ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/329548658

Emotion in a Century: A Review of Emotion Recognition

Conference Paper · December 2018

DOI: 10.1145/3291280.3291788

CITATIONS
27

READS
4,674

4 authors, including:

Thanyathorn Thanapattheerakul
King Mongkut's University of Technology Thonburi
8 PUBLICATIONS **32** CITATIONS
SEE PROFILE

Jacqueline Amoranto
University of Toronto
1 PUBLICATION **27** CITATIONS
SEE PROFILE

All content following this page was uploaded by Jacqueline Amoranto on 01 October 2019.

The user has requested enhancement of the downloaded file.

Emotion in a Century: A Review of Emotion Recognition

Thanyathorn Thanapattheerakul
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
thanyathorn.tha@mail.kmutt.ac.th

Katherine Mao
University of Toronto
Toronto, Canada
katherine.mao@mail.utoronto.ca

Jacqueline Amoranto
University of Toronto
Toronto, Canada
j.amoranto@mail.utoronto.ca

Jonathan H. Chan
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
jonathan@sit.kmutt.ac.th

ABSTRACT

Emotion plays an important role in our daily lives. Ever since the 19th century, experimental psychologists have attempted to understand and explain human emotion. Despite an extensive amount of research conducted by psychologists, anthropologists, and sociologists over the past 150 years, researchers still cannot agree on the definition of emotion itself and have continued to try and devise ways to measure emotional states. In this paper, we provide an overview of the most prominent theories in emotional psychology (dating from the late 19th century to the present day), as well as a summary of a number of studies which attempt to measure certain aspects of emotion. This paper is organized chronologically; first with an analysis of various uni-modal studies, followed by a review of multi-modal research. Our findings suggest that there is insufficient evidence to neither prove nor disprove the existence of coherent emotional expression, both within subjects and between subjects. Furthermore, the results seem to be heavily influenced by both experimental conditions as well as by the theoretical assumptions that underpin them.

CCS CONCEPTS

• **Applied computing** → **Psychology**; • **Computing methodologies** → *Machine learning*; • **Human-centered computing** → Human computer interaction (HCI);

KEYWORDS

Emotion, Theory of Emotion, Measure of Emotion, Emotion Recognition

ACM Reference Format:
Thanyathorn Thanapattheerakul, Katherine Mao, Jacqueline Amoranto, and Jonathan H. Chan. 2018. Emotion in a Century: A Review of Emotion Recognition. In *The 10th International Conference on Advances in Information Technology (IAIT2018), December 10–13, 2018, Bangkok, Thailand*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3291280.3291788

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
IAIT 2018, December 10–13, 2018, Bangkok, Thailand
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-6568-0/18/12…$15.00
https://doi.org/10.1145/3291280.3291788

1 INTRODUCTION

Emotion plays an important role in our daily lives. Many researchers attempt to measure different aspects of emotion covering broad areas, noticeably in Human-Computer Interaction (HCI). In particular, measurements of emotion have been used to improve designs, human interaction, support decision making, and so on [15], e.g. using speech to recognize learners in online learning to study how they respond to a course [8]; brainwave was used to detect players' emotion while they are playing game to learn their affective states [11]. However, the question of whether emotional states can be empirically measured remains a hotly debated topic [7]. The concept of emotion itself presents its own problem. Ever since early theories attempted to give a definitive answer to the question "what is/are emotion(s)?", the debate surrounding the exact meaning of emotion has prevailed within the psychology community [4]. Unable to measure emotions directly, researchers studying emotion have elected to measure everything surrounding it [4]. This paper is a preliminary attempt to review (recent) works which attempt to define and measure different aspects. We focused on studies which attempted to map components of emotional experience (empirically measured) to discrete emotional states. In addition to summarizing the results of these studies, we briefly analyze/evaluate their methodologies and their implications.

To investigate the different methods of emotion measurement or recognition, we begin by providing an overview of the history of theories of emotion that spans over a century (1800s - present). We believe this part is important to provide adequate background knowledge. Then, we present a broad overview of methodologies that covers single measurements, or unimodal studies, as well as multi-measurements, or multi-modal researches. We end the paper with discussions and conclusions.

2 HISTORY OF THEORIES OF EMOTION

In this section, we present a broad overview of theories of emotion, beginning in the late 1800's up to the present day. Figure 1 shows timeline of theories of emotion.

2.1 Pre-1950's

Before the 1950's, there were roughly three leading theories of emotion. Perhaps the earliest was that of Charles Darwin. In Darwin's third major work of evolutionary theory, *The Expression of the Emotions in Man and Animals,* he contends emotional expression is a

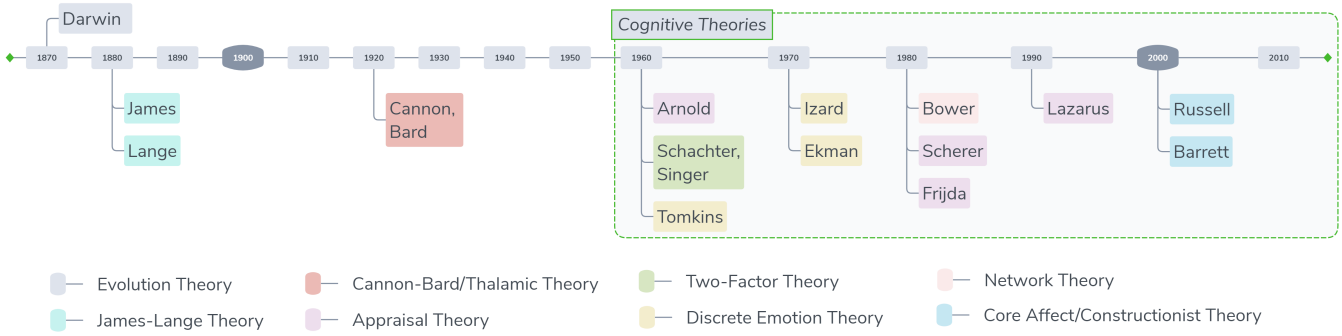

Figure 1: Timeline of Theories of Emotion

derived human character [12]. He suggests that emotional expressions are inherited, habitual and reflect the state of mind of the person affected. In addition, he argues that emotional expression provides a social function, serving as an important communicative tool. While his ideas were largely based on observation and anecdotal evidence, they helped shape contemporary emotional expression research. About a decade after Darwin published his theory, two theorists, William James and Carl Lange, independently developed a new theory, known today as the James-Lange theory. They hypothesized that emotions are the experience of physiological responses to stimuli [34]. These are somatic or motor responses and they precede the feeling aspect of emotion. In other words, the conscious experience of a physical sensation, such as increased heart rate or crying, or a combination of them, equates to the experience of an emotion. Although the James-Lange theory remained prominent during their time, it has since been heavily criticized by a number of theorists. Most notably, Walter Cannon and Philip Bard criticized this theory due to its lack of supporting empirical evidence. Cannon and Bard proposed an alternative theory, known as the Cannon-Bard theory or Thalamic theory, which states that physiological responses to emotional stimuli and the experience of emotion occur simultaneously and independently of one another [6]. They believed that the thalamic region of the brain, where sensory information is processed, was responsible for the experience of emotion. Critics of this theory argue that the feeling portion of emotion cannot be separated completely from the physiological components [38]. It was from these critiques that new theories continued to develop throughout the century.

2.2 Post-1950's

With the rise of cognitive psychology in the 1950's, cognitive theories of emotion became increasingly popular, supplanting feeling-behavioral theories. Today, cognitive theories are widely accepted; nevertheless, much variation exists within them. Cognitive theories suggest that thought process plays a crucial role in a multi-component process of emotion. The question is how and in what capacity does cognition affect the entire process or individual components. There are a number of ways to categorize the differing theories; some address the causation of emotion while others address whether emotions are hardwired or learned. A few of the prominent cognitive theories of emotion are the Two-Factor theory, Appraisal theory, Affect Program theory, Network theory, and Conceptual Act theory.

2.2.1 Two-Factor Theory. The Two-Factor theory combines elements of both the James-Lange and Cannon-Bard theories, addressing the main pitfalls of both [50]. Proposed in 1962 by Stanley Schachter and Jerome Singer, this theory argues that there are two steps to emotion. The first step is the physiological component, such as from Automatic Nervous System (ANS) responses [60]. The second step, also known as cognitive labelling, involves the interpretation of these responses given the circumstances surrounding the individual. This two-step process addresses the issue of similar ANS responses for different emotions that the James-Lange theory does not [50]. It also maintains the connection between visceral reactions and emotional experience - which the Cannon-Bard theory fails to do [57].

2.2.2 Appraisal Theory. Appraisal theory refers to a branch of cognitive theories that suggest that the cognitive aspect of emotion is largely unconscious [50]. This category of emotion theory was pioneered by Magda Arnold in 1960 who coined the term appraisal to mean the cognitive act of assessing the situation. Her theory differs from the Two-Factor theory in that the appraisal occurs before visceral and motor responses, which in turn happen as a result of the appraisal. The unconscious appraisal, which called intuitive appraisal, assesses whether the situation is positive or negative, which then activates a motivation or action tendency of approach or avoidance [1]. Another pioneer of appraisal theory is Richard Lazarus who proposed a multi-level structural model of appraisal. There is the primary appraisal, or the immediate reactions to the situation and its significance, and secondary appraisal, in which the person assesses the ways of coping with the situation [40]. Secondary appraisal is further divided into two categories: direct action (physical ways to alter the situation such as running away), and cognitive reappraisal (altering the way a person feels about the situation by reassessing the situation from a different perspective). More recently, leading appraisal theories include Nico Frijda's Action Readiness theory and Klaus Scherer's Componential theory. The Action Readiness theory, as the name suggests, defines emotion as an awareness of state of action readiness [23]. Similar to

Magda's theory, the core of emotional appraisal is assessing the situation along the axis of pleasantness or unpleasantness which causes an action tendency, however, he expands these tendencies beyond approach or avoidance. In Scherer's Component Process model, he defines emotion as an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism [61]. Thus, the five emotional components are cognitive, neurophysiological, motivational, motor expression, and subjective feeling. Scherer further stresses emotions as being event focused, appraisal driven, and able to elicit a synchronized response of the organism's subsystems to differentiate it from other affective phenomena. He suggests that to measure emotion, an assessment of every component is necessary.

2.2.3 Network Theory. Network Theory is a branch of emotion psychology created by Gordon Bower in the 1980's that focuses on the complex neurological networks associated with memory [50]. He believes that emotions exist as nodes in the brain [5]. As emotional episodes occur, connections are made between the emotion and specific nodes relating to the stimulus, action tendencies, and responses to the event. Eventually these patterns keep building and networks are formed surrounding each emotion node. When such an emotion occurs again, the existing network of responses is activated, and the new experience is added to the network. Variations within this branch of theories deal with the specific ways the connections between these nodes are formed, and whether emotion nodes are innately hardwired or learned [50].

2.2.4 Discrete Emotion Theory. Another way of categorizing emotion theories centers around the debate of the existence of discrete, basic emotions that act as building blocks to all human emotional experience. Theories discussed previously may agree with many or most aspects of the discrete emotion theory [50]. Different theorists propose different sets of basic emotions (in Tomkins case, he describes the basic units as affects rather, see [16, 33, 66]), however, it is agreed that each basic emotion can have different levels of intensity which would elicit slightly more nuanced emotions and allows for more precise language (i.e. interest vs. excitement, anger vs. rage) [66]. Additionally, theorists agree that each emotion triggers a specific neural circuit that elicits a specific pattern of responses in the body's various systems [50]. These emotions are intrinsic and biologically wired from the time we are born and are consistent across cultures [16]. This category of emotion theory is heavily cited in the field of emotion recognition as it lends itself nicely to mapping various bodily responses, in a one-to-one relationship, to discrete emotions. However, this has been hard to prove with empirical data, and will be discussed more in depth later in this paper.

2.2.5 Core Affect/Constructionist Theory. In direct contradiction to the Discrete Emotion theory is the theory of Core Affect. This theory places the human state on a two or three-dimensional scale of level of arousal (high to low intensity), valance (positive or negative), and sometimes motivation tendency (approach or avoid). James Russell, who proposed the Core Affect theory, suggests that our understanding of basic emotions is merely a labelling of regions of core affect [59]. Lisa Barrett added to this theory with specific interest to the categorization of core affect, a process she calls the conceptual act [3]. She believes that emotion is constructed from past experiences and socialization. In her more recent publications, she describes the perception of emotion as conceptual synchrony between two or more individuals [24]. Additionally, she argues that language and culture play major roles in emotion categorization, namely that people of similar backgrounds are more likely to perceive emotions the same way. In recent years, Barrett has done much work in the discreditation of the search for unique and universal responses to discrete emotions and pushes for a more personalized approach to emotion recognition [44, 62].

The overview presented above provides a summary of some of the leading theories and theorists of emotion. There are of course, many more papers and theorists not mentioned in this review (see [50] for a more comprehensive discussion). Regardless, cognitive theories have been at the forefront of theories of emotion in the past few decades. This suggests a heavier emphasis on cognition rather than visceral responses and expressive behavior on the theoretical front. Empirically, however, there seems to be a persistent effort to classify specific somatic and motor responses to specific emotions, particularly in areas of ANS and facial recognition. The relationship between these different elements is still unclear, as is suggested by the vast variation in approaches to emotion recognition, empirical data, and theory. The rest of this paper will review the current landscape and direction of emotion recognition, discussing particularly the different methods explored and the advantages and limitations of each.

3 METHODOLOGY

To identify appropriate content, we surveyed major scientific research article databases (Research Gate, Science Direct, Google Scholar, NCBI etc.) using keywords, which included: *emotion recognition, measurement, Automatic Nervous System or ANS, etc.* First, articles with narrow scopes were excluded. By narrow scope, we refer to studies which attempted to measure emotions in very specific contexts (i.e. measuring emotional response to a particular type of food), or studies that measured the emotional experience of a particular type of person (i.e. only including subjects with specific psychological conditions). Then we identified common modes of emotion recognition, including facial recognition, Autonomic Nervous System (ANS) responses, brain scans, and voice recognition. Within each modality and using existing meta-analytic reviews as guides, we looked for studies that employed similar methodologies and reported results in comparable ways. Given the diverse range of experimental methods used in psychological research, selecting studies with similar (but not identical) general approaches allowed us to better compare, and synthesize results (across multiple studies). While we acknowledge that our paper reviews a mere subset of the existing literature, we believe that we scanned through sufficient reputable databases to make our review fairly representative of the overall progress/landscape of emotion recognition. Additional criteria included availability of the literature in English as well as the year it was published. The landscape of psychological research is constantly changing, and as such, we chose to focus the majority of our review on studies which were published within the last decade.

3.1 Uni-modal Consistency

First, we shall discuss the various uni-modal methods of emotion recognition. The largest, most researched categories are neuroimaging, ANS, facial expression, and speech.

3.1.1 *Neuroimaging.*

Functional Magnetic Resonance Imaging (fMRI) and PET. Developments in imaging technology during the 1990's - most notably, the introduction of the modern fMRI in 1991 - led to a surge in research investigating the relationship between neural structures and emotion [64]. Several recent neuroimaging (PET, fMRI) studies as well as some meta-analyses offer evidence in favour of the neurobiological existence of basic emotions [7]. For instance, three separate meta-analyses [51, 55, 70], have suggested a correlation between certain discrete brain regions and specific emotions or emotional tasks. More recent findings [35, 70], however, suggest that neural profiles for each basic emotion exist, and that a multi-system model with distributed networks should be adopted to differentiate emotions rather than the traditional locationist approach. One of the challenges in meta-analysis is determining how to usefully compare heterogeneous data [20]. While the raw data from individual studies may seem to be similar, meta-analyses combine studies conducted under different experimental conditions (e.g. differences in: type of emotion stimuli, statistical methods used during analysis, theoretical frameworks, etc.). Some argue that meta-analysis methods help to dampen the effects of methodological inconsistencies; however, others claim that due to these differences in methodologies, studies investigate different aspects and functions of emotional phenomena, and therefore, it is impossible to conclude (through meta-analysis) whether emotional categories are associated with unique prototypical brain activity patterns.

Electroencephalogram (EEG). While fMRI and PET technologies can locate activation in specific brain regions, their low time resolution limits researchers' ability to investigate temporal aspects of emotional states [46]. Though some have proposed that neuroimaging methods may be better suited than EEG to reveal emotion specificity in the brain [54], recent developments in EEG technology and the advent of deep neural networks has helped to make results of EEG-based experiments more precise - both from a biological and statistical perspective. It must be noted, however, that due to the recency of said developments, only a limited number of researchers have used machine learning methods to explore the correspondence between emotional states and EEG spectral changes (across the whole brain). While these inconsistencies leave many hypotheses regarding EEG-based emotion recognition unanswered, lack of agreement within the psychological community should not diminish the validity of more recent works.

To this end, we have decided to review a few very recent papers, which used machine learning approaches in EEG-based emotion research. [43] applied machine-learning algorithms to categorize EEG dynamics according to subject self-reported emotional states during music listening. The study evaluated two classifiers, multilayer perceptron (MLP) and support vector machine (SVM), and systematically compared the effects of four feature types on the accuracy of each classifier. By applying SVM, they were able to identify four discrete emotional states (joy, anger, sadness, and pleasure) of participants during music listening with a maximum classification accuracy of 82.29% ± 3.06%. When compared to similar works [9, 29, 32, 65, 74], Lin et al.'s classification method produced the highest reported accuracy. Nevertheless - as the authors themselves acknowledged - this may have been a result of external factors such as experimental conditions, stimulus types, and, the number of induced emotions [43]. An even greater source of concern (not discussed in the paper) is how an individual researcher evaluated participants self-reports. Participants were asked to report their emotional states using FEELTRACE tool - which represents emotion in 2-dimensional (activation-evaluation) space. The classifiers, however, were trained to identify four discrete emotions. This paper did not fully report how the researchers mapped dimensional reporting to discrete emotional states, and this abates the validity of their results. Another recent study [30], explored the EEG correlates of ten discrete positive emotions. Participants were invited to watch short film clips, and then rate their emotional experience according to 14 items, included the ten positive emotions, arousal level, valence, familiarity, and liking. Like Yuan Pin-Lin et al. [43], they similarly employed an SVM classification method and obtained mean classification accuracies between 79.1 ± 5.1% (for Inspiration) and 82.1 ± 4.9% (for Amusement). They also found that using grand-average features (when compared to individual-based features) produced consistently higher classification accuracies. This may suggest that particular emotional stimuli elicited similar neural responses across subjects. Curiously enough, when compared to several works that used EEG signals from the DEAP database [36] to identify affective levels (valence vs. arousal) [2, 13, 14, 36, 52, 67, 75], the two aforementioned studies achieved significantly higher classification accuracy. This discrepancy is particularly notable, as some of the modality used the same SVM classification method [10]. This may indicate an oversimplified categorization of discrete emotions - especially since both (discrete emotion) researcher groups failed to disclose how they came up with their discrete emotion categories. Once again, however, we reach the unsatisfying conclusion that present research neither confirms nor denies the existence of basic emotions or universal emotional expression.

3.1.2 *Autonomic Nervous System (ANS).*

The Autonomic Nervous System (ANS) has been investigated extensively in the area of emotion recognition. Despite this, the complexity of raw data from physiological signals has made discerning clear relationships between discrete emotions and specific ANS patterns a difficult and inconclusive venture. This section will discuss the most common modes of ANS measurement used for emotion recognition and their correlations to certain emotions. The ANS is the part of the nervous system that controls bodily functions. It is divided into two branches, the sympathetic branch, which consumes energy, and parasympathetic branch, which replenishes energy [62]. The three major types of responses measured are cardiovascular, respiratory, and electrodermal (skin conductivity). The following section will focus on the two most commonly studied ANS responses, heart rate (HR) and skin conductance levels (SCL), and their relationship to the most commonly studied emotions (anger, fear, sadness, disgust, happiness). For a more comprehensive review of ANS responses in emotion recognition, see [37, 62].

Heart Rate (HR). Heart rate (HR) is in the most frequently studied ANS measurement across studies [62]. Generally speaking, anger is associated with an increase in HR. However, it seems that the mode of sensory input substantially affects HR. One study found that approach-oriented anger had no effect on HR, while withdrawal-oriented anger resulted in decreased HR [63]. In anxiety, HR usually increases, except when the subject is shocked, shown images, or played anxiety inducing music. In those cases, HR decreased or remained the same. Disgust generally increases or decreases HR according to whether the stimulus is contamination oriented (increases) or mutilation oriented (decreases). Fear mostly triggers an increase in HR except in situations where subjects were shown images and videos [37]. Sadness that induced crying increased HR but otherwise decreased HR. With joy and happiness, HR increased. It seems from the variability in HR within each emotion listed above, intensity, context, and stimulus have greater impact on HR than specific emotions themselves.

Skin Conductance Level (SCL). SCL is the second most commonly studied ANS response to emotion [37]. This ANS response seems to have more consistency within emotions than HR. Anger elicited an increase in SCL [37]. Disgust showed an increase in SCL as did fear. However, certain methods of inducing fear, especially in more realistic settings, resulted in a decrease in SCL. Similar to HR, changes in SCL in sadness depended on whether or not the subject cried. If they did, SCL were unchanged while if they did it decreased. With happiness, results were varied [37].

While some meta-analytic papers suggest correlation between ANS responses and specific emotions, a closer reading indicates a high degree of variation and many exceptions to supposed ANS patterns [37, 62]. The [62] could not find any ANS changes specific to any one emotion category. They also used a multivariate pattern classification analysis (MPCA) to assess multiple ANS responses together. Due to the imbalance of empirical data that exists across literature on emotions studied and ANS responses measured, they used binary classification to find correlations. Ultimately, however, this method did not result in any strong multivariate patterns for emotion categories. While some argue that the next steps would be to gather even more data, divide emotions further into subcategories, and measure more modes of ANS [37], eventually the list of factors become so extensive that attempting to generalize seems to lack purpose. Additionally, the activation components, level of arousal, stimulus, and laboratory conditions seemed to have significant impact on results. For example, there seems to be a tendency for the use of picture or video as stimuli to result in responses that may or may not fit with other trends, posing the question of whether these emotions are genuine. Inconsistencies in experimental methods across studies cause a lot of noise, thus, standardized methods or parameters in emotion recognition needs to be created in order to yield more conclusive results.

3.1.3 Facial Expression - Facial Action Coding System (FACS). Up until the 1960's the dominant perspective in psychology was that facial expressions were culturally specific. This view was dismantled by the work of Tomkins, Ekman, and Izard (what is known today as the "universality studies" [45]), which provided strong evidence of universality of some facial expressions of emotion. In addition to spurring, also known as the basic emotions theory, these findings also inspired the development of the Facial Action Coding System (FACS) - arguably, the most extensively used observational coding system [17]. FACS describes facial activity on the basis of 44 action units (AUs), as well as head and eye positions and movements. Each AU has a numeric code; discrete facial expressions (named events) are described by one or many AUs. This measurement tool was revised again in 2003, and since has been applied in the study of computer based facial measurements. Although Ekman himself has said that facial expressions alone do not equal emotion, they are thought to be a distinctive characteristic of emotion (expression).

3.1.4 Speech - Speech Emotion Recognition (SER). Speech emotion recognition (SER) has attracted a considerable amount of interest over the past two decades - particularly from the machine learning community. Several recent studies have proposed machine-learning-based SER systems. For instance, [27] proposed a method using Deep Neural Network (DNN) architecture with convolutional, pooling and fully-connected layers. Using raw audio data from the German Corpus (Berlin Database of Emotional Speech), their trained model achieved an overall test accuracy of 96.97% on whole-file classification into three classes of angry, neutral, or sad. Another notable database is the Interactive Emotional Dyadic Motion Capture (IEMOCAP) database from USC, which has been used by a number of researchers to evaluate different SER approaches [22, 26, 31, 41, 47, 49, 69]. The most successful models came from Fayek and Tzinis [22], whose Convolutional Neural Network (CNN) architecture and Bi-directional Long Short-Term Memory-Extreme Learning Machine (LSTM-ELM) model yielded accuracies of 64.78% and 64.16%, respectively. Another group of researchers, [68] undertook a similar approach, combining CNNs with LSTM networks, and achieved similar prediction accuracies when applied to the REmote COLaborative and Affective (RECOLA) dataset [58] (68.6% for arousal, 26.1% valence). It is important to note, the audio tracks from these datasets were recorded by actors - and therefore, are not entirely reflective of natural speech. That is not to diminish the achievements of previous works - more research needs to be done particularly on organic speech.

3.2 Multi-modal

In the last decade, multi-modal approaches have been proposed to better recognize emotions. With rapid advancement in computational power, the use of Machine Learning (ML) and Deep Learning (DL) techniques has become increasingly attractive for this purpose. The DL is a subfield of ML inspired by human brain neurons. It allows more complex problems to be solved with higher accuracy, given sufficient training samples. In addition, it can be applied to a variety of problems, for example, classification and segmentation. In the emotional researches, ML and DL have addressed various tasks, including face detection and tracking, speech recognition, voice-activity detection, and emotion classification from face and voice [39].

In [69], they propose an emotion recognition system which uses an end-to-end deep learning multi-modal model. Before training the multi-modal model, each modality specific network was trained separately as follows:

- ***Visual Network:*** To extract features, they used a deep residual network (ResNet) [28] of 50 layers. Screen captures were taken from videos of subjects, and pixel intensities of cropped subjects' faces were used as input data
- ***Speech Network:*** A Convolutional Neural Network (CNN) model was utilized to extract features from the raw audio signal.

The features from each network were combined and then fed to a 2-layer Long-Short Term Memory (LSTM) with 256 cells in each stack. The LSTM was used to consider contextual information. They then used audio and visual data from the REmote COLaborative and Affective (RECOLA) Database [58] to evaluate their model. The dataset contains four modalities: audio, video, electro-cardiogram (ECG) and electro-dermal activity (EDA). To assess the accuracy of their model, [25] used a framework that takes into account multiple sensory inputs: speech, facial movements, and everyday activities. The machine learning classification model was employed to combine sensor inputs and predict the emotional state at a given time. The framework was underpinned on the Componential Emotion Theory and Scherer's Emotional Semantic Space (ESS) [19, 53, 61], as the authors believed that adopting a holistic view of emotions would allow them to draw significant conclusions about emotional states. The emotions were divided in eight dimensions: tension, frustration, dissatisfaction, anxiousness, seriousness, gladness, enthusiasm, and excitation. Their model utilized different machine learning algorithms, namely support vector machine (SVM) and decision trees, as classifiers. Their results suggest that the combination of outputs generated by multiple sensors provides more accurate assessment of emotional states than when the sensors are treated individually.

Physiological signal data has been used as indicators of human emotions as we have mentioned earlier. Some researchers have asserted that models might suffer as a result of inappropriate selection of model structure and/or oversimplification of multi-modal features fusion [73]. [72] addressed those issues by proposing a multiple-fusion-layer based ensemble classifier of stacked autoencoder (MESAE). The stacked autoencoder (SAE) consisted of three hidden layers to reduce noise in physiological signal data. An additional deep model was used to achieve the SAE ensembles. The derived SAE abstractions were combined according to the physiological modality to create six sets of encodings, and then fed to a three-layer, adjacent-graph-based network for feature fusion. The fused features are used to recognize binary arousal or valence states. DEAP multi-modal database [36] was employed to validate the MESAE. Robust physiological signal data is available in the DEAP database, including electroencephalogram (EEG), galvanic skin response (GSR), respiration amplitude, skin temperature, electrocardiogram (ECG), blood volume, electromyograms (EMG), and electrooculogram (EOG). The results showed that the mean classification rate and F-score improved by 5% approximately when compared to the methods of [36] and [13]. Many researches have also attempted to analyze multiple physical modalities of expression, such as face, voice or speech, body posture and walking gait [18, 48]. Recently, [56] designs an experiment using smart watch sensor data to measure emotions. 50 participants were asked to wear smart watch (Samsung Gear 2) while watching two types of stimuli; audio-visual and audio. The participants also rated their current mood state using the Positive Affect and Negative Affect (PANAS) schedule