least negative emotional impact on the future. This strategy is also used in cognitive behavioral therapy, where the interventions increase a person's exposure to positive state-inducing activities. However,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

Fig. 4. *Steps involved in the emotion regulation process [36]. An individual can regulate their emotions at various stages through their choices. They can select situations that may have a lesser negative emotional impact or modify the existing situation to avoid negative emotions. In addition, they can choose which aspects of a situation to focus on and which meaning to attach to an aspect. Once the emotion is generated, they can choose to modulate the responses by suppressing or expressing them differently.*

interventions for situation selection are challenging since it is hard to gauge one's intrinsic feelings about different situations, mainly when driven by an impact bias. Another strategy is situation modification, where an individual can physically alter an existing situation, such as moving away from a negative emotion-eliciting scene, person, or object. In addition, one can choose to focus on a certain favorable aspect of the given situation. This strategy is known as attention deployment. When facing a situation and a particular aspect that elicit negative emotions, one can choose to attach a meaning to that aspect that may elicit more positive emotions. This strategy is known as cognitive change, and one way to achieve it is through cognitive reappraisal. Once the emotion has been evoked, one can modulate one or more of the behavioral, experiential, and physiological response tendencies, such as using physical exercise as an intervention. While adaptive forms of emotion regulation are vital for the successful functioning of humans in daily life, the autonomic and behavioral responses due to regulation may overlap with those of emotion expression. Hence, it necessitates the consideration of emotion regulation while detecting emotions. For example, studies have shown that emotion regulation strategies, such as suppression through facial expressions, result in decreased facial activity [37] but an increase in sympathetic nervous system (SNS) activity, such as increased blood pressure [38]. However, the self-reported subjective experiences remained unchanged. In contrast, the regulation strategy of cognitive reappraisal decreased HR and corrugator muscle activity [39]. Therefore, understanding the impact of various regulation strategies potentially aids better emotion recognition, provided that such interference in physiological responses to emotions is carefully modeled.

Emotion elicitation and regulation theories provide interrelated components that explain or predict characteristics of human emotions and corresponding behavior by specifying relations among different modalities [40]. Theories also help researchers accurately hypothesize and model the antecedents and consequences of emotion. While emotion theories help explain a model's predictions, they also assist in reasoning the variance in the predictions. Hence, theories act as a foundation for prediction models.

III. PHYSIOLOGICAL CORRELATES OF EMOTIONS

An emotional experience constitutes changes in the psychological and physiological states in response to a stimulus. Early studies reported specific physiological and behavioral patterns for different emotions [26]. Later, the studies investigated how the human brain, which hosts the emotion-processing center of the human body and also regulates the organs of the human body that it innervates. The human nervous system mainly comprises the central and peripheral nervous systems. The central nervous system includes the brain, its stem, and the spinal cord, whereas the peripheral nervous system includes the network of nerves passing through different types of muscles. The peripheral system is further divided into autonomic and somatic nervous systems. These two systems play a primary role in regulating the physiological and behavioral responses to emotions. The autonomic nervous system (ANS) is further divided into two branches: the SNS, responsible for stimulative functions, and the parasympathetic nervous system (PNS), responsible for restorative functions in the body. The ANS traverses the end effectors, including smooth cardiac muscles and glands that are predominantly involuntary. The somatic nervous system makes up the nerves in the skeletal muscles that are often voluntary. The following paragraphs describe, with the help of previous research, the manifestation of emotions through different physiological responses in coordination with ANS.

A. Brain Responses

The brain, along with serving several necessary functions in our daily lives, plays a significant role in emotion expression and regulation. Various regions of the brain are involved in emotion processing, such as the amygdala, prefrontal cortex, insula, and cingulate cortex. Amygdala is known to be involved in a negative emotional response. When the emotion processing regions are active, several neurons located in the cerebral cortex communicate by generating electric potentials synchronously. This neural activity collectively results in electric activity, which can be measured by placing electrodes on the scalp. Studies show that individuals exhibited relatively higher neural activity in the left prefrontal cortex for positive emotions and higher right prefrontal cortex activity for negative emotions [41], [42]. The neural activity is measured in terms of the asymmetry in the neural activation of the left and right prefrontal cortices [43]. Furthermore, such

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

neural activity in active regions of the brain demands more oxygen and nutrients, and this results in increased blood flow to that region.

B. Cardiac Responses

The cardiovascular system comprises the heart and blood vessels. The heart is responsible for pumping blood to all parts of the body. It has specialized muscle cells that generate electrical impulses that initiate the heart contractions or heartbeat. Activation of SNS due to negative emotions or various stress stimuli results in the release of substances called neurotransmitters that bind to the cardiac muscles stimulating an increase in the heartbeat rate or HR, whereas activation of PNS due to relaxation or positive emotions results in a decrease in the rate and force of heart contractions. Any change in contractions is associated with a change in the electrical activity of the heart. Since the cardiovascular system is dually innervated, i.e., simultaneously controlled by both sympathetic and parasympathetic branches of ANS, the end response measurements will not reveal the activity of individual branches due to the reciprocal control due to dual innervation. For instance, an increase in HR could be influenced by increased activity of the sympathetic branch or decreased activity of the parasympathetic branch, or a combination of both where either of the activities dominates [44]. Hence, specific features, such as HR variability (HRV), should be considered to differentiate the two types of activation.

C. Skin Responses

The outer layer of the skin is capable of conducting electricity but offers a certain level of resistance. The middle, dermis layer of our skin, comprises the blood vessels and sweat glands. Sweat glands, mainly innervated by the SNS, produce moisture to facilitate grasping during the fight–flight reaction. When SNS is activated due to emotional stimuli, the emitted neurotransmitters induce changes in the resistance (or conductance) of the skin. According to secretion theory [45], the changes in skin conductance are triggered by sweat gland activity. Furthermore, due to the evaporation of sweat, the skin temperature (ST) reduces as well.

D. Muscle Responses

ANS activity elicited by emotions can lead to changes in muscle activity, both voluntary and involuntary. Involuntary muscle movements include tensing up of shoulders and twitching due to the activation of the SNS [46]. Voluntary muscle movements may include facial expressions. Even though the boundary between involuntary and voluntary muscle movements is not always clear, both types of movements can generate electrical activity in the muscles. The skeletal muscle fibers that make up the muscle tissue are innervated by the motor neurons. Motor neurons are a part of the somatic nervous system that send and receive muscle contraction information. The depolarization of motor neurons upon contraction results in electrical activity measurable from the skin surface.

E. Respiration Responses

Respiratory organs, mainly lungs, are dually innervated as well. Respiratory sinus arrhythmia (RSA), a phenomenon where the heart contracts and relaxes as a function of respiration due to the inherent coupling between breathing and blood pumped by the heart, is a noninvasive index of parasympathetic activity as it arises from the fluctuations in the vagal control [47]. The chemoreceptors in the arteries detect small decreases in the amount of oxygen or increases in carbon dioxide and trigger respiration activity. Negative emotions, such as anger, trigger a higher respiration rate than positive emotions [48].

F. Behavioral Responses

Although closely tied to the ANS-mediated responses described above, we categorize action tendencies or expressions driven by underlying changes in the physiological state as behavioral responses. Behavioral changes resulting from emotions include changes in facial expressions, gait, speech properties, body postures, gestures, and so on. For instance, speech is influenced by respiration rate. A variation in respiration rate triggered by SNS affects the air pressure below the larynx. The variation in air pressure affects the opening and closing of the vocal folds, thereby resulting in variations in voice intensity [49]. Emotion-specific variations in speech have been studied [50]. Furthermore, underlying emotions are found to activate various facial muscles, resulting in facial expressions. Ekman et al. [26] conducted a pioneering study on facial expressions and autonomic responses, discovering that the activation of prototypical facial muscles or action units is associated with corresponding changes in the ANS activity. A specific combination of action units is involved in particular emotions [27]. For instance, negative emotions, such as sadness, activate the action units near the eyebrows, whereas positive emotions, such as happiness, activate the action units of the cheek.

IV. MEASUREMENT OF EMOTION RESPONSES

This section is dedicated to exploring the measurement techniques of the modalities that were discussed in Section III. We will examine the features of each modality and how they are utilized for detecting emotions (see Table 1).

A. Brain Activity

Electroencephalogram (EEG) involves recording and amplifying the collective electrical signals generated by billions of nerve cells through the use of electrodes and wires attached to the scalp. Despite its ability to offer researchers

Table 1 Activity Types and Corresponding Measurement Types

Activity Type	Measurement Type	Brief Description
Brain Activity	Electroencephalography (EEG)	Measures the total electrical activity of neurons in the brain using electrodes placed on the scalp.
Heart Activity	Electrocardiogram (ECG)	Measures the heart's electrical activity by placing electrodes in specific regions.
	Photoplethysmogram (PPG)	Measures the changes in blood volume, i.e., Blood Volume Pulse (BVP) optically.
Muscle Activity	Electromyography (EMG)	Measures the changes in electrical activity of muscles of interest.
Skin Activity	Electrodermal activity (EDA)	Measures the resistance between two electrodes (on fingers, palm or wrist) on the skin.
	Skin temperature	Measures the temperature of the epidermis, skin's outermost layer.

tremendous time resolution, the spatial resolution of EEG is relatively low, and it requires multiple electrodes to be placed at various locations on the head. Nonetheless, EEG remains a valuable tool for investigating phase transitions in response to emotional stimuli [51]. Functional neuroimaging techniques, including positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), have been utilized to investigate the impact of emotion on the limbic system [52]. Researchers discovered emotion-related increases in cerebral blood flow or blood-oxygen-level-dependent signals in cortical, limbic, and paralimbic regions. This suggested that specific brain regions have specialized functions for emotional operations. To investigate this specificity, researchers induced visual, auditory, and recall-based stimuli to recognize emotions by analyzing the activated regions using PET and fMRI technologies [53], [54]. In addition, EEG is noninvasive, fast, and cost-effective, making it a preferred method for investigating the brain's responses to emotional stimuli [55]. EEG is commonly combined with speech [56] and facial expression [57] data to improve the robustness of emotion recognition systems. Recently, new EEG devices have emerged in the market, which offers several advantages, such as unobtrusiveness, affordability, portability, and ease-of-use. These devices, such as the Emotiv Epoch 14-channel, the Emotiv Insight 5-channel, and the Omnifit Brain 2-channel headsets, are typically equipped with 10–20 electrodes and can be utilized to capture raw EEG data.

1) Preprocessing: There are two types of artifacts that can affect EEG data: technical (extrinsic) artifacts and physiological (intrinsic) artifacts [58]. Technical artifacts include noise from electrode misplacement, powerline interference, and other electromagnetic interferences, while physiological artifacts include eye movements and blinks (electrooculogram artifacts), muscle activities (electromyogram artifacts), and cardiac activities. Frequency-domain filters, such as a bandpass filter between 0.5 and 60 Hz, can remove most technical artifacts. However, removing physiological artifacts is more complex and requires the use of threshold-based time-domain filters and independent component analysis techniques [59].

2) Feature Extraction: EEG features can be divided into two groups: time and frequency domains. Time-domain features can be listed as mean, median, variance, standard deviation, skewness, kurtosis, zero crossing rate, wave duration, peak amplitude, instantaneous frequency, complexity, and energy [60]. In frequency-domain analysis, brain rhythms are very well established. Gamma waves can be found over 30 Hz and related to activity in fronto-central areas. They have the highest frequencies and can be used to monitor regions related to voluntary movements, cognitive functioning, learning, memory, and processing information [61]. Beta waves are between 14 and 30 Hz and are related to activity in the parietal, somatosensory, frontal, and motor areas. They are seen during awakened states, and they are correlated with memory, focus, and problem-solving functions. Alpha rhythms are between 8 and 13 Hz and are related to occipital and parietal regions. Alpha rhythms are made up of the subconscious activity of the brain, and they are related to relaxed and mediated mind states. Another rhythm is the theta rhythm and related to the hippocampus region. They are commonly observed under drowsy, daydreaming, and sleep states. The last rhythm is delta waves, and they are the slowest brain waves. They can be observed during deep sleep states. Frequency-domain features are mostly built up on well-established brain rhythms. $\delta, \theta, \alpha, \beta, \gamma, \theta/\alpha, \beta/\alpha, (\theta + \alpha)/\beta, \theta/\beta, \gamma/\delta$, mean, median, variance, standard deviation, and reflection coefficients are commonly used frequency-domain features.

B. Electrical Activity of Heart

There are two methods for measuring heart activity: electrocardiography (ECG) and photoplethysmography (PPG). ECG sensors use multiple electrodes placed symmetrically on specific areas of the body to measure the heart's electrical activity, resulting in an ECG signal with essential information, including the R peak, which is commonly used for extracting emotion-specific features [62]. On the other hand, PPG sensors measure the changes in blood volume by measuring the extent of reflection absorbed by skin-reflected infrared light initially emitted by a light-emitting diode, resulting in a PPG signal that can be used to estimate R peaks from the peaks of blood volume (refer Fig. 5). Although PPG data have lower quality and are more susceptible to motion artifacts under physically active situations, they offer greater unobtrusiveness and can be used without interrupting users during long experiments in daily life. Therefore, sensors should be selected based on the performance requirement, experiment duration, and environment of the study. Devices

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

providing raw ECG data include BIOPAC's MP150, MP35, Shimmer Sensing 3, Polar H9, Polar H10, Firstbeat Bodyguard 2 and 3, Zephyr HxM, and Bitalino (r)evolution. Wristbands such as Empatica E3 and E4, Samsung Galaxy S1 and S2, Angel, Polar Verity Sense, and finger sensors, such as CorSense, UFI model 1020, and BIOPAC BioNomadix PPGED-R, provide raw PPG data.

1) Preprocessing: Robust artifact detection and removal algorithms are applied before processing the PPG data. In the literature, several frequency- and time-domain filters have been used. Generally, every data point is compared with the local average for time-domain filters. A data point is labeled as an artifact if the percentage of difference is greater than a certain threshold (approximately 20% [63]). The commonly used frequency-domain filters include Butterworth high-pass filters with a cutoff frequency of 1 Hz to eliminate baseline wander, lowpass filters with a cutoff around 25 Hz to eliminate high-frequency artifacts (also from other sensors, such as EMG), and band rejection filters to eliminate power line interference between 50 and 60 Hz [64]. The removed artifact data points can be replaced using different interpolation techniques. The cubic spline interpolation is one of the most commonly used techniques since it has a structure similar to the heart activity signal.

2) Feature Extraction: HR is commonly used to estimate the degree of emotions. It can be calculated by counting the number of heartbeats per minute. Alternatively, the time interval between consecutive R peaks called the RR interval or interbeat interval (IBI) is used. IBI has an inverse relationship with HR. HRV is another widely used measure for heart activity, and it can be computed from the distribution of RR intervals over a time interval. Variation in HRV corresponds to SNS and PNS activities.

HRV features can be extracted from time and frequency domains. Mean HR, standard deviation of IBI, mean RR, root mean square of successive differences (RMSSD) of respiration rate (RR) intervals, and the percentage of successive RR intervals that differ from the previous RR interval by more than 50 ms (pNN50) are considered the most distinctive time-domain features. The IBI data should be converted to the frequency domain to extract frequency-domain features. Since R-peaks are not equidistant, either the IBI signal needs to be resampled to obtain equidistant samples in order to use fast Fourier transform or methods such as Lomb–Scargle periodogram [65] can be used. After the conversion to the frequency domain, powers in very low, low, prevalent low, high, and prevalent high-frequency ranges and the ratio of power in low- to high-frequency ranges are commonly extracted.

Several nonlinear features of HRV [66] are evaluated using various state-space domain entropy-related measures. The most commonly used measures are the standard deviations of the Poincare plots, approximate and sample entropy, correlation dimension, recurrence, and fluctuation slopes [67].

C. Muscle Activity

Electromyography (EMG) utilizes electrodes to quantify the electrical activity changes in muscles as a result of contraction. The facial and trapezius muscles are the most extensively examined muscles for emotional responses [68]. Facial muscle activity is commonly employed for emotion recognition and is recognized via the facial action coding system (FACS) [69]. While the visual inspection is subjective in nature and has potential coding errors, facial EMG is an objective method with fewer true negatives than [46]. However, facial EMG measurement may be intrusive and alter the participant's natural behavior. Facial expressions resulting from muscle activity will be discussed in greater depth in Section IV-G. Yet, the importance of bodily expressions of emotions is currently being investigated as they have been found to correlate with facial expressions during social interactions [70].
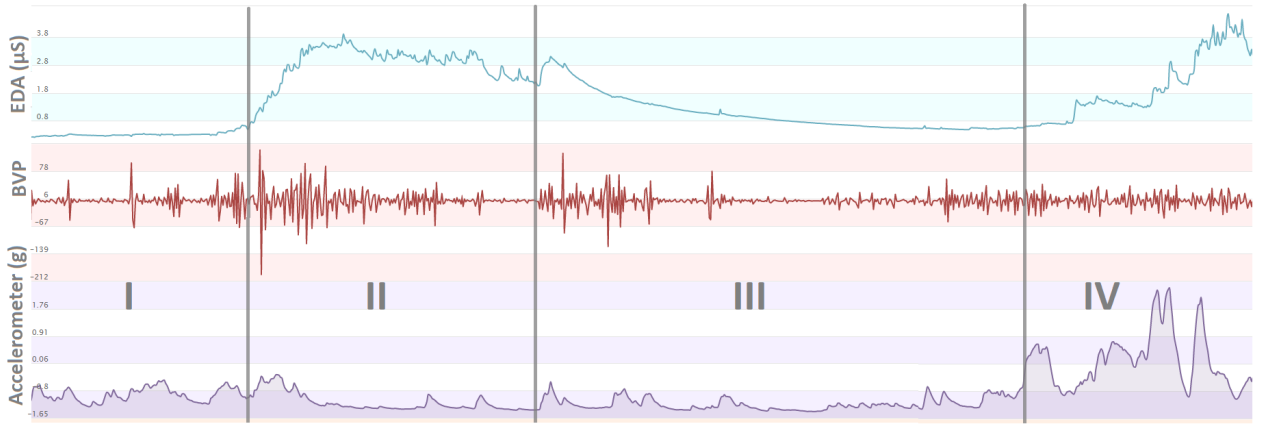
1) Preprocessing: The EMG signal is often affected by noise. The possible noises include the motion artifacts arising from user motion or cable and electrode interfaces, inherent device noise, and ambient noise [71]. Frequency-domain filters are applied to remove artifacts in specific frequency bands [72]. In addition, adaptive prediction error filters have been proposed for eliminating nonstationary artifacts affected by factors such as stimulation intensity, fatigue, and the contraction level of the muscle [71].

2) Feature Extraction: Muscle activity signals obtained from the EMG sensor include the superposition of actions of numerous motor units. Therefore, they need to be decomposed to reveal the mechanisms of muscle and nerve control. The decomposition is commonly performed using wavelet spectrum matching and principle component analysis of wavelet coefficients [71]. Commonly extracted features include wavelet-based features [73], Mel-frequency cepstral coefficients [72], and statistical features, such as mean, standard deviation, rms, peak loads, and gaps per minute. Furthermore, muscle tremors are known to be signs of different emotions [74], and they can be detected around 11 Hz using frequency-domain analysis.

D. Skin Activity

Electrodermal activity (EDA) is the activity of the skin where the electrical properties change based on the emotion a person experiences. EDA is measured in terms of change in skin conductance estimated by passing a small amount of current through silver–silver chloride electrodes. An instantaneous surge in skin conductance constitutes the phasic component of EDA. Darrow [75] found a correlation between the sweat gland activity and the phasic skin conductance response (SCR) upon exposure to an emotional stimulus; however, there is a delay of a few seconds between the two. The dc component

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

Fig. 5. *Recorded signals from a laboratory experiment comprising four phases. In the first phase, the baseline is shown. In the second phase, participants are induced with mental stress using TSST. The changes in BVP and EDA signals can be observed in the stress phase. The third is a recovery phase using breathing exercises. The last phase is a physical activity phase with increased acceleration, EDA, and BVP signal activities.*

of EDA is the skin conductance level (SCL) and is either low or high in resting and activated states, respectively. Although EDA is a good approximation of SNS activity and an easy yet inexpensive way to measure, it is unreliable when the subject moves or the external temperature conditions vary. Furthermore, researchers must be cautious while measuring the EDA signal due to the factors such as the contact between the electrodes and the measurement area, the salinity of the electrolyte, skin area preparation, the controllability of the stimulus, and respiration matter. EDA is a promising signal for emotion recognition along with the heart activity signal. Measuring instruments, such as Shimmer 3 GSR+, ProComp Infiniti, Bitalino (r)evolution, BIOPAC MP150, and wrist devices such as Movisens EDAMove 4, Empatica E3, and E4, are widely used to measure EDA, which provides raw data [68].

1) Preprocessing: EDA increases with physical activity and changes in temperature as they cause sweating. Therefore, a multimodal approach with physical sensors is required to isolate the effect of emotional state changes on EDA. Physical activity measured using accelerometer sensors and external temperature changes inferred by ST sensors can be useful. There are several preprocessing tools for cleaning the EDA signal. Though wavelet-based artifact removal techniques are common in the literature [76], [77], supervised machine learning-based techniques [78] for artifact removal exist. Manually annotated data labeled by experts for artifacts are used to train supervised models.

2) Feature Extraction: The EDA signal has two components: SCL and SCRs. SCL is a slow-changing dc component, also called the tonic component. In contrast, SCR is an event-related and short-term component of the EDA and is also called the phasic component. There are open-source tools for analyzing the EDA signal, such as cvxEDA [79] (a convex optimization-based EDA analysis

tool) and pyEDA [80]. The tonic component is used for long-term baseline measurement using statistical features, such as the mean, minimum, maximum, standard deviation, quartile deviation, 20th percentile and 80th percentile of values over an interval, and first and second derivative features. For short-term arousal detection, features from the phasic component, such as the peaks count over a specific duration, the total number of peaks above a certain high threshold (one micro Siemens) over a duration, the delay between stimulus and peak response, peak amplitude, and rise and recovery times, are measured.

a) Skin temperature: Besides emotions, STs are affected by various factors, such as weather and physical activity. Previous research has shown that increased blood flow due to arousal induces about 0.1 °C–0.2 °C change in ST [81]. With controlled external factors, such subtle changes in the ST resulting from an emotional response can be measured. Often, ST is combined with additional biosignals to get a more robust recognition performance. Standard time-domain statistical features of ST signals are used in the literature.

E. Blood Pressure

High-arousal negative emotions cause an increase in blood pressure levels, whereas low-arousal positive emotions can decrease them [82]. Recently, commercial wearables have been equipped with blood pressure sensors, namely, ASUS VivoWatch BP (HC-A04) and Omron HeartGuide. These devices make it possible to monitor blood pressure levels continuously. Systolic and diastolic components of blood pressure can be used as features.

F. Respiration

Furthermore, respiration data are used to decouple the EDA data from the effects of breathing. Respiration measurement is inexpensive as it involves a simple belt

containing a piezoelectric device. However, one should beware of possible issues during the measurement. For example, the tightness of the chest strap may lead to either ceiling effect or inaccurate recordings, the discomfort caused by the strap leading to new breathing patterns, or voluntary controlled breathing. Breathing rate and amplitude can be indirectly measured using transducer-based sensors that rely on chest cavity expansion [83], [84]. PPG data from wearable devices can be used to derive respiration rate [85]. Statistical features, such as minimum, maximum, mean, and standard deviation of respiration rate, mean and standard deviation of the first and second derivatives, and frequency-domain features such as spectral power [16], are extracted. In addition, nonlinear features are extracted using recurrence quantification analysis, deterministic chaos, and detrended fluctuation techniques [86].

G. Measurement of Behavioral Responses

Behavioral responses are best suited for noncontact measurement. Behavioral responses are commonly combined with physiological signals to obtain a more accurate emotion recognition system. Yang et al. [87] combined several behavioral (facial expression, speech, and keystroke) and physiological (blood volume, EDA, and ST) modalities and achieved 89% accuracy for binary emotion recognition. One of the advantages of deep learning approaches is their ability to effectively utilize multimodal data, which includes information from physiology, facial expressions, and speech. Moreover, facial muscle activity has independently aided emotion recognition. The measurement has started with facial EMG, but, recently, RGB cameras have been used more commonly to capture emotions from facial muscle activity. The discovery of action units by Ekman et al. [26] led to the development of the FACS. This system represents facial expression prototypes in terms of the location of action units on the face [69], and geometry-based facial feature extraction approaches that involve the position, size, and shape of facial landmarks were developed to detect these action units [88]. In addition, appearance-based approaches that utilize color intensity and texture of facial features, such as spatial filters and local binary patterns [89], have also been explored. Early approaches to facial emotion recognition primarily relied on traditional emotion classification methods that utilized these extracted features from facial expressions. However, with the availability of large datasets and advancements in computing technology, recent research has introduced deep learning approaches that can inherently capture the nuances of facial expressions from images and directly classify them into emotions [90], [91].

Speech signal has been widely combined with physiological signals [92], [93] and improved emotion recognition performance. Emotion-specific variations in the speech were identified several decades ago [50]. While the electrical activity of the vocal cords can be measured through electroglottography, it is easier to capture emotion-related patterns of speech in microphone audio data. Recent advances in machine learning have resulted in learning emotional feature representations from speech data [94]. Furthermore, transformer-based speech emotion models have led to improved recognition of positive and negative emotions, with good generalization and robustness across different domains, speakers, and genders [95].

In addition, recent research has demonstrated that alterations in body posture can indicate changes in affective states [96], [97]. Consequently, numerous studies have investigated the utilization of body postures and movements for emotion recognition [98]. Specific body postures, such as head tilts and clenched fists, have been linked to the expression of specific emotions [99], [100], suggesting their involvement in nonverbal communication and emotion perception. Moreover, recent studies have revealed that body movements [101] including measures such as the velocity of joints, acceleration, and jerk, and other gesture-specific features such as height, angle, and movement direction of the hands and arms, body movement trends, head movement, symmetry [102], and gait [103] can carry information relevant to emotions.

H. Contextual Information

Context influences an emotional experience but is challenging to obtain in an uncontrolled setting, such as daily life. Nevertheless, the system's robustness can be increased by adding contextual information to the physiological signals. Activity-associated context actively acquired from the user in combination with HRV significantly increased the stress detection performance in the wild (around 25% increase in F1-score) [104]. Since the active acquisition of context from the users may interrupt them, passive acquisition using smartphone data may provide helpful insights. Passive context based on physical activity and location, smartphone activity (calls, SMS, applications, battery status, and screen usage), and ambient conditions (light and weather) can detect stress independent of physiological signals [105]. Context based on smartphone activity has been used in addition to physiological data, such as EDA [106]. A 10%–15% increase in stress recognition accuracy was reported when weather data (air temperature, humidity, and air pressure), activity information, and physical activity intensity were added to HRV and EDA signals [107].

I. Combination of Multiple Physiological Signals and Interdependencies With Other Modalities

Emotion recognition studies often combine multimodal physiological signals to obtain a more comprehensive view of emotional states [74]. Adding more modalities can eliminate the drawbacks of individual signals and develop more robust systems. Soleymani et al. [108] investigated

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

the interactions between EEG signals and facial expressions for emotion recognition. In particular, they show that informative features of EEG signals originated to a large extent from facial expressions. Insights on potential artifacts in channels of affect-related information could be deployed when designing fusion processes and, thus, contribute to a more reliable emotion recognition process.

The multimodal fusion process is of three different types: early, intermediate, and late [16].

1) Early Fusion: This type of fusion occurs at the feature level by selecting the features from multiple signals and combining them to form a single input for feature extraction or classification. Fabiano and Canavan [109] used a feature-level fusion and showed a 10%–15% improvement in valence, arousal, and dominance recognition. However, this fusion is suitable for synchronized input signals.

2) Intermediate Fusion: This kind of fusion can overlook synchronization issues by leveraging feature extraction from different time lengths. Furthermore, by comparing previous instances with the current ones, probabilities for imperfect instances can be statistically predicted [16]. Methods using hidden Markov models and Bayesian networks are practical for dealing with these situations. Shin et al. [110] used a Bayesian network to fuse features from EEG and ECG for recognizing comic, fear, sadness, joy, anger, and disgust emotions and increased the accuracy by 35.78%.

3) Late Fusion: This type involves aggregating results generated by different classifiers to obtain a final result, often through voting. The classifiers can be trained separately on each modality, hence not requiring synchronization [16]. Wang et al. [111] applied three SVM (RBF kernel) classifiers to power spectral, Higuchi fractal dimension, and Lempel–Ziv complexity features. They integrated these classifiers by employing a weighted fusion strategy that computes confidence estimation on each class by each classifier. They evaluated their approach on the DEAP dataset (on EEG data) and showed that this late fusion method outperforms the performance of individual classifiers and the early fusion methods.

J. Insights

Multimodality has advantages such as increased redundancy, i.e., when one signal fails to detect emotion in a specific situation, thereby improving prediction performance. Furthermore, specific signals can be used to detect and remove artifacts from other signals. For example, EDA is very sensitive to physical activity and increased room temperature. Under such conditions, changes in EDA could be falsely regarded as increased arousal or valence. Acceleration and ST data can be used for cleaning the artifacts in EDA data [78]. In addition, some signals, such as ST and respiration, achieve better results when combined with additional biosignals [16]. The selection of modalities depends mostly on application type and environment.

The behavioral (i.e., speech and body movements) and muscle-based responses (such as facial expressions) are more robust in controlled environments than physiological signals. Moreover, researchers obtain robust performance with EEG signals, especially in laboratory or controlled environments. In more controlled situations, they can be preferred. However, it is challenging to monitor speech and facial expressions in the wild, and users will be reluctant to wear EEG devices in daily life although they are more reliable. Therefore, the story is different for daily life emotion recognition. User's self-reports reflecting the issues such as comfort and utility are more important for daily life [112]. Wrist-worn devices have advantages in these aspects, but they have lower data quality [113]. Therefore, the selection of modalities and wearable devices is a multivariate problem, and researchers should make a tradeoff by evaluating their application-specific requirements.

K. Unexpected Observations

In some of the experiments, researchers observed unexpected phenomena while analyzing the data. The most common ones are observed during the emotion elicitation phases of experiments. Wagner et al. [114] observed that all classification algorithms had particular problems in separating pleasure and sadness that they found surprising. After further analysis, it is revealed that listening to sad music may elicit positive feelings [115]. Emotions are complex phenomena, and assumptions made while designing experiments for emotion elicitation might not hold on some participants. In another case, researchers noticed that some participants did not report stress in the arithmetic phase of the trier social stress test (TSST) in the questionnaires [116]. They recruited the participants from a university, and they saw that students from mathematics or computer science departments tend to report low stress in the arithmetic phase of TSST. Therefore, to detect these unexpected observations, perceived emotion self-reports can be collected and cross-referenced with the elicited emotional context (whether the participant watches a sad video or stress is induced) to validate whether the experienced emotion is the same as the elicited one. Moreover, although multimodality yields generally better results, it is not always the case [117]. Sometimes, research using only ECG or EEG data showed better or sometimes worse than the multimodal approaches. One signal can be dominantly better than others for a task or all signals can be noisy in similar intervals, and they could not contribute to the performance of others. In these cases, multimodality does not necessarily improve performance. In the context of interpersonal differences, a study found that women and men do not react the same way and also showed different patterns in physiological (skin conductance) recording. Women were found to be more emotionally expressive than men [118]. Individuals using an emotional regulation strategy, such as suppression, yield different physiological responses to emotions than those [119].

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

V. EXPERIMENTAL DESIGN FOR PHYSIOLOGICAL DATA COLLECTION AND EXISTING DATASETS

Data collection in emotion research has no consensus on emotion elicitation and measurement methods owing to the highly subjective nature of emotions. However, several measures can be adopted during data collection to capture emotions reliably and facilitate effective emotion recognition. The most important factors to be considered during data collection are sample population, emotion stimulus, modalities measured, emotion annotation possibilities, and sensing equipment [132]. Below are some points to consider.

1) *Emotional stimuli:* Stimuli are characterized by categorical emotions, such as happy and sad, and are employed for emotion elicitation. Appraisal-based theories of emotion elicitation have emphasized that the emotion elicited in an individual is specific to the stimulus and its appraisal. Therefore, due to individual differences, perception of the stimulus may vary from the intended emotion or its intensity. Research has progressed from eliciting strong emotions in a laboratory to measuring emotions in real life, thereby dealing with low-intensity emotions. Therefore, it is crucial to ensure: 1) the stimulus for a specific category of emotion should be verified to elicit the intended emotion; 2) to the required intensity; and 3) no other overlapping emotion is elicited. An example of a verified stimulus for inducing stress is the TSST [133]. It is a method consisting of a public interview and arithmetic tests to induce stress and is widely used for stress response elicitation. TSST has been clinically validated to induce a stress response in most of the population and is characterized by novelty, uncontrollability, unpredictability, and socio-evaluative threat [134].

2) *Emotion regulation:* It is an innate process that may take place alongside emotion expression [135]. Participants may use different emotion regulation strategies to modify their subjective emotional experiences or responses during measurement. This may result in inaccurate physiological responses. Depending on age, culture, and personality, participants may adopt different regulation strategies [136], [137]. Different regulation strategies may influence physiology differently. Research has shown that participants who used suppression to regulate their emotions, in contrast to reappraisal as a regulation strategy, showed higher physiological responses to emotional stimuli [138]. Therefore, it is important to instruct the participants not to adopt an emotion regulation strategy during the experiment.

3) *Sample population:* Research has shown that cultural differences influence physiological emotion responses [139]. Depending on the emotion regulation strategy used, age is also a factor influencing physiological responses [137]. Therefore, results from an emotion recognition study involving participants from a specific age group or cultural background may not be generalized to other populations. Depending on the context of the study, a larger and more diverse sample population size is crucial to overcome the interindividual variability and the effect of confounding variables. Sample size can be obtained using appropriate statistical tools, such as G*Power [140], Krejcie and Morgan's formula [141], or Cochran's sample size formula [142], by specifying the allowed margins of error.

4) *Measurement:* As described in Section III, the physiological manifestation of emotions makes it possible to identify emotions through multiple modalities. Depending on measurement convenience, the selected physiological measures should include major ANS responses. Cardiac and electrodermal responses are helpful for autonomic activity estimation. Research suggests seeking convergent evidence across multiple responses for a particular emotion [143]. Furthermore, since emotions are short-lived, the timing of physiological measurement is important. This is especially true when emotional stimuli produce a less intense emotional response. Measuring devices play a crucial role in data collection. Medical-grade devices are often not suited for real-life data collection. Therefore, researchers are directed toward more unobtrusive and easy-to-use devices. However, scientifically validated devices should be chosen for the data collection.

5) *Annotation:* Most often used means of self-reporting are Likert scales of valence and arousal. While self-reported data are the closest reflection of an individual's emotion, it is prone to several errors, such as inaccurate understanding of scales or negligence in reporting. Timing of self-reporting is also crucial as the reports may be affected by failure to recall events or the obtrusiveness of prompts.

6) *Context:* Models built on artificially elicited emotions in laboratories cannot be generalized to the real-life environment. The psychophysiological responses to artificial stimuli do not represent those in real life. Although real-life data collection has more issues when compared to collecting data in a controlled laboratory environment, the research direction is toward developing real-life and daily emotion recognition systems. However, data collection in real life poses several challenges. First, the subtlety of the emotional responses is a hurdle for annotation in real life. Unlike the ability to control the stimuli in the lab, a real-life scenario requires considerable contextual information to be recorded. Individual-specific information, such as personality, demographics, and health conditions that potentially impact emotional responses, are likely to yield better confidence in the computed emotion recognition models. The self-reported and sensor-based contextual information about the participant

Table 2 Comparison of Physiological Datasets Collected for Emotion Recognition. A Stands for Arousal, V Stands for Valence, and D Stands for Dominance

Dataset	Environment	Signals	Participants	Emotion Inducing Method	Labels
DEAP [120]	Laboratory	EEG, PPG, EDA, EMG, EOG and Respiration	32	Music videos	A,V, D
MAHNOB [121]	Laboratory	EEG, ECG, EDA, Skin Temperature, Face and Body Videos	27	Movies	A, V, D
DREAMER [122]	Laboratory	EEG and ECG	23	Movies	A, V, D
WESAD [123]	Laboratory	ECG, EDA, EMG, PPG, respiration, Acceleration and Skin Temperature	15	Trier Social Stress Test	Amusement, Stress, Recovery
AMIGOS [124]	Laboratory	EEG, ECG, EDA, face and body videos	40	Emotional Videos	V, A, control, familiarity, liking, and basic emotions
CASE [125]	Laboratory	ECG, BVP, EMG, EDA, respiration and skin temperature	30	Emotional Videos	A, V
ASCERTAIN [126]	Laboratory	ECG, EDA and face videos	58	Emotional Videos	A, V
EMDB [127]	Laboratory	Heart Rate and EDA	32	Emotional Film Clips	A, V, D
RWDADW [128]	Automobile	ECG, EDA, and skin temperature	10	Real World Driving Task	Perceived Workload
DSDRWDT [129]	Automobile	ECG, EMG, EDA, and skin temperature	17	Real World Driving Task	Perceived Stress
EMOTIONS [130]	In the wild	PPG, EDA, EMG and respiration	1	N/A (daily emotions)	Perceived 8 Emotions
DAPPER [131]	In the wild	PPG, EDA and acceleration	88	N/A (daily emotions)	A, V , 10 emotions

and the experimental conditions, such as physical activity type and intensity, location, and ambient conditions, is necessary to reason for the anomalies in the emotion recognition models as they tend to influence the physiological modalities. For example, an increase in the EDA signal could result from physical activity, environment and weather changes, or emotional stimuli. Furthermore, collecting data in the laboratory and in real life from each participant could increase the robustness of the systems. Responses to emotional stimuli can be more accurately modeled in a controlled environment, and these personalized models could be adapted to a real-life environment.

A. Existing Datasets for Emotion Recognition

In this section, we provide the prominent emotion recognition datasets that consist of physiological signals (see Table 2). Although most of these datasets are recorded in laboratory environments, recently, new studies created datasets recorded in real-life environments [131], which would help researchers to improve emotion research in real life or the wild.

1) DEAP [120]: The Database for Emotion Analysis Using Physiological Signals (DEAP) dataset[1] was collected from 32 participants in a laboratory environment. Participants were asked to watch annotated 1-min music videos and evaluate them on arousal, valence, dominance, likability, and familiarity scales. EEG, PPG, EDA, EMG, electrooculography, respiration, and temperature signals were collected. In addition, frontal face videos were recorded for 22 participants.

2) MAHNOB-HCI [121]: Similar to the DEAP dataset, the MAHNOB-HCI dataset[2] was also recorded in a laboratory environment. Twenty-seven participants watched

video segments from commercial movies and assessed them on valence, arousal, and dominance scales. EEG, ECG, EDA, and ST were collected. In addition, face and body videos were recorded using six cameras.

3) DREAMER [122]: The DREAMER dataset was collected from 23 participants in a controlled environment. Scenes from commercial movies were selected to induce different emotions. EEG and ECG signals were recorded. The participants assessed arousal, valence, and dominance levels on a scale from 1 to 5. The dataset was collected using portable and low-cost wearable devices, which are viable options for real-life data collection. However, the dataset has restricted access and is available upon request.

4) WESAD [123]: The Wearable Stress and Affect Detection (WESAD) dataset[3] was collected from 15 participants in the laboratory environment. The experiment included amusement, stress, meditation, and recovery conditions. Positive and negative affect schedule (PANAS), state-trait anxiety inventory (STAI), and additional Likert scale questions (stress, frustration, happy, and sad) were used as self-reports. The physiological signals recorded were ECG, EDA, EMG, PPG, respiration, accelerometer, and ST. The experiment duration was about 2 h.

5) AMIGOS [124]: A dataset for Multimodal research of affect, personality traits, and mood on the Individuals and GrOupS (AMIGOS) dataset[4] was gathered in two experimental settings. First, 40 participants watched 16 short emotional videos (50–150 s) in the laboratory environment. Second, the participants watched four longer videos individually and in groups. EEG, ECG, and EDA signals were recorded. High-quality frontal face and body videos were also recorded. Participants reported their valence,

---

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

Table 3 Performance of Varying Machine Learning Algorithms for Recognizing Emotions. The Accuracies Are Two-Class by Default If Not Reported Otherwise. LALV Is Low Arousal Low Valence, LAHV Is Low Arousal High Valence, HALV Is High Arousal Low Valence, and HAHV Is High Arousal High Valence. A: Arousal; V: Valence; Acc: Accuracy; and RBC: Radial-Basis Classifier

Study	Signals	Classifier	Accuracy
[126]	ECG, EEG, EDA, facial activity	SVM	0.6 - 0.7 F_score A and V.
[144]	ECG	SVM	70-80% acc. for A and V.
[145]	EDA	SVM	80% acc. (stress, cognitive load and relax)
[146]	ECG	LDA	80% acc for A and V.
[147]	EDA and blood oxygen levels	RF	74% 4 class acc.(amusement, fear, grief and fear)
[148]	EDA, HRV	SVM,RF	70% acc. for A and V
[114]	ECG, EDA, EMG, respiration	MLP	88,64% and 94,32% acc. for V and A
[149]	ECG, EDA, Blood Pressure, EMG and ST	MLP, kNN, RBC	∼80% acc.
[113]	PPG, EDA, ACC	MLP	92.15% acc. (Binary Stress)
[13]	EDA	CNN	95% acc. (5 class V)
[150]	EEG, ECG, EDA	Graph CNN	70% acc. for A and V
[151]	DEAP dataset	CNN + LSTM	78.72% and 79.03% acc. for V and A.
[152]	AMIGOS, DREAMER datasets	CNN + LSTM	99% acc. for AMIGOS and 90.8% acc. for DREAMER datasets (4 class LALV, LAHV, HALV, HAHV)
[153]	CASE dataset	LSTM (raw data)	∼95% acc. for Amusement, Boredom, Relax and Fear
[154]	EDA, ST, ACC, mobile phone usage and location data	LSTM (features)	83% acc. (Binary Stress Level Prediction)
[155]	ECG (AMIGOS dataset)	CNN + Transformer	83% and 88% acc. for V and A.
[156]	PPG, EDA and ST	CNN + Transformer	77.37% and 79.42% acc. for V and A.

arousal, control, familiarity, liking, and basic emotions and were also evaluated externally. They also collected the big five questionnaires for personality-related information and PANAS questionnaire for mood-related data.

6) CASE [125]: The Continuously Annotated Signals of Emotion (CASE) dataset[5] consists of real-time annotated emotions of participants while watching videos in the laboratory environment. Twenty videos whose emotional content is verified by previous studies were selected. ECG, BVP, EMG, EDA, respiration, and ST signals were recorded from 30 participants. In addition, valence and arousal levels were reported by the participants.

7) ASCERTAIN [126]: the databASe for impliCit pERsonaliTy and Affect recognition (ASCERTAIN) dataset[6] includes big-five personality scales and emotional self-ratings of 58 participants. EEG, ECG, EDA, and facial activity data were recorded, while the participants watched audiovisual clips. Arousal, valence, and personality were collected using self-reports.

8) EMDB [127]: The Emotional Movie DataBase (EMDB) dataset[7] was recorded in a laboratory environment. Thirty-two participants provided psychological data during watching 52 emotional film clips, which took around 40 s each. HR and EDA data were recorded. Arousal, valence, and dominance were recorded as the ground truth.

9) RWDADW [128]: The Real World Driving to Assess Driver Workload (RWDADW) dataset[8] was recorded in an automobile environment. Ten participants provided psychological data during real-word driving tasks under the 30-km zone, the 50-km zone, highway, freeway, and tunnel conditions. At the end of the driving task, they filled perceived workload questionnaire. ECG, EDA, and ST data were recorded.

10) DSDRWDT [129]: The Detecting Stress During Real-World Driving Tasks (DSDRWDT) dataset[9] was recorded in an automobile environment. 17 participants provided psychological data during watching real-world driving tasks. The duration of sessions was between 54 and 93 min. HR and EDA data were recorded. Perceived stress scores were collected for each session.

11) EMOTIONS [130]: The EMOTIONS dataset[10] was recorded once a day, in a session lasting around 25 min, for over twenty days. It was recorded by one participant. Eight emotions (neutral, anger, hate, grief, joy, platonic love, romantic love, and reverence) were annotated for each session. PPG, EDA, EMG, and respiration data were recorded.

12) DAPPER [131]: The DAPPER dataset[11] was recorded in an ambulatory environment, unlike the abovementioned ones collected in a laboratory; 142 participants provided psychological recordings, whereas only 88 provided physiological recordings over five days. Emotions were annotated using the experience sampling method (ESM), and detailed descriptions of everyday emotional experiences were obtained using the day reconstruction method. ESM comprises arousal and valance ratings and

[5]CASE access link: https://gitlab.com/karan-shr/case_dataset/tree/ver_SciData_0
[6]ASCERTAIN access link: https://ascertain-dataset.github.io/
[7]EMDB access request at: EMDB@psi.uminho.pt
[8]RWDADW access link: https://www.hcilab.org/wp-content/uploads/hcilab_driving_dataset.zip

[9]DSDRWDT access link: https://physionet.org/content/drivedb/1.0.0/
[10]EMOTIONS access link: https://dam-prod2.media.mit.edu/x/2022/01/05/SetA.tar.gz
[11]DAPPER access link: https://www.synapse.org/#!Synapse:syn22418021/files/

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

PANAS questions for ten selected emotions. HR, EDA, and acceleration data were recorded.

VI. MACHINE LEARNING APPROACHES

Emotion recognition systems are based on supervised learning and consist of binary or multiclass classifiers. The inputs to these classifiers are various signals, and the output class labels correspond to an emotional state (i.e., different emotion types and levels). Early studies employed traditional classifiers to recognize emotions. Classification tools can be listed as linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), k-nearest neighbor (kNN), random forest (RF), and support vector machine (SVM). With the advancements in deep learning algorithms, multilayer perceptron (MLP), convolutional neural networks (CNNs), and long short-term memory (LSTM) techniques are also tested for recognizing emotions (see Table 3).

A. Traditional Machine Learning Approaches

Traditional algorithms and their advantages and disadvantages can be described briefly as follows: the SVM algorithm defines a hyperplane that separates data points belonging to different classes with the largest spatial margin. Although originally designed as a linear classifier, SVM can be scaled to perform nonlinear classification using different kernel functions efficiently. However, it is used predominantly for binary classification of emotions [16]. kNN is an algorithm that assigns a class to a new data point based on the classes of its k closest data points and is rather straightforward to implement. Nevertheless, kNN requires storing all training data, which causes increased complexities in time and space. Kernel SVM and kNN being nonlinear classifiers compute the decision boundary accurately depending on their hyperparameters, which can cause overfitting and decrease the generalization capability. The generalization capability of LDA is better when compared with the mentioned nonlinear classifiers [16]. It assigns instances to classes with a projection of the feature values to a new subspace. The classification performance of RF is typically higher for high-dimensional data. However, the decision tree classifier has an issue of overfitting, which can be alleviated by assigning class labels with the results of several decision trees in the RF classifier.

SVM is commonly used for recognizing emotions. It is applied to different public datasets. Around 0.6–0.7 F1-scores for recognizing arousal and valence in two-class classification in the ASCERTAIN dataset [126] and 45%–50% accuracy for three-class arousal and valence classification in the MAHNOB-HCI dataset [144] are reported. LDA is another widely used classifier for recognizing emotions. It achieved around 80% accuracy for differentiating stress from the cognitive load by analyzing the EDA signal [145] and around 80% accuracy for recognizing two-class valence and arousal levels from

the ECG signals. Due to its suitability to high dimension data, RF was also tested for emotion recognition, and it achieved around 70% accuracy for two-class arousal and valence classification, and outperformed other traditional methods [148]. Wen et al. [147] applied RF to recognize emotional states, such as baseline, amusement, anger, grief, and fear using heart activity, EDA, and blood oxygen saturation signals. They achieved 74% accuracy for quinary classification on their dataset consisting of 477 cases of 101 subjects while watching emotional videos.

B. Deep Learning Approaches

After the improvements in deep learning algorithms, they are also widely used for emotion recognition. The researchers first tested MLP, an artificial neural network that generally outperformed other traditional algorithms. It was among one of the best-performing classifiers [149]. In one of the preliminary works, Wagner et al. [114] applied MLP and compared the results with LDF and kNN. MLP classifier achieved better results than the other classifiers when applied to ECG, EDA, EMG, and respiration data for emotions such as joy and anger (88,64% for valence detection and 94.32% for arousal detection). However, the best-performing classifier changed with the selected emotion and feature selection technique. The MLP classifier was also applied to PPG, EDA, and ACC data for stress level detection and achieved better results (92.15% accuracy for binary stress classification) than LDA, SVM, kNN, logistic regression, and RF [113].

CNN is another type of deep, feed-forward neural network. They achieved significant success in the image domain [16], and recently, researchers have applied them to physiological signals, such as EEG, EMG, and ECG. In one of the preliminary studies, Martinez et al. [157] tested several CNN architectures on BVP and EDA signals for recognizing four emotional states (relaxation, anxiety, excitement, and fun) and achieved better results than using traditional techniques (70% accuracy for fun and excitement and 60% accuracy for relaxation and anxiety). CNNs were also used for automatically extracting high-level features from physiological signals. Kanjo et al. [13] extracted features from the EDA signal using CNN architecture and achieved 95% accuracy and outperformed the usage of handcrafted features (which has 83%) for five-class valence detection. Graph CNNs are also used for recognizing emotions from physiological signals. They are appropriate for the irregular structure of EEG data and can discover the intrinsic relationship between various EEG channels. The graph CNN algorithm achieved higher accuracies with the EEG signals of the SEED dataset reaching 94.24% [158]. After the success of graph CNNs with EEG signals, they were also applied to a combination of physiological signals. Wierciński et al. [150] reported that they achieved 70% accuracy for valence and arousal detection when the graph CNN algorithm was applied to EEG, ECG, and EDA signals

on the AMIGOS dataset. They further stated that EEG alone achieved better accuracy (75% accuracy for arousal and valence detection) compared to the multimodal approach. However, it can be inferred that the performance of graph CNN algorithms for recognizing emotions using physiological signals (except for EEG) is not investigated comprehensively.

In recent years, recurrent neural networks (RNNs) have had remarkable success in various areas, such as speech recognition, language modeling, translation, and image captioning, due to their structure being suitable for time series. LSTM is a special type of RNN capable of learning long-term dependencies and overcoming the vanishing gradients problem of RNN. LSTM is commonly applied to the output of CNN for recognizing emotions using CNN as an automatic feature extractor [151], [152]. In these studies, Kim and Jo [151] achieved 78.72% and 79.03% for recognizing valence and arousal on the DEAP dataset, and Dar et al. [152] achieved 99.0% accuracy for the AMIGOS dataset and 90.8% for the DREAMER dataset in four class classification (high-arousal, high-valence, low-valence, and low-arousal areas). In some cases, LSTM is directly applied to the raw physiological data [153] and handcrafted feature sets [154]. Awais et al. [153] applied LSTM to a combination of raw signals (i.e., ECG, EMG, BVP, EDA, ST, and respiration) and achieved 97%, 94.2%, 93.9%, and 95.2% accuracies for detecting amusement, boredom, relax, and scared emotions, respectively, on the CASE dataset. On the other hand, Umematsu et al. [154] achieved 83% accuracy in predicting the next day's stress level by applying LSTM to the features obtained from EDA, ST, ACC, mobile phone usage, and location data in their local dataset. RNN variants are the most common classifiers for recognizing emotion levels from physiological signals.

Another important issue for processing time-series signals using deep learning methods is aggregating information from the raw signal by giving more importance to the most relevant parts [155]. The attention mechanism technique employs attention weights to restrict processing to relevant information independent of their distances. Transformers can be regarded as one of the most prosperous attention-based techniques. They have been first implemented for natural language processing (NLP) and employ attention mechanisms to analyze sequences of words and are appropriate for use in other applications, such as time-series forecasting, medical, physiological signal analysis, and human activity recognition [155]. Recent studies also use these architectures for recognizing emotions from physiological signals (see [156] and [155]). Yang et al. [156] combined CNN architectures with conformer blocks and tested them on PPG, EDA, and ST data from the K-Emocon dataset. They achieved 77.37% and 79.42% accuracies for detecting valence and arousal levels. Vazquez et al. [155] tested a transformer model (by combining it with a 1-D CNN) on ECG data of the AMIGOS dataset and achieved 83% accuracy for valence

and 88% accuracy for arousal detection. The transformer architectures achieved promising results on these public datasets.

C. Insights

Deep learning approaches improved the emotion recognition results by analyzing physiological signals on prominent public datasets. However, it is important to note that deep learning approaches require a huge amount of data compared to traditional classifiers. Therefore, if the dataset size and the number of data points are limited, it is advised to use traditional approaches. CNN-based techniques automatize the feature extraction phase, and RNN-based techniques use previous and current data for enhanced predictions. The performance of providing raw data to classifiers, usage of handcrafted features, and automatic feature extraction with CNN depend on the application and data. As an example, CNN requires a larger amount of data for automatically extracting features. If the data are limited, handcrafted features can be used instead of automatically extracted features. Architecture and hyperparameter selection are other challenging tasks for researchers that change with varying applications. It is also important to note that other metrics, such as privacy and explainability, are as crucial as classification performance. To protect the users' privacy, researchers applied differential privacy (DP) [159] and federated learning (FL) [160] approaches with a tradeoff in the performance.

Another issue is the lack of information about the decision-making process of deep learning. By automatizing the feature extraction process with CNNs and using deep learning for classification, the emotion recognition systems have turned into black boxes with high accuracy. Although several studies applied explainable methods for face [161] and speech-based [162] emotion recognition systems, there are only a few explainable AI works for recognizing emotions from physiological signals. For example, Liew et al. [163] evaluated and analyzed contributions of individual features and feature interactions for representing human emotions by employing the Shapley additive explanation values method on multimodal DEAP, DREAMER, and AMIGOS datasets.

VII. PRACTICAL APPLICATIONS OF EMOTION RECOGNITION STUDIES

Emotion recognition systems have a wide range of applications in various fields, such as the workplace, education, automobile, healthcare, and other areas. By continuously monitoring physiological signals in real time, these systems can detect and interpret emotions, and adapt their responses and actions accordingly.

A. Workplace and Office

Researchers aimed to recognize emotions in workplaces, considering that individuals spend a significant amount of time in these settings, and given that emotion recognition

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

Table 4 Summary of Practical Applications Using Physiological Signals That Use Emotion Recognition Systems. V: Valence and A: Arousal

Study	Practical Application	Signals	Classifier	Accuracy
[164]	Workplace Emotion Recognition	BVP, EDA, ST and respiration	RF	98% V
[165]	Worker Anxiety in Dangerous Line of Works	EDA	Statistical Methods	N/A
[166]	Worker Perceived Risk Evaluation	PPG, EDA, ST	SVM	81.2%
[167]	Worker Emotion Recognition during Human Robot Collaboration	ECG, EMG, EDA	kNN, Regression Tree, Bayesian Network, SVM	∼80%
[154]	Student Stress Recognition	EDA, ST, ACC, mobile phone usage, location	LSTM	83%
[168]	Emotion Aware E-Learning System	EDA, PPG and EEG	SVM	86%
[169]	Anxiety Detection in Phobic People	BVP	SVM	76%
[170]	Emotion Aware Virtual Therapist	ECG, EEG and EDA	SVM, LDA, kNN, RF	85-98%
[171]	Pain Recognition	ECG, EDA	Multi Task NN, SVM, RF	∼80%
[172]	Pain Recognition	ECG, EDA	CNN-LSTM	95%
[173]	Depression Detection	EDA, BVP, HRV	Statistical Analysis	N/A
[174]	Depression Detection	EEG	kNN, DT, SVM, NN	79%
[175]	Emotion Recognition of Children with Autism	ECG, EDA, respiration and ST	kNN, SVM, LDA	∼80%
[128]	Driver Stress Level Detection	ECG, EMG, EDA and respiration	LDA	97 %
[176]	Driver Drowsiness Detection	EEG and ECG	RF	81.4%
[177]	Emotion Aware Video Game Design	ECG, EDA, EMG, respiration, ACC, facial and game screen video	SVM, DT, RF	∼70%
[178]	Emotion Aware Video Game Design	ECG, EDA, EMG, BVP and respiration	Deep NN	∼80% V, ∼80% A
[179]	Equine-assisted therapy (EAT)	ECG, EMG, EDA	SVM, LDA, kNN	78% V, 65% A
[180]	Emotion Aware Tourism Experience	EDA	SPSS Analysis	N/A

systems have the potential to improve workers' well-being, reduce work-related accidents, and enhance productivity (see Table 4). In a study by Al Jassmi et al. [164], the researchers explored the relationship between workers' emotions and their productivity, discovering a moderate positive correlation. This prompted them to develop an automated emotion recognition system for construction workers. By utilizing blood volume pulse (BVP), RR, galvanic skin response (GSR), skin temperature (TEMP), and HR data, they were able to accurately detect workers' positive and negative emotions with a 98% accuracy rate, using an RF classifier. The authors conducted a four-day field experiment at a prefabricated stone construction factory to collect data for their study. Using virtual reality technology, Sun et al. [165] designed environments with varying heights, including ground level, 4 m, and 8 m. The researchers found a statistically significant difference in anxiety levels as indicated by EDA signals in response to the different heights. In a subsequent study, Lee et al. [166] utilized PPG, EDA, and ST signals to determine workers' perceived risk levels in hazardous occupations. They applied an SVM classifier and obtained an 81.2% accuracy rate for binary classification.

With the increasing adoption of robotics technology in factories, there has been a significant focus on developing and improving the accuracy of these systems. However, researchers have also explored the emotions of workers during human–robot interactions, given that this is a relatively new experience for workers with a fear of robots potentially replacing them. Liu et al. [167] used various classification models, including kNN, regression tree, Bayesian network, and SVM, to analyze physiological signals (ECG, EDA, and EMG) and recognize five distinct emotions (anxiety, engagement, boredom, frustration, and anger) during interactions, achieving an accuracy of around 80%.

B. Automotive Environment

Given that people spend a significant amount of time in their cars, monitoring their emotions and intervening when necessary could help reduce accidents, injuries, and fatalities. Emotion research in automotive environments has focused on identifying and mitigating conditions, such as fatigue, confusion, nervousness, distraction, and stress that can impact drivers in automotive environments [181]. Nonintuitive user interfaces, complex navigation systems, ambiguous traffic signs, and intricate routing can cause confusion. Nervousness is another affective state characterized by heightened arousal levels and can negatively impact decision-making processes. Li and Ji [182] proposed a method based on dynamic Bayesian networks to detect fatigue, confusion, and nervousness from physiological signals, facial features, and gaze data from both synthetic and real-world environments.

Earlier stages of fatigue can impact driving performance by reducing physiological vigilance/arousal, slowing down sensorimotor processes, and impairing information processing, leading to slower reaction times and decreased

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

ability to respond to urgent situations, ultimately increasing the risk of accidents. As a result, fatigue has been extensively studied in the automotive environment. Crawford [183] suggested that physiological signals are the most reliable indicators of driver fatigue, which has been corroborated by numerous studies (e.g., [184], [185], and [176]) that use physiological signals to estimate driver fatigue and drowsiness.

Research has shown that increased driver stress, whether short term or long term, can have negative effects on decision-making ability, driver awareness, and reaction times in automotive environments [181]. As a result, there is a growing interest in developing methods to detect stress levels in drivers. In one pioneer study, Healey and Picard [129] presented a method that employed HR, EEG, and respiration data to assess drivers' stress levels. EDA signals were also employed with an LDA classifier to detect driver stress, and around 80% accuracies were obtained [186].

C. Education and e-Learning

Emotion recognition research has found another important application in the field of education, particularly in improving e-learning technologies compared to traditional learning methods. By monitoring the emotions of both teachers and students, emotion-aware e-learning systems have the potential to enhance receptiveness and productivity. Umematsu et al. [154] detected student stress utilizing LSTM classifiers on physiological signals, mobile phone usage, location, and behavioral surveys, achieving 83% accuracy for daily stress level detection. In another study, Shen et al. [168] identified four emotions that commonly arise during learning engagement (confusion, boredom, hopefulness, and engagement) and employed SVM on EDA, PPG, and EEG signals to detect them with 86% accuracy. The performance of the emotion-aware e-learning system was compared with a baseline e-learning scheme. Their experiment prototype offered appropriate interventions based on the emotional state of the learner. The emotion-aware e-learning system was found to be effective in reducing the number of required interventions and improving the effectiveness of the e-learning system.

D. Healthcare

The use of physiological data analysis has demonstrated potential in the identification of mental disorders, such as depression, panic disorder, anxiety, and phobias. Researchers have been focused on detecting fear and phobia automatically using physiological data. In one study, Handouzi et al. [169] exposed participants to anxiogenic (the environment that causes anxiety and fear) virtual environments to identify anxiety levels in phobic individuals. They applied the SVM classifier to BVP data and achieved 76% accuracy in detecting anxiety levels. In another study, Bălan et al. [170] developed an automatic emotion recognition model using SVM, LDA, kNN,

and RF classifiers on the DEAP dataset. The researchers created a smart virtual therapist that recognizes human emotions using physiological signals (EEG, ECG, and EDA) and provides encouragement, suggestions, and adapts its voice parameters to the scenario accordingly.

Pain is a combination of sensory and emotional experiences. It can be difficult for infants, anesthetized patients, and people with speech impairments to communicate their pain. Self-reports have been the traditional method of gathering data from patients with serious illnesses or those who have undergone surgery. Nevertheless, these reports have a subjective nature and may not always be feasible to obtain in real time, such as during surgical procedures. Automated pain assessment can be helpful in alleviating suffering, but more improvements are needed before it can be clinically adopted. Researchers have developed various machine-learning techniques to detect pain and mental illnesses. For example, Lopez-Martinez and Picard [171] attempted detecting pain using a MultiTask Neural Network classifier along with SVM and RF classifiers using ECG and EDA data from the BioVid Heat Pain Database [188] and achieved around 80% accuracy. Subramaniam and Dass [172] achieved 95% accuracy using a CNN-LSTM classifier on the same dataset. Depression is another frequently researched mental illness. Chen et al. [173] investigated the physiological signals of depression patients and control groups while inducing emotions in the laboratory. They computed and presented a significant statistical difference between these groups. Cai et al. [174] produced a physiological dataset that included 213 participants (92 of whom had depression and 121 were normal controls). EEG signals were recorded during the resting state and sound stimulation. They applied kNN, decision tree, SVM, and NN classifiers and obtained a maximum of 79% accuracy for detecting depression. In addition, emotion recognition systems have the potential to enhance the quality of life for individuals with various genetic disorders, such as autism, by aiding in the perception and expression of emotions. Sarabadani et al. [175] induced emotions using images on 15 children diagnosed with autism disorder and collected ECG, EDA, respiration, and ST. They detected binary arousal and valence with around 80% accuracy using an ensemble of kNN, LDA, and SVM classifiers. After detecting the emotions of children with autism disorder, some studies also try to intervene with social robots to teach them to perceive and express emotions better [189]. Another interesting application is the detection of emotion during equine-assisted therapy (EAT), which is a therapy type that uses horse-related activities to alleviate mental health issues. Althobaiti et al. [179] applied SVM, LDA, and kNN classifiers to ECG, EMG, and EEG signals recorded during horse-related activities (looking, grooming, and leading) and achieved an F1-score of 78.27% for valence and 65.49% for arousal detection.

When it comes to emotion regulation, individuals often regulate their emotions and other affective states passively.
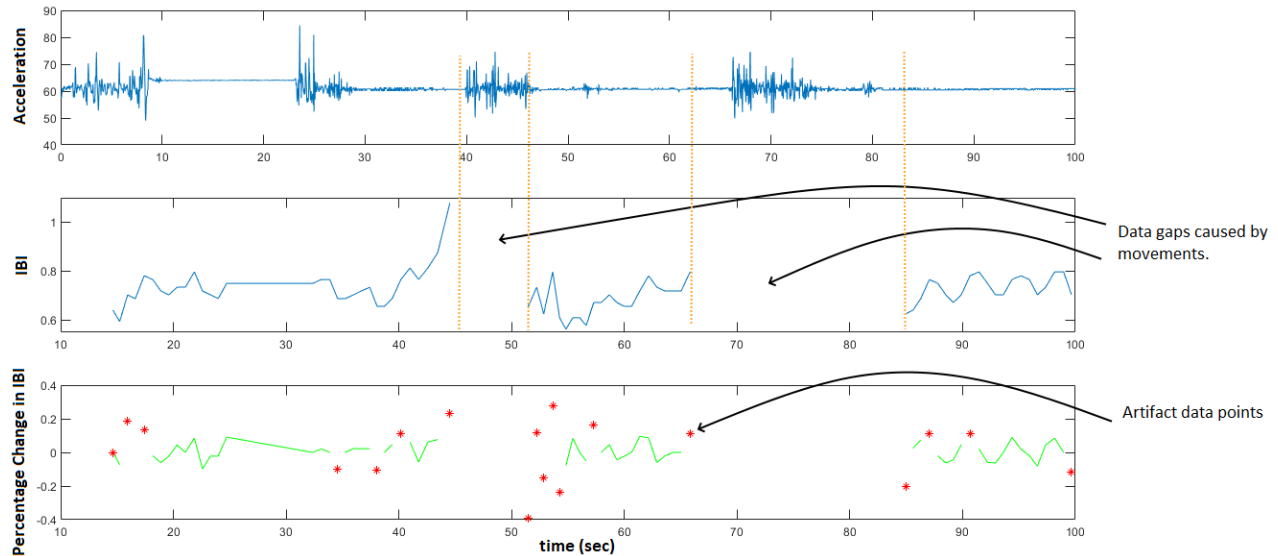
Fig. 6. *Heart activity signal obtained from a PPG sensor during a study in the wild. Artifacts and data gaps in heart activity signal can be seen when the subject moves (during an increased activity in the acceleration signal) [187].*

However, certain regulation strategies, such as emotion suppression [36], are known to have a more negative impact than a positive impact. Technology can help people identify appropriate strategies through experimentation. While research has shown that emotion regulation is often hard to detect with a visual inspection, physiological modalities are promising in validating the efficacy of the interventions for regulation. Slow, controlled breathing has been known to regulate affect positively. Several vibrotactile methods, such as Doppel [190], ambienBeat [191], and BoostMeUp [192], have been introduced as the means for affect regulation. They provide heartbeat-like stimulation on the wrist. Physiological measurements of respiration and HRV due to controlled breathing induced by these devices are measured. There are more applications to monitor breathing and encourage slower breathing during daily activities, such as Just Breathe [193] and Calm Commute [194]. Furthermore, skin conductance can measure the extent of regulation using such applications. However, more studies are required for assessing the effectiveness and validity of such technological interventions and the affect regulation strategies adopted by the individuals [195].

E. Other Applications

The application of emotion recognition is not restricted to industries such as the workplace, automotive, healthcare, and education. It also has a significant role in enhancing user experience, such as in the field of affective gaming, where emotions are detected to enhance the gaming experience of players. Yang et al. [177] detected anger, boredom, frustration, happiness, and fear emotions during the FIFA2016 video game by analyzing ECG, EDA, EMG, respiration, and body movement with

a three-axis accelerometer, facial recording, and game screen recording, and achieved around 70% accuracy with SVM, decision tree, and RF classifiers. In another study, AlZoubi et al. [178] applied deep neural networks to ECG, EDA, EMG, BVP, and respiration signals collected during PlayerUnknown's Battlegrounds (PUBG) gameplay. They achieved around 80% for detecting arousal and valence levels. Emotions were also analyzed during touristic travels to design and manage tourism experiences better. Kim and Fesenmaier [180] monitored the EDA signals of two travelers during their touristic visit to Philadelphia (the United States of America) and demonstrated the changes in signals in different activities.

VIII. RESEARCH ISSUES FOR EMOTION RECOGNITION IN THE WILD

Emotion recognition in the wild or real-world settings involves detecting and identifying emotions in uncontrolled and unpredictable environments. However, several challenges and limitations must be overcome to achieve accurate emotion recognition in such scenarios, including device limitations, data quality concerns (as depicted in Fig. 6), labeling difficulties, privacy considerations, and more. A few of the challenges are described in the following.

A. Issues Related to Devices

1) Selection of Unobtrusive Devices and Access to Raw Data: In order to develop an emotion recognition system suitable for everyday use, one should employ unobtrusive devices, such as smart bands, watches, or straps that can be worn without much discomfort (refer Fig. 7 for examples of unobtrusive wrist-worn devices). However, most of the renowned commercial smart band/watch providers, such

Fig. 7. *Medical-grade devices are shown in the top row. At the bottom, unobtrusive wrist-worn devices are demonstrated [68].*

as Apple Watch, Fitbit, and Microsoft Band 2 [Microsoft ceased support for software development kit (SDK)], do not provide access to raw data for research purposes. After the release of Samsung Galaxy Gear S3, Samsung stopped providing IBI data, which was used for HRV feature calculation, and instead started providing only HR data. Often, the devices provide processed data and insights related to the user's health via their proprietary algorithms and applications rather than providing raw data for research purposes. When researchers aim to develop a multimodal system that includes multiple physiological modalities, such as HRV, EDA, ACC, ST, and BVP, the options for unobtrusive smart bands become more limited. As a result, researchers are often directed toward expensive, research-oriented bands, such as Empatica E3, E4, and Q sensor instead of off-the-shelf commercial bands.

2) Battery Life: Continuous data from sensors are necessary for monitoring the mental health of individuals in their daily life. However, unobtrusive smart bands or watches have limitations when it comes to battery life. When all sensors are active, the state-of-the-art batteries of these devices can only endure for a few hours. In our tests with devices that provide raw data, Samsung Gear S, S2, and S3 lasted around 4 h, while Microsoft Band 2 (with the latest SDK before support ceased) lasted approximately 8 h [113]. Empatica E4 wristband (a research-oriented band with no display) lasted longer than these commercial devices, with a duration of about 48 h as stated on the website [196]. Commercial devices need to be charged at least once a day, which makes users hesitant to use them in everyday life. This limitation forces researchers to develop more energy-efficient emotion monitoring methods.

3) Data Quality and Artifacts: Unobtrusive smart bands offer lower data quality and lower sampling frequencies compared to medical-grade systems. They are more susceptible to artifacts, which can complicate the decision process of affect recognition systems. To develop a robust system, modality-specific artifact detection and removal algorithms should be developed. Furthermore, since the movement of the wrist is almost unrestricted, data gaps can occur during intense activity. To address this issue, researchers need to investigate the characteristics of

modalities and select the most appropriate interpolation technique (i.e., one that captures the modality characteristics) to fill in the gaps.

B. Issues Related to Data Annotation

1) Reliability of Self-Report Questionnaires and Emotion Awareness: To train supervised machine learning algorithms, the physiological data require the ground truth depicting emotions and their intensity. In laboratory experiments, researchers may establish the intended emotion and intensity level of the stimulus as the ground truth. The ground truth for emotions outside the laboratory is typically obtained through ecological momentary assessment, such as self-report questionnaires, as the context and induced emotion level of participants in their daily lives are unknown to the experimenter. However, the reliability of self-reports is questionable because they are subjective and dependent on factors such as the individual, culture, and gender, as described in Section V. In addition, some individuals may try to conceal their true emotional state in experiments, or they may have difficulty accessing and expressing their own emotions. When considering a general model capable of recognizing the emotions of all people, subjective self-reports can decrease accuracy. Furthermore, self-reports are challenging to obtain frequently in real time as the emotions occur, leading to delays in labeling. This can result in the loss of valuable information and affect the accuracy of the emotion recognition model.

2) Necessity for a Substantial Amount of Labeling: Emotion recognition studies in the wild rely on self-reports collected from users as the ground truth. Although more frequent and correct labels can result in better-trained models, it is challenging for participants to provide self-reports frequently and accurately during their daily routines as this process is time-consuming and demands increased compliance from the participants. Therefore, researchers try to balance this out by finding optimal intervals for collecting self-reports without causing significant inconvenience to users. Machine learning methods involving deep learning generally outperform traditional methods, but they require a significant amount of labeled training data for robust models. This further increases the demand for annotated data. Recently, semisupervised methods (SSMs) have been proposed for decreasing the need for labels. These methods can generate labels for unannotated data points by making use of the existing labeled data. Although researchers recently started using SSM techniques for emotion recognition [197], their use in research is still limited.

C. Issues Related to Emotion Classes

1) Division of Self-Report Scales Into Classes: Self-report collection in the wild involves Likert or Self-Assessment Manikin (SAM) scales with different resolutions. After the data collection, the scale is divided into a number of emotion levels or classes for emotion recognition. However,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

defining a general threshold for dividing low and high levels of emotions is challenging, given the subjectivity of self-reports and the potential for variation in baselines across individuals. A fixed threshold might decrease the performance of affect recognition systems. In the literature, a common technique is to use a fixed threshold, which can be calculated as the number of scales/the number of classes. Suppose that we used a 10 Likert scale for emotion detection. We want to detect two-class emotion levels. If we use a fixed threshold of "5" and decide the emotion level accordingly, we might misclassify the people with enduring high emotion levels and classify all their data as high emotion. However, investigating the baseline with questionnaires and increasing or decreasing their baselines dynamically will improve the performance. In other words, person-specific thresholds might increase the accuracy. Automatic clustering methods, such as K-means clustering, can also be employed to assign self-reports to the desired number of affect levels.

2) Data Sparsity: As mentioned previously, especially, deep learning algorithms require a huge amount of data for training. Otherwise, they may overfit, learn the noise in the data, and cannot be generalized to other applications. In order to overcome this issue, researchers first try to increase the amount of data synthetically. In a recent study, Nita et al. [198] augmented an ECG dataset with a considerable amount of representative ECG samples that were created by randomizing, concatenating, and resampling realistic ECG signals in the DREAMER dataset. By applying a seven-layer CNN classifier, they achieved an accuracy of 95.16% to detect valence, 85.56% for arousal, and 77.54% for dominance and increased the baseline (without data augmentation) drastically. When the local dataset size is relatively small, another technique is applying deep transfer learning (DTL) techniques from prominent large datasets. In DTL, parameters are learned from a relatively large dataset, and they are adapted to the local dataset. In the literature, DTL techniques were applied from the SEED dataset to the DREAMER dataset, and it is reported that DTL is beneficial in comparison to traditional machine learning techniques. Another problem occurs when data are imbalanced in terms of class labels. Especially in the wild, datasets have fewer negative labels than positive labels. In this case, machine learning algorithms have the tendency to classify data points as the majority classes. In order to avoid this issue, researchers can randomly undersample the majority class and balance the dataset. Another technique is called Synthetic Minority Oversampling Technique (SMOTE), and it increases the size of the minority class by creating synthetic data points.

D. Privacy and Ethical Concerns

Collecting and processing physiological signals require careful consideration as they carry sensitive, health-related information about individuals. Privacy and ethical concerns must be addressed in two stages. The first stage involves the data collection process, which demands specific procedures. Ethical approval for the experiment protocol and informed consent must be obtained from the ethical boards before collecting data from participants. During the data collection, informed consent must be obtained from the participants by clarifying the purpose of the study, the data that will be collected, and the rights of participants with respect to their data and contact persons, both verbally and in writing. Another crucial ethical element related to the experiments is the emotional stimulus. Inducing negative effects (i.e., anger, stress, and sadness) can be challenging because of the ethical constraints [199]. Generally, researchers use low-intensity emotion induction techniques, namely, IAPS images, movie clips, emotional videos, and music, which are approved by the ethical committees. However, this can create a problem when the models cannot learn high-intensity responses as in daily lives since they are not present in the training data [11].

Furthermore, privacy must be ensured during storing and processing of the data. The most important step is the anonymization of the information. Instead of anonymization, researchers sometimes also applied pseudonymization in which data without personal information are stored along with a table that maps the subjects to their identity. However, without accessing this table, it is impossible to get the identity of the subjects. The following example can be provided to clarify the difference between anonymization and pseudonymization. In pseudonymization, P32's physiological data and a table that maps P32 to the participant's real name are stored separately. On the other hand, in anonymization, it is stated that a patient has the corresponding physiological data, and there is no way to get the identity of this patient. Both techniques are allowed in different privacy protection laws, such as General Data Protection Regulation.

The second stage pertains to the implementation of emotion recognition technologies in real life. A crucial concern is the access rights to physiological data and outcomes. For instance, if employers can access their employees' stress, anxiety, and workload data, they may exploit it unethically. Potential misuse may include assigning more tasks to workers with low mental workloads or terminating those with intense anxiety or stress. Another instance is that health insurance companies can determine the likelihood of mental health disorders and charge higher contribution premiums to those affected. In addition, the presence of hidden biases in the training data used for these systems can lead to unfair or discriminatory outcomes. These examples highlight the significance of ensuring user data privacy and addressing ethical concerns.

E. Privacy Preserving Machine Learning for Affect Recognition From Physiological Signals

Researchers proposed FL and DP approaches for addressing privacy concerns that occurred during machine learning processes. The FL approach uploads the model

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

parameters obtained from the sensitive physiological data instead of the data itself [160]. Although FL has been widely applied for facial features and speech for affect recognition [200], [201], [202], it is rarely used for recognizing affects from physiological signals. Can and Ersoy [160] applied FL for predicting binary perceived stress using heart activity. Each client trained an MLP classifier on local data and shared model parameters for each update. The parameters were then averaged by using the FedAvg algorithm. FL was also applied to multimodal physiological signals. Nandi and Xhafa [203] developed an FL-based Fed-ReMECS framework for recognizing arousal and valence levels. They validated their neural network-based FL approach on EDA and respiration data from the DEAP dataset. In these studies, researchers applied FL without sacrificing the affect recognition performances.

Although FL improved the process of training models in terms of privacy, the privacy vulnerabilities of the stochastic gradient descent (SGD) algorithm remain unsolved. The DP mechanism can be explained as injecting noise into each model client or server, perturbing the updates, and restricting gradient leakage between client and server [204]. DP can be applied alone without FL settings. In a physiological signal-based activity recognition case, the noise is added to the data directly so that personal information is lost, but activity data can still be used by compromising on the performance to an extent [159]. It further improved the privacy vulnerabilities when applied together with FL on speech emotion recognition tasks [205]. However, a combination of FL and DP has not been applied to the physiological data for recognizing emotions yet.

F. Generalizability

Another issue is the generalizability of the emotion recognition research. Unfortunately, most of the studies are published on private datasets, which makes it difficult to apply new techniques to these datasets and creates a question about repeatability. On the other hand, as previously mentioned in Section V, many of the current datasets were collected in controlled laboratory settings with artificial stimuli, such as watching movie clips or listening to music. It is widely known that emotional responses in such laboratory environments can differ from those in natural daily life situations where the stimuli may be more personal and subjectively appraised with greater intensity [11]. Furthermore, since most of the research is conducted at universities, participants are generally college students of a certain age. However, if these algorithms are applied to the general population, the participant should be selected from different ages, cultures, gender, and social status homogeneously. Liapis et al. [206] examined the effect of gender on stress recognition using EDA signals. They trained gender-specific models and achieved high accuracy for detecting stress (94.80% accuracy for

males and 98.85% for females). They reported that there is a significant difference in how both genders communicate their emotions using the arousal self-reports. On the other hand, they also stated that gender does not have an effect on the EDA signal during subtle human–computer interaction tasks. However, more comprehensive experiments are needed for more accurate conclusions. The research community should encourage people to create more open real-life datasets with this homogeneity. Another state-of-the-art solution to the generalizability and transferability problem of traditional machine learning algorithms (statistical models) is causal representation learning [207]. Although causal representation learning has several possible real-world applications in different fields, such as health care, marketing, political science, and online advertising, and has achieved promising performances, it has not been applied to physiological signals for emotion recognition, but it can solve the abovementioned problems.

The development of accurate and reliable emotion recognition systems for real-world environments is a complex and challenging task. It demands interdisciplinary collaboration and encourages the development of new techniques and methodologies.

IX. CONCLUSIONS AND FUTURE PERSPECTIVES

The purpose of this tutorial was to provide guidance for new researchers entering the field of emotion recognition. It covered the essential steps of developing an emotion recognition system, including understanding the theories of emotion and their regulation, the physiological and psychological basis of emotions, designing scientific experiments for studying emotions, utilizing wearable devices for capturing physiological modalities, identifying prominent features of each modality, and applying both traditional machine learning and deep learning methods for analyzing physiological data.

Emotion elicitation and regulation theories have provided a framework for understanding the factors that contribute to the experience of emotions and their expressions, which can aid in the development of more accurate emotion recognition models. Research has demonstrated that emotions are expressed through various psychological, physiological, and behavioral modalities. Multimodality has been shown to enhance the performance of emotion recognition systems. We emphasize the importance of multimodality and selecting appropriate ones considering their advantages and disadvantages of each modality for specific environments and application goals.

Another crucial consideration is the choice of machine learning techniques. While many studies prioritize performance and accuracy, other important factors, such as privacy and explainability, also need to be taken into account when designing emotion recognition systems. Unfortunately, many existing research works overlook these factors, and it is essential to explicitly address and

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

discuss them during the development and deployment of such systems.

As research progresses toward real-life emotion data collection and recognition, there are several open challenges that need to be addressed, including selecting good-quality unobtrusive devices, handling low-quality data, and using subjective self-reports as ground truth. This tutorial aims to provide the necessary information for future research in addressing these challenges.

In summary, this tutorial covers various aspects from theoretical foundations to practical implementation of emotion recognition systems, especially using physiological signals. By considering the aspects of emotions, utilizing multimodality, and addressing ethical considerations, researchers can develop more robust and effective emotion recognition systems that can contribute to a wide range of applications in fields such as psychology, healthcare, human–computer interaction, and social robotics. ∎

Acknowledgment

REFERENCES

[1] A. Rowe and J. Fitness, "Understanding the role of negative emotions in adult learning and achievement: A social functional perspective," *Behav. Sci.*, vol. 8, no. 2, p. 27, Feb. 2018.

[2] R. Erickson and W. Grove, "Why emotions matter: Age, agitation, and burnout among registered nurses," *Online J. Issues Nursing*, vol. 13, no. 1, pp. 1–12, Oct. 2007.

[3] M. R. Kamdar and M. J. Wu, "PRISM: A data-driven platform for monitoring mental health," in *Proc. Pacific Symp., Biocomputing*, Singapore: World Scientific, Jan. 2016, pp. 333–344.

[4] *2015 Motor Vehicle Crashes: Overview, Traffic Safety Facts: Research Note*, National Highway Traffic Safety Administration, Washington, DC, USA, 2016, pp. 1–9.

[5] L. C. De Silva, T. Miyasato, and R. Nakatsu, "Facial emotion recognition using multi-modal information," in *Proc. Int. Conf. Inf., Commun. Signal Process. (ICICS), Theme, Trends Inf. Syst. Eng. Wireless Multimedia Commun.*, 1997, pp. 397–401.

[6] B. Schuller, G. Rigoll, and M. Lang, "Hidden Markov model-based speech emotion recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 2, Apr. 2003, p. 2.

[7] IDC. (Mar. 10, 2020). *Wearables Unit Shipments Worldwide by Vendor From 2014 to 2019 (in Millions) [Graph]*. Accessed: Feb. 25, 2023. [Online]. Available: https://www.statista.com/statistics/515634/wearables-shipments-worldwide-by-vendor/

[8] B. Zhao, Z. Wang, Z. Yu, and B. Guo, "EmotionSense: Emotion recognition based on wearable wristband," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Oct. 2018, pp. 346–355.

[9] B. Nakisa, M. N. Rastgoo, A. Rakotonirainy, F. Maire, and V. Chandran, "Long short term memory hyperparameter optimization for a neural network based emotion recognition framework," *IEEE Access*, vol. 6, pp. 49325–49338, 2018.

[10] S. Saganowski et al., "Emotion recognition using wearables: A systematic literature review—Work-in-progress," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2020, pp. 1–6.

[11] R. W. Picard, "Automating the recognition of stress and emotion: From lab to real-world impact," *IEEE Multimedia Mag.*, vol. 23, no. 3, pp. 3–7, Jul. 2016.

[12] S. Saganowski, B. Perz, A. Polak, and P. Kazienko, "Emotion recognition for everyday life using physiological signals from wearables: A systematic literature review," *IEEE Trans. Affect. Comput.*, early access, May 20, 2022, doi: 10.1109/TAFFC.2022.3176135.

[13] E. Kanjo, E. M. G. Younis, and C. S. Ang, "Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection," *Inf. Fusion*, vol. 49, pp. 46–56, Sep. 2019.

[14] F. A. Alskafi, A. H. Khandoker, and H. F. Jelinek, "A comparative study of arousal and valence dimensional variations for emotion recognition using peripheral physiological signals acquired from wearable sensors," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 1104–1107.

[15] P. Schmidt, R. Dürichen, A. Reiss, K. Van Laerhoven, and T. Plötz, "Multi-target affect detection in the wild: An exploratory study," in *Proc. 23rd Int. Symp. Wearable Comput.*, Sep. 2019, pp. 211–219.

[16] L. Shu et al., "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, p. 2074, Jun. 2018.

[17] Y. S. Can, N. Chalabianloo, D. Ekiz, J. Fernandez-Alvarez, G. Riva, and C. Ersoy, "Personal stress-level clustering and decision-level smoothing to enhance the performance of ambulatory stress detection with smartwatches," *IEEE Access*, vol. 8, pp. 38146–38163, 2020.

[18] W. James, "What is an emotion?" *Mind*, vol. 9, no. 34, pp. 188–205, 1884. [Online]. Available: http://www.jstor.org/stable/2246769

[19] W. James, *The Principles of Psychology*, vol. 1. New York, NY, USA: Henry Holt, 1890.

[20] W. B. Cannon, "The James–Lange theory of emotions: A critical examination and an alternative theory," *Amer. J. Psychol.*, vol. 39, nos. 1–4, p. 106, Dec. 1927.

[21] S. Schachter, *The Interaction of Cognitive and Physiological Determinants of Emotional State* (Advances in Experimental Social Psychology).

[22] M. B. Arnold, *Emotion and Personality*. New York, NY, USA: Columbia Univ. Press, 1960.

[23] K. R. Scherer, "Appraisal considered as a process of multilevel sequential checking," in *Appraisal Processes in Emotion: Theory, Methods, Research*, vol. 92, no. 120. 2001, p. 57.

[24] R. S. Lazarus, *Psychological Stress and the Coping Process*. New York, NY, USA: McGraw-Hill, 1966.

[25] I. J. Roseman, A. A. Antoniou, and P. E. Jose, "Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory," *Cogn. Emotion*, vol. 10, no. 3, pp. 241–277, 1996.

[26] P. Ekman, R. W. Levenson, and W. V. Friesen, "Autonomic nervous system activity distinguishes among emotions," *Science*, vol. 221, no. 4616, pp. 1208–1210, Sep. 1983.

[27] P. Ekman, "An argument for basic emotions," *Cogn. Emotion*, vol. 6, nos. 3–4, pp. 169–200, May 1992.

[28] R. S. Lazarus and B. N. Lazarus, *Passion and Reason: Making Sense of Our Emotions*. New York, NY, USA: Oxford University Press, 1994.

[29] P. Ekman, *Basic Emotions*. Hoboken, NJ, USA: Wiley, 1999, ch. 3, pp. 45–60. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/0470013494.ch3

[30] A. S. Cowen and D. Keltner, "Self-report captures 27 distinct categories of emotion bridged by continuous gradients," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 38, pp. E7900–E7909, Sep. 2017.

[31] J. A. Russell, "A circumplex model of affect," *J. Pers. Social Psychol.*, vol. 39, no. 6, pp. 1161–1178, Dec. 1980.

[32] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychol., J. Diverse Perspect. Diverse Psychol. Issues*, vol. 14, no. 4, pp. 261–292, Dec. 1996.

[33] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *Amer. Scientist*, vol. 89, no. 4, pp. 344–350, 2001.

[34] D. Sander, D. Grandjean, and K. R. Scherer, "A systems approach to appraisal mechanisms in emotion," *Neural Netw.*, vol. 18, no. 4, pp. 317–352, May 2005.

[35] Y. S. Can and C. Ersoy, "Smart affect monitoring with wearables in the wild: An unobtrusive mood-aware emotion recognition system," *IEEE Trans. Affect. Comput.*, early access, Dec. 27, 2022, doi: 10.1109/TAFFC.2022.3232483.

[36] J. J. Gross, "Emotion regulation: Affective, cognitive, and social consequences," *Psychophysiology*, vol. 39, no. 3, pp. 281–291, May 2002.

[37] H. A. Demaree, B. J. Schmeichel, J. L. Robinson, J. Pu, D. E. Everhart, and G. G. Berntson, "Up- and down-regulating facial disgust: Affective, vagal, sympathetic, and respiratory consequences," *Biol. Psychol.*, vol. 71, no. 1, pp. 90–99, Jan. 2006.

[38] C. R. Harris, "Cardiovascular responses of embarrassment and effects of emotional suppression in a social setting," *J. Pers. Social Psychol.*, vol. 81, no. 5, pp. 886–897, 2001.

[39] J. Zaehringer, C. Jennen-Steinmetz, C. Schmahl, G. Ende, and C. Paret, "Psychophysiological effects of downregulating negative emotions: Insights from a meta-analysis of healthy adults," *Frontiers Psychol.*, vol. 11, p. 470, Apr. 2020.

[40] J. Gratch and S. Marsella, "A domain-independent framework for modeling emotion," *Cogn. Syst. Res.*, vol. 5, no. 4, pp. 269–306, Dec. 2004.

[41] R. J. Davidson, P. Ekman, C. D. Saron, J. A. Senulis, and W. V. Friesen, "Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology: I," *J. Pers. Social Psychol.*, vol. 58, no. 2, pp. 330–341, 1990.

[42] R. J. Davidson, "Anterior cerebral asymmetry and the nature of emotion," *Brain Cogn.*, vol. 20, no. 1, pp. 125–151, Sep. 1992.

[43] J. A. Coan and J. J. B. Allen, "Frontal EEG asymmetry as a moderator and mediator of emotion," *Biol. Psychol.*, vol. 67, nos. 1–2, pp. 7–50, Oct. 2004.

[44] J. T. Larsen, G. G. Berntson, K. M. Poehlmann, T. A. Ito, and J. T. Cacioppo, "The psychophysiology of emotion," in *Handbook of Emotions*, M. Lewis, J. M. Haviland-Jones, and L. F. Barrett, Eds. The Guilford Press, 2008, pp. 180–195.

[45] J. Tarchanoff, "Galvanic phenomena in the human skin during stimulation of the sensory organs and during various forms of mental activity," *Pflugers Arch. fur die Gesamte Physiologiedes Menschen und*

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

der Tiere, vol. 46, no. 1, pp. 46–55, 1890.

[46] J. T. Cacioppo and R. E. Petty, "Electromyograms as measures of extent and affectivity of information processing," *Amer. Psychol.*, vol. 36, no. 5, pp. 441–456, 1981.

[47] G. G. Berntson, J. T. Cacioppo, and K. S. Quigley, "Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications," *Psychophysiology*, vol. 30, no. 2, pp. 183–196, Mar. 1993.

[48] S. E. Taylor, "Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis.," *Psychol. Bull.*, vol. 110, no. 1, pp. 67–85, Jul. 1991.

[49] P. J. Davis, S. P. Zhang, A. Winkworth, and R. Bandler, "Neural control of vocalization: Respiratory and emotional influences," *J. Voice*, vol. 10, no. 1, pp. 23–38, Jan. 1996.

[50] K. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Commun.*, vol. 40, nos. 1–2, pp. 227–256, Apr. 2003.

[51] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 374–393, Jul. 2019.

[52] K. L. Phan, T. Wager, S. F. Taylor, and I. Liberzon, "Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI," *NeuroImage*, vol. 16, no. 2, pp. 331–348, Jun. 2002.

[53] A. R. Damasio et al., "Subcortical and cortical brain activity during the feeling of self-generated emotions," *Nature Neurosci.*, vol. 3, no. 10, pp. 1049–1056, Oct. 2000.

[54] A. R. Hariri, S. Y. Bookheimer, and J. C. Mazziotta, "Modulating emotional responses: Effects of a neocortical network on the limbic system," *NeuroReport*, vol. 11, no. 1, pp. 43–48, Jan. 2000.

[55] C. P. Niemic, "Studies of emotion: A theoretical and empirical review of psychophysiological studies of emotion," *J. Undergraduate Res.*, vol. 1, no. 1, pp. 15–18, Fall 2002.

[56] Y. Jiang, W. Li, M. S. Hossain, M. Chen, A. Alelaiwi, and M. Al-Hammadi, "A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition," *Inf. Fusion*, vol. 53, pp. 209–221, Jan. 2020.

[57] X. Huang et al., "Multi-modal emotion analysis from facial expressions and electroencephalogram," *Comput. Vis. Image Understand.*, vol. 147, pp. 114–124, Jun. 2016.

[58] A. Khosla, P. Khandnor, and T. Chand, "A comparative analysis of signal processing and classification methods for different applications based on EEG signals," *Biocybern. Biomed. Eng.*, vol. 40, no. 2, pp. 649–690, Apr. 2020.

[59] I. Winkler, S. Haufe, and M. Tangermann, "Automatic classification of artifactual ICA-components for artifact removal in EEG signals," *Behav. Brain Functions*, vol. 7, pp. 1–15, Aug. 2011.

[60] I. Stancin, M. Cifrek, and A. Jovic, "A review of EEG signal features and their application in driver drowsiness detection systems," *Sensors*, vol. 21, no. 11, p. 3786, May 2021.

[61] P. A. Abhang, B. W. Gawali, and S. C. Mehrotra, "Technical aspects of brain rhythms and speech parameters," in *Introduction to EEG- and Speech-Based Emotion Recognition*. New York, NY, USA: Academic, 2016, ch. 3, pp. 51–79.

[62] S. Greene, H. Thapliyal, and A. Caban-Holt, "A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health," *IEEE Consum. Electron. Mag.*, vol. 5, no. 4, pp. 44–56, Oct. 2016.

[63] B. Cinaz, B. Arnrich, R. La Marca, and G. Tröster, "Monitoring of mental workload levels during an everyday life office-work scenario," *Pers. Ubiquitous Comput.*, vol. 17, no. 2, pp. 229–239, Feb. 2013.

[64] M. Adnane and A. Belouchrani, "Heartbeats classification using QRS and T waves autoregressive features and RR interval features,"

Expert Syst., vol. 34, no. 6, Dec. 2017, Art. no. e12219.

[65] T. Ruf, "The Lomb–Scargle periodogram in biological rhythm research: Analysis of incomplete and unequally spaced time-series," *Biol. Rhythm Res.*, vol. 30, no. 2, pp. 178–201, Apr. 1999.

[66] V. Houshyarifar and M. C. Amirani, "Early detection of sudden cardiac death using Poincaré plots and recurrence plot-based features from HRV signals," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 25, no. 2, pp. 1541–1553, 2017.

[67] J. Bolea, E. Pueyo, M. Orini, and R. Bailón, "Influence of heart rate in non-linear HRV indices as a sampling rate effect evaluated on supine and standing," *Frontiers Physiol.*, vol. 7, p. 501, Nov. 2016.

[68] Y. S. Can, B. Arnrich, and C. Ersoy, "Stress detection in daily life scenarios using smart phones and wearable sensors: A survey," *J. Biomed. Informat.*, vol. 92, Apr. 2019, Art. no. 103139.

[69] P. Ekman and W. V. Friesen, "Facial action coding system (FACS) [database record]," APA PsycTests, 1978, doi: 10.1037/t27734-000.

[70] B. D. Gelder, "Why bodies? Twelve reasons for including bodily expressions in affective neuroscience," *Phil. Trans. Roy. Soc. B, Biol. Sci.*, vol. 364, no. 1535, pp. 3475–3484, Dec. 2009.

[71] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of EMG signal analysis: Detection, processing, classification and applications," *Biol. Procedures Online*, vol. 8, no. 1, pp. 11–35, Dec. 2006.

[72] J. Chen, T. Ro, and Z. Zhu, "Emotion recognition with audio, video, EEG, and EMG: A dataset and baseline approaches," *IEEE Access*, vol. 10, pp. 13229–13242, 2022.

[73] M. Sharma, J. Darji, M. Thakrar, and U. R. Acharya, "Automated identification of sleep disorders using wavelet-based features extracted from electrooculogram and electromyogram signals," *Comput. Biol. Med.*, vol. 143, Apr. 2022, Art. no. 105224.

[74] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 440–460, Jan. 2022.

[75] C. W. Darrow, "Sensory, secretory, and electrical changes in the skin following bodily excitation," *J. Exp. Psychol.*, vol. 10, no. 3, pp. 197–226, Jun. 1927.

[76] J. Shukla, M. Barreda-Ángeles, J. Oliver, and D. Puig, "Efficient wavelet-based artifact removal for electrodermal activity in real-world applications," *Biomed. Signal Process. Control*, vol. 42, pp. 45–52, Apr. 2018.

[77] W. Chen, N. Jaques, S. Taylor, A. Sano, S. Fedor, and R. W. Picard, "Wavelet-based motion artifact removal for electrodermal activity," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 6223–6226.

[78] S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard, "Automatic identification of artifacts in electrodermal activity data," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 1934–1937.

[79] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "CvxEDA: A convex optimization approach to electrodermal activity processing," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 797–804, Apr. 2016.

[80] S. A. H. Aqajari, E. K. Naeini, M. A. Mehrabadi, S. Labbaf, N. Dutt, and A. M. Rahmani, "PyEDA: An open-source Python toolkit for pre-processing and feature extraction of electrodermal activity," *Proc. Comput. Sci.*, vol. 184, pp. 99–106, Jan. 2021.

[81] C. Kappeler-Setz, *Multimodal Emotion and Stress Recognition*. Zürich, Switzerland: ETH Zürich, 2012.

[82] A. M. Gordon and W. B. Mendes, "A large-scale study of stress, emotions, and blood pressure in daily life using a digital platform," *Proc. Nat. Acad. Sci. USA*, vol. 118, no. 31, Aug. 2021,

Art. no. e2105573118.

[83] M. Chu et al., "Respiration rate and volume measurements using wearable strain sensors," *NPJ Digit. Med.*, vol. 2, no. 1, pp. 1–9, Feb. 2019.

[84] T. Nguyen, H. Okada, Y. Takei, A. Takei, and M. Ichiki, "A band-aid type sensor for wearable physiological monitoring," in *Proc. 21st Int. Conf. Solid-State Sens., Actuators Microsyst. (Transducers)*, Jun. 2021, pp. 1432–1435.

[85] A. A. Alian and H. K. Shelley, "Photoplethysmography," *Best Pract. Res. Clin. Anaesthesiol.*, vol. 28, no. 4, pp. 395–406, 2014.

[86] G. Valenza, A. Lanata, and E. P. Scilingo, "The role of nonlinear dynamics in affective valence and arousal recognition," *IEEE Trans. Affect. Comput.*, vol. 3, no. 2, pp. 237–249, Apr. 2012.

[87] K. Yang et al., "Behavioral and physiological signals-based deep multimodal approach for mobile emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 14, no. 2, pp. 1082–1097, Apr./Jun. 2023.

[88] M. Valstar and M. Pantic, "Fully automatic facial action unit detection and temporal analysis," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop (CVPRW)*, Jun. 2006, p. 149.

[89] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image Vis. Comput.*, vol. 27, no. 6, pp. 803–816, May 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885608001844

[90] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1195–1215, Jul. 2022.

[91] F. Z. Canal et al., "A survey on facial emotion recognition techniques: A state-of-the-art literature review," *Inf. Sci.*, vol. 582, pp. 593–617, Jan. 2022, doi: 10.1016/j.ins.2021.10.005.

[92] M. Ali, A. H. Mosa, F. A. Machot, and K. Kyamakya, "Emotion recognition involving physiological and speech signals: A comprehensive review," in *Recent Advances in Nonlinear Dynamics and Synchronization* (Studies in Systems, Decision and Control), vol. 109, K. Kyamakya, W. Mathis, R. Stoop, J. Chedjou, and Z. Li, Eds. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-319-58996-1_13.

[93] J. Kim and E. André, "Emotion recognition using physiological and speech signal in short-term observation," in *Perception and Interactive Technologies*. Kloster Irsee, Germany: Springer, Jun. 2006, pp. 53–64.

[94] M. B. Akçay and K. Oğuz, "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers," *Speech Commun.*, vol. 116, pp. 56–76, Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167639319302262

[95] J. Wagner et al., "Dawn of the transformer era in speech emotion recognition: Closing the valence gap," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Mar. 31, 2023, doi: 10.1109/TPAMI.2023.3263585.

[96] B. D. Gelder, "Towards the neurobiology of emotional body language," *Nature Rev. Neurosci.*, vol. 7, no. 3, pp. 242–249, Mar. 2006, doi: 10.1038/nrn1872.

[97] A. Mehrabian, *Nonverbal Communication*. Evanston, IL, USA: Routledge, 2017.

[98] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," *IEEE Trans. Affect. Comput.*, vol. 4, no. 1, pp. 15–33, Jan. 2013.

[99] Z. Witkower and J. L. Tracy, "Bodily communication of emotion: Evidence for extrafacial behavioral expressions and available coding systems," *Emotion Rev.*, vol. 11, no. 2, pp. 184–193, May 2018, doi: 10.1177/1754073917749880.

[100] L. D. Lopez, P. J. Reschke, J. M. Knothe, and E. A. Walle, "Postural communication of emotion: Perception of distinct poses of five discrete emotions," *Frontiers Psychol.*, vol. 8, p. 710, May 2017, doi: 10.3389/fpsyg.2017.00710.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

[101] F. Ahmed, A. S. M. H. Bari, and M. L. Gavrilova, "Emotion recognition from body movement," *IEEE Access*, vol. 8, pp. 11761–11781, 2020, doi: 10.1109/ACCESS.2019.2963113.

[102] J. Wu, Y. Zhang, S. Sun, Q. Li, and X. Zhao, "Generalized zero-shot emotion recognition from body gestures," *Appl. Intell.*, vol. 52, pp. 8616–8634, Jun. 2022.

[103] S. Xu et al., "Emotion recognition from gait analyses: Current research and future directions," *IEEE Trans. Computat. Social Syst.*, early access, Dec. 5, 2022, doi: 10.1109/TCSS.2022.3223251.

[104] V. Mishra et al., "Investigating the role of context in perceived stress detection in the wild," in *Proc. ACM Int. Joint Conf. Int. Symp. Pervasive Ubiquitous Comput. Wearable Comput.*, Oct. 2018, pp. 1708–1716.

[105] P. Kostopoulos, A. I. Kyritsis, M. Deriaz, and D. Konstantas, "Stress detection using smart phone data," in *eHealth 360°*. Berlin, Germany: Springer, 2017, pp. 340–351.

[106] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *Proc. Hum. Assoc. Conf. Affect. Comput. Intell. Interact.*, Sep. 2013, pp. 671–676.

[107] Y. S. Can et al., "Real-life stress level monitoring using smart bands in the light of contextual information," *IEEE Sensors J.*, vol. 20, no. 15, pp. 8721–8730, Aug. 2020.

[108] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of EEG signals and facial expressions for continuous emotion detection," *IEEE Trans. Affect. Comput.*, vol. 7, no. 1, pp. 17–28, Jan. 2016.

[109] D. Fabiano and S. Canavan, "Emotion recognition using fused physiological signals," in *Proc. 8th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Sep. 2019, pp. 42–48.

[110] D. Shin, D. Shin, and D. Shin, "Development of emotion recognition interface using complex EEG/ECG bio-signal for interactive contents," *Multimedia Tools Appl.*, vol. 76, no. 9, pp. 11449–11470, May 2017.

[111] S. Wang, J. Du, and R. Xu, "Decision fusion for EEG-based emotion recognition," in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, vol. 2, Jul. 2015, pp. 883–889.

[112] M. Umair, N. Chalabianloo, C. Sas, and C. Ersoy, "HRV and stress: A mixed-methods approach for comparison of wearable heart rate sensors for biofeedback," *IEEE Access*, vol. 9, pp. 14005–14024, 2021.

[113] Y. S. Can, N. Chalabianloo, D. Ekiz, and C. Ersoy, "Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study," *Sensors*, vol. 19, no. 8, p. 1849, Apr. 2019.

[114] J. Wagner, J. Kim, and E. André, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2005, pp. 940–943.

[115] D. Huron and J. K. Vuoskoski, "On the enjoyment of sad music: Pleasurable compassion theory and the role of trait empathy," *Frontiers Psychol.*, vol. 11, p. 1060, May 2020.

[116] Y. S. Can, D. Gokay, D. R. Kılıç, D. Ekiz, N. Chalabianloo, and C. Ersoy, "How laboratory experiments can be exploited for monitoring stress in the wild: A bridge between laboratory and daily life," *Sensors*, vol. 20, no. 3, p. 838, Feb. 2020.

[117] E. Grande, "From physiological signals to emotions: An integrative literature review," Bachelor Thesis, Universität Stuttgart, Stuttgart, Germany, 2022, no. 44.

[118] A. M. Kring and A. H. Gordon, "Sex differences in emotion: Expression, experience, and physiology," *J. Pers. Social Psychol.*, vol. 74, no. 3, pp. 686–703, 1998.

[119] J. J. Gross and R. W. Levenson, "Hiding feelings: The acute effects of inhibiting negative and positive emotion," *J. Abnormal Psychol.*, vol. 106, no. 1, pp. 95–103, Feb. 1997.

[120] S. Koelstra et al., "DEAP: A database for emotion analysis using physiological signals," *IEEE Trans.*

[121] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, Jan. 2012.

[122] S. Katsigiannis and N. Ramzan, "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 1, pp. 98–107, Jan. 2018.

[123] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in *Proc. 20th ACM Int. Conf. Multimodal Interact.*, Oct. 2018, pp. 400–408.

[124] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "AMIGOS: A dataset for affect, personality and mood research on individuals and groups," *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 479–493, Apr. 2021.

[125] K. Sharma, C. Castellini, E. L. van den Broek, A. Albu-Schaeffer, and F. Schwenker, "A dataset of continuous affect annotations and physiological signals for emotion analysis," *Sci. Data*, vol. 6, no. 1, pp. 1–13, Oct. 2019.

[126] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe, "ASCERTAIN: Emotion and personality recognition using commercial sensors," *IEEE Trans. Affect. Comput.*, vol. 9, no. 2, pp. 147–160, Apr. 2018.

[127] S. Carvalho, J. Leite, S. Galdo-Álvarez, and Ó. F. Gonçalves, "The emotional movie database (EMDB): A self-report and psychophysiological study," *Appl. Psychophysiol. Biofeedback*, vol. 37, no. 4, pp. 279–294, Dec. 2012.

[128] S. Schneegass, B. Pfleging, N. Broy, F. Heinrich, and A. Schmidt, "A data set of real world driving to assess driver workload," in *Proc. 5th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.*, Oct. 2013, pp. 150–157.

[129] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.

[130] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, 2001.

[131] X. Shui, M. Zhang, Z. Li, X. Hu, F. Wang, and D. Zhang, "A dataset of daily ambulatory psychological and physiological recording for emotion research," *Sci. Data*, vol. 8, no. 1, pp. 1–12, Jun. 2021.

[132] B. Mahesh, T. Hassan, E. Prassler, and J. Garbas, "Requirements for a reference dataset for multimodal human stress detection," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2019, pp. 492–498.

[133] B. M. Kudielka, D. H. Hellhammer, and C. Kirschbaum, "Ten years of research with the trier social stress test–revisited," in *Social Neuroscience: Integrating Biological and Psychological Explanations of Social Behavior*, E. Harmon-Jones and P. Winkielman, Eds. The Guilford Press, 2007, pp. 56–83.

[134] S. S. Dickerson and M. E. Kemeny, "Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research," *Psychol. Bull.*, vol. 130, no. 3, pp. 355–391, 2004.

[135] J. J. Gross, "The emerging field of emotion regulation: An integrative review," *Rev. Gen. Psychol.*, vol. 2, no. 3, pp. 271–299, Sep. 1998.

[136] S. Scheibe, G. Sheppes, and U. M. Staudinger, "Distract or reappraise? Age-related differences in emotion-regulation choice," *Emotion*, vol. 15, no. 6, pp. 677–681, 2015.

[137] M. N. Shiota and R. W. Levenson, "Effects of aging on experimentally instructed detached reappraisal, positive reappraisal, and emotional behavior suppression.," *Psychol. Aging*, vol. 24, no. 4, pp. 890–900, Dec. 2009.

[138] J. J. Gross, "Antecedent- and response-focused

emotion regulation: Divergent consequences for experience, expression, and physiology," *J. Pers. Social Psychol.*, vol. 74, no. 1, pp. 224–237, 1998.

[139] J. L. Tsai, R. W. Levenson, and K. McCoy, "Cultural and temperamental variation in emotional response," *Emotion*, vol. 6, no. 3, pp. 484–497, 2006.

[140] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, "G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behav. Res. Methods*, vol. 39, no. 2, pp. 175–191, May 2007.

[141] R. V. Krejcie and D. W. Morgan, "Determining sample size for research activities," *Educ. Psychol. Meas.*, vol. 30, no. 3, pp. 607–610, Sep. 1970.

[142] W. Cochran, W. Cochran, and A. Bouclier, *Sampling Techniques* (Wiley Series in Probability and Statistics). Hoboken, NJ, USA: Wiley, 1977.

[143] R. W. Levenson, "The autonomic nervous system and emotion," *Emotion Rev.*, vol. 6, no. 2, pp. 100–112, Mar. 2014.

[144] Y.-L. Hsu, J.-S. Wang, W.-C. Chiang, and C.-H. Hung, "Automatic ECG-based emotion recognition in music listening," *IEEE Trans. Affect. Comput.*, vol. 11, no. 1, pp. 85–99, Jan. 2020.

[145] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 410–417, Mar. 2010.

[146] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "ECG pattern analysis for emotion detection," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 102–115, Jan. 2012.

[147] W. Wen, G. Liu, N. Cheng, J. Wei, P. Shangguan, and W. Huang, "Emotion recognition based on multi-variant correlation of physiological signals," *IEEE Trans. Affect. Comput.*, vol. 5, no. 2, pp. 126–140, Apr. 2014.

[148] D. Ayata, Y. Yaslan, and M. E. Kamasak, "Emotion based music recommendation system using wearable physiological sensors," *IEEE Trans. Consum. Electron.*, vol. 64, no. 2, pp. 196–203, May 2018.

[149] V. Kolodyazhniy, S. D. Kreibig, J. J. Gross, W. T. Roth, and F. H. Wilhelm, "An affective computing approach to physiological emotion specificity: Toward subject-independent and stimulus-independent classification of film-induced emotions," *Psychophysiology*, vol. 48, no. 7, pp. 908–922, Jul. 2011.

[150] T. Wierciński, M. Rock, R. Zwierzycki, T. Zawadzka, and M. Zawadzki, "Emotion recognition from physiological channels using graph neural network," *Sensors*, vol. 22, no. 8, p. 2980, Apr. 2022.

[151] B. H. Kim and S. Jo, "Deep physiological affect network for the recognition of human emotions," *IEEE Trans. Affect. Comput.*, vol. 11, no. 2, pp. 230–243, Apr. 2020.

[152] M. N. Dar, M. U. Akram, S. G. Khawaja, and A. N. Pujari, "CNN and LSTM-based emotion charting using physiological signals," *Sensors*, vol. 20, no. 16, p. 4551, Aug. 2020.

[153] M. Awais et al., "LSTM-based emotion detection using physiological signals: IoT framework for healthcare and distance learning in COVID-19," *IEEE Internet Things J.*, vol. 8, no. 23, pp. 16863–16871, Dec. 2021.

[154] T. Umematsu, A. Sano, S. Taylor, and R. W. Picard, "Improving students' daily life stress forecasting using LSTM neural networks," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, May 2019, pp. 1–4.

[155] J. Vazquez-Rodriguez, G. Lefebvre, J. Cumin, and J. L. Crowley, "Transformer-based self-supervised learning for emotion recognition," 2022, *arXiv:2204.05103*.

[156] K. Yang et al., "Mobile emotion recognition via multiple physiological signals using convolution-augmented transformer," in *Proc. Int. Conf. Multimedia Retr.* New York, NY, USA: Association for Computing Machinery, Jun. 2022, pp. 562–570.

[157] H. P. Martinez, Y. Bengio, and G. N. Yannakakis,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Can et al.: Approaches, Applications, and Challenges in Physiological Emotion Recognition

"Learning deep physiological models of affect," *IEEE Comput. Intell. Mag.*, vol. 8, no. 2, pp. 20–33, May 2013.

[158] T. Zhang, X. Wang, X. Xu, and C. L. P. Chen, "GCB-Net: Graph convolutional broad network and its application in emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 379–388, Jan. 2022.

[159] N. Saleheen et al., "mSieve: Differential behavioral privacy in time series of mobile sensor data," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.* New York, NY, USA: Association for Computing Machinery, Sep. 2016, pp. 706–717.

[160] Y. S. Can and C. Ersoy, "Privacy-preserving federated deep learning for wearable IoT-based biomedical monitoring," *ACM Trans. Internet Technol.*, vol. 21, no. 1, pp. 1–17, Feb. 2021.

[161] P. Prajod, T. Huber, and E. André, "Using explainable AI to identify differences between clinical and experimental pain detection models based on facial expressions," in *Proc. Int. Conf. Multimedia Modeling*. Berlin, Germany: Springer, 2022, pp. 311–322.

[162] W. Zhang and B. Y. Lim, "Towards relatable explainable AI with the perceptual process," in *Proc. CHI Conf. Hum. Factors Comput. Syst.* New York, NY, USA: Association for Computing Machinery, Apr. 2022, pp. 1–24.

[163] W. S. Liew, C. K. Loo, and S. Wermter, "Emotion recognition using explainable genetically optimized fuzzy ART ensembles," *IEEE Access*, vol. 9, pp. 61513–61531, 2021.

[164] H. A. Jassmi, S. Ahmed, B. Philip, F. Al Mughairbi, and M. Al Ahmad, "E-happiness physiological indicators of construction workers' productivity: A machine learning approach," *J. Asian Archit. Building Eng.*, vol. 18, no. 6, pp. 517–526, Nov. 2019.

[165] G. Sun, F. Pang, Q. Liu, Y. Lin, L. Xu, and Y. Meng, "Research on recognizable physiological signals of workers working at heights," in *Proc. Int. Conf. Man-Mach.-Environ. Syst. Eng.* Berlin, Germany: Springer, 2020, pp. 75–82.

[166] B. G. Lee, B. Choi, H. Jebelli, and S. Lee, "Assessment of construction workers' perceived risk using physiological data from wearable sensors: A machine learning approach," *J. Building Eng.*, vol. 42, Oct. 2021, Art. no. 102824.

[167] C. Liu, P. Rani, and N. Sarkar, "An empirical study of machine learning techniques for affect recognition in human–robot interaction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Aug. 2005, pp. 2662–2667.

[168] L. Shen, M. Wang, and R. Shen, "Affective e-learning: Using 'emotional' data to improve learning in pervasive learning environment," *J. Educ. Technol. Soc.*, vol. 12, no. 2, pp. 176–189, 2009.

[169] W. Handouzi, C. Maaoui, A. Pruski, and A. Moussaoui, "Objective model assessment for short-term anxiety recognition from blood volume pulse signal," *Biomed. Signal Process. Control*, vol. 14, pp. 217–227, Nov. 2014.

[170] O. Bălan, A. Moldoveanu, and M. Leordeanu, "A machine learning approach to automatic phobia therapy with virtual reality," in *Modern Approaches to Augmentation of Brain Function*. Berlin, Germany: Springer, 2021, pp. 607–636.

[171] D. Lopez-Martinez and R. Picard, "Multi-task neural networks for personalized pain recognition from physiological signals," in *Proc. 7th Int. Conf. Affect. Comput. Intell. Interact. Workshops Demos (ACIIW)*, Oct. 2017, pp. 181–184.

[172] S. D. Subramaniam and B. Dass, "Automated nociceptive pain assessment using physiological signals and a hybrid deep learning network," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3335–3343, Feb. 2021.

[173] Y. Chen, I. Hung, M. Huang, C. Hou, and K. Cheng, "Physiological signal analysis for patients with depression," in *Proc. 4th Int. Conf. Biomed. Eng. Informat. (BMEI)*, vol. 2, Oct. 2011, pp. 805–808.

[174] H. Cai et al., "A pervasive approach to EEG-based depression detection," *Complexity*, vol. 2018, pp. 1–13, Feb. 2018.

[175] S. Sarabadani, L. C. Schudlo, A. A. Samadani, and A. Kushski, "Physiological detection of affective states in children with autism spectrum disorder," *IEEE Trans. Affect. Comput.*, vol. 11, no. 4, pp. 588–600, Oct. 2020.

[176] J. Gwak, M. Shino, and A. Hirao, "Early detection of driver drowsiness utilizing machine learning based on physiological signals, behavioral measures, and driving performance," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 1794–1800.

[177] W. Yang, M. Rifqi, C. Marsala, and A. Pinna, "Physiological-based emotion detection and recognition in a video game context," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.

[178] O. AlZoubi, B. AlMakhadmeh, M. B. Yassein, and W. Mardini, "Detecting naturalistic expression of emotions using physiological signals while playing video games," *J. Ambient Intell. Hum. Comput.*, vol. 14, pp. 1133–1146, Feb. 2023.

[179] T. Althobaiti, S. Katsigiannis, D. West, and N. Ramzan, "Examining human-horse interaction by means of affect recognition via physiological signals," *IEEE Access*, vol. 7, pp. 77857–77867, 2019.

[180] J. Kim and D. R. Fesenmaier, "Measuring emotions in real time: Implications for tourism experience design," *J. Travel Res.*, vol. 54, no. 4, pp. 419–429, Jul. 2015.

[181] C. D. Katsis, G. Rigas, Y. Goletsis, and D. I. Fotiadis, "Emotion recognition in car industry," in *Emotion Recognition: A Pattern Analysis Approach*. New York, NY, USA: Wiley, 2015, pp. 515–544.

[182] X. Li and Q. Ji, "Active affective state detection and user assistance with dynamic Bayesian networks," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 35, no. 1, pp. 93–105, Jan. 2005.

[183] A. Crawford, "Fatigue and driving," *Ergonomics*, vol. 4, no. 2, pp. 143–154, 1961.

[184] Z. Li, K. Jiao, M. Chen, Y. Yang, C. Wang, and S. Qi, "Spectral analysis of heart rate variability as a quantitative indicator of driver mental fatigue," *SAE Trans.*, vol. 111, pp. 249–253, Jan. 2002.

[185] G. Yang, Y. Lin, and P. Bhattacharya, "A driver fatigue recognition model using fusion of multiple features," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, vol. 2, Oct. 2005, pp. 1777–1784.

[186] Y. Liu and S. Du, "Psychological stress level detection based on electrodermal activity," *Behav. Brain Res.*, vol. 341, pp. 50–53, Apr. 2018.

[187] D. Ekiz, Y. S. Can, and C. Ersoy, "Long short-term memory network based unobtrusive workload monitoring with consumer grade smartwatches," *IEEE Trans. Affect. Comput.*, vol. 14, no. 2, pp. 895–905, Apr./Jun. 2021.

[188] S. Walter et al., "The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system," in *Proc. IEEE Int. Conf. Cybern. (CYBCO)*, Jun. 2013, pp. 128–131.

[189] K. D. Bartl-Pokorny et al., "Robot-based intervention for children with autism spectrum disorder: A systematic literature review," *IEEE Access*, vol. 9, pp. 165433–165450, 2021.

[190] R. T. Azevedo, N. Bennett, A. Bilicki, J. Hooper, F. Markopoulou, and M. Tsakiris, "The calming effect of a new wearable device during the anticipation of public speech," *Sci. Rep.*, vol. 7, no. 1, p. 2285, May 2017.

[191] K. Y. Choi and H. Ishii, "AmbienBeat: Wrist-worn mobile tactile biofeedback for heart rate rhythmic regulation," in *Proc. 14th Int. Conf. Tangible, Embedded, Embodied Interact.* New York, NY, USA: Association for Computing Machinery, Feb. 2020, pp. 17–30.

[192] J. Costa, F. Guimbretière, M. F. Jung, and T. Choudhury, "BoostMeUp: Improving cognitive performance in the moment by unobtrusively regulating emotions with a smartwatch," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 2, pp. 1–23, Jun. 2019.

[193] P. E. Paredes et al., "Just breathe: In-car interventions for guided slow breathing," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–23, Mar. 2018.

[194] S. Balters, M. L. Mauriello, S. Y. Park, J. A. Landay, and P. E. Paredes, "Calm commute: Guided slow breathing for daily stress management in drivers," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 1, pp. 1–19, Mar. 2020.

[195] P. Miri et al., "Challenges in evaluating technological interventions for affect regulation," *IEEE Trans. Affect. Comput.*, early access, May 19, 2022, doi: 10.1109/TAFFC.2022. 3175687.

[196] (2022). *Empatica Website*. Accessed: Aug. 2022. [Online]. Available: https://www.empatica.com/

[197] J. Luo, Y. Tian, H. Yu, Y. Chen, and M. Wu, "Semi-supervised cross-subject emotion recognition based on stacked denoising autoencoder architecture using a fusion of multi-modal physiological signals," *Entropy*, vol. 24, no. 5, p. 577, Apr. 2022.

[198] S. Nita, S. Bitam, M. Heidet, and A. Mellouk, "A new data augmentation convolutional neural network for human emotion recognition based on ECG signals," *Biomed. Signal Process. Control*, vol. 75, May 2022, Art. no. 103580.

[199] F. Larradet, R. Niewiadomski, G. Barresi, D. G. Caldwell, and L. S. Mattos, "Toward emotion recognition from physiological signals in the wild: Approaching the methodological issues in real-life data collection," *Frontiers Psychol.*, vol. 11, p. 1111, Jul. 2020.

[200] K. Somandepalli et al., "Federated learning for affective computing tasks," in *Proc. 10th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Oct. 2022, pp. 1–8.

[201] S. Latif, S. Khalifa, R. Rana, and R. Jurdak, "Federated learning for speech emotion recognition applications," in *Proc. 19th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, Apr. 2020, pp. 341–342.

[202] P. Chhikara, P. Singh, R. Tekchandani, N. Kumar, and M. Guizani, "Federated learning meets human emotions: A decentralized framework for human–computer interaction for IoT applications," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6949–6962, Apr. 2021.

[203] A. Nandi and F. Xhafa, "A federated learning method for real-time emotion state classification from multi-modal streaming," *Methods*, vol. 204, pp. 340–347, Aug. 2022.

[204] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014.

[205] T. Feng, R. Peri, and S. Narayanan, "User-level differential privacy against attribute inference attack of speech emotion recognition in federated learning," 2022, *arXiv:2204.02500*.

[206] A. Liapis, C. Katsanos, D. Sotiropoulos, M. Xenos, and N. Karousos, "Stress recognition in human–computer interaction using physiological and self-reported data: A study of gender differences," in *Proc. 19th Panhellenic Conf. Informat.*, Oct. 2015, pp. 323–328.

[207] B. Schölkopf et al., "Toward causal representation learning," *Proc. IEEE*, vol. 109, no. 5, pp. 612–634, May 2021.

ABOUT THE AUTHORS

Yekta Said Can received the B.Sc., M.Sc., and Ph.D. degrees from Boğaziçi University, Istanbul, Turkey, in 2012, 2014, and 2020, respectively.

He has worked as a Teaching Assistant at Boğaziçi University for six years during his Ph.D. degree. After obtaining his Ph.D. degree, he worked as a Postdoctoral Researcher in a European Union's Horizon 2020 ERC project (UrbanOccupations) for applying computer vision techniques to retrieve information from historical documents for two years. He is currently working on recognizing emotions and stress at the University of Augsburg, Augsburg, Germany, as a Postdoctoral Researcher. His research interests include biometrics, document analysis, physiological signal processing, affective and wearable computing, and machine learning.

Bhargavi Mahesh received the M.Sc. degree in autonomous systems from the Bonn-Rhein-Sieg University of Applied Sciences, Sankt Augustin, Germany, in 2019. She is currently working toward the Ph.D. degree at the Chair of Human-Centered Artificial Intelligence, University of Augsburg, Augsburg, Germany.

From 2019 to 2022, she worked as a Researcher at the Smart Sensing Electronics Division, Fraunhofer Institute for Integrated Circuits, Erlangen, Germany. Her research interests include affective computing and social signal processing.

Elisabeth André (Senior Member, IEEE) is currently a Full Professor of computer science and the Founding Chair of human-centered artificial intelligence with the University of Augsburg, Augsburg, Germany, where she is also a Co-Speaker of the Bavarian Research Association ForDigitHealth. She has a long track record in multimodal human–machine interaction, embodied conversational agents, social robotics, affective computing, and social signal processing.

Dr. André is a member of the prestigious Academy of Europe, the German Academy of Sciences Leopoldina, and the CHI Academy. Her work has won many awards, including the Gottfried Wilhelm Leibniz Prize, the most important research funding award in Germany. In 2013, she was awarded a European Association for Artificial Intelligence (EurAI) fellowship. In 2019, she was named one of the ten most influential figures in the history of AI in Germany by the National Society for Informatics (GI). From 2019 to 2022, she has served as the Editor-in-Chief of IEEE Transactions on Affective Computing.