IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 19, 2019, accepted September 15, 2019, date of publication September 26, 2019, date of current version October 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2944001

A Review, Current Challenges, and Future Possibilities on Emotion Recognition Using Machine Learning and Physiological Signals

PATRÍCIA J. BOTA[1,2], CHEN WANG[3], ANA L. N. FRED[1,2], (Member, IEEE), AND HUGO PLÁCIDO DA SILVA[1], (Senior Member, IEEE)

[1]Instituto Superior Técnico, Instituto de Telecomunicações, 1049-001 Lisbon, Portugal
[2]Department of Bioengineering, Instituto Superior Técnico, 1049-001 Lisbon, Portugal
[3]Future Media & Convergence Institute (FMCI), Xinhuanet, Beijing 100000, China

Corresponding author: Patrícia J. Bota (patricia.bota@tecnico.ulisboa.pt)

This work was supported in part by the Xinhua Net Future Media Convergence Institute under Project S-0003-LX-18, in part by the Ministry of Economy and Competitiveness of the Spanish Government co-founded by the ERDF (PhysComp project) under Grant TIN2017-85409-P, and in part by the Instituto de Telecomunicações (IT) in the scope of program under Grant UID/EEA/50008/2019.

ABSTRACT The seminal work on Affective Computing in 1995 by Picard set the base for computing that relates to, arises from, or influences emotions. Affective computing is a multidisciplinary field of research spanning the areas of computer science, psychology, and cognitive science. Potential applications include automated driver assistance, healthcare, human-computer interaction, entertainment, marketing, teaching and many others. Thus, quickly, the field acquired high interest, with an enormous growth of the number of papers published on the topic since its inception. This paper aims to (1) Present an introduction to the field of affective computing though the description of key theoretical concepts; (2) Describe the current state-of-the-art of emotion recognition, tracing the developments that helped foster the growth of the field; and lastly, (3) point the literature take-home messages and conclusions, evidencing the main challenges and future opportunities that lie ahead, in particular for the development of novel machine learning (ML) algorithms in the context of emotion recognition using physiological signals.

INDEX TERMS Affective computing, emotion recognition, machine learning, physiological signals, signal processing.

I. INTRODUCTION

Affective computing was defined by Rosalind Picard as the computing that relates to, arises from, or influences emotions [1]. This emerging field focuses on better understanding the psychophysiological phenomena underlying the ways in which humans recognise, interpret and simulate emotional states [2]. Therefore, it is a multidisciplinary field of research spanning the areas of computer science, psychology, and cognitive science. Emotions possess a nuclear role in human behaviour, exerting a powerful influence in mechanisms such as perception, attention, decision making and learning. Thus, understanding emotional states is essential to understand human behaviour, cognition and intelligence [1].

The field of affective computing presents applications in many areas, including automated driver assistance- through alert systems monitoring the user sentic state by means of physiological signals. The system could be capable of warning the user if he is sleepy, unconscious or unhealthy to drive, lowering the speed or stopping the car if necessary, towards a more safe and secure driving experience. In a driving setting, the user's physiological signals could be read unobtrusively and pervasively through non-intrusive techniques integrated into components with which the driver naturally interacts with, such as the steering wheel [3].

In healthcare, through wellness monitoring, one can envision the ability to create an individual profile identifying causes of stress, anxiety, depression or chronic diseases. The profile could be kept private or shared with a professional. Another possible application includes teaching, enhancing the human-computer interaction through the adaptation of the study material and teaching velocity to the subject response in each exercise, its personality and current mood; Recommendation Systems- adapting the movie, TV series or music

The associate editor coordinating the review of this manuscript and approving it for publication was Abdel-Hamid Soliman.

recommendation to the user likes and preferences according to its pre-emotional responses.

Humans often communicate emotions and their current sentic state via extraneous body expressions such as with a smile or more physiological expressions, such as an increase in heart rate (HR). These body expressions occur naturally and subconsciously. Several theorists argue that each emotion provokes its own unique somatic response [1]. Therefore, the modulation of the motor system expressions, sentic modulation, can be used to infer the individual emotional state.

Physical manifestations are easily collected; however, they present low reliability since they depend on the user social environment, cultural background (if he is alone or in a group setting), their personality, mood, and can be easily faked, becoming compromised [4]. On the other hand, these constraints do not apply to physiological signals, such as the HR, perspiration, pupil dilation, among others. Alterations in the physiological signals are not easily controlled by the subject, presenting a more authentic look into the subject emotional experience. For this reason, in this paper, we will focus on the recognition of human emotions based on physiological signals.

This survey aims to showcase the evolution and current landscape of emotion recognition systems based on physiological signals. We start by including a solid overview of fundamental theoretical concepts, providing a review of state-of-the-art publications and current researchers (see Fig. 1, where a histogram of the number of publications surveyed for this document per year of publication is displayed). Then, from the surveyed papers, the most relevant results, challenges, and the most promising future possibilities are discussed as a guide to new and current researchers with a focus on each part of an emotion recognition system from emotion elicitation to decision.

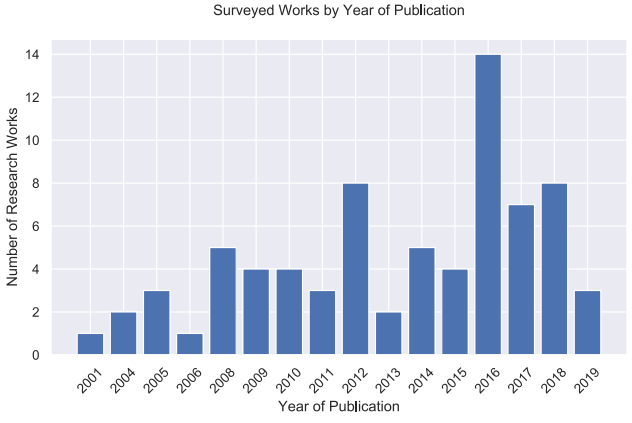

FIGURE 1. Histogram of the number of publications surveyed for this document per year of publication.

The remaining of this paper is organised as follows: In Section II the theoretical concepts needed to contextualise the reader are presented, with essential principles applied in the field of affective computing. Then, in Section III, we describe the key main steps required for the development of a novel Machine Learning (ML) algorithm for emotion assessment. In Section IV, the emotion recognition state-of-the-art literature is discussed and its main take-home messages identified. Lastly, the main conclusions and challenges derived from literature are described, along with recommendations for further work on the field.

II. THEORETICAL BACKGROUND

In this section we provide essential theoretical background for the concepts needed to develop novel algorithms for emotion recognition. We start by introducing the concept of emotion and its two main forms of characterisation: continuous and discrete in sub-sections II-A.1 and II-A.2, respectively. Then, in sub-section II-B, the Autonomic Nervous System (ANS) is introduced and its correlation with emotion generation explained. Next, in sub-section II-C, common state-of-the-art sensors used in affective computing and their correlation to sentic state assessment is described. Benchmark datasets used in emotion recognition are presented in sub-section II-F and, lastly, literature assessment methods are presented in sub-section II-E.

A. EMOTION MODELS

The first question in order to recognise emotions should be to define the concept of emotion. What is an emotion? Theorists from multidisciplinary fields such as neuroscience, philosophy and computer science have tried to answer this question and define a universal definition of emotion. However, with discord, thus, there is no single widely acknowledged definition. In ML, a definition of emotion is especially important since it is necessary to establish the targeting criteria of success. A common approach to mitigate this problem is to define emotions according to two models. One decomposes emotions in continuous dimensions and the second in discrete categories [1].

1) DISCRETE EMOTION SPACES

Since the ancient of times, emotion and feelings have been the thought of many philosophers. Per example, Cicero and Graver [5] back in the Roman empire organised emotions in four basic categories: Fear, Pain, Lust, Pleasure. Meanwhile, for Darwin [6], along with the theory of natural selection, emotions have an evolutionary history and are shared across cultures. Ekman [7] continued his work and argued that emotions are shared between cultures, thus, able to be universally recognised. Ekman described emotions as discrete, measurable and physio-related, arising from evolutionary evolved physiological and communicative functions. Hence, physiological expressions deriving from emotions functioned as a warning, sometimes separating life and death scenarios [8]. Ekman enumerated six basic emotions: Happy, Sad, Anger, Fear, Surprise and Disgust, with more complex emotions created as a combination of these basic emotions.

Plutchik [9] proposed a taxonomy to classify emotions in the form of a wheel model (Fig. 2) incorporating eight basic emotions: Joy, Trust, Fear, Surprise, Sadness, Anger,

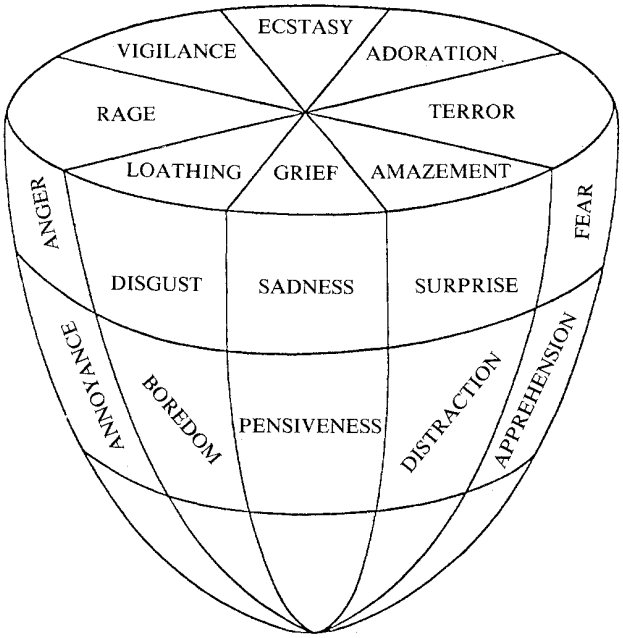

FIGURE 2. Plutchik wheel theory of emotion [9].

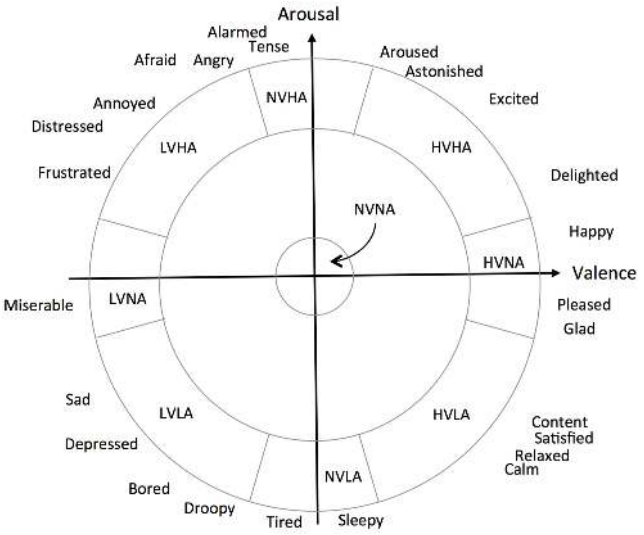

FIGURE 3. Valence-arousal model of emotion [14].

Disgust and Anticipation. In his taxonomy, once again, emotions can be mixed to form complex forms, personality traits and psychopathology. However, his taxonomy differs through the incorporation of intensity levels, as shown in Fig. 2, where stronger emotions occupy the centre, while weaker emotions occupy the extremities. Then, in [10], [11], Izard suggested 10 basic emotions: Interest, Joy, Surprise, Sadness, Fear, Shyness, Guilt, Angry, Disgust and Contempt. Izard advocated that emotions are the result of human evolution, and each emotion is correlated to a simple brain circuit where a complex cognitive component is not involved.

Lastly, Damasio [12] defined emotion as a neutral reaction to a certain stimulus, which can be categorised as primary (deriving from innate fast and responsive "flight-or-fight" behaviour) or secondary (deriving from cognitive thoughts).

In all the aforementioned theories of emotion, human emotional experiences are described in words. However, a discrete qualification of emotions can present difficulties, since complex mixed emotions can be difficult to precise and different individuals/cultures may describe a similar experience with different words. In order to overcome these difficulties, many authors have adopted the concept of continuous multi-dimensional space models. In a continuous multi-dimensional space model, emotions are measured along a defined axis, thus, simplifying the process of comparison and emotion discrimination.

2) CONTINUOUS DIMENSIONS

A continuous description of emotion must address two issues: The possibility to describe correlation among different emotional states, for example, Grief vs Sadness, Admiration vs Trust; and the quantification in a given state, for example, very sad vs sad vs not sad. A first approach was proposed by Schmidt *et al.* [8], where emotions were described as a single point in a pleasure-displeasure, excitement-inhibition, tension-relaxation three-dimensional space.

Following Wundt, Schmidt *et al.* [8] suggested a valence-arousal two-dimensional model where different emotions are described. The valence axis denotes how positive (pleasant) versus negative (unpleasant) the emotion is, while the arousal axis indicates its activation/intensity level.

Similarly, Lang [13] described emotions in two-dimensions: Valence (negative/positive) and Arousal (calm/excited). Fig. 3 displays the mapping of several emotions on the two-dimensional valence-arousal space.

Mehrabian [15] added a new dimension to describe the consciousness of the emotion, denoted as dominance (Fig. 4), facilitating the discrimination between emotions such as Fear vs Anger [4]. From all the aforementioned models, the valence-arousal is the most commonly applied due to its low simplicity of integration into an emotion assessment questionnair and low complexity in the modelling of ML algorithms, attaining overall good results.

B. AUTONOMIC NERVOUS SYSTEM

According to Levenson [17], emotions were preserved across natural selection due to the need for an efficient mechanism able to mobilise and organise the selection of quick responses from highly differentiate and disparate systems when environmental stimuli pose a threat to survival. Thus, in order to give a quick response to life-threatening situations, the emotion system is able to override high cortex functionalities, with quick, automated responses coordinated by the ANS. Levenson's theory has been supported by many theorists, however, they lack consensus on how many different emotional states are associated with distinct patterns of the ANS [17].

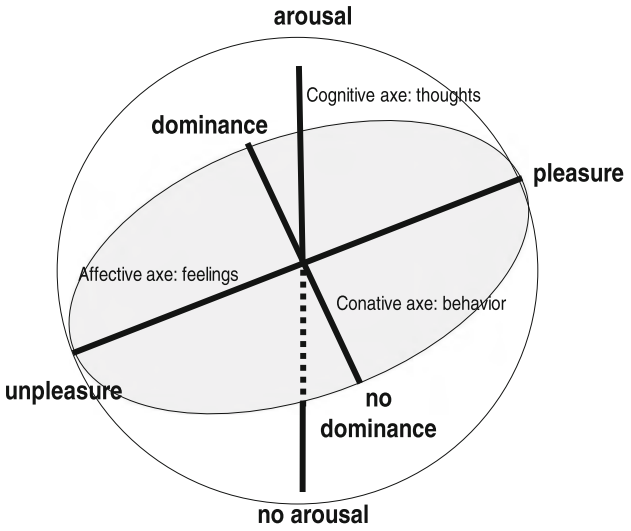

FIGURE 4. Mehrabian three-dimensional space theory of emotion [16].

On one hand, a few theorists defend that there are only two ANS patterns: The 'on' status and the 'off' status. Others theorists postulate that there is a large number of patterns of ANS activation, each associated with a different emotion [17].

The ANS is mediated by the two branches of the ANS. The 'on' status is mediated globally by the Sympathetic Nervous System (SNS) and the 'off' by the Parasympathetic Nervous System (PNS) [17]. The SNS is activated during physically or mentally stressful situations, thus, controlling the body responses to threats. The SNS is responsible for the increase in HR due to the increase in the sino-atral (SA) stimulation and for increasing the strength of contractions due to an increase in the propagation velocity of the depolarisation wave that travels through the heart, bronchial tubes dilation, muscles contraction, pupils dilation, decrease in stomach movement and secretions, decrease in saliva production and lastly, release of adrenaline. On the other hand, the PNS is responsible for homeostasis, i.e. the maintenance of the internal bodily milieu while at rest: slowing down the HR, decrease in the blood pressure and increase in the digestive system activity, bronchial tubes constriction, muscles relaxation, pupils constriction, increase in stomach movement and secretions, increase in saliva production and increase in urinary output. The SNS is often referred to as the ''fight-or-flight'' response, i.e. the activating and energising system; while the PNS is referred to as the ''rest-and-digest'' system. However, although this metaphor might fit the heart, for the remaining systems of the ANS the same is not verified, since the PNS causes increased activation in salivary glands, tear ducts, and the stomach and intestinal activity [17].

The ANS is responsible for many functions: serving as regulator, through the homeostasis, maintaining our internal bodily milieu within strict limits so as to minimise damage and maximise functioning; as an activator, allocating body resources in order to better respond to internal or external stimulus; as an coordinator, organising a continuous bidirectional flow of data between the somatic and brain systems; and, lastly as a communicator, through body responses with discernible dynamic variations for conspecifics [17].

This multi-dimensional functionality of the ANS increases notably the difficulty of correlating a subject emotional state with their current physiological signals, since, when a certain change in a physiological signal occurs, such as increase of the HR or respiration, it is more likely to have resulted from one of the several ANS non-emotional functionalities than from an emotional one [17].

C. PHYSIOLOGICAL SIGNALS

As stated, emotional states are associated with discernible ANS physiological responses. These responses can be read through body-worn physiological sensors such as the ECG, EEG, EDA, and BVP, which are briefly described below. The figures displayed in this section were obtained using the BioSPPy library [18], a Python library for physiological signals processing.

(a) **Electrocardiography (ECG)**: is a numerical recording of the potential differences that are propagated to the skin surface resulting from the electrical activity of the heart (arising from the contraction and relaxation of the cardiac muscle when electrically stimulated). The heart's contraction and relaxation rate is the result of three main components: (a) The action of the SA node, localised in the right atrium at the superior vena cava, which receives inputs from both branches of the ANS to initiate the cardiovascular activity with an intrinsic frequency of 100-120 bpm [17]; (b) PNS fibres modulated by the vagal nerve, slowing down the HR to approximately 70 bpm; (c) SNS fibres modulated by the post-ganglionic fibre, increasing the HR during an emotional episode or non-emotional ANS modulation [17]. This complementary modulation between the two branches of the ANS system is known as sympathovagal balance. Thus, the HR is a function of the ANS activity, which in turn is dependent on emotional stimuli, therefore, information about the emotional state can be inferred from the ECG data [19]. The ECG should be sampled with high-frequency rate (500-1000 Hz) in order to be possible to more accurately determine the instants when the heartbeats occur and use these instants to calculate and modulate the HR. The ECG signal presents amplitudes between 10uV (fetal) to 5mV (adult). Fig. 5 displays an ECG and some of its main characteristics, namely, the R peaks and HR.

(b) **Electrodermal Activity (EDA)/ Galvanic Skin Response (GSR)**: provides a measure of the resistance of the skin by passing a negligible current or voltage through the body and measuring the voltage or current variation between the two sensor leads, respectively. Thus, the skin is considered to be equivalent to a variable resistor. When given a known voltage or current, the other is measured and the skin conductance level

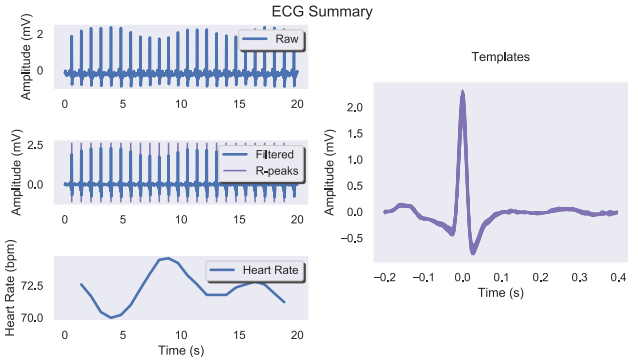

FIGURE 5. Raw Electrocardiography (ECG) sensor signal, filtered signal, heartbeats waveform templates and its main data characteristics (R-peaks and heart rate (HR)).

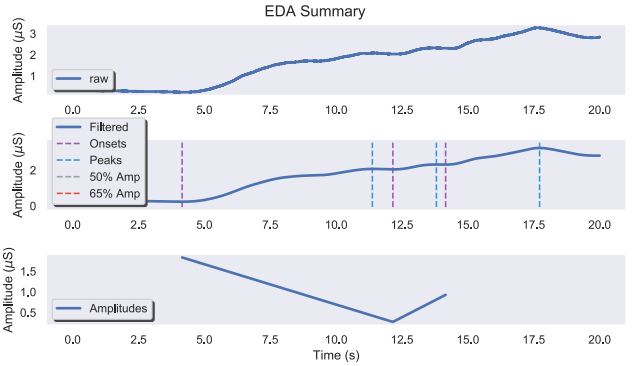

FIGURE 6. Raw Electrodermal Activity (EDA) sensor signal, filtered signal, Electrodermal Response (EDR) amplitudes and its main data characteristics (onsets, peaks, recovery rates).

derived from $G = 1/R; R = V/I$. Fig. 6 displays a common EDA signal and its main characteristics. As observed, the EDA signal is characterised by a baseline, from which, phasic perturbations arise in response to certain events. Thus, an EDA signal can be decomposed in two main components: a baseline tonic component of low bandwidth (f < 3Hz) expressing the thermal regulation activities denoted as Electrodermal level (EDL), and an Electrodermal Response (EDR) phasic component expressing psychological-related responses when an SNS regulatory activity occurs. The mean value of the EDA signal enables to infer the level of arousal and activation of SNS system since the EDR response is usually observable in a stressful or surprise event when an increase of perspiration decreases the skin resistance [20], [21]. Ionic sweat is more conductive than dry skin, hence, causes an increase in conductivity proportional to the amount the glands have filled coordinated by the sympathetic activation due to external sensory or a cognitive stimuli [19]. Thus, the EDA provides a non-intrusive look into ANS activity through the EDR psychological response to certain stimuli. The EDA electrodes are usually placed at areas of high sweat glad density, such as on the 2^{nd} phalanx of the index and middle fingers, the index and ring fingers, the hand or feet palms [8]. The EDA data has been used to study emotion-related PNS activity with applications such as deception, stress, frustration, arousal and anxiety detection [8].

(c) **Photoplethysmography (PPG) or Blood Volume Pulse (BVP)**: a photodiode measures the amount of backscattered light by a skin voxel. Thus, in a BVP signal, the amount of light that returns or passes through the finger to a BVP sensor is proportional to the volume of blood in the tissue. Hence, it is possible to detect the heartbeats through the pulse local maximum by the passage of blood, indicating each cardiac cycle from which the HR can be inferred. The sensor is usually placed in the subject index finger.

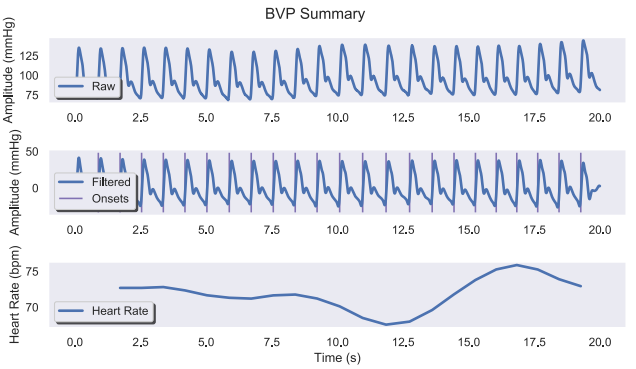

FIGURE 7. Raw Blood Volume Pulse (BVP) sensor signal, filtered signal and its main data characteristics (onsets and HR).

The BVP sensor data is highly prone to noise with its quality depending on the sensor location, motion, external light artefacts and subject dependent physiological characteristics, such as: level of tan, skin absorption properties, skin structure, the blood oxygen saturation, blood flow rate, skin temperatures and the measuring environment [8], [22]. Fig. 7 displays a BVP signal and some of its main characteristics.

As stated in Section II-B, the ANS is responsible for dilating or contracting the blood vessels diameter. Hence, changes in BVP amplitude reflect instantaneous sympathetic activation such as in high arousal and pleasant situations where the SNS increases the blood pressure and heart rate variability (HRV), both possible metrics to be deduced from the BVP signal in order to modulate the user sentic state. For example, when a person relaxes, vasodilatation usually occurs which is reflected as an increase in the blood flow volume, consequently affecting the BVP amplitude; when anxious or fearful, the opposite is verified [23]. Generally, the BVP sensor data is recorded using sampling rates below 100Hz [8].

(d) **Respiration (RESP)**: usually in the form of a chest belt worn in the thorax or the abdominal area, it is used to measure the respiration pattern, namely, how

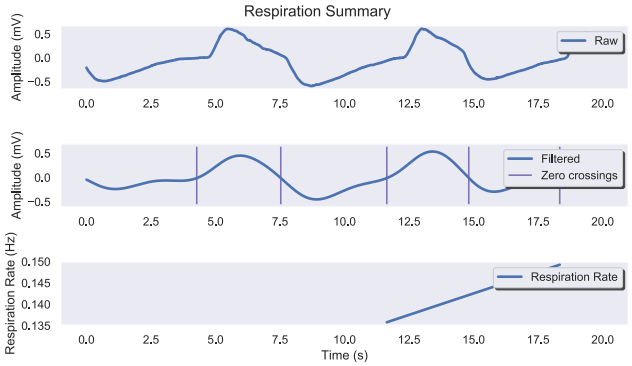

FIGURE 8. Raw Respiration (RESP) sensor signal, filtered signal and its main data characteristics (zero-crossings, respiration rate).

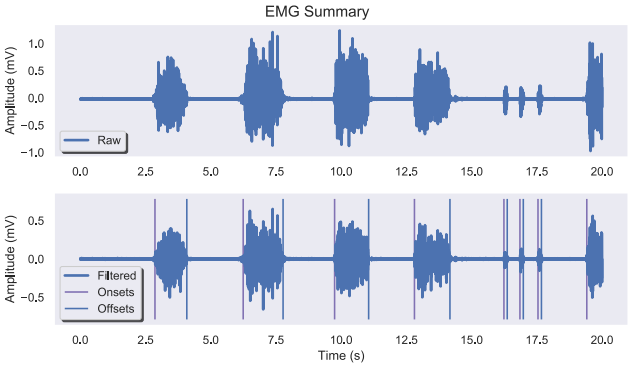

FIGURE 9. Raw Electromyography (EMG) sensor signal, filtered signal and event onsets.

deep and fast a subject is breathing [8]. During a respiration cycle, the thorax expands and constricts in the inhalation and exhalation of air resulting in the stretching and de-stretching of the chest belt. From this movement, the respiration rate and volume can be derived [8]. Regarding emotion recognition, the respiration rate with fast and deep breathing can indicate high arousal such as anger, fear, or joy, rapid shallow breathing can indicate tense anticipation, such as panic, fear or concentration, slow and deep breathing indicates a relaxed resting state while slow and shallow breathing can indicate states of withdrawal, passive like depression or calm happiness [24]. In the literature, a few papers have proposed ECG-derived respiration techniques allowing to obtain the respiration waveform from an ECG signal, namely from the RS-decline quantified by central moments, respiratory sinus arrhythmia, R-wave amplitude, QRS area, RS-distance and maximum RS-slope [25], [26]. Fig. 8 displays a RESP signal and some of its main characteristics.

(e) **Skin Temperature (Temp) or (SKT)**: can be measured using an infrared thermopile or a temperature-dependent resistor at the skin surface. The temperature of the human skin can change for numerous reasons correlated with main functions of the ANS such as physical exercise, physiological conditions, environmental conditions and emotional reactions through mechanisms such as sweating, shivering, vasoconstriction or vasodilatation. For example, sweating and vasoconstriction decrease the body temperature, while vasodilation and shivering, increase the heat production in the muscles, thus, increasing the skin surface temperature [23]. When the muscle contracts or relaxes, vasoconstriction or dilatation occur, respectively. The smooth muscle contraction is regulated by the SNS, which is linked to emotion. Hence, SKT can provide a look into the ANS system. For example, in a 'fight-or-flight' response mediated by the SNS, the muscles under strain show high vascular resistance and increase the arterial flow. The blood flow to the extremities becomes restricted in favour of increased blood supply to the vital organs, decreasing the temperature of the extremities [8], [19]. The literature describes the use of temperature, however, in theory, it should take several minutes for a change in the body temperature to be noticeable, displaying overall small amplitude variations.

(f) **Electromyography (EMG)**: measures the skeletal muscle electric activity with a skin surface electrode or with a needle electrode. Upon a muscle contraction, there is an amplitude rise in the EMG signal from an electrical potential difference that appears between the interior and the exterior of the muscle cell. The difference is short-lasted and is denoted as an action potential. The EMG signal presents amplitudes between 50uV-30mV and bandwidth between 2-500Hz. The surface EMG sensors placement can be directed for emotion recognition of both facial or body expressions in order to capture the subject facial expression or stress. Common placements are the on trapezoid and the Zygomaticus major to modulate head movements and tension; laugh or a smile, respectively. Fig. 9 displays a raw and filtered EMG signal.

(g) **Electroencephalography (EEG)**: is a measurement of the electrical field from currents that flow during neurons synaptic excitation in the cerebral cortex when these are activated. All of the aforementioned sensors record changes in the physiology of various organs as a result of ANS adaptations; in contrast, the EEG records the aggregate potential differences from active neurons, thus, capturing an electrical perspective on the local source of the ANS activity from the CNS. The EEG signal presents amplitudes between 2-100uV on the scalp and a dynamic range between 0.5-60Hz. The brain is the main control unit for all the functions of the organism, including the control of the body movement, sensory processing, language and communication, memory and emotions. Therefore, EEG can be used to correlate emotion generation and brain regions. Fig. 10 displays a raw and filtered EEG signal.

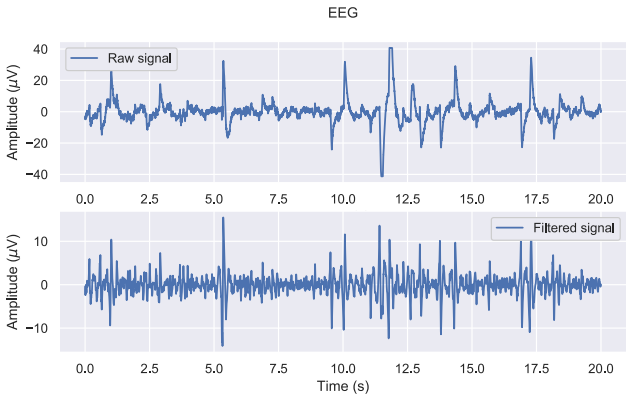

FIGURE 10. Raw Electroencephalography (EEG) sensor signal and filtered signal.

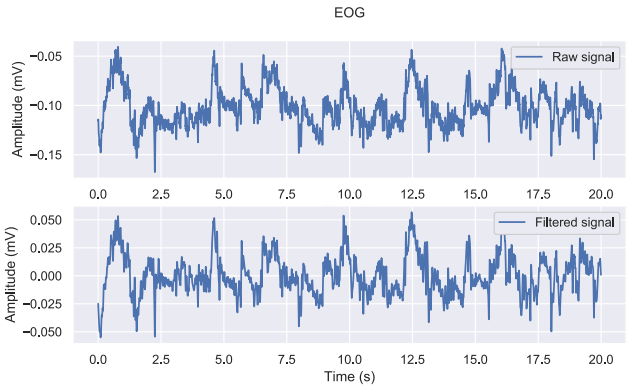

FIGURE 11. Raw Electrooculography (EOG) sensor signal and filtered signal.

(h) **Eye Gaze**: Measured through Electrooculography (EOG), Infrared Reflection Oculography IROG) or photoelectric techniques; the EOG measures the resting potential of the eye and its variations derived from horizontal and vertical eye movements. The EOG works based on the fact that the eye act as an electrical dipole between the positive potential of the cornea and the negative potential of the retina, maintained by means of active ion transport. Therefore, an electrode placed in the vicinity of the eye will become more positive when the eye rotates towards it, and less positive when it rotates in the opposite direction [27]. Other common techniques are the Infrared Reflection Oculography IROG) and the photoelectric techniques, which rely on the fact that the white sclera reflects more light than the pupil and the iris. Hence, when the eye moves to one side, less infrared light is reflected to the detector on one side of the eye and vice-versa. These approaches are based on videooculography and Purkinje eye-trackers, which uses head-mounted miniaturised video cameras to track the image of the pupil or of the light reflexes [27]. The amount of light entering the eyes is regulated by radial and circulatory fibres innervated by the PNS and SNS systems, respectively, regulating the dilation and constriction of the muscle fibres. Thus, by an EOG or using an eye tracker, information about the ANS can be deduced. The literature shows that the pupillary responses, frowns and blinks have distinct patterns according to different human emotional states, however, with conflicting results. Per example, the rate of blinks and saccades is found to provide information regarding fatigue or anxiety, while the focus on a point indicates high attention. Fig. 11 displays a raw and filtered EOG signal.

Fig. 12 displays a histogram with the number of sensors used in each publication surveyed for this document. As it is possible to observe the GSR, ECG, and Resp sensors are the three most commonly applied in literature; in contrast to the EOG, BVP and ACC. The ECG, EDA, EMG, SKT

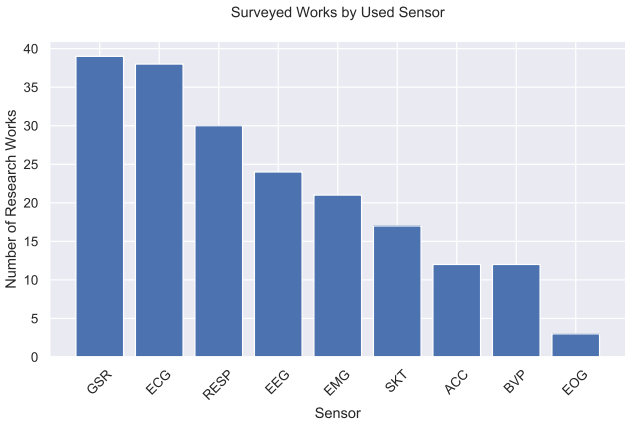

FIGURE 12. Histogram of the number of publications surveyed for this document per sensor.

and RESP present the advantage of being easily introduced in wearable systems with great comfort to the user. Lastly, inertial sensors such as the Accelerometer (ACC), Gyroscope (GYR), Barometer (BAR) and Magnetometer (MAG), generally used for human activity recognition, could be used to correlate each emotional state with certain activities and derive contextual information about the daily living, likes-dislikes, and preferences of the subject.

D. EMOTION ELICITATION MATERIAL

Due to the high subjectivity and variability in emotion elicitation, it is important to use a set of pre-validated emotional stimuli in order to ensure the expression of a wide spectrum of emotions and of high intensity each. In literature, this is performed by selecting the elicitation material from different affective categories presenting the most consensual and reliable self-ratings across different subjects.

Within the state-of-the-art, affect elicitation is commonly performed via pictures [28], films [29], VR videos [30], games [31]–[36], music videos [21], sound [37], [38], words [39], recall [40]–[43] or in well controlled settings, although real-world scenarios have started to be explored [44].

(a) **Images:** The International Affective Picture System (IAPS) [28] by University of Florida Center for the Study of Emotion and Attention (CSEA), provides a large set of standardised colour photographs for the elicitation of attention and a wide range of emotional experiences rated in terms of pleasure, arousal, and dominance. Additionally, the Geneva affective picture database (GAPED) [45] contains 730 pictures with negative (spiders, snakes, immoral and illegal scenes), positive (human and animal babies, nature sceneries), and neutral (objects) content annotated in terms of valence-arousal and congruence with moral and legal norms. Thirdly, the Museum of Modern and Contemporary Art of Trento and Rovereto (MART), contains 500 images of abstract paintings [46] labelled according to their positive or negative content. Similarly, the deviantArt dataset [46] contains 500 amateur artworks of abstract art labelled according to the positive or negative emotions evoked by the artworks. Further image-based datasets are the Flickr dataset [47], Artistic dataset (ArtPhoto) [48], Abstract dataset (Abstract) [48], Emotion6 dataset [49], and the Image-Emotion-Social-Net (IESN) dataset [50]. The features that can be derived from images in order to infer its emotional content can be categorised as low, mid and high-level features. Low-level features include colour (saturation, brightness, hue, hue, intensity, and colourfulness), value (lightness or darkness), line (amounts and lengths of static and dynamic lines), texture (wavelet-based features, Tamura features, grey level co-occurrence matrix and LBP features), shape (roundness, angularity, simplicity, and complexity), and space (distance or area between, around, above, below or within things). Mid-Level features, in contrast to low-level features, are more interpretable by Humans and include the image materials, surface properties, functions or affordances, spatial attributes and the objects present in the image. Lastly, high-level features are the semantic contents of the image, such as facial expressions. For further information on image content analysis, the authors refer the reader to [48], [51].

(b) **Video:** In [52], the authors created a validated catalogue of film clips stimuli for emotion elicitation covering 24 articles and 295 film clips from four decades of research. In addition, the LIRIS-ACCEDE Database [53] is composed of 9800 video clips (8-12 seconds long) extracted from 160 movies annotated according to the valence-arousal scale. The authors in [54] present 70 film excerpts from 1 to 7 minutes long to elicit emotions. The excerpts were validated according to 24 classification criteria: subjective arousal, positive and negative affect (derived from the Positive and Negative Affect Schedule (PANAS)), a positive and a negative affect scores (derived from the Differential Emotions Scale (DES)), 6 emotional discreteness scores (anger, disgust, sadness, fear, amusement and tenderness), and 15 mixed feelings scores assessing the effectiveness of each film excerpt to produce blends of specific emotions. For the benchmark of violence content, the Violent Scenes Detection (VSD2014) [55] contains 31 movies and 86 web video clips (6 seconds to 6 minutes long) retrieved from YouTube. The videos are annotated according to their violent content and to 10 high-level concepts for the visual and audio modalities such as the presence of blood, fights, gunshots, screams, etc. Further available public databases containing validated videos are the HUMAIN [29], DEAP [21], EMDB [56] and MAHNOB-HCI [20], explained in more detail in the next section. In order to validate the video content for emotion elicitation, the authors in [21], developed a method for video affective content analysis using retrieval by affective tags from the last.fm website, video highlight detection, and an online assessment tool to extract videos lying closest to the extreme corners in the arousal-valence quadrants. Audio and visual cues are usually tool elements used by movie directors to elicit certain emotions in the viewers. Therefore, in [21], for the detection of 1-minute video highlight: low-Level video and audio features were extracted. Regarding the former, some of the extracted features were colour variance, shadow proportion, visual excitement, greyness and 20-bin colour histogram of hue and lightness values in the HSV space. Additionally, fast-moving scenes can be an indicator for exciting scenes, thus, the average shot change rate and shot length variance were extracted to characterise the video rhythm. The literature has shown that certain speech features such as speech energy, pitch, timing, voice quality, duration, fundamental frequency, and format, capture emotional information to discriminate high and low valence, while loudness (speech energy) and speech rate, are related to arousal [57]. For discrete emotions, features like pitch levels can indicate feelings such as astonishment, boredom, or puzzlement, while speech volume is generally representative of emotions such as fear or anger. Thus MFCC, energy, gormants, time-frequency, pitch, zero-crossing rate and silence ratio features were extracted. For further information on video affective content analysis, the authors refer the reader to [57].

(c) **Sound:** The International Affective Digitized Sound system (IADS) [37] contains sets of standardised acoustic stimuli across a wide range of affective categories rated in terms of pleasure, arousal, and dominance. In the AuBT dataset [38], the authors used four music songs targeting the emotion classes: joy, anger, sadness and pleasure, handpicked by each subject to induce special memories.

(d) **Words/Text:** The Affective Norms for English Text (ANET) [58] and the Affective Norms for English Text (ANET) [58] dataset contain sets of English words

and brief texts, respectively, rated in terms of pleasure, arousal, and dominance.

(e) **Recall/Acting:** The Interactive emotional dyadic motion capture database (IEMOCAP) contains data from ten actors in dyadic sessions with markers on the face, head, and hands, which provide detailed information about their facial expression and hand movements during scripted and spontaneous spoken communication scenarios. The data was annotated using the valence-arousal-dominance scale. The authors in [59], with the help of pre-selected images, attempted to feel and express the emotional states of no emotion, anger, hate, grief, platonic love, romantic love, joy and reverence.

(f) **Video Games:** In [60], it is presented a multi-modal database containing peripheral physiological signals (ECG, EDA, RESP, EMG, SKT), ACC and facial data acquired by 50 subjects as they played FIFA 2016 (a football video-game). The data is self-reported according to the arousal-valence scale and categorically rated according to happiness, frustration, proud, curious, angry, fear, boredom and sadness. In [36] the authors acquired HR, BVP and GSR data from 36 subjects playing a Maze-Ball game to study the emotional states of fun, challenge, boredom, frustration, excitement, anxiety and relaxation.

(g) **Social-evaluative and Cognitive Stressors:** Stress can be defined as a non-specific response of the organism to any pressure or demand, displaying a physiological, psychological and behavioural response when demands exceed the individual's ability to cope [61]. According to [8], stress-inducing events can be categorised as social-evaluative, cognitive, or physical. A social-evaluative standardised stressor protocol is the Trier Social Stress Test (TSST) [62], where the subjects are asked to deliver a free speech and perform mental arithmetic in front of an audience. To induce cognitive load, the Stroop Color and Word Test (SCWT) is commonly applied [63]. In the SCWT, the subjects are asked to read three different tables as fast as possible. In two, the subjects are asked to read the names of the colours printed in black ink and name the different colour patches. On the other hand in the latter table, colour-words are printed in inconsistent colour ink, and the subjects are asked to name the colour of the ink. In [21], cognitive and social-evaluative stressors were applied with the authors simulating a job interview where each subject was asked to perform a 5-minute speech on their personality traits, and count from 2023 to 0 doing steps of 17 in front of a three members panel.

(h) **Standardised ANS Clinic Tests:** As stated in Subsection II-B, the ANS is responsible for many functions to maintain the body homeostasis, thus, several tests have been developed and standardised in routine clinical evaluations for the diagnosis of many diseases. These entitle per example: the Deep breathing test, Isometric handgrip test, Cold pressor test, Active standing (orthostatic test) and Head-up tilt test [64].

To conclude, the use of images as elicitation material presents the advantages of being user-friendly, low cost, easy and fast to execute in a laboratory. However, images might not be enough to evoke impactful, strong-lasting emotions, enough to be consciously perceived by the user and physiologically observable [65]. On the other hand, music or music-videos, although simple and low cost, might be constrained to the evocation of a limited range of positive-negative emotions, highly correlated with the subjects' music taste and the memories it invokes. Thus, films or short-duration audiovisual video clips are the most applied methodology in emotion-recognition [65], and have shown to be the most reliable material for emotion elicitation. Under those circumstances, in this paper, we will focus on emotion recognition using movies or films. A new approach for emotion elicitation that has emerged in the last two decades, is the usage of Virtual Reality (VR), as in the dataset obtained by [30]. VR allows a deeper immersion by the subject in the activity, thus, increasing the reliability of the research study.

E. ASSESSMENT METHODS

The annotation of the individuals' emotions can be accomplished resorting to internal or external methodologies. In internal annotation (self-assessment), the subject directly assesses its affective state. Although this seems to be a very easy method to implement and replicate, it is not a direct task since it is very difficult to infer one's emotional state and describe it into words. Additionally, self-assessment can be felt to the subject as an intrusive process, evoking a defence-mechanism with the subject filling an unreliable report of their emotions both unconsciously or consciously to preserve their privacy. A solution could be the implementation of data protection measures ensuring the subject's privacy and confidentially.

On the other hand, in an external annotation (implicit-assessment), an external subject assesses the subject affective state through the analysis of their externally observable behaviour and physiological responses [66]. An external annotator can be just as easily deceived, being dependent on the user ability to externalising its emotional experiences, which often correlated with the subject personality, environment and culture. Thus, external annotation is dependent on multiple factors.

Given these considerations, self-assessment is widely the most commonly applied methodology in affect recognition state-of-the-art research, with self-reporting questionnaires being presented to the user. To visualise the scales of emotion dimensions, Self-Assessment Manikins (SAM) [67] provides a graphical interface for cross-cultural measurement of emotional response through manikins along a continuous nine-point scale. As displayed in Fig. 13, for the pleasure axis, SAM ranges from a smiling manikin, denoting happiness

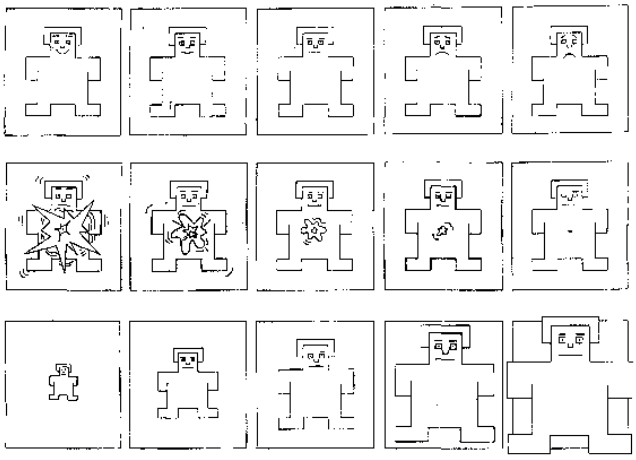

FIGURE 13. Self-Assessment Manikin (SAM) self-report questionnaire for valence (first row), arousal (second row) and dominance (last row) [67].

to a frowning figure, denoting unhappiness. Likewise, for the arousal axis, SAM's manikins range from an SLeepy figure with eyes closed to an excited figure with eyes open, denoting low activation and high activation, respectively. Lastly, the dominance scale shows a manikin ranging from a very small figure, representing a feeling of being controlled or submissive, to a very large figure, representing in-control or a powerful feeling.

A further technique is the Ecological Momentary Assessment or Experience Sampling Method (EMA), where subjects are asked to self-report their thoughts, feelings, behaviours, and/or environment context questions. The subject reports can be either scheduled a few hours apart or event-triggered. For a correct annotation, a trade-off between the frequency of the questions, the length and complexity of the questionnaires should be optimised. Additionally, the EMA should be directed to the study goal and as brief as possible in order to attain a good interpretation of the subject emotional state with comfort without interrupting its daily routine.

A few applications have been developed to assist the continuous annotation of the subjects' instantaneous mood, such as in [68]–[70].

Affective experiences and, consequently, the elicited emotional state in an experiment are modulated by both internal and external factors. The first entitles the individuals' mood, personality, age and culture in which it was raised; the latter if the individual is alone or in a group setting. Hence, an elicited emotional state might differ from subject to subject, presenting high subject-dependency, day-dependency and localisation-dependency. These factors increase the challenge of mapping and correlating physiological signals to universal emotional states in pre-defined emotional classes [65].

To access the subject personality, the Big-Five factor model is majorly utilised in literature [66], [71]. The Big-Five model describes personality in terms of five dimensions: Extraversion (sociable vs reserved), Agreeableness, (compassionate vs dispassionate and suspicious), Conscientiousness (dutiful vs easy-going), Emotional stability (nervous vs confident) and Openness to experience (curious vs cautious). Common questionnaires measuring these dimensions are the Neuroticism, Extraversion and Openness Five Factor Inventory (NEO-FFI), the Big-Five Marker Scale (BFMS) and the Big Five Inventory (BFI).

On the other hand, to examine the subject mood, a positive-negative scale is often inferred from the PANAS [72]. The PANAS questionnaire consists of two 10-item mood scales, in which the positive affect axis reflects enthusiasm, activation and alertness, while the negative affect axis reflects distress and unpleasant engagement [66].

Regarding stress-detection, current methods consist on the measurement of cortisol levels, involving an invasive, laborious, not immediate process, or assessment questionnaires, such as the Perceived Stress Scale (PSS) and the Stress Response Inventory (SRI). Similarly, for anxiety and depression, commonly applied questionnaires are the Strait-Trait Anxiety Inventory (STAI), the Hospital Anxiety and Depression Scale (HADS) [73] (see Fig. 14) and Patient Health Questionnaire (Patient Health Questionnaire (PHQ-9)). The aforementioned tests and further mental-health assessment questionnaires are presented in Table 1. These rely on biased responses from the individuals or the recognition and interpretation of visual patterns by an expert. Worldwide, more than 300 million people suffer from depression, with 20% of the adult working population suffering from a mental health problem [74]. Depression is highly correlated with anxiety and stress, being the leading cause of disability and greatly affecting the individual wellness and quality of life. Therefore, effort must be performed in order to detect stress as early as possible through accessible, objective, impartial, and prompt methodologies, preventing it to reach its highest level with serious implications to the individual's wellness and quality of life.

F. EMOTION ELICITATION DATASETS

Several datasets relying on movies for emotion recognition using physiological signals are publicly available. These allow for the benchmark of emotion recognition algorithms, facilitating a direct comparison of the results of different methodologies:

(a) **DEAP** [21]: contains EEG, GSR, RESP, SKT, EMG, EOG, BVP data from 32 volunteers watching 40 one-minute-long music videos. For 22 of the participants, frontal face video was also recorded. The dataset was self-annotated after each video in terms of arousal-valence, like-dislike, familiarity and dominance by the volunteers.

(b) **MAHNOB-HCI** [20]: contains facial video recordings, audio, eye gaze, ECG, EEG, SKT and GSR data from 30 volunteers watching 20 film excerpts from 35 to 117 seconds long. The dataset was annotated according to emotion keywords and arousal-valence-dominance.

TABLE 1. State-of-the-art questionnaires grouped by domain, name, brief description, number of questions (#), range and Cronbach's α reliability (α). Table adapted from [8].

Domain	Name	Description	#	Range	α
Emotion	**Positive and Negative Affect Schedule (PANAS)** [74]	Measurement of positive affect (PA) and negative affect (NA)	10	Distressed, upset; hostile, irritable; scared, afraid; ashamed, guilty; and nervous, jittery	PA: 0.86-0.90; NA: 0.84-0.87
Emotion	**Photo Affect Meter (PAM)** [77]	User selects image that best fits their mood from 16 images mapped into the PANAS space	16	Distressed, upset, hostile, irritable, scared, afraid, ashamed, guilty, nervous, jittery	
Emotion	**Differential Emotions Scale (DES)** [78]	Asks the subjects to consider the experience they described and to rate how often they experienced each emotion item during the experience	30	Anger, disgust, contempt, interest, joy, surprise, sadness, fear, shyness, guilt	0.81
Well-being	**Flourishing Scale (FS)** [79]	Provides a success score on relationships, self-esteem, purpose, and optimism	8	8-58	0.87
Well-being	**Scale of Positive and Negative Experience (SPANE)** [79]	Measurement of positive-negative experiences	12	-24-24	0.81-0.90
Stress	**Perceived Stress Scale (PSS)** [80]	Queries on how unpredictable, uncontrollable, and overloaded respondents find their lives	10/14	0-40	0.82
Stress	**Stress Response Inventory (SRI)** [81]	Emotional, somatic, cognitive, and behavioural stress responses	39	0-156	0.96
Stress	**Global Assessment of Recent Stress (GARS)** [82]	Capture amount of stress or change associated with stressful events that have occurred over the past six to twenty-four months.	8	0-9	0.74-0.92
Psychopathology	**Symptom Checklist-90-Revised (SCL-90-R)** [83]	Evaluate a broad range of psychological problems and symptoms of psychopathology	90	Nine scores along primary symptom dimensions and three scores among global distress indices	0.77-0.90
Depression	**Patient Health Questionnaire-9 (PHQ)** [84]	Depression diagnostic instrument	9	0-27	0.86-0.89
Anxiety and Depression	**Hospital Anxiety and Depression Scale (HADS)** [75]	Assess the contribution of mood disorder with focus on anxiety and depression	14	0-21	0.68-093
Anxiety	**State-Trait Anxiety Inventory (STAI)** [85]	Measures anxiety about an event and as a personality characteristic	40	20-80	0.86
Loneliness	**UCLA Loneliness Scale** [86]	Assesses how often a person feels disconnected from others	20	20-80	0.88
Panic	**Panic Disorder Severity Scale** [87]	Assesses the severity of seven dimensions of panic disorder and associated symptoms	15	0-8	0.92-0.94
Sleep Quality	**Pittsburgh Sleep Quality Index (PSQI)** [88]	Measure the quality and patterns of sleep in the older adult	19	0 to 21	0.83
Personality	**Big Five Marker Scale (BFMS)** [89]	Measures an individual on the Big Five dimensions of personality	50	Extraversion, agreeableness, conscientiousness, emotional stability and openness	0.88-0.93
Personality	**Neuroticism, Extraversion and Openness Five Factor Inventory (NEO-FFI)** [90]	Measures an individual on the Big Five dimensions of personality	60	Extraversion, agreeableness, conscientiousness, emotional stability and openness	0.75-0.83
Personality	**Big Five Inventory (BFI)** [91]	Measures an individual on the Big Five dimensions of personality	44	Extraversion, agreeableness, conscientiousness, emotional stability and openness	0.79-0.88

(c) **ASCERTAIN** [71]: contains EEG, ECG, GSR and video facial activity data while 58 volunteers watched 36 movie clips between 51-127s long. Each clip was self-annotated in the form of arousal-valence, liking, engagement and familiarity ratings. Additionally, a big-five marker scale questionnaire was filled by each volunteer allowing to correlate different personality traits and affective states with physiological responses.

(d) **Eight-Emotion** [90]: contains BVP, EMG, EDA data of one subject eliciting eight states: neutral, anger, hate, grief, love, romantic love, joy, and reverence.

(e) **EMDB** [56]: contains Skin Conductance Level (SCL) and HR data collected from 32 volunteers watching 52 pre-selected and edited film clips without auditory content. Each film was validated and rated across different quadrants of affective space by 113 participants.

(f) **AMIGOS** [66]: contains full-body and depth videos, EEG, ECG and GSR data from 40 volunteers watching 16 short videos; and 37 volunteers watching 4 long-videos. The dataset was both annotated by self-assessment of affective levels (valence-arousal, control, familiarity, like-dislike, and selection of basic emotions) and external assessment of participants' levels of valence. Additionally, for the study of the participants' emotions correlation with their personality and mood, the participants were asked to fill forms with Personality Traits and PANAS questionnaires. The dataset was created to study the users' affect, personality traits and mood on an individual and group settings elicited by short and long videos.

(g) **DECAF** [91]: contains near-infra-red (NIR) facial videos, horizontal EOG (hEOG), ECG, and trapezius-EMG (tEMG) peripheral physiological responses and Magnetoencephalogram (MEG) sensor data of the emotional responses of 30 participants to 40 one-minute music video segments, and 36 movie clip. The dataset contains the users' self-assessment of valence-arousal-dominance, and time-continuous emotion annotations for movie clips from seven users, which were used to demonstrate dynamic emotion prediction.

Tick the box beside the reply that is closest to how you have been feeling in the past week.
Don't take too long over you replies: your immediate is best.

D	A		D	A	
		I feel tense or 'wound up':			**I feel as if I am slowed down:**
	3	Most of the time	3		Nearly all the time
	2	A lot of the time	2		Very often
	1	From time to time, occasionally	1		Sometimes
	0	Not at all	0		Not at all
		I still enjoy the things I used to enjoy:			**I get a sort of frightened feeling like 'butterflies' in the stomach:**
0		Definitely as much		0	Not at all
1		Not quite so much		1	Occasionally
2		Only a little		2	Quite Often
3		Hardly at all		3	Very Often
		I get a sort of frightened feeling as if something awful is about to happen:			**I have lost interest in my appearance:**
	3	Very definitely and quite badly	3		Definitely
	2	Yes, but not too badly	2		I don't take as much care as I should
	1	A little, but it doesn't worry me	1		I may not take quite as much care
	0	Not at all	0		I take just as much care as ever
		I can laugh and see the funny side of things:			**I feel restless as I have to be on the move:**
0		As much as I always could		3	Very much indeed
1		Not quite so much now		2	Quite a lot
2		Definitely not so much now		1	Not very much
3		Not at all		0	Not at all
		Worrying thoughts go through my mind:			**I look forward with enjoyment to things:**
	3	A great deal of the time	0		As much as I ever did
	2	A lot of the time	1		Rather less than I used to
	1	From time to time, but not too often	2		Definitely less than I used to
	0	Only occasionally	3		Hardly at all
		I feel cheerful:			**I get sudden feelings of panic:**
3		Not at all		3	Very often indeed
2		Not often		2	Quite often
1		Sometimes		1	Not very often
0		Most of the time		0	Not at all
		I can sit at ease and feel relaxed:			**I can enjoy a good book or radio or TV program:**
	0	Definitely	0		Often
	1	Usually	1		Sometimes
	2	Not Often	2		Not often
	3	Not at all	3		Very seldom

Please check you have answered all the questions

Scoring:
Total score: Depression (D) _____ Anxiety (A) _____
0-7 = Normal
8-10 = Borderline abnormal (borderline case)
11-21 = Abnormal (case)

FIGURE 14. Hospital Anxiety and Depression Scale (HADS) self-report questionnaire [73].

(h) **Detecting Stress During Real-World Driving Tasks (DSDRWDT)** [92]: contains ECG, EMG (right trapezius), GSR measured on the hand and foot, and RESP data measured of 24 volunteers in rest position and driving in city streets and highways. Self-report questionnaires were used to map the subject experiment into low, medium, and high-stress levels.

(i) **Real World Driving to Assess Driver Workload (RWDADW)** [93]: contains SKT, GSR and HR data measured from 10 volunteers driving different road types and obstacles in a real-world scenario. The driving test took around 20 minutes and posteriorly the driver workload was annotated using post-hoc video analysis.

(j) **WESAD** [94]: contains BVP, ECG, EDA, EMG, RESP, SKT and ACC data recorded from both a wrist and a chest-worn device from 15 subjects during a lab study experience of emotional and stress stimuli. The dataset contains three affective states (neutral, stress, amusement) self-reported using state-of-the-art questionnaires: SAM, PANAS, Short Stress Scale Questionnaire (SSSQ), a shorter version of SSQ, and STAI.

The aforementioned information regarding the surveyed public datasets for emotion and stress recognition based on physiological signals is summarised in Table 2. The literature review allowed to verify that most datasets are obtained in a laboratory setting, with the few real-world scenarios [59], taking place in a specific constrained scenario such as driving [92], [93]. The selection of in laboratory studies versus in-real-world requires a trade-off between external validity in the natural environment versus higher certainty in the self-reported validation of the user emotion.

Regarding an in-laboratory setting, a few environment parameters should be controlled namely: (1) The subject should be kept isolated from the outside environment in order not to be able to see the examiners or be disturbed from the outside for an immersing experience eliciting a deep response; (2) The room illumination and temperature should be controlled to avoid the depolarisation of a non-emotional ANS response adding bias to the physiological signals; (3) As it was aforementioned, videos have been the preferred material for emotion elicitation due to their good results; however, a few parameters should be paid attention to, namely, their duration. The videos should be short enough to facilitate their annotation and prevent boredom and mixed emotions, but long enough to elicit the desired emotion. Thus, the literature recommends video lengths of 1-10 minutes for the elicitation of a single emotion [65]. In [21], the authors recommend the use of videos validated to induce single, primary emotions; (4) To avoid bias in the recognition of the different emotions, an equal duration of emotional eliciting videos and in-between non-stimuli phases is preferred; (5) The implementation of rewards has shown to be a good practice, leading to the participants motivation for a detailed description of their emotional experiences and lasting enrolment; (6) Most datasets rely on ECG and GSR data, sensors correlated to the discrimination of high arousal states, so further modalities should be tested for valence discrimination.

Lastly, it would also be very beneficial to have a publicly available collection of signals acquired from the same users at various interdependent sessions separated by significant large time intervals over time, thus, allowing a contextual longitudinal analysis of the user emotional experiences. In addition, a research line to explore concerns the exploration of large-scale groups (in large audiences [95], [96]) versus individual experiences, since individuals tend to respond differently if they are alone or in groups.

III. METHODS

This section describes the overall main steps required in the development of an ML system for emotion recognition, summarised in the diagram in Fig. 15.

A. SIGNAL PRE-PROCESSING

During the data acquisition protocol, many events may occur causing the degradation of the sensor signal with noise and external interference, namely, subject movement, electrodes disconnection, unstable ambient temperature and humidity, subject-dependent physiological dysfunctions, electrostatic artefacts and other non-related user movements. Therefore, signal pre-processing methodologies should be implemented on the raw signal and it is usually the first step in the

TABLE 2. Summary of the publicly available datasets for emotion and stress recognition specificities, namely, domain (Dom), number of subjects (#S), number of stimuli (#Stimuli), location (Loc): lab(L); constrained (C); daily-living (F), purpose, annotations and acquired modalities. Table adapted from [8].

Dom	Name	#S	#Stimuli	Loc	Purpose	Annotations	Modalities
E	**DEAP** [22]	32	40 1min long music videos	L	Analysis of human affective states	Arousal, valence, liking, dominance, familiarity	ECG, EDA, EEG, EMG, EOG, RESP, SKT, face video
E	**MAHNOB-HCI** [21]	27	20 35-117s long film excerpts	L	Emotion recognition and implicit tagging research	Valence, arousal, dominance, predictability, emotional keywords	ECG, EDA EEG, RESP, SKT, face and body video, eye gaze tracker, audio
E	**ASCERTAIN**	58	36, 51-128s long movie clips	L	Implicit personality and affect recognition	Valence, arousal, liking, engagement, familiarity, Big Five	ECG, EDA, EEG, facial video
E	**Eight-Emotion** [92]	1	8	L	Test for the presence of unique physiological patterns for the emotion set	Neutral, anger, hate, grief, joy, platonic love, romantic love, reverence	ECG, EDA, EMG, RESP
E	**EMDB** [57]	32	52 40s long non-auditory film clips	L	Study of affective film clips without auditory content	Arousal, valence, dominance	SCL, HR
E	**AMIGOS** [68]	40	1st exp: 40 participants watched 16 250s long videos. 2nd exp: 4 14min long videos alone and in groups	L	Research of affect, personality traits and mood on Individuals and groups	Big-Five personality traits and PANAS. Valence, arousal, dominance, liking, familiarity and basic emotions. External annotation of valence and arousal	Audio, visual, depth, EEG, GSR and ECG
E	**DECAF** [93]	30	40, 1min long music video segments used in DEAP and 36 movie clips	L	Decoding user physiological responses to affective multimedia content	Valence, arousal, dominance	ECG, EMG, EOG, MEG, near-infrared face video
S	**DSDRWDT** [94]	24	24 drives of 50min-1.5 h	FC	Analysing physiological data during real-world driving tasks to determine a driver's relative stress level	Low, medium, high stress levels	ECG, EDA, EMG, RESP
S	**RWDADW** [95]	10	30mins drives	FC	Assessing the driver's workload	Five different road types, discrete scale of driver workload	ACC, ECG, EDA, GPS, SKT, brightness level
E + S	**WESAD** [96]	15	20mins baseline + funny video clips for 392 seconds + TSST (5min speech on the subject personal traits and mental arithmetic count from 2023 to 0, doing steps of 17) + guided meditation for 7mins	L	Wearable stress and affect detection	Neutral, stress, amusement using SAM, PANAS, SSSQ and STAI	Chest: ACC, ECG, EDA, EMG, RESP, SKT; wrist: ACC, EDA, PPG, SKT

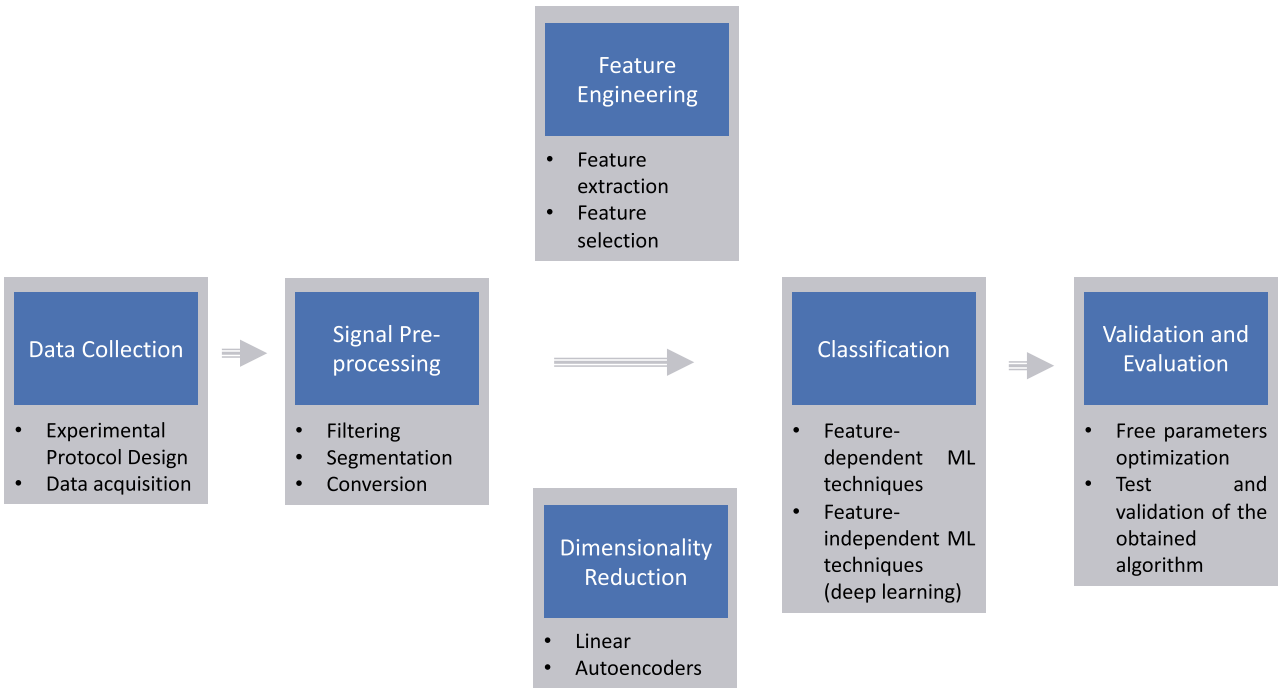

FIGURE 15. Schematic representation of a machine learning process for emotion recognition.

elaboration of an ML system. It consists of: synchronisation of the different sensor's signals; removal of data-loss and null values; (generally, through linear interpolation); filtering, noise, and outlier removal. The type of filter and its characteristics depends on the type of sensor and the study goal, as described next:

(a) **Electrocardiography (ECG)**: ECG data is commonly contaminated with powerline interference (50-60 Hz), electrode contact noise, motion artefacts, muscle contraction, baseline drift, instrumentation noise generated by electronic devices and electrosurgical noise [97]. According to [98], the desirable bandpass to maximise the QRS energy is approximately 5-15Hz. In [99], a 3^{rd} order Butterworth filter with 0.002 Hz and 100 Hz cut-off frequencies was applied to remove the effects of noise and baseline wander from the ECG data. In the works of [20], [21], the trend of the ECG and GSR signals was removed by subtracting the temporal low-frequency drift. The low-frequency drift was computed for each sensor by smoothing the signals with a 256 points moving average filter. In [100], a 300 order bandpass Finite Impulse Response (FIR) filter with a Hamming window, and cut-off frequencies of 5Hz and 20Hz is developed. In [101], a 4^{th} order bandpass Infinite Impulse Response (IIR) Butterworth filter between the frequencies of 2Hz and 30Hz. This filter removes EMG activity noise and 50Hz noise that could be induced by some badly filtered power supply. The pre-processing step aims to clean the signal to enhance its main characteristics, such as the R peaks and QRS complex, from which features, such as HR can be extracted. The easiest event to detect on an ECG signal is the R wave component since it presents the higher amplitude component. To detect the R-peaks many algorithms can be applied, namely: Pan and Tompkin's algorithm [98], Hamilton's [102], Christov [103], Engelse and Zeelenberg [104], ECG Slope Sum Function [105]. A detailed overview comparing the aforementioned methods can be found in [100]. In [19], a Teager Energy Operator (TEO) was used to detect the R-peak in the raw ECG signal. Posteriorly, if the baseline drift was prohibitively high, a median filter was used to estimate the baseline wander (low-frequency flotation) to generate a baseline-removed signal. After the R-peaks had been detected, the spikes train could be transformed into a continuous-time signal called HRV by interpolation and downsampling from which frequency domain features were extracted.

(b) **Electrodermal Activity (EDA)/ Galvanic Skin Response (GSR)**: The GSR ANS signal data presents low-frequency physiological patterns, thus, a low-pass filter is often applied to remove high-frequency noise. After noise removal, the filtered GSR signal can be de-trended by smoothing the signal over a given interval [8]. A further pre-processing technique is the separation of the SCL and SCR components. In [19], the SCL and SCR were separated by downsampling the signal to 20 samples/s, differentiation and subsequent convolution with a 20-point Bartlett window. The occurrence of the SCR was detected by finding two consecutive zero-crossings, from negative to positive and positive to negative. The amplitude of the SCR was obtained by finding the maximum value between these two zero-crossings. The detected SCRs with amplitude smaller than 10% of the maximum were excluded. Further techniques can be found in [2], [106].

(c) **Photoplethysmography (PPG) or Blood Volume Pulse (BVP)**: is usually prone to low-frequency motion artefacts caused by poor contact with the fingertip photosensor, variations in temperature and bias in the instrumentation amplifiers causing baseline drift. Therefore, motion artefacts can be removed using a high-pass filter [8], [22]. Additionally, high-frequency powerline interference artefacts caused by mains power sources interference can be induced onto the PPG recording probe or cable and removed using a low-pass filter. In [107], a 4^{nd} order Butterworth bandpass filter with 1-8Hz cut-off frequencies is used. After cleaning the signal from noise to enhance its characteristics, its maximum peaks are identified, corresponding to the heartbeats, used to extract relevant features [22]. In [23], the derivative is computed and a threshold is applied to determine the signal maximum peaks. Zong *et al.* [105] developed an algorithm using a windowed and weighted slope sum function to extract arterial blood pressure waveform features. Then, adaptive thresholding and search strategies are applied to the weighted slope sum function of the signal to detect arterial blood pressure pulses and to determine their onsets.

(d) **Respiration (RESP)**: Low-pass noise removal filters are often applied. Additionally, if necessary the filtered signal can be de-trended by smoothing the signal. According to [108], The dominant frequency of the respiratory component is about 0.25Hz. In [109], a 30 order FIR filter with frequency cut off 0.15 Hz was applied. In [18] it is used a 4^{th} order Butterworth bandpass filter with a pass-band from 0.1-0.35Hz [94]. Similarly, as the previous signals, the RESP signal can be de-trended using a moving average. From the filtered signal, the respiration rate can be computed from the zeros of the signal first derivative [18].

(e) **Electromyography (EMG)**: The EMG signal amplitude varies from a couple of mV up to tenths of mV and its spectrum contains frequencies in the 10-500Hz range. Thus, a high-pass filter with cutoff frequency superior to 100Hz can be applied to remove noise and heart-related information. The EMG data noise, on the whole, consists of: inherent electronics equipment noise (low-frequency noise), ambient noise (power line interference), motion artefacts (1-10Hz), inherent instability of signal (for signals with frequency components ranging between 0-20 Hz), ECG artefacts, cross talk, electrode-electrolyte-skin contact and baseline shifts. In [94], the signal DC component was removed by applying a high-pass filter. The filtered signal was segmented into 5 seconds, windows from which statistical and frequency-domain features were extracted.

Additionally, a low-pass filter with a cut-off frequency of 50 Hz was applied to the raw EMG signal. The signal was segmented into 60 seconds-long windows, from which temporal features were extracted.

(f) **Skin Temperature (Temp) or (SKT)**: In SKT data, low-pass filters are generally applied in order to remove high-frequency noise. In [110], the signal was down-sampled at 50Hz and then passed through a low-pass filter to remove noise. In [94], the raw SKT signal was used to extract the signal features. Likewise in [19], with the signal being posteriorly segmented in 50 seconds-long windows.

(g) **Electroencephalography (EEG)**: In EEG data common noise sources consist of poor contact of the electrodes, perspiration of the patient (affecting the electrode impedance with low-frequency artefacts), baseline drift caused by variations in temperature and bias in the instrumentation and amplifiers, power-line noise from wires, fluorescents light and other equipment. Many methodologies can be applied to remove noise in EEG data and the signal-to-noise ratio, namely Adaptive Filters, LMSAlgorithm or NLMSAlgorithm [111]. In [66], the EEG data was acquired using a sampling frequency of 128Hz. The signals were average-referenced and high-pass filtered with a 2Hz cut-off frequency. The eye artefacts were removed with a blind source separation technique. In [21], the EEG data was downsampled to 256Hz, and then a high-pass filter with a 2Hz cut-off frequency was applied.

(h) **Eye Gase**: The EOG signal information is mainly contained in the low frequencies, degraded with the subject movement and the equipment high-frequency noise. In [112], the EOG signals were filtered in the band of interest using a notch filter at 50Hz, and a 4^{th} order Butterworth bandpass filter with cut-off frequencies of 0.2Hz and 30Hz.

Additionally, as previously stated in Sub-section II-C, inertial sensors such as ACC and GYR, ambient sensors such as BAR and MAG, usually applied in human activity recognition, can provide a deeper insight into the user context and daily-living. These are usually at disposal in wearable fitness bracelets and watches and can enable a deeper insight into the subject goals, behaviour, cognition and emotions according to their response to a certain activity. Daily-living activates are generally below 20Hz, therefore, a low-pass filter with 15-20Hz cutoff frequency is often applied to inertial sensors. In the ACC sensor, the gravitational and body component can be separated using a high-pass filter with around 0.3Hz cutoff frequency. After the signal is filtered, it is segmented in static or sliding windows. The time between a stimulus and its physiological response depends on the subject and the signal modality and requires a trade-off between the window resolution and the algorithm time and computational complexity. Thus, the definition of the size of the windows is a difficult task [8]. According to [113], emotional physiological responses are generally averaged over 60- or 30-s intervals, 1/2- or 10- seconds intervals and 120-, 180-, or 300- seconds intervals. In [94], segmentation of the sensor signals was done for all signals but statistical and feature domain EMG features, using a 60s sliding window with a 0.25 seconds shift. The ACC features were computed with a window size of 5 seconds. The aforementioned, statistical and feature domain EMG features, were computed using a 5 seconds window. In [19] 50s windows were used.

In a subject-independent algorithm and ML distance-based algorithms, in order to diminish the individual physiological responses differences among the training subjects signals, the data should be normalised [114]. Data normalisation is usually performed with respect to the maximum and minimum values of the respective participant data or through the subtraction of the mean and division by its standard deviation (STD) [110].

B. DATA REPRESENTATION

The recognition of emotional states can be performed based on two different methodologies: (1) Feature-class representation feature-dependent ML techniques; (2) Feature-independent ML methodologies such as deep learning (DL) approaches. Sub-section III-B.1 focus on the former, being divided in Feature Extraction, Feature Selection and Feature Fusion. On the other hand, Sub-section III-B.2 focus on Feature-independent ML models.

1) FEATURE-CLASS REPRESENTATION

In general, when designing a traditional ML system, after the signal is pre-processed, a feature engineering step is implemented, in an attempt to maximize the informative content of the subject physiological signals. After feature engineering, the returned input is ready to be introduced into a classifier from which an output identifying the subject emotion class label is returned.

a: FEATURE EXTRACTION

Once the raw signal is cleaned and segmented in windows, metrics describing the physiological signals can be extracted. These metrics denoted as features, characterise the signal in a compact manner and allow the comparison between different signals in transformed dimensions enhancing informative signals characteristics. The characteristics can belong to temporal, statistical or spectral-domain, be linear or non-linear features, and unimodal or multimodal features [8]. Table 3 presents some of the most commonly extracted features from time-series grouped by their domain.

Most papers focus on the use of temporal, statistical and spectral domain characteristics. However, physiological signals present non-linear characteristics, hence, at present, the focus is on developing and extracting non-linear features. A deeper insight into non-linear features can be found in [115]–[118]. In the work of [116], a method using a recursive graph and recursive quantitative analysis is used to extract non-linear features from the EMG, SKT and RESP. In the work, the extracted features achieved a superior classification

TABLE 3. Overview of features commonly extracted from time-series signals grouped by their domain.

Domain	Features
Temporal	Maximum, minimum, centroid, median/mean absolute deviation/difference, zero-crossing rate, linear regression, range, absolute integral
Statistical	Mean, median, standard devidation (STD), variance, interquartile range, root mean square, skewness, kurtosis, histogram
Spectral	Total energy, spectral centroid, spectral spread, spectral skewness, spectral kurtosis, spectral slope, spectral decrease, spectral roll-on/off, spectral variation
Non-Linear	Lyapunov exponent, SD1 and SD2 from poincaré plot, SD1/SD2 ratio, sample entropy, approximate entropy, correlation dimension, Short-term and long-term fluctuations of detrended fluctuation analysis, mean line length of diagonal lines in recurrence plot (RP), maximum line length of diagonal lines in RP, Recurrence rate (percentage of recurrence points in RP), determinism (percentage of recurrence points which form diagonal lines in RP), Shannon entropy of diagonal line lengths' probability distribution

in comparison to traditional temporal and statistical features. In [117], both traditional time and frequency domain The HRV analysis together with nonlinear/complexity analysis features from ECG data were combined to recognise states of panic and pre-panic. The research concluded that the models that combined domains via data fusion achieved the greatest accuracy. Additionally, modality-dependent features can be extracted, as displayed in Table 4, where modality-based features are presented and grouped by their domain.

From the ECG data, as described in Sub-section III-A, the R-peaks are generally identified, from which inter-beat intervals, HR and HRV can be computed. The HRV describes the temporal variation between consecutive heartbeats. The HRV is modulated by the two branches of the ANS: PNS and SNS in a cooperation known as sympathovagal balance. Thus, HRV allows a deeper insight into the ANS system correlation with emotional changes. In [21], it was observed that pleasantness of stimuli can increase peak HR response, with HRV decreasing with fear, sadness, and happiness. Additionally, spectral features derived from HRV were shown to be a useful feature in emotion assessment [21]. From the R-peaks, a new time-series signal can be interpolated, from which various temporal, statistical and spectral HRV features can be derived [8]. The NN20/50 and pNN20/50 metrics denote the number and percentage of successive RR intervals differing by more than 20ms and 50ms, respectively. The spectral HRV features are computed from the Fourier transform of the interpolated R-peaks signal, reflecting the SNS and PNS responses of the ANS. The frequency-domain signal can be decomposed in four different frequency bands: the ultra-low frequency (ULF) (0 to 0.003Hz), very-low frequency (VLF) (0.003 to 0.03Hz), low-frequency (LF) (0.03 and 0.15Hz) and high-frequency (HF) (0.15 to 0.4Hz) bands. The LF band is primarily correlated with the SNS with moderate PNS activity influence. The HF band, on the other hand, is correlated with the PNS. Thus, the LF/HF ratio is usually computed to gather SNS to PNS influence on the cardiac activity [8]. From the ECG histogram, additional geometric metrics can be extracted, namely, the triangular interpolation index (TINN) providing the baseline width of the RR interval histogram, HRV triangular index, consisting of the integral of the RR interval histogram divided by the height of the histogram, log and differential index, the difference between the widths of the histogram of differences between adjacent [115]–[118].

Regarding the EDA signal, temporal, statistical and spectral features are generally extracted with the mean value of the GSR signal being correlated to the level of arousal [119]. In [21], the following features were extracted from EDA data: average skin resistance, average of derivative, average of derivative for negative values only, proportion of negative samples in the derivative vs all samples, number of local minimum average rising time, ten spectral power in the $[0-2.4]$Hz bands, zero-crossing rate of skin conductance slow response (SCSR) $[0-0.2]$Hz, zero-crossing rate of the skin conductance very slow response (SCVSR) $[0-0.08]$Hz, SCSR and SCVSR mean of peaks magnitude. As described in Sub-section II-C, the EDA signal can be decomposed in two components: SCL and SCR. The SCL is the baseline tonic component, whose degree of linearity has been proved to be a useful feature for emotion recognition [8], [106]. On the other hand, the SCR consists of the ANS response to a stimulus. From the SCR further temporal, statistical, spectral and morphologic modality-based features can be extracted, namely, frequency-phasic response rate, amplitude (onset-peak amplitude difference), latency between stimulus and onset, rise time (onset-peak time difference), half-rise time (time between onset and 50% amplitude), 50/63% recovery time (time between peak and 50/63% amplitude), respectively, number of SCR events, sum of SCR startle magnitudes and response duration, SCR/SCL ratio, std of SCR/SCL ratio [92], [94], area under the identified SCR events. Additionally, in [21], 10 spectral power values in the $0-2.4$Hz frequency bands features were extracted from the EDA signal.

For the respiration data, the authors in [94], [120] extracted the signal breathing rate, inhalation (I) and exhalation (E) duration, the ratio between I/E, stretch (the difference between the peak and the minimum amplitude of a respiration cycle), and the volume of air inhaled/exhaled. These

TABLE 4. Overview of modality-based features commonly extracted from physiological signals grouped by their domain and modality [8].

Modality	Domain	Features
ECG/PPG	Temporal	Temporal features on signal and r-peak interpolation, number and percentage of successive RR intervals differing by more than 20 ms (NN20, pNN20) or 50 ms (NN50, pNN50), pNN50/pNN20 ratio, integral of the RR interval histogram divided by the height of the histogram, baseline width of the RR interval histogram,
	Statistical	Statistical features on signal, HR, HRV
	Spectral	Total energy, spectral centroid, spectral spread, spectral skewness, spectral kurtosis, spectral slope, spectral decrease, spectral roll-on/off, spectral variation applied to: ultra low (ULF, 00.003Hz), very low (VLF, 0.0030.03Hz), low (LF,0.030.15Hz), and high (HF, 0.150.4Hz) frequency bands of HRV, normalised LF and HF, LF/HF ratio, Total spectral power
	Non-linear:	Non-linear on signal and R-peaks interpolation
	Geometrical	Triangular interpolation of R-peaks intervals, histogram (TINN)
	Multimodal	Respiratory sinus arrhythmia (respiratory sinus arrhythmia (RSA)), respiration-based HRV decomposition
EDA	Temporal	SCL degree of linearity, temporal features on SCR signal, number of SCR events, sum of SCR startle magnitudes and response durations, area under the SCR events, temporal features on SCR amplitudes, rise and 50%/60%recovery times
	Statistical	Statistical features applied to: SCR signal, amplitudes, rise and 50%/60% recovery times
	Spectral	Spectral features applied to SCR signal, 10 spectral power in the 02.4Hz bands
RESP	Temporal	Temporal features applied to: breathing rate, inhalation and exhalation duration, ratio between I/E, stretch, volume of air inhaled/exhaled
	Statistical	Statistical features applied to: breathing rate, inhalation (I) and exhalation (E) duration, ratio between I/E, stretch, volume of air inhaled/exhaled
	Spectral	Spectral features applied on (00.1 Hz, 0.10.2 Hz, 0.20.3 Hz and 0.30.4 Hz) bands
	Multimodal	RSA

metrics allowed to reach a deeper insight into the subject breathing cycles [8]. In addition, the authors in [92] calculated four spectral power features by summing the energy in the sub-bands (0-0.1Hz, 0.1-0.2Hz, 0.2-0.3Hz and 0.3-0.4Hz). Lastly, the respiratory sinus arrhythmia (RSA) can be extracted. In [21], the authors extracted the following features: band energy ratio (logarithm of the energy between the lower (0.05-0.25)Hz and the higher bands (0.25-5)Hz, average resp signal, mean of derivative, STD, range or greatest breath, breathing rhythm (spectral centroid), breathing rate, 10 spectral power in the bands from 0 to 2.4Hz, average peak-to-peak time, and median peak-to-peak time. Similar features were extracted in [20].

For the EMG signal, traditional temporal and statistical features are often extracted. In [94], mean and STD of the EMG signal dynamic range, absolute integral, median of the EMG signal 10^{th} and 90^{th} percentile mean, median and peak frequency, energy in seven bands, number of peaks, mean and STD of the peak amplitudes and normalised sum of peak amplitudes were computed. The spectral energy was computed in seven evenly spaced frequency bands from 0 to 350Hz. Most of the power in the spectrum of an EMG during muscle contraction is in the frequency range between 4 and 40Hz. Thus, in [21] the muscle spectral activity was obtained from the energy of EMG signals in this frequency range, along with the signal mean and variance.

The SKT is a low-frequency slowly varying signal, hence, traditional temporal and statistical features can be extracted providing useful information [65]. In [94], mean, STD of the SKT minimum and maximum dynamic range and slope. Similarly, in [20], the average, average of its derivative and spectral power in the bands [0-0.1]Hz and [0.1-0.2]Hz.

From the EOG sensor, the authors in [20], [21] extracted the eye-blinking rate, energy, mean and variance of the signal. For further eye gaze data, we refer the reader to [20], where features based on the pupil diameter, gaze distance, eye linking, and gaze coordinates were extracted.

Regarding EEG data, in [21], power spectral features were extracted, namely, the logarithms of the spectral power from theta (4-8Hz), slow alpha (8-10Hz), alpha (8-12Hz), beta (12-30Hz), and gamma (30+Hz) bands, and additionally, the difference between the spectral power of all the symmetrical pairs of electrodes on the right and left hemisphere in

order to measure the asymmetry in the brain activities due to emotional stimuli.

b: FEATURE SELECTION AND DIMENSIONALITY REDUCTION

After feature extraction, ideally, the resulting feature vector expresses the data quality and will modulate the classification performance. If the emotions were not induced, the degree of emotion elicitation is distinct between different subjects, the data presents high sensor noise or motion artefacts, and/or the data contains outliers [110], resulting in a poor classification performance.

The recognition of human emotions is usually a multi-modal problem, thus, prone to the curse of dimensionality, since the feature vector will have a high dimension due to the high number of features. In order to solve this issue, feature selection techniques are generally applied to decrease the data dimensionality. These can be divided into wrapper, filter and embedded methods. Filter methods select variables independently of the model and choose subsets of variables to detect independence among all the features. Thus, in filter methods, first the features are ranked according to a criterion and then the features with highest rankings are selected for the next ML step. Common criteria for the feature selection in emotion recognition are the Fisher Score [121] methods based on mutual information [122] and ReliefF [123], [124]. Per example, in [121], an algorithm using Fisher Score was used to select optimal features for Thai Speech Emotion Recognition. The algorithm was able to reduce the use of 14 to 7 features with a high reduction of computing time.

However, filter methods ignore the effects of the selected features subset on the performance of the ML classification algorithm. Wrappers methods solve this issue by using a classifier to evaluate the quality of the different features and the resulting different features subsets. Example of wrapper methods are forward selection and backward elimination. In [125], a recursive feature elimination and margin-maximising feature elimination feature selection methods were performed. In [126], the interactive Feature Selection method based on reinforcement learning was developed and compared against random selection and Sequential Forward Selection (SFS) and Genetic Algorithm Feature Selection (GAFS), showing an SLight increment in the performance.

To implement a wrapper method it is necessary to perform an exhaustive search over all features, which can result in high computational complexity and become impractical for a reliable real-life solution. Embedded methods solve this issue, being more computationally efficient through the combination of both aforementioned techniques. In Embedded methods, first, it is implemented the filter method, decreasing the size of the search space and then the wrapper method is implemented on a lower dimensionality space in comparison to its raw form. For further information on feature selection, we refer the reader to [127].

The aforementioned methods are highly time consuming and prone to overfitting, thus, many authors prefer to use dimensionality reduction transformations such as Principal Component Analysis (PCA) [128] (projects the data into a reduced dimensionality space preserving the large variability of the data) or Fisher Linear Discriminant [38], [66], [71], [92] (projects the data into higher variance between different classes and smallest within each class). For further information on dimension reduction algorithms, the authors refer the reader to [129]. In [128], the introduction of PCA to an SVM classifier as a pre-processing step enabled to increase the emotion recognition performance by 3.1%.

c: FEATURE FUSION

The literature has shown that the classification performance improves with the simultaneous exploitation of different signal modalities [21], [130]. Modality fusion can be performed at two main levels: feature fusion [23], [131], [132] and classifier fusion [21], [71], [130], [133]. In the former, features are extracted from each modality and latter concatenated to form a single feature vector space, to be used as input for the ML model. On the other hand, in decision fusion, from each modality, a feature vector is extracted to form a classifier prediction. Hence, with n modalities, n classifiers will be created and n predictions obtained and combined to yield a final result. Additionally, feature fusion requires normalisation of features; on the other hand, the merging of classifiers can be done with parallel processing architectures, reducing the computation time.

In [71], decision fusion was implemented estimating a per-sample weighting α for the different modalities where the final decision is a weighted sum of the outputs from the classification of the individual modalities. In [131], a feature fusion approach is proposed and compared with decision-level fusion and non-fusion approaches used as input to hidden Markov models (HMM) for predicting emotions using the DEAP dataset. The developed fusion approach showed significant improvements in the model's accuracy. The authors in [134], combined both methodologies, using first feature fusion to independently combine the features, and then decision fusion to combine the results of each classifier for a final recognition classification.

Thereupon, while feature fusion is simpler and has lower computational complexity to compute, it is unable to deal with poor data and requires for the different modalities to be synchronous, which decision fusion does not. A further difference is that decision fusion allows the use of weights to adjust the contribution of each modality to the final prediction output [21], while feature fusion is constrained to an use-ignore method.

2) DEEP LEARNING IN EMOTION RECOGNITION

On a second methodology, instead of representing an object by a feature-based representation, the input for the ML model can be a distance matrix or a cleaner/raw version of the physiological signal.

On the latter methodology, a common approach is to use an autoencoder as a signal pre-processing methodology before it is provided to a model classifier. An autoencoder is an unsupervised learning (UL) approach trained by back-propagation with an input layer, hidden layers and an output layer. The input layer is equal to the output layer and the hidden layer usually has smaller dimensionality than the input layer. The hidden layer minimises the reconstruction error between the data input and the data output reconstruction. Thus, an autoencoder forces a data dimensionality reduction and enhances the most salient characteristics of the input data. Many different variants of the general autoencoder architecture exist with the overall objective of obtaining cleaned, meaningful, information of the input data. The authors of [35] assessed the emotional manifestations of relaxation, anxiety, excitement, and fun, embedded in GSR and BVP data to compare a Convolutional Neural Network (CNN) approach against ad-hoc feature extraction methodologies. An autoencoder is used to denoise the signal and lower its dimensionality. The experimental results showed that the model outperforms the standard feature extraction across all affective states examined.

In [60], an effective pre-processing method is proposed as an alternative to traditional feature extraction methodologies. In the proposed method, a hybrid neural network combines a CNN and a Recurrent neural network (RNN) to discriminate emotion states by learning a compositional spatial-temporal representation of raw EEG data. The experimental results show that the proposed pre-processing method increases the emotion recognition accuracy by approximately 32%, attaining a mean accuracy of 90.80% and 91.03% on valence and arousal classification, respectively.

The autoencoders can be used having as input the physiological signal or a set of extracted features as in [135], [136]. For example, in [135], the authors applied a Bi-modal Deep AutoEncoder to extract shared representations of EEG and eye movement data. The proposed model was able to reach a mean accuracy of 91.01% and 83.25% on the SEED and DEAP datasets, respectively.

An autoencoder can be an alternative to PCA and the aforementioned data reduction methodologies, presenting the advantage of being able to learn non-linear complex data representations.

C. CLASSIFICATION

Traditional model-based ML methodologies are divided in supervised, unsupervised and semi-supervised methodologies. In supervised learning (SL), a model is created from a training set mapping the physiological signal features to its labels. Examples of SL algorithms are Naive-Bayes (NB) [137], k-Nearest Neighbour (K-NN) [137], [138], Support Vector Machine (SVM) [137], [139], Linear Discriminant Analysis (LDA) [138], Quadratic Discriminant Analysis (QDA) [138] and many more. The SVM classifier is the most commonly applied in the literature (see Fig. 16). For example, in [139], an SVM method on ECG and RESP

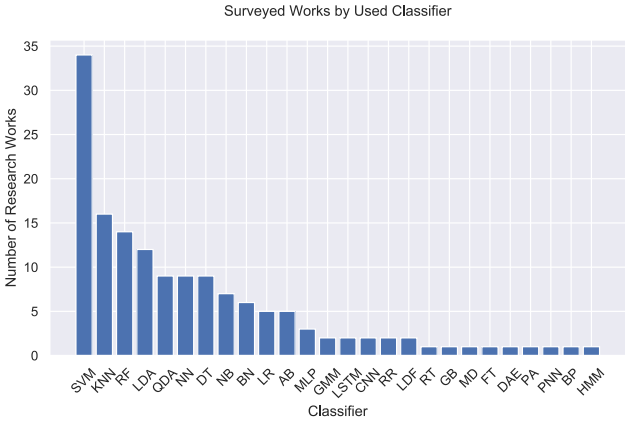

FIGURE 16. Histogram of the number of publications surveyed for this document per classifier.

data was applied to recognise joy, sadness, anger, and pleasure, achieving a recognition accuracy of 81.82%, 63.64%, 54.55%, and 30.00%, respectively. In [137], the recognition performance of a Random Tree (RT), Decision Tree (DT) J48, NB, K-NN, SVM, and Multilayer Perceptron Neural Networks (MPNN) classifier using HR, GSR and SKT data was tested. The K-NN classifier attained the best performance with an average accuracy of all the personalised models of 97.78%.

In [138] a K-NN, LDA, QDA, and Radial Basis Function Network (RBFN) classifier was implemented on respiratory and facial muscle activity data for the assessment of the fearful, sad and neutral emotion states. The experimental results showed that the K-NN model was the best for both subject and stimulus-dependent, and subject and stimulus-independent classification; while the RBFN model was the best for subject-independent classification, and the LDA model for stimulus-independent classification. In [117], Passive Aggressive (pA), Gradient Boosting (GB), DT, Ridge, SVM, Random Forest (RF), K-NN, Logistic Regression (LR) classifiers were analysed for ECG data for panic and no-panic assessment. The RF classifier achieved an accuracy of 97.2% and 90.7% for panic and pre-panic recognition, respectively. For further information on SL algorithms, the authors refer the reader to [140].

On the other hand, in UL, only the signal features are provided and a model is created from the unlabelled data structure. Usually, the data's structure is found in the form of clusters maximising intra-class similarity and minimising inter-class similarity. Example of UL algorithms are the k-means, affinity propagation, spectral clustering, hierarchical clustering, density-based spatial clustering of applications with noise (DBSCAN), Gaussian mixture models (GMM) and many more.

For example, in [141], the authors used a GMM-based model to classify the EDA data into arousal-not arousal, presenting an accuracy of 74.3%. In [142], EDA, HR and SPO2 data was used to create an unsupervised GMM model

able to accurately separate relaxation, physical, emotional and cognitive stress status with an accuracy greater than 84%. For further information on UL algorithms, the authors refer the reader to [143].

Lastly, Semi-Supervised Learning (SSL) is a hybrid form between both aforementioned methodologies, creating an SL classifier on labelled data and posteriorly incorporating further information from the unlabelled data. Examples of SSL algorithms are Self-Training (ST) and Active Learning (AL). In [144], an SSL on EEG data for affective state recognition using DL. The experimental results show that the proposed model surpasses extensive baselines in classification and the proposed reinforced process outpaces random annotation. For further information on SSL algorithms, the authors refer the reader to [145].

Additionally, many works have applied DL feature-independent methodologies such as CNN [35], [146]–[149], RBM [135], [150], [151], autoencoder [35], [135], [152], [153], and deep belief networks (DBNs) [151], [154], [155], Long short-term memory (LSTM) [136], [156]–[158], probabilistic neural network (PNN) [124] and many others. In [151], a DBN was applied for emotion recognition using EEG, GSR, EMG and EOG data achieving an accuracy of 78.28%, 70.33%, 70.16% for valence, arousal and dominance, respectively. In the presented methodology, features are extracted using a DBN and are used as input to a Restricted Boltzmann Machine (RBM). The presented methodology is an SSL approach, thus, able to reduce significantly the amount of labelled data required for learning the model and uses a DBN instead of feature engineering techniques. In [146], a CNN is used to extract features from ECG and GSR data, then, through fully connected network layers, the emotional assessment on arousal and valence is obtained. The CNN in comparison with the classic algorithms of ML demonstrated a better performance in the emotion detection and a large number of instances showed to directly influence the emotion prediction performance. For further information on DL, the authors refer the reader to [159], [160].

To conclude, the high accuracy results support the hypothesis of the correlation between emotional states and physiological data. The SL methodologies although achieving great results, present the disadvantage of requiring a high amount of labelled data to train the model. The UL methods allow to solve the issue of data annotation, however, at the cost of lower prediction accuracy and lost of the class labels information. Additionally, traditional model-based ML classifiers require the data to be previously pre-processed and transformed into a feature vector which can present high complexity in a multi-modal approach as emotion recognition. The DL approaches remove the requirement for signal pre-processing and feature engineering, the latter, being one of the most time-consuming parts of an ML system. Instead, DL uses denoising and dimensionality reduction techniques such as auto-enconders, which have been applied for emotion recognition with great results. The literature suggests that the DL methodologies are highly appropriate for affective modelling and ad-hoc feature extraction can be redundant for physiology-based modelling [35]. However, the DL approaches present the disadvantage of behaving like a black box, which, once applied, do not show the relationship between the physiological signals and each emotion [110], require large amounts of data, and are extremely computationally expensive to train.

D. VALIDATION

Once defined the classifier, and trained (learned its hyperparameters) on the input data, the final step in an ML framework is the validation of the model in order to obtain an overall view of how the model will perform on never-before-seen data, as in a real-world scenario outside of laboratory constraints, i.e. the model must be able to generalise into new unseen data and avoid overfitting on the training set data. Hence, the model must find an equilibrium, fitting the training data well but with relative variability so it avoids overfitting and is able to generalise. A solution to have both input data and new never-before-seen data is to divide the data into a training set and an independent test set; or in a training set (to train a model), a validation set (to tune the model hyperparameters on unseen data) and a test set (to obtain evaluate the model performance). The test set should yield some characteristics to return meaningful results, namely: should take a considerable size, be representative of the entire data, and not be repeated in the train set.

A common methodology found in the literature to ensure a meaningful validation and is to perform k-fold Cross-Validation (CV). In k-CV, k iterations are performed with the data being partitioned into k equally-sized folds. In each iteration, k-1 folds are used for training and 1 fold for testing so a fold is used for testing only once. The results of the k iterations are then averaged and a final overall performance computed. The Leave-one-subject-out (LOSO) and leave-one-out (LOO) techniques are a specification of k-fold CV where k is set to one user and one sample, respectively. The LOO technique introduces the highest variability and returns the most pessimist classifier in comparison with k-CV. However, at the cost of high computational power, therefore it is used generally when there is a small amount of data. On the other hand, the LOSO validation tests on an independent subject from the training set, thus, returns user-independent generalised results. To obtain a measurable evaluation of the model performance, the metrics presented in Table 6 are often applied: Accuracy- percentage of correctly classified samples; Precision- proportion of actual positives instances among the classified positive instances; Recall- proportion of positives correctly identified among the existing ground truth positives; Specificity- proportion of actual negatives correctly identified among the existing ground truth negatives; F1-score- the harmonic mean of precision and recall; mean square error-average squared loss per example over the whole dataset. For a visual interpretation and comparison between the classification results for each class, a confusion matrix is

TABLE 5. Illustration of a Confusion Matrix. (Nomenclature: TP: True-Positive samples, FN: False-Negative, FP: False-Positive and TN: True-Negative samples.

		Predicted	
		Positive	Negative
Real	Positive	TP	FN
	Negative	FP	TN

TABLE 6. Evaluation metrics for a binary classification.

Metrics	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision (P)	$\frac{TP}{TP+FP}$
Recall (R)	
Sensitivity	$\frac{TP}{TP+FN}$
TPR	
Specificity	$\frac{TN}{TN+FP}$
TNR	
F-Score	$\frac{2\times(P\times R)}{P+R}$

often used (see Table 5). A confusion matrix is a n×n matrix, *n* being the total number of classes. Each cell *Cij* is filled with the total number of predicted samples belonging to the class label *i* and predicted with the label *j*.

IV. DISCUSSION

Table 7 displays a summary of state-of-the-art research studies in the field of emotion recognition. Comparing the performance of the research papers is a difficult task since they often differ in the classifiers, the datasets used to train and test the model, form of validation and the extracted features and signal modalities used. Notwithstanding, an overall analysis of the current emotion recognition state-of-the-art can be performed, as follows:

(a) **Elicitation Material:** As observed in Fig. 17, displaying the number of publications surveyed for this document per elicitation material, video and films are the most commonly used elicitation material, namely the DEAP (music videos) and the IAPS (images) datasets.

(b) **Constrained vs. Unconstrained setting**: Most studies are performed in a lab setting, and these, on average, achieve higher accuracy. This observation arises from the fact that in-lab experiments are devised to elicit specific emotions, pre-validated and easily acquired and replicated in an elevated number of subjects with quality ground-truth annotation. Additionally, often the subjects are asked to remain still, thus, minimising movement artefacts (high source of error).

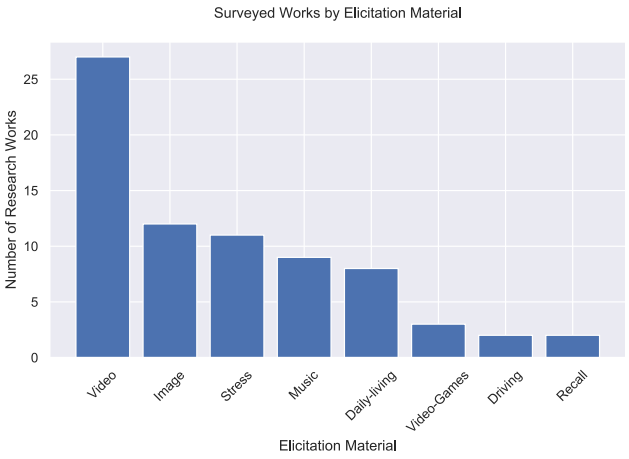

FIGURE 17. Histogram of the number of publications surveyed for this document per elicitation material.

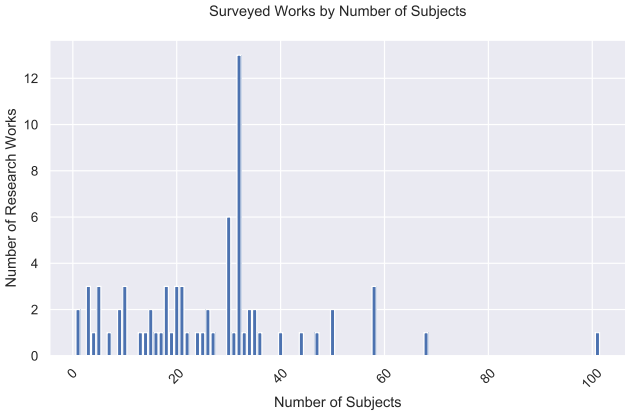

FIGURE 18. Histogram of the number of publications surveyed for this document per number of subjects in the datasets used.

(c) **Number of Subjects**: As it can be seen in Fig. 18, the vast majority of the surveyed publications reported the use of data between 1 and 50 subjects.

(d) **Subject-dependent vs. subject-independent:** Subject-dependent algorithms achieve on average higher results than subject-independent since subjects show high inter-dissimilarity with their elicitation emotions biased by the subject's physiological internal and external factors.

(e) **Emotion Models:** Most works focus on the implementation of binary classification techniques, separating arousal from valence and stress from no-stress activities.

(f) **Modalities:** Most research studies agreed that the classification performance increased with the increment of the number of signal modalities.

(g) **Classifiers:** Most works focus on using SL methodologies, namely SVM, kNN, DT, RF, AB, LDA, QDA, LR, NB and BN. However, SL algorithms rely on annotated data. Data annotation is time expensive, costly and very difficult since the users show difficulty describing the

TABLE 7. Summary of state-of-the-art research studies in the field of affect recognition and its main characteristics in terms of: author, year, stimulus, localisation, number of subjects (#), subject dependency (SD), emotion labels (Labels), modalities, classifier, validation method (Val) (leave-1-out (LOO); leave-1-subject-out (LOSO); 10-fold cross-validation (CV) and recognition rate (Rec Rate). Table adapted from [4], [8].

Author	Year	Stim	Loc	#	SD	Labels	Modalities	Classifier	Val	Rec Rate
Petrantonakis P C, et al. [165]	2010	IAPS	L	16	No	Happiness, surprise, anger, fear, disgust, sadness	EEG	KNN, QDA, MD, SVM	LNO	85.17%
Samara A, et al. [166]	2016	DEAP	L	32	Yes	Arousal, valence	EEG	SVM	10-fold	79.83%; 60.43%
Jianhai Zhang et al. [126]	2016	DEAP	L	32	Yes	Arousal, valence	EEG	PNN, SVM	10-fold	PNN: 81.76%; SVM: 82.00%
Ping Gong et al. [136]	2016	Music	L		Yes	Joy, anger, sadness, pleasure	ECG, EMG, RSP, GSR	DT		92%
Gyanendra K. Verma [167]	2014	DEAP	L	32	Yes	Terrible, love, hate, sentimental, lovely, happy, fun, shock, cheerful, Depressing, exciting, melancholy, mellow	EEG, ECG, GSR, EMG, EOG, RESP, SKT, face video	SVM, MLP, KNN, MMC	10-fold	EEG: 81%; peripheral signals: 78%
Vitaliy Kolodyazhniy et al. [140]	2011	Film clips	L	34	Both	Fear, sadness, neutral	ECG, GSR, RSP, SKT, EMG	KNN, MLP, QDA, LDA, RBNF		subj-dep: 81.90%; subj-indep: 78.9%
Dongmin Shin et al. [168]	2017	Videos	L	30	Yes	Amusement, fear, sadness, joy, anger, and disgust	EEG, ECG	BN		98.06%
Foteini Agrafioti et al. [169]	2012	IAPS, video game	L	44	No	Valence, arousal	ECG	LDA	LOO	Arousal: bipartition 76.19%; C.36% valence: 52%-89%
Wanhui Wen et al. [170]	2014	Videos	L	101	No	Amusement, grief, anger, fear, baseline	OXY, GSR, ECG	RF	LOO	74%, LOO
Jonghwa Kim et al. [171]	2008	Music	F	3	Both	Valence, arousal	ECG, EMG, RSP, SC	pLDA	LOO	Subj-dep: 95%; subj-ind: 77%
Cong Zong et al. [172]	2009	Music (AuBT)	1		Yes	Joy, anger, sadness and pleasure	ECG, EMG, SC, RSP	SVM	10-fold	76%
Valenza et al. [173]	2012	IAPS	L	35	No	Valence, arousal	ECG, EDR, RSP	QDA	40-fold	>90%
Wee Ming Wong et al. [174]	2010	Music (AuBT)	L		Yes	Joy, anger, sadness, pleasure	ECG, EMG, SC, RSP	PSO of synergetic neural classifer (PSO-SNC)	LOO	SBS: 86%; SFS: 86%
Leila Mirmohamadsadeghi et, al. [175]	2016	DEAP	L	32	Yes	Valence, arousal	EMG, RSP	SVM	LOVO	Valence: 74%, arousal: 74%, liking: 76%
Chi-Keng Wu et al. [176]	2012	Film clips	L	33	Yes	Love, sadness, joy, anger, fear	RSP	KNN5	LOO	88%
Xiang Li et al. [159]	2016	DEAP	L	32	Yes	Valence, arousal	EEG	LSTM	5-fold	Valence: 72.06%, arousal: 74.12
Zied Guendil et al. [177]	2015	Music (AuBT)	L		Yes	Joy, anger, sadness, pleasure	EMG, RESP, ECG, SC	SVM		95%
Yuan-Pin Lin et al. [178]	2010	Music	L	26	No	Joy, anger, sadness, pleasure	EEG	MLP, SVM	10-fold	82.29%
Bo Cheng et al. [179]	2008	Music (AuBT)	L		Yes	Joy, anger, sadnes, pleasure	EMG	BP		75%
Saikat Basu et al. [180]	2015	IAPS	L	30	Yes	Valence, arousal (HVHA, HVLA, LVHA, LVLA)	GSR, HR, RESP, SKT	LDA, QDA	LOO	HVHA: 98%, HVLA: 96%, LVHA: 93%, LVLA: 97%
Ingxin Liu et al. [181]	2016	DEAP	L	32	Yes	Valence, arousal	EEG	KNN, RF	10-fold	Valence: 69.9%, arousal: 71.2%
Mahdis Monajati et al. [182]	2012	Shock test	L	13	Yes	Negative, neutral	GSR, HR, RSP	Fuzzy adaptive resonance theory		94%
Lan Z et al. [183]	2016	IADS	L	5	Yes	Positive, negative	EEG	SVM	5-fold	73.10%
Zheng W L et al. [184]	2018	DEAP + video	L	47	Yes	HAHV HALV LAHV LALV	EEG	G extreme Learning Machine	5-fold	DEAP: 69.67%, SEED: 91.07%
Picard et al. [185]	2001	Clynes protocol	L	1	Yes	Neutral, anger, hate, grief, joy, platonic/romantic love, reverence	EDA, EMG, PPG, RESP	KNN	LOO	81%
Haag et al. [186]	2004	IAPS	L	1	Yes	Low/medium/high arousal and positive/negative valence	ECG, EDA, EMG, SKT, PPG, RESP	NN	3-fold	AR: <96%, VA: <90%
Lisetti and Nasoz [116]	2004	Movie clips and difficult mathematics questions	L	14		Sadness, anger, fear, surprise, frustration, amusement	ECG, EDA, TEMP	kNN; LDA; NN	LOO	72%; 75%; 84%

TABLE 7. *(Continued.)* **Summary of state-of-the-art research studies in the field of affect recognition and its main characteristics in terms of: author, year, stimulus, localisation, number of subjects (#), subject dependency (SD), emotion labels (Labels), modalities, classifier, validation method (Val) (leave-1-out (LOO); leave-1-subject-out (LOSO); 10-fold cross-validation (CV) and recognition rate (Rec Rate). Table adapted from [4], [8].**

Author	Year	Stim	Loc	#	SD	Labels	Modalities	Classifier	Val	Rec Rate
Healey and Picard [94]	2005	Driving	FC	24		3 stress levels	ECG, EDA, EMG, RESP	LDF	LOO	97%
Leon et al. [187]	2017	IAPS	L	9	No	Neutral/positive/negative valence	EDA, HR, BP	NN	LOSO	71%
Zhai and Barreto [188]	2006	Paced stroop test	L	32		Relaxed and stressed	EDA, PD, PPG, TEMP	NB; DT; SVM	20-fold	79%; 88%; 90%
Kim et al. [189]	2008	Natural	FC	68		Distinguish high/low stress group of individuals	PPG	LR	5-fold	áĹij 63%
Kim and André [171]	2008	Music	L	3	Both	HAHV HALV LAHV LALV	ECG, EDA, EMG, RESP	LDA	LOO	subj-dep: 95%, subj-indep: 70%
Katsis et al. [190]	2008	Simulated driving	L	10		High-low stress, disappointment, euphoria	ECG, EDA, EMG, RESP	SVM; ANFIS	10-fold	79%;77%
Calvo et al. [191]	2009	Clynes protocol	L	3	Both	Neutral, anger, hate, grief, joy, platonic/romantic love, reverence	ECG, EMG	FT; NB; BN; NN; LR, SVM	10-fold	one subject: 37%-98%, all subjects: 23%-71%
Chanel et al. [42]	2009	Recall	L	10		Positively/negatively excited, calm-neutral (in valence-arousal space)	BP, EEG, EDA, PPG, RESP	LDA, QDA, SVM	LOSO	<50%: <47%; <50%, binary: <70%
Khalili and Moradi [192]	2009	IAPS	L	5		Positively/negatively excited, calm (in valence-arousal space)	BP, EEG, EDA, RESP, TEMP	QDA	LOO	66.66%
Healey et al. [193]	2010	Daily-living	F	19		Points in valence arousal space, moods	ACC, EDA, HR, audio	BN; NB; AB; DT	10-fold	
Plarre et al. [122]	2011	Public speaking, mental arithmetic, cold pressor	L/F	21/17		Baseline, different types of stress (social, cognitive, and physical), perceived stress	ACC, ECG, EDA, RESP, TEMP	DT; AB; SVM/HMM	10-fold	82%, 88%, 88%
Hernandez et al. [194]	2011	Calls	F	9	Both	Detect stressful calls	EDA	SVM	LOSO	73%
Valenza et al. [195]	2012	IAPS	L	35	No	5 classes of arousal and five valence levels	ECG, EDA, RESP	QDA	40-fold	>90%
Koelstra et al. [22]	2012	DEAP	L	32	No	HAHV HALV LAHV LALV	ECG, EDA, EEG, EMG, EOG, RESP, TEMP, facial video	NB	LOSO	AR/VA/LI: 57%/63%/59%
Soleymani et al. [21]	2012	MAHNOB-HCI	L	27	No	Neutral, anxiety, amusement, sadness, joy, disgust, anger, surprise, fear	ECG, EDA, EEG, RESP, TEMP	SVM	LOSO	VA: 46%, AR: 46%
Sano and Picard [196]	2013	Daily-living	F	18	Yes	Stress, neutral	ACC, EDA, phone usage	SVM, kNN	10-fold	<88%
Martínez et al. [36]	2013	Video-game (maze-ball)	L	36	No	Relaxation, anxiety, excitement, fun	EDA, PPG	NN	10-fold	learned features: <75%, hand-crafted: <69%
Valenza et al. [195]	2014	IAPS	L	30	Yes	HAHV HALV LAHV LALV	ECG	SVM	LOO	Valence: 79%, arousal: 84%
Adams et al. [143]	2014	Daily-living	F	7	Yes	Stress, neutral (aroused, non-aroused)	EDA, audio	GMM		74%
Hovsepian et al. [197]	2015	Socioevaluative, cognitive, and physical challenges	L/F	26/20	No	Stress, neutral	ECG, resp	SVM/BN	LOSO	92%/>40%
Abadi et al. [93]	2015	DECAF	L	30		High/Low valence, arousal, and dominance	ECG, EOG, EMG, near-infrared face video, MEG	NB, SVM	LOTO	VA/AR/DO: 50-60%
Rubin et al. [119]	2016	Daily-living	F	10		Panic attack	PA: GB; DT; RR; SVM; RF; kNN; LR	ACC, ECG, RESP	10-fold	bin. panic: 73%-97% bin. pre-panic: 71% - 91%
Jaques et al. [198]	2016	Daily-living	F	30	No	Stress, happiness, health values	SVM; LR; NN;	EDA, TEMP, ACC, phone usage	5-fold	<76%: <86%; <88%
Zenonos et al. [199]	2016	Daily-living	F	4	No	Excited, happy, calm, tired, bored, sad, stressed, angry	ACC, ECG, PPG, SKT	kNN, DT, RF	LOSO	58%; 57%; 62%

TABLE 7. ***(Continued.)*** **Summary of state-of-the-art research studies in the field of affect recognition and its main characteristics in terms of: author, year, stimulus, localisation, number of subjects (#), subject dependency (SD), emotion labels (Labels), modalities, classifier, validation method (Val) (leave-1-out (LOO); leave-1-subject-out (LOSO); 10-fold cross-validation (CV) and recognition rate (Rec Rate). Table adapted from [4], [8].**

Author	Year	Stim	Loc	#	SD	Labels	Modalities	Classifier	Val	Rec Rate
Gjoreski et al. [200]	2017	Daily-living	L/F	21/5	No	Lab: no/low/high stress; field: stress, neutral	ACC, GSR, BVP, SKT	SVM, RF, AB, kNN, BN, DT	LOSO	*<73%/ <90%*
Mozos et al. [201]	2017	TSST	L	18		Stress, neutral	ACC, GSR, BVP, audio	AB, SVM, kNN	CV	94%; 93%; 87%
Schmidt et al. [96]	2018	WESAD	L	15	No	Neutral, fun, stress	ACC, ECG, GSR, EMG, RESP, SKT, BVP	DT, RF, kNN, LDA, AB	LOSO	*<80%/<93%*
Hao Tang et al. [138]	2017	DEAP	L	32		Arousal, valence	EEG, ECG, GSR, EMG, EOG, RESP, SKT, face video	Bimodal-LSTM	10-fold	Arousal: 83.23%, valence:83.83%
Wei Liu et al. [137]	2016	DEAP	L	32		Positive, neutral, negative	EEG	BDAE		83.25%
Tripathi et al. [202]	2017	DEAP	L	32		Valence, arousal	EEG	DNN, CNN		CNN: (V)81.406%, 73.36%(A); (DNN) valence: 75.78%, arousal: 73.125%
Wenqian Lin et al. [203]	2017	DEAP	L	32	Yes	Valence, arousal	EEG, ECG, GSR, EMG, EOG, RESP, SKT, face video	CNN	10-fold	Arousal: 87.30%, valence: 85.50%
Santamaria-Granados et al. [148]	2019	AMIGOS	L	40		Valence, arousal	EEG + ECG	DCNN		Arousal: 0.76, valence: 0.75
Subramanian et al. [73]	2018	ASCERTAIN	L	58		Arousal, Valence	EEG, ECG, GSR, facial activity data	SVM, NB	LOO	(GSR,NB) Arousal: 0.68, valence: 0.68
Lee et al. [112]	2018	Movie	L	50	Yes	Negative, neutral emotions	ECG, SKT, EDA	NN, LDA, QDA	LOO	NN: 92.5%
Yang et al. [32]	2018	Video game	L	58		Arousal, valence	ECG, EDA, RESP, EMG, ACC	SVM, RBF SVM, DT, RF	10-fold	Arousal: 0.559, valence: 0.524
Li et al. [204]	2019	DEAP, stroop test	L	32 + 20	Both	Low, medium and high stress	BVP, GSR	LR, eSVR, CNN, ST-SVR	CV	F1-score between 0.943, 0.970 and 0.984
Zhao et al. [205]	2018	ASCERTAIN	L	58	No	Arousal, valence	GSR, EEG, ECG, facial landmarks	Vertex-weighted Multi-modal Multi-task Hypergraph Learning, SVM, NB, hypergraph	10-fold	(VM2HL) Valence: 74.34, arousal: 79.46
Anusha et al. [206]	2018	TSST, Stroop Color Word test, Mental Arithmetic test	L	34	No	Baseline, stress	EDA, ECG, SKT	LDA, QDA, SVM, 3-NN,	LOSO	(EDA+SKT) 97.13%
Sirisha Devi et al. [207]	2019	DEAP	L	50		Valence, arousal	EEG, HR, GSR, RESP	LDA		93.8%
Xia et al. [208]	2018	Stress	L	22		Stress, control	EEG, ECG	SVM-sigmoid, SVM-RF	10-fold	79.54%
Agrafioti et al. [169]	2012	IAPS	L	31	Yes	Neutral, gore, fear, disgust, excitement, erotica, game elicited mental arousal	ECG	LDA	LOO	active/pas AR: 78/52% positive/neg VA: <62%
Han-Wen Guo et al. [209]	2016	Movie clips	L	25	Yes	Positive, negative	ECG	SVM		71.40%
Hernan F. Garcia et al. [210]	2016	DEAP	L	32	Yes	Valence, arousal	EEG, EMG, EOG, GSR, RSP, T, BVP	SVM		Valence: 88.33%, arousal: 90.56%
Liu et al. [211]	2005	Cognitive tasks.	L	15		Anxiety, boredom, engagement, frustration, anger	ECG, EDA, EMG	kNN; RT; BN; SVM	LOO	75%; 84%; 74%; 85%
Wagner et al. [39]	2005	Music (AuBT)	L	1	Yes	Joy, anger, pleasure, sadness	ECG, EDA, EMG, RESP	kNN; LDF; NN	LOO	81%; 80%; 81%
Zhu et al. [212]	2016	Daily-living	F	18	No	Angle in valence arousal space	ACC, phone context	RR	LOSO	
Birjandtalab et al. [144]	2016	Physical, cognitive, emotional stress	L	20		Relaxation, physical, emotional, cognitive stress	ACC, EDA, TEMP, HR, SpO2	GMM		<85%

emotion they are feeling, thus, introducing bias to the recognition results. The SVM method is the most commonly applied algorithm showing good results and low computational complexity. In literature, non-traditional DL techniques use mainly EEG data.

(h) **Dimensionality Reduction:** Data representation highly influences the classification performance. Several works, suffering from the curse of dimensionality apply feature selection and data dimensionality reduction algorithms. These allow to increase the classification performance, however, at the expense of increased time and computational cost.

(i) **Validation Techniques:** To avoid overfitting many works apply k-CV, LOO, LOSO-CV techniques and subsequents. However, CV techniques lead to subject-dependent evaluations, thus, a LOSO-CV should be applied for generalised results.

(j) **Evaluation Metrics:** Accuracy is the most commonly applied metric to evaluate the model's performance. Metrics such as F1-score, Precision and Recall can also be found in the literature.

(k) **Arousal vs. Valence:** Physiological signals are directly the output from the SNS, as well as the arousal dimension, thus, generally classified with higher accuracy than the valence axis.

V. CONCLUSION AND FUTURE WORK

Although relatively young, the field of affective computing has experienced enormous growth and accumulation of knowledge since its inception in 1995, with many papers published in the field (more than 2k according to IEEE Xplore search results). This paper starts by introducing theoretical background key concepts needed to understand the concept of emotion and the connection between the ANS and physiological data. We present benchmarked datasets for emotion recognition using physiological signals, validated elicitation materials and assessment methodologies. Thirdly, we describe the main steps required for the development of a novel ML algorithm for emotion recognition. Lastly, we analyse the current state-of-the-art of emotion recognition, pointing its main achievements, take-home messages, challenges and possible future opportunities. Within the state-of-the-art, several challenges and opportunities have been identified, towards the development of a framework for emotion recognition which must be addressed [4], [8]:

(a) **Experimental Design:** The design of the experimental setup for emotion elicitation can be both induced or obtained genuinely, i.e. in a constrained lab setting or a daily-living unconstrained scenario. Daily-living solutions present additional variables and challenges, such as an increase in the difficulty of emotion awareness, ground-truth data annotation, and a decrease in the signal-to-noise ratio due to uncontrolled subject movements. Thus, daily-living algorithms generally present lower prediction scores. Further variables show impact in the subject response such as its environment, i.e. if the subject is alone or in a group setting (audience), its current mood, personality, gender, background, age and culture. Thus, in order to ensure the study validity, reliability, and generalizability, the experiment should focus on unconstrained scenarios and be performed by a large number of individuals with all of the aforementioned characteristics in order to approximate the algorithm to a reliable real-life solution. Moreover, the subject's mood changes throughout the day, hence, further research should focus on a continuous evaluation of the subject response to different stimuli along the day, correlating the elicitation material to the subject environmental context, personality, mood, and whether the content can influence its emotions, cognition and behaviour. To reduce the volunteers' bias in the ground truth annotation, it might be beneficial to inform the goal of the study to the subject only after the study is performed.

(b) **Elicitation Material:** Short films are low cost, easily scalable, goal-oriented, and able to trigger high-intensity emotions in the subjects, thus, have shown to be a reliable elicitation material. These should focus on the elicitation of a single emotion and should be validated in order to ensure its reliability. To validate the elicitation material further research should focus on the definition of metrics to assess the elicitation material emotional content, i.e. what kind of narrations, images and sounds will arouse the subject attention and interest and how to measure it.

(c) **Emotion Dimensions:** Current studies focus primarily on the recognition of emotions in a binary valence-arousal, stress-no stress scenario. Additionally, the arousal axis has been classified with high accuracy, however, the valence axis, still lacks a reasonable performance. Further research lines could focus on the development of new metrics and dimensions for emotion assessment and classification of complex emotions.

(d) **EMAs:** The subject self-annotated reports should be simple, quick, goal-oriented and include a reward system. The latter has shown to maintain the subject motivated and more likely to deliver quality ground truth data.

(e) **Person-independent vs. Person-Dependent:** Person-independent algorithms are generally outperformed by subject-dependent algorithms since even for the same emotion-eliciting materials, the elicited emotions depend on the subject environment, culture, current mood, personality and perception. Thus, further research must be implemented to obtain more generalised algorithms.

(f) **Data Annotation:** The SL algorithms rely on annotated data. Data annotation is a costly, difficult, time consuming and error-prone task, since awareness and description into works of emotions at all times is

challenging, requiring high-quality ground truth data. Thus, focus should be given to UL algorithms or in facilitating the ground truth annotation in an in-loco all-times annotation setting.

(g) **Classifier:** The feature and modality-dependent nature of previous works resulted in the increase of the model time and computational cost for a real-life reliable solution. The DL approaches and dimensionality data reduction algorithms can be a viable solution to this problem, both as data denoising and as a feature-independent modality to learn unobservable data information. Current works using deep-learning for emotion recognition focus mostly on EEG data, therefore further work can be explored using DL in a multi-modal setting. A further approach, to the best knowledge of the authors, yet to be applied to Emotion Recognition in a multi-modal physiological signal-based context, is Dissimilarity-based Classification, based on the hypothesis that objects that are similar present close representations. For the calculation of the object's similarity, many similarity metrics can be used namely the Euclidean distance, Cosine similarity, and many others.

(h) **Sensor Modalities:** Most papers have reported an increase of recognition rate with the increase of the number of data modalities, namely EEG, ECG, EDA, EMG, RESP and SKT data, however, there is still no clear evidence of which feature combinations of which physiological signals are the most relevant. Moreover, there are limited public datasets for emotion recognition considering all possible modalities in unconstrained daily-living scenarios.

To conclude, in the past 13 years, improvements in the fields of affective science and emotion science, computer science and electronics have endured the growth for affective computing theory and research through a deeper knowledge of emotion theory, the development of accurate ML algorithms, and the creating of ubiquitous, fast and pervasive wearable technology [161], [162]. These new technologies have become a part of our daily life, contributing to continuously improved life quality, and allow the acquisition of high amounts of data that can be used for the development of complex ML models for reliable emotion recognition algorithms. Over these years, many methodologies have been developed in affective computing, culminating in the emergence of new research questions, challenges and opportunities, bringing the recognition and knowledge of emotion one step further.

REFERENCES

[1] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 1997.

[2] H. Gamboa, H. Silva, and A. Fred, "HiMotion: A new research resource for the study of behavior, cognition, and emotion," *Multimedia Tools Appl.*, vol. 73, pp. 345–375, Nov. 2014.

[3] H. Silva, A. Lourenço, and A. Fred, "In-vehicle driver recognition based on hand ECG signals," in *Proc. ACM Int. Conf. Intell. User Interfaces (IUI)*, 2012, pp. 25–28.

[4] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, and X. Yang, "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, p. 2074, 2018.

[5] M. T. Cicero and M. R. Graver, "Cicero on the emotions: Tusculan disputations 3 and 4," Bibliovault OAI Repository, Univ. Chicago Press, Chicago, IL, USA, Tech. Rep., Jan. 2002. [Online]. Available: https://chicago.universitypressscholarship.com/view/10.7208/chicago/9780226305196.001.0001/upso-9780226305776

[6] C. Darwin, *The Expression of the Emotions in Man and Animals* (Cambridge Library Collection—Darwin, Evolution and Genetics), 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2009.

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

[8] P. Schmidt, A. Reiss, R. Duerichen, and K. V. Laerhoven, "Wearable affect and stress recognition: A review," 2018, *arXiv:1811.08854*. [Online]. Available: https://arxiv.org/abs/1811.08854

[9] R. Plutchik, "A psychoevolutionary theory of emotions," *Social Sci. Inf.*, vol. 21, nos. 4–5, pp. 529–553, 1982.

[10] C. E. Izard, "Basic emotions, natural kinds, emotion schemas, and a new paradigm," *Perspectives Psychol. Sci.*, vol. 2, no. 3, pp. 260–280, 2007.

[11] C. E. Izard, "Emotion theory and research: Highlights, unanswered questions, and emerging issues," *Annu. Rev. Psychol.*, vol. 60, no. 1, pp. 1–25, 2009.

[12] A. R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY, USA: G.P. Putnam's, 1994.

[13] P. J. Lang, "The emotion probe: Studies of motivation and attention," *Amer. Psychol.*, vol. 50, pp. 372–385, Jun. 1995.

[14] O. Alemi, W. Li, and P. Pasquier, "Affect-expressive movement generation with factored conditional restricted boltzmann machines," in *Proc. Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Sep. 2015, pp. 442–448.

[15] A. Mehrabian, "Comparison of the PAD and PANAS as models for describing emotions and for differentiating anxiety from depression," *J. Psychopathol. Behav. Assessment*, vol. 19, pp. 331–357, Dec. 1997.

[16] I. Bakker, T. van der Voordt, P. Vink, and J. de Boon, "Pleasure, arousal, dominance: Mehrabian and Russell revisited," *Current Psychol.*, vol. 33, no. 3, pp. 405–421, 2014.

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

[18] C. Carreiras *et al.* (2018). *BioSPPy: Biosignal Processing in Python (2015–)*. [Online]. Available: https://github.com/PIA-Group/BioSPPy

[19] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Med. Biol. Eng. Comput.*, vol. 42, pp. 419–427, Jun. 2004.

[20] 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.

[21] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan./Mar. 2012.

[22] M. Elgendi, "On the analysis of fingertip photoplethysmogram signals," *Current Cardiol. Rev.*, vol. 8, no. 1, pp. 14–25, Feb. 2012.

[23] F. Canento, "Affective mouse electrophysiological signal processing for affective computing," M.S. thesis, Inst. Superior Técnico, Univ. Técnica Lisboa, Lisbon, Portugal, 2011.

[24] E.-H. Jang, B.-J. Park, M. Park, S. Kim, and J. Sohn, "Analysis of physiological signals for recognition of boredom, pain, and surprise emotions," *J. Physiol. Anthropol.*, vol. 34, no. 1, p. 25, 2015.

[25] M. Schmidt, A. Schumann, J. Müller, K.-J. Bär, and G. Rose, "ECG derived respiration: Comparison of time-domain approaches and application to altered breathing patterns of patients with schizophrenia," *Physiol. Meas.*, vol. 38, no. 4, pp. 601–615, 2017.

[26] G. B. Moody, R. G. Mark, M. A. Bump, J. S. Weinstein, A. D. Berman, J. E. Mietus, and A. Goldberger, "Clinical validation of the ECG-derived respiration (EDR) technique," *Comput. Cardiol.*, vol. 13, no. 1, pp. 507–510, 1986.

[27] W. Heide, E. Koenig, P. Trillenberg, D. Kömpf, and D. S. Zee, "Electrooculography: Technical standards and applications," *Electroencephalogr. Clin. Neurophysiol. Suppl.*, vol. 52, pp. 223–240, 1999. [Online]. Available: https://jhu.pure.elsevier.com/en/publications/electrooculography-technical-standards-and-applications-the-inter-4

[28] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual," Univ. Florida, Gainesville, FL, USA, Tech. Rep. A-8, 2008.

[29] E. Douglas-Cowie, R. Cowie, I. Sneddon, C. Cox, O. Lowry, M. Mcrorie, J.-C. Martin, L. Devillers, S. Abrilian, A. Batliner, N. Amir, and K. Karpouzis, "The HUMAINE database: Addressing the collection and annotation of naturalistic and induced emotional data," in *Proc. 2nd Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, 2007, pp. 488–500.
[30] B. J. Li, J. N. Bailenson, A. Pines, W. J. Greenleaf, and L. M. Williams, "A public database of immersive VR videos with corresponding ratings of arousal, valence, and correlations between head movements and self report measures," *Frontiers Psychol.*, vol. 8, p. 2116, Dec. 2017.
[31] W. Yang, M. Rifqi, C. Marsala, and A. Pinna, "Physiological-based emotion detection and recognition in a video game context," in *Proc. Int. Conf. Neural Netw. (IJCNN)*, Rio, Brazil, Jul. 2018, pp. 194–201.
[32] K. Karpouzis, G. N. Yannakakis, N. Shaker, and S. Asteriadis, "The platformer experience dataset," in *Proc. Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Sep. 2015, pp. 712–718.
[33] P. P. A. B. Merkx, K. Truong, and M. Neerincx, "Inducing and measuring emotion through a multiplayer first-person shooter computer game," in *Proc. Comput. Games Workshop*, 2007, pp. 1–7.
[34] S. Huynh, Y. Lee, T. Park, and R. K. Balan, "Jasper: Sensing gamers' emotions using physiological sensors," in *Proc. 3rd Workshop Mobile Gaming (MobiGames)*, 2016, pp. 1–6.
[35] H. P. Martinez, Y. Bengio, and G. N. Yannakakis, "Learning deep physiological models of affect," *IEEE Comput. Intell. Mag.*, vol. 8, no. 2, pp. 20–33, May 2013.
[36] G. N. Yannakakis, H. P. Martínez, and A. Jhala, "Towards affective camera control in games," *User Model. User-Adapt. Interact.*, vol. 20, no. 4, pp. 313–340, 2010.
[37] M. M. Bradley and P. J. Lang, "The international affective digitized sounds (IADS): Affective ratings of sounds and instruction manual," Univ. Florida, Gainesville, FL, USA, Tech. Rep. B-3, 2007.
[38] J. Wagner, J. Kim, and E. André, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," in *Proc. Int. Conf. Multimedia Expo*, Jul. 2005, pp. 940–943.
[39] M. M. Bradley and P. J. Lang, "Affective norms for english words (ANEW): Instruction manual and affective ratings," UF Center Study Emotion Attention, Gainesville, FL, USA, Tech. Rep. C-3, 2017.
[40] T. W. AlHanai and M. M. Ghassemi, "Predicting latent narrative mood using audio and physiologic data," in *Proc. AAAI*, 2017, pp. 1–7.
[41] G. Chanel, J. J. M. Kierkels, M. Soleymani, and T. Pun, "Short-term emotion assessment in a recall paradigm," *Int. J. Hum.-Comput. Stud.*, vol. 67, no. 8, pp. 607–627, Apr. 2009.
[42] S. Dobrišek, R. Gajšek, F. Mihelič, N. Pavešić, and V. Štruc, "Towards efficient multi-modal emotion recognition," *Int. J. Adv. Robot. Syst.*, vol. 10, no. 1, p. 53, 2013.
[43] G. Castellano, L. Kessous, and G. Caridakis, "Emotion recognition through multiple modalities: Face, body gesture, speech," in *Affect and Emotion in Human-Computer Interaction*. Berlin, Germany: Springer, 2008, pp. 92–103.
[44] J. A. Healey, "Affect detection in the real world: Recording and processing physiological signals," in *Proc. 3rd Int. Conf. Affect. Comput. Intell. Interact. Workshops*, Sep. 2009, pp. 1–6.
[45] E. S. Dan-Glauser and K. R. Scherer, "The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance," *Behav. Res. Methods*, vol. 43, no. 2, p. 468, 2011.
[46] A. Sartori, V. Yanulevskaya, A. A. Salah, J. Uijlings, E. Bruni, and N. Sebe, "Affective analysis of professional and amateur abstract paintings using statistical analysis and art theory," *ACM Trans. Interact. Intell. Syst.*, vol. 5, pp. 8:1–8:27, Jun. 2015.
[47] Y. Yang, J. Jia, S. Zhang, B. Wu, Q. Chen, J. Li, C. Xing, and J. Tang, "How do your friends on social media disclose your emotions?" in *Proc. 28th AAAI Conf. Artif. Intell. (AAAI)*, 2014, pp. 306–312.
[48] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in *Proc. ACM Int. Conf. Multimedia (MM)*, 2010, pp. 83–92.
[49] K.-C. Peng, T. Chen, A. Sadovnik, and A. Gallagher, "A mixed bag of emotions: Model, predict, and transfer emotion distributions," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 860–868.
[50] S. Zhao, H. Yao, Y. Gao, G. Ding, and T.-S. Chua, "Predicting personalized image emotion perceptions in social networks," *IEEE Trans. Affect. Comput.*, vol. 9, no. 4, pp. 526–540, Oct./Dec. 2016.
[51] S. Zhao, G. Ding, Q. Huang, T.-S. Chua, B. W. Schuller, and K. Keutzer, "Affective image content analysis: A comprehensive survey," in *Proc. Int. Conf. Artif. Intell. (IJCAI)*, Jul. 2018, pp. 5534–5541.
[52] T. L. Gilman, R. Shaheen, K. M. Nylocks, D. Halachoff, J. Chapman, J. J. Flynn, L. M. Matt, and K. G. Coifman, "A film set for the elicitation of emotion in research: A comprehensive catalog derived from four decades of investigation," *Behav. Res. Methods*, vol. 49, no. 6, pp. 2061–2082, 2017.
[53] Y. Baveye, J. Bettinelli, E. Dellandréa, L. Chen, and C. Chamaret, "A large video database for computational models of induced emotion," in *Proc. Humaine Assoc. Conf. Affect. Comput. Intell. Interact.*, 2013, pp. 13–18.
[54] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers," *Cogn. Emotion*, vol. 24, no. 7, pp. 1153–1172, 2010.
[55] M. Schedi, M. Sjöberg, I. Mironică, B. Ionescu, V. L. Quang, Y. Jiang, and C. Demarty, "VSD2014: A dataset for violent scenes detection in hollywood movies and Web videos," in *Proc. 13th Int. Workshop Content-Based Multimedia Indexing (CBMI)*, Jun. 2015, pp. 1–6.
[56] S. Carvalho, J. Leite, S. Galdo-Álvarez, and O. Gonçalves, "The emotional movie database (EMDB): A self-report and psychophysiological study," *Appl. Psychophysiology Biofeedback*, vol. 37, no. 4, pp. 279–294, 2012.
[57] S. Wang and Q. Ji, "Video affective content analysis: A survey of state-of-the-art methods," *IEEE Trans. Affect. Comput.*, vol. 6, no. 4, pp. 410–430, Oct./Dec. 2015.
[58] M. M. Bradley and P. J. Lang, "Affective norms for english text (ANET): Affective ratings of text and instruction manual," Univ. Florida, Gainesville, FL, USA, Tech. Rep. D-1, 2007.
[59] W. Picard, J. A. Healey, and J. A. Healey, "Wearable and automotive systems for affect recognition from physiology," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2000.
[60] Y. Yang, Q. Wu, M. Qiu, Y. Wang, and X. Chen, "Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network," in *Proc. Int. Conf. Neural Netw. (IJCNN)*, 2018, pp. 1–7.
[61] M. Milczarek, E. Schneider, and E. R. González, "OSH in figures: Stress at work—Facts and figures," EU-OSHA, Bilbao, Spain, Tech. Rep., Jan. 2009. [Online]. Available: https://osha.europa.eu/en/tools-and-publications/publications/reports/TE-81-08-478-EN-C_OSH_in_figures_stress_at_work
[62] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, "The 'trier social stress test'—A tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, nos. 1–2, pp. 76–81, 1993.
[63] F. Scarpina and S. Tagini, "The stroop color and word test," *Frontiers Psychol.*, vol. 8, p. 557, Apr. 2017.
[64] A. Zygmunt and J. Stanczyk, "Methods of evaluation of autonomic nervous system function," *Arch. Med. Sci.*, vol. 6, pp. 11–18, Mar. 2010.
[65] M. Ali, F. Al Machot, A. H. Mosa, M. Jdeed, E. Al Machot, and K. Kyamakya, "A globally generalized emotion recognition system involving different physiological signals," *Sensors*, vol. 18, no. 6, p. 1905, 2018.
[66] 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," Feb. 2017, *arXiv:1702.02510*. [Online]. Available: https://arxiv.org/abs/1702.02510
[67] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behav. Therapy Exp. Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
[68] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, "Studentlife: Assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Proc. Int. Conf. Pervas. Ubiquitous Comput. (UbiComp)*, 2014, pp. 3–14.
[69] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong, "MoodScope: Building a mood sensor from smartphone usage patterns," in *Proc. Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, 2013, pp. 389–402.
[70] A. Muaremi, B. Arnrich, and G. Tröster, "Towards measuring stress with smartphones and wearable devices during workday and sleep," *BioNanoScience*, vol. 3, no. 2, pp. 172–183, 2013.
[71] 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./Jun. 2018.
[72] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The PANAS scales," *J. Pers. Social Psychol.*, vol. 54, no. 6, pp. 1063–1070, 1988.

[73] R. P. Snaith, "The hospital anxiety and depression scale," *Health Quality Life Outcomes*, vol. 1, no. 1, p. 29, 2003.

[74] G. Harnois and P. Gabriel, "Mental health and work: Impact, issues and good practices," World Health Org., Geneva, Switzerland, Tech. Rep., 2000. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.893.1141&rep=rep1&type=pdf

[75] J. P. Pollak, P. Adams, and G. Gay, "PAM: A photographic affect meter for frequent, in situ measurement of affect," in *Proc. Int. Conf. Hum. Factors Comput. Syst. (CHI)*, 2011, pp. 725–734.

[76] G. J. Boyle, "Reliability and validity of Izard's differential emotions scale," *Pers. Individual Differences*, vol. 5, pp. 747–750, Jan. 1984.

[77] E. Diener, D. Wirtz, W. Tov, C. Kim-Prieto, D. Choi, S. Oishi, and R. Biswas-Diener, "New well-being measures: Short scales to assess flourishing and positive and negative feelings," *Social Indicators Res.*, vol. 97, no. 2, pp. 143–156, 2010.

[78] S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," *J. Health Social Behav.*, vol. 24, pp. 385–396, Jan. 1984.

[79] K. B. Koh, J. K. Park, C. H. Kim, and S. Cho, "Development of the stress response inventory and its application in clinical practice," *Psychosomatic Med.*, vol. 63, no. 4, pp. 668–678, 2001.

[80] M. W. Linn, "A global assessment of recent stress (GARS) scale," *Int. J. Psychiatry Med.*, vol. 15, no. 1, pp. 47–59, 1986.

[81] L. R. Derogatis and R. Unger, "Symptom checklist-90-revised," in *The Corsini Encyclopedia of Psychology*. Atlanta, GA, USA: American Cancer Society, 2010, pp. 1–2.

[82] K. Kroenke, R. L. Spitzer, and J. B. Williams, "The PHQ-9: Validity of a brief depression severity measure," *J. Gen. Internal Med.*, vol. 16, no. 9, pp. 606–613, 2001.

[83] C. D. Spielberger, "State-trait anxiety inventory," in *The Corsini Encyclopedia of Psychology*. Atlanta, GA, USA: American Cancer Society, 2010, p. 1.

[84] D. W. Russell, "UCLA loneliness scale (version 3): Reliability, validity, and factor structure," *J. Pers. Assessment*, vol. 66, no. 1, pp. 20–40, 1996.

[85] K. Shear, T. A. Brown, B. David, R. Money, D. E. Sholomskas, S. Woods, J. Gorman, and L. A. Papp, "Multicenter collaborative panic disorder severity scale," *Amer. J. Psychiatry*, vol. 154, pp. 1571–1575, Nov. 1997.

[86] D. J. Buysse, C. F. Reynolds, T. H. Monk, S. R. Berman, and D. J. Kupfer, "The pittsburgh sleep quality index: A new instrument for psychiatric practice and research," *Psychiatry Res.*, vol. 28, pp. 193–213, May 1989.

[87] M. Perugini and L. Blas, *Big Five Marker Scales (BFMS) and the Italian AB5C Taxonomy: Analyses From an Etic–Emic Perspective*. Göttingen, Germany: Hogrefe & Hube, 2002, pp. 281–304.

[88] P. Costa and R. R. McCrae, "The revised NEO personality inventory," in *The SAGE Handbook of Personality Theory and Assessment: Personality*, vol. 2. SAGE Publications, Jan. 2008, pp. 179–198. [Online]. Available: https://jhu.pure.elsevier.com/en/publications/the-revised-neo-personality-inventory-neo-pi-r

[89] O. P. John, L. Naumann, and C. J. Soto, "Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues," in *Handbook of Personality: Theory and Research*. New York, NY, USA: Guilford Press, Jan. 2008, pp. 114–158.

[90] E. Vyzas and R. W. Picard, "Offline and online recognition of emotion expression from physiological data," MIT Media, Cambridge, MA, USA, Tech. Rep. 483, 1999, pp. 135–142. [Online]. Available: https://www.media.mit.edu/publications/offline-and-online-recognition-of-emotion-expression-from-physiological-data-3/

[91] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, "DECAF: MEG-based multimodal database for decoding affective physiological responses," *IEEE Trans. Affect. Comput.*, vol. 6, no. 3, pp. 209–222, Jul. 2015.

[92] 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.

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

[94] 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. ACM Int. Conf. Multimodal Interact.*, 2018, pp. 400–408.

[95] C. Wang and P. S. C. Garcia, "The play is a hit-but how can you tell? Measuring audience bio-responses towards a performance," in *Proc. ACM Int. Conf. Creativity Cogn.*, Jun. 2017, pp. 336–347.

[96] C. Wang, J. Wong, T. Röggla, J. Jansen, and P. Cesar, "Quantifying audience experience in the wild: Heuristics for developing and deploying a biosensor infrastructure in theaters," in *Proc. Int. Conf. Qual. Multimedia Exper. (QoMEX)*, Jun. 2016, pp. 1–6.

[97] M. Chavan, R. Agarwala, and M. Uplane, "Suppression of baseline wander and power line interference in ECG using digital IIR filter," *Int. J. Circuits, Syst. Signal Process.*, vol. 2, no. 2, pp. 356–365, 2008.

[98] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.

[99] M. Murugappan, S. Murugappan, and B. S. Zheng, "Frequency band analysis of electrocardiogram (ECG) signals for human emotional state classification using discrete wavelet transform (DWT)," *J. Phys. Therapy Sci.*, vol. 25, no. 7, pp. 753–759, Jul. 2013.

[100] F. Canento, A. Lourenço, H. Silva, and A. L. N. Fred, "Review and comparison of real time electrocardiogram segmentation algorithms for biometric applications," in *Proc. Proc. 6th Int. Conf. Health Inform. (HEALTHINF)*, 2012, pp. 1–9.

[101] H. Gamboa, "Multi-modal behavioral biometrics based on HCI and electrophysiology," Ph.D. dissertation, Inst. Superior Técnico, Univ. Técnica Lisboa, Lisbon, Portugal, 2008.

[102] P. Hamilton, "Open source ECG analysis," in *Proc. Comput. Cardiol.*, Sep. 2002, pp. 101–104.

[103] I. I. Christov, "Real time electrocardiogram QRS detection using combined adaptive threshold," *Biomed. Eng. Online*, vol. 3, no. 1, p. 28, 2004.

[104] A. Lourenço, H. Silva, P. Leite, R. Lourenço, and A. L. N. Fred, "Real time electrocardiogram segmentation for finger based ECG biometrics," in *Proc. Int. Conf Bio-Inspired Syst. Signal Process.*, 2012, pp. 49–54.

[105] W. Zong, T. Heldt, G. B. Moody, and R. G. Mark, "An open-source algorithm to detect onset of arterial blood pressure pulses," in *Proc. Comput. Cardiol.*, Sep. 2003, pp. 259–262.

[106] J. Choi, B. Ahmed, and R. Gutierrez-Osuna, "Development and evaluation of an ambulatory stress monitor based on wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 2, pp. 279–286, Mar. 2012.

[107] M. S. N. M. dos Santos, "Biometrical and psychophysiological assessment through biosensors," M.S. thesis, Instituto Superior Técnico, Univ. Técnica Lisboa, Portugal, 2012.

[108] S. Liu, R. X. Gao, D. John, J. Staudenmayer, and P. Freedson, "Tissue artifact removal from respiratory signals based on empirical mode decomposition," *Ann. Biomed. Eng.*, vol. 41, no. 5, pp. 1003–1015, 2013.

[109] P. S. Wardana, "Processing of respiration signals using FIR filter for analyze the condition of lung," in *Proc. Int. Electron. Symp. Eng. Technol. Appl. (IES-ETA)*, Sep. 2017, pp. 229–233.

[110] J. Lee and S. K. Yoo, "Design of user-customized negative emotion classifier based on feature selection using physiological signal sensors," *Sensors*, vol. 18, no. 12, p. 4253, 2018.

[111] A. Guruvareddy and S. Narava, "Artifact removal from EEG signals," *Int. J. Comput. Appl.*, vol. 77, no. 13, pp. 1–3, Sep. 2013.

[112] A. Cruz, D. Garcia, G. Pires, and U. Nunes, "Facial expression recognition based on EOG toward emotion detection for human-robot interaction," in *Proc. Int. Joint Conf. Biomed. Eng. Syst. Technol. (BIOSTEC)*, 2015, pp. 31–37.

[113] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, 2010.

[114] C. L. Lisetti and F. Nasoz, "Using noninvasive wearable computers to recognize human emotions from physiological signals," *EURASIP J. Adv. Signal Process.*, vol. 2004, no. 11, Dec. 2004, Art. no. 929414.

[115] M. P. Tarvainen, J.-P. Niskanen, J. A. Lipponen, P. O. Ranta-Aho, and P. A. Karjalainen, "Kubios HRV—Heart rate variability analysis software," *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 210–220, 2014.

[116] C. Li, N. Ye, H. Huang, R. Wang, and R. Malekian, "Emotion recognition of human physiological signals based on recursive quantitative analysis," in *Proc. Int. Conf. Adv. Comput. Intell. (ICACI)*, 2018, pp. 217–223.

[117] J. Rubin, R. Abreu, S. Ahern, H. Eldardiry, and D. G. Bobrow, "Time, frequency & complexity analysis for recognizing panic states from physiologic time-series," in *Proc. Int. Conf. Pervasive Comput. Technol. Healthcare (PervasiveHealth)*, 2016, pp. 81–88.

[118] Task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability. Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 17, no. 3, pp. 354–381, 1996.

[119] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm, "Looking at pictures: Affective, facial, visceral, and behavioral reactions," *Psychophysiology*, vol. 30, no. 3, pp. 261–273, 1993.

[120] K. Plarre, A. Raij, S. M. Hossain, A. A. Ali, M. Nakajima, and M. Al'absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott, D. Siewiorek, A. Smailagic, and L. E. Wittmers, "Continuous inference of psychological stress from sensory measurements collected in the natural environment," in *Proc. Int. Conf. Inf. Process. Sensor Netw.*, 2011, pp. 97–108.
[121] P. Boonthong, P. Kulkasem, S. Rasmequan, A. Rodtook, and K. Chinnasarn, "Fisher feature selection for emotion recognition," in *Proc. Int. Comput. Sci. Eng. Conf. (ICSEC)*, Nov. 2015, pp. 1–6.
[122] Y. Cui, S. Luo, Q. Tian, S. Zhang, Y. Peng, L. Jiang, and J. S. Jin, "Mutual information-based emotion recognition," in *The Era of Interactive Media*. New York, NY, USA: Springer, 2013, pp. 471–479.
[123] J. Zhang, M. Chen, S. Zhao, S. Hu, Z. Shi, and Y. Cao, "ReliefF-based EEG sensor selection methods for emotion recognition," *Sensors*, vol. 16, no. 10, p. 1558, 2016.
[124] J. Zhang, M. Chen, S. Hu, Y. Cao, and R. Kozma, "PNN for EEG-based emotion recognition," in *Proc. Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2016, pp. 2319–2323.
[125] C. Torres-Valencia, M. Álvarez-López, and Á. Orozco-Gutiérrez, "SVM-based feature selection methods for emotion recognition from multimodal data," *J. Multimodal User Interfaces*, vol. 11, no. 1, pp. 9–23, 2017.
[126] C.-H. Park and K.-B. Sim, "The novel feature selection method based on emotion recognition system," in *Computational Intelligence and Bioinformatics*, D. Huang, K. Li, and G. W. Irwin, Eds. Berlin, Germany: Springer, 2006, pp. 731–740.
[127] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," in *Data Classification: Algorithms and Applications*. Boca Raton, FL, USA: CRC Press, Jan. 2014, pp. 37–64.
[128] C. Quan, D. Wan, B. Zhang, and F. Ren, "Reduce the dimensions of emotional features by principal component analysis for speech emotion recognition," in *Proc. Int. Symp. Syst. Integr.*, 2013, pp. 222–226.
[129] C. Sorzano, J. Vargas, and A. P. Montano, "A survey of dimensionality reduction techniques," Mar. 2014, *arXiv:1403.2877*. [Online]. Available: https://arxiv.org/abs/1403.2877
[130] W. Wei, Q. Jia, Y. Feng, and G. Chen, "Emotion recognition based on weighted fusion strategy of multichannel physiological signals," *Comput. Intell. Neurosci.*, vol. 2018, Jul. 2018, Art. no. 5296523.
[131] J. Chen, B. Hu, L. Xu, P. Moore, and Y. Su, "Feature-level fusion of multimodal physiological signals for emotion recognition," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2015, pp. 395–399.
[132] X. Zhang, C. Xu, W. Xue, J. Hu, Y. He, and M. Gao, "Emotion recognition based on multichannel physiological signals with comprehensive nonlinear processing," *Sensors*, vol. 18, no. 11, p. 3886, 2018.
[133] J. Xie, X. Xu, and L. Shu, "WT feature based emotion recognition from multi-channel physiological signals with decision fusion," in *Proc. Asian Conf. Affect. Comput. Intell. Interact. (ACII Asia)*, May 2018, pp. 1–6.
[134] P. Gong, H. T. Ma, and Y. Wang, "Emotion recognition based on the multiple physiological signals," in *Proc. Int. Conf. Real-Time Comput. Robot. (RCAR)*, Jun. 2016, pp. 140–143.
[135] W. Liu, W.-L. Zheng, and B.-L. Lu, "Multimodal emotion recognition using multimodal deep learning," 2016, *arXiv:1602.08225*. [Online]. Available: https://arxiv.org/abs/1602.08225
[136] H. Tang, W. Liu, W.-L. Zheng, and B.-L. Lu, "Multimodal emotion recognition using deep neural networks," in *Neural Information Processing*. Cham, Switzerland: Springer, 2017, pp. 811–819.
[137] B. Myroniv, C. Wu, Y. Ren, A. Christian, E. Bajo, and Y. C. Tseng, "Analyzing user emotions via physiology signals," *Data Sci. Pattern Recognit.*, vol. 1, no. 2, pp. 11–25, 2017.
[138] 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, 2011.
[139] C. He, Y.-J. Yao, and X.-S. Ye, "An emotion recognition system based on physiological signals obtained by wearable sensors," in *Wearable Sensors and Robots*. 2017, pp. 15–25. [Online]. Available: https://link.springer.com/chapter/10.1007%2F978-981-10-2404-7_2#citeas
[140] S. Kotsiantis, "Supervised machine learning: A review of classification techniques," *Informatica*, vol. 31, no. 3, pp. 249–268, Oct. 2007.
[141] P. Adams, M. Rabbi, T. Rahman, M. Matthews, A. Voida, G. Gay, T. Choudhury, and S. Voida, "Towards personal stress informatics: Comparing minimally invasive techniques for measuring daily stress in the wild," in *Proc. Int. Conf. Pervasive Comput. Technol. Healthcare (PervasiveHealth)*, 2014, pp. 72–79.
[142] J. Birjandtalab, D. Cogan, M. B. Pouyan, and M. Nourani, "A non-EEG biosignals dataset for assessment and visualization of neurological status," in *Proc. Int.Workshop Signal Process. Syst. (SiPS)*, pp. 110–114, 2016.
[143] M. Khanam, T. Mahboob, W. Imtiaz, H. A. Ghafoor, and R. Sehar, "A survey on unsupervised machine learning algorithms for automation, classification and maintenance," *Int. J. Comput. Appl.*, vol. 119, pp. 34–39, Jan. 2015.
[144] X. Jia, K. Li, X. Li, and A. Zhang, "A novel semi-supervised deep learning framework for affective state recognition on eeg signals," in *Proc. IEEE BIBE*, Nov. 2014, pp. 30–37.
[145] X. Zhu, "Semi-supervised learning literature survey," Dept. Comput. Sci., Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep., 2005. [Online]. Available: http://pages.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf
[146] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-González, E. Abdulhay, and N. Arunkumar, "Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS)," *IEEE Access*, vol. 7, pp. 57–67, 2019.
[147] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, to be published.
[148] S. Salari, A. Ansarian, and H. Atrianfar, "Robust emotion classification using neural network models," in *Proc. Iranian Joint Congr. Fuzzy Intell. Syst. (CFIS)*, Feb. 2018, pp. 190–194.
[149] R. Qiao, C. Qing, T. Zhang, X. Xing, and X. Xu, "A novel deep-learning based framework for multi-subject emotion recognition," in *Proc. Int. Conf. Inf., Comput. Social Syst. (ICCSS)*, Jul. 2017, pp. 181–185.
[150] H. Chao, H. Zhi, L. Dong, and Y. Liu, "Recognition of emotions using multichannel EEG data and DBN-GC-based ensemble deep learning framework," in *Computational Intelligence and Neuroscience*. 2018, pp. 1–11. [Online]. Available: https://www.hindawi.com/journals/cin/2018/9750904/cta/
[151] P. Kawde and G. K. Verma, "Deep belief network based affect recognition from physiological signals," in *Proc. Int. Conf. Elect., Comput. Electron. (UPCON)*, Oct. 2017, pp. 587–592.
[152] Z. Yin, Y. Wang, W. Zhang, L. Liu, J. Zhang, F. Han, and W. Jin, "Physiological feature based emotion recognition via an ensemble deep autoencoder with parsimonious structure," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 6940–6945, 2017.
[153] B. Yang, X. Han, and J. Tang, "Three class emotions recognition based on deep learning using staked autoencoder," in *Proc. Int. Cong. Image Signal Process., Biomed. Eng. Inform. (CISP-BMEI)*, Oct. 2017, pp. 1–5.
[154] W.-L. Zheng, J.-Y. Zhu, Y. Peng, and B.-L. Lu, "EEG-based emotion classification using deep belief networks," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2014, pp. 1–6.
[155] J. Huang, X. Xu, and T. Zhang, "Emotion classification using deep neural networks and emotional patches," in *Proc. Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2017, pp. 958–962.
[156] J. Liu, Y. Su, and Y. Liu, "Multi-modal emotion recognition with temporal-band attention based on LSTM-RNN," in *Proc. Adv. Multimedia Inf. Process.*, B. Zeng, Q. Huang, A. El Saddik, H. Li, S. Jiang, and X. Fan, Eds. Harbin, China: Springer, 2018, pp. 194–204.
[157] X. Li, D. Song, P. Zhang, G. Yu, Y. Hou, and B. Hu, "Emotion recognition from multi-channel eeg data through convolutional recurrent neural network," in *Proc. Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2016, pp. 352–359.
[158] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 10, pp. 355–358, 2017.
[159] S. Pouyanfar, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Comput. Surv.*, vol. 51, no. 5, p. 92, 2018.
[160] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 96–108, Nov. 2017.
[161] X. Zhu, Y. Wang, C. Wang, H. Yang, X. Wang, and M. Yang, "Developing a driving fatigue detection system using physiological sensors," in *Proc. 29th Austral. Conf. Comput.-Hum. Interact. OZCHI*, 2017, pp. 566–570.
[162] M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti, "Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisitionn," in *Proc. 4th Int. Conf. Wireless Mobile Commun. Healthcare-Transforming Healthcare Through Innov. Mobile Wireless Technol. (MOBIHEALTH)*, Nov. 2014, pp. 39–42.
[163] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis," *IEEE Trans. Affect. Comput.*, vol. 1, no. 2, pp. 81–97, Jul. 2010.

[164] A. Samara, M. L. R. Menezes, and L. Galway, "Feature extraction for emotion recognition and modelling using neurophysiological data," in *Proc. Int. Conf. Ubiquitous Comput. Commun. Int. Symp. Cyberspace Secur. (IUCC-CSS)*, Dec. 2016, pp. 138–144.

[165] G. K. Verma and U. S. Tiwary, "Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals," *NeuroImage*, vol. 102, pp. 162–172, Nov. 2014.

[166] 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, 2017.

[167] 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./Mar. 2012.

[168] 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.

[169] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2067–2083, Feb. 2008.

[170] C. Zong and M. Chetouani, "Hilbert-Huang transform based physiological signals analysis for emotion recognition," in *Proc. Int. Symp. Signal Process. Inf. Technol.*, Dec. 2009, pp. 334–339.

[171] 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, Sep. 2012.

[172] W. M. Wong, A. W. Tan, C. K. Loo, and W. S. Liew, "PSO optimization of synergetic neural classifier for multichannel emotion recognition," in *Proc. 2nd World Congr. Nature Biol. Inspired Comput. (NaBIC)*, Dec. 2010, pp. 316–321.

[173] L. Mirmohamadsadeghi, A. Yazdani, and J. Vesin, "Using cardio-respiratory signals to recognize emotions elicited by watching music video clips," in *Proc. Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2016, pp. 1–5.

[174] C. K. Wu, P. C. Chung, and C. J. Wang, "Representative segment-based emotion analysis and classification with automatic respiration signal segmentation," *IEEE Trans. Affect. Comput.*, vol. 3, no. 4, pp. 482–495, Fourth 2012.

[175] Z. Guendil, Z. Lachiri, C. Maaoui, and A. Pruski, "Emotion recognition from physiological signals using fusion of wavelet based features," in *Proc. Int. Conf. Modeling, Identificat. Control (ICMIC)*, Dec. 2015, pp. 1–6.

[176] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "EEG-based emotion recognition in music listening," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798–1806, Jul. 2010.

[177] B. Cheng and G. Liu, "Emotion recognition from surface EMG signal using wavelet transform and neural network," in *Proc. Int. Conf. Bioinf. Biomed. Eng.*, May 2008, pp. 1363–1366.

[178] S. Basu, N. Jana, A. Bag, J. Mukherjee, S. Kumar, and R. Guha, "Emotion recognition based on physiological signals using valence-arousal model," in *Proc. Int. Conf. Image Inf. Process. (ICIIP)*, Dec. 2015, pp. 50–55.

[179] J. Liu, H. Meng, A. Nandi, and M. Li, "Emotion detection from EEG recordings," in *Proc. Int. Conf. Natural Comput., Fuzzy Syst. Knowl. Discovery (ICNC-FSKD)*, Aug. 2016, pp. 1722–1727.

[180] M. Monajati, H. Abbasi, F. Shabaninia, and S. Shamekhi, "Emotions states recognition based on physiological parameters by employing of fuzzy-adaptive resonance theory," *Int. J. Intell. Sci.*, vol. 2, pp. 166–175, Jan. 2012.

[181] Z. Lan, O. Sourina, L. Wang, and Y. Liu, "Real-time EEG-based emotion monitoring using stable features," *Vis. Comput.*, vol. 32, no. 3, pp. 347–358, 2016.

[182] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 417–429, Jul./Sep. 2019.

[183] 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, Oct. 2003.

[184] A. Haag, S. Goronzy, P. Schaich, and J. Williams, "Emotion recognition using bio-sensors: First steps towards an automatic system," in *Affective Dialogue Systems*, E. André, L. Dybkjær, W. Minker, and P. Heisterkamp, Eds. Berlin, Germany: Springer, 2004, pp. 36–48.

[185] E. Leon, G. Clarke, V. Callaghan, and F. Sepulveda, "A user-independent real-time emotion recognition system for software agents in domestic environments," *Eng. Appl. Artif. Intell.*, vol. 20, no. 3, pp. 337–345, 2007.

[186] J. Zhai and A. Barreto, "Stress detection in computer users through non-invasive monitoring of physiological signals," *Biomed. Sci. Instrum.*, vol. 42, pp. 495–500, Feb. 2006.

[187] D. Kim, Y. Seo, J. Cho, and C.-H. Cho, "Detection of subjects with higher self-reporting stress scores using heart rate variability patterns during the day," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 682–685.

[188] C. D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis, "Toward emotion recognition in car-racing drivers: A biosignal processing approach," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 38, no. 3, pp. 502–512, May 2008.

[189] R. A. Calvo, I. Brown, and S. Scheding, "Effect of experimental factors on the recognition of affective mental states through physiological measures," in *AI 2009: Advances in Artificial Intelligence*, A. Nicholson and X. Li, Eds. Berlin, Germany: Springer, 2009, pp. 62–70.

[190] Z. Khalili and M. H. Moradi, "Emotion recognition system using brain and peripheral signals: Using correlation dimension to improve the results of EEG," in *Proc. Int. Conf. Neural Netw.*, Jun. 2009, pp. 1571–1575.

[191] J. Healey, L. Nachman, S. Subramanian, J. Shahabdeen, and M. Morris, "Out of the lab and into the fray: Towards modeling emotion in everyday life," in *Pervasive Computing*, P. Floréen, A. Krüger, and M. Spasojevic, Eds. Berlin, Germany: Springer, 2010, pp. 156–173.

[192] J. Hernandez, R. R. Morris, and R. W. Picard, "Call center stress recognition with person-specific models," in *Affective Computing and Intelligent Interaction*, S. D'Mello, A. Graesser, B. Schuller, and J. Martin, Eds. Berlin, Germany: Springer, 2011, pp. 125–134.

[193] G. Valenza, L. Citi, A. Lanatá, E. P. Scilingo, and R. Barbieri, "Revealing real-time emotional responses: A personalized assessment based on heartbeat dynamics," *Sci. Rep.*, vol. 4, May 2014, Art. no. 4998.

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

[195] K. Hovsepian, M. Al'absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar, "cStress: Towards a gold standard for continuous stress assessment in the mobile environment," in *Proc. ACM Int. Conf. Ubiquitous Comput. (UbiComp)*, 2015, pp. 493–504.

[196] N. Jaques, S. Taylor, E. Nosakhare, A. Sano, and R. W. Picard, "Multi-task learning for predicting health, stress, and happiness," in *Proc. NIPS Workshop ML Health*, Barcelona, Spain, Dec. 2016, pp. 1–5.

[197] A. Zenonos, A. Khan, G. Kalogridis, S. Vatsikas, T. Lewis, and M. Sooriyabandara, "HealthyOffice: Mood recognition at work using smartphones and wearable sensors," in *Proc. Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2016, pp. 1–6.

[198] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *J. Biomed. Inform.*, vol. 73, pp. 159–170, Sep. 2017.

[199] O. M. Mozos, V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, R. Dobrescu, and J. M. Ferrandez, "Stress detection using wearable physiological and sociometric sensors," *Int. J. Neural Syst.*, vol. 27, no. 2, 2017, Art. no. 1650041.

[200] S. Tripathi, S. Acharya, R. D. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2017, pp. 4746–4752.

[201] W. Lin, C. Li, and S. Sun, "Deep convolutional neural network for emotion recognition using EEG and peripheral physiological signal," in *Proc. 9th Int. Conf. Image Graph. (ICIG)*, Y. Zhao, X. Kong, and D. Taubman, Eds. Shanghai, China: Springer, 2017, pp. 385–394. [Online]. Available: https://link.springer.com/book/10.1007/978-3-319-71589-6

[202] M. Li, L. Xie, and Z. Wang, "A transductive model-based stress recognition method using peripheral physiological signals," *Sensors*, vol. 19, no. 2, p. 429, 2019.

[203] S. Zhao, G. Ding, J. Han, and Y. Gao, "Personality-aware personalized emotion recognition from physiological signals," in *Proc. 27th Int. Conf. Artif. Intell.*, Jul. 2018, pp. 1660–1667.

[204] A. S. Anusha, J. Jose, S. P. Preejith, J. Jayaraj, and S. Mohanasankar, "Physiological signal based work stress detection using unobtrusive sensors," *Biomed. Phys. Eng. Express*, vol. 4, no. 6, 2018, Art. no. 065001.

[205] S. Devi and S. Nandyala, "Electroencephalography and physiological signals for emotion analysis," *Int. J. Innov. Technol. Exploring Eng.*, vol. 8, pp. 293–297, Jan. 2019.

[206] L. Xia, A. S. Malik, and A. R. Subhani, "A physiological signal-based method for early mental-stress detection," *Biomed. Signal Process. Control*, vol. 46, pp. 18–32, Sep. 2018.

[207] H. Guo, Y. Huang, C. Lin, J. Chien, K. Haraikawa, and J. Shieh, "Heart rate variability signal features for emotion recognition by using principal component analysis and support vectors machine," in *Proc. Int. Conf. Bioinf. Bioeng. (BIBE)*, 2016, pp. 274–277.
[208] H. F. García, M. A. Álvarez, and A. A. Orozco, "Gaussian process dynamical models for multimodal affect recognition," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 850–853.
[209] C. Liu, P. Rani, and N. Sarkar, "An empirical study of machine learning techniques for affect recognition in human-robot interaction," in *Proc. Int. Conf. Intell. Robots Syst.*, Aug. 2005, pp. 2662–2667.
[210] Z. Zhu, H. F. Satizábal, U. Blanke, A. Perez-Uribe, and G. Tröster, "Naturalistic recognition of activities and mood using wearable electronics," *IEEE Trans. Affect. Comput.*, vol. 7, no. 3, pp. 272–285, Oct. 2016.

PATRÍCIA J. BOTA received the M.Sc. degree in biomedical engineering from the Faculty of Sciences and Technology, NOVA University of Lisbon. She is currently pursuing the Ph.D. degree with the Instituto Superior Técnico (IST-UL). She was a Scientist with Fraunhofer AICOS focusing on the development of human activity recognition algorithm based on the smartphone's built-in sensors. She is currently a Research Member with the Pattern and Image Analysis (PIA-Lx) Research Group, Instituto Telecomunicações (IT). At IT, she involved in the study of the media's impact on human emotions, cognition and behaviour in unconstrained scenarios, and the development of machine learning algorithms for the recognition of affect and emotion through multi-source physiological data collected in long-term. Her main research interests include machine learning, affective computing, artificial intelligence, and signal processing.

CHEN WANG received the B.Sc. degree in telecommunication engineering from the Nanjing University of Posts and Telecommunications, China, the M.Sc. degree in electrical engineering from the Delft University of Technology, The Netherlands, and the Ph.D. degree from the National Research Center of Mathematics and Computer Science, The Netherlands. During the Ph.D. degree, she shifted to physiological computing on user experience evaluation. She designed and developed the different versions of physiological sensors, which are suitable for simultaneously measuring group user experience in field studies, e.g., theaters, museum visitors, and learning environment. She is currently the Vice Director of the Future Media & Convergence Institute (FMCI), Xinhuanet China.

ANA L. N. FRED received the M.S. and Ph.D. degrees in electrical and computer engineering from the Instituto Superior Técnico (IST), Technical University of Lisbon, Portugal, in 1989 and 1994, respectively. She has been a Faculty Member of IST, since 1986, where she has also been a Professor with the Department of Electrical and Computer Engineering and, more recently, with the Department of Biomedical Engineering. She is currently a Researcher with the Pattern and Image Analysis Group, Instituto de Telecomunicações. She has published over 200 articles in international refereed conferences, peer-reviewed journals, and book chapters. Her main research interests include pattern recognition, both structural and statistical approaches, with application to data mining, learning systems, behavioral biometrics, and biomedical applications. She has done pioneering work on clustering, namely on cluster ensemble approaches.

HUGO PLÁCIDO DA SILVA was born in Vila Franca de Xira, Portugal, in 1979. He received the Ph.D. degree in electrical and computers engineering from the University of Lisbon, Lisbon, Portugal, in 2015. Since 2004, he has been a Researcher with the Instituto de Telecomunicações (IT). He has also been a Professor with the Polytechnic Institute of Setúbal, since 2016, and a Co-Founder and the Chief Innovation Officer at PLUX-Wireless aBiosignals, S.A., since 2007. His main interests include biosignal research, system engineering, signal processing, and pattern recognition. His work has been distinguished with several academic and technical awards.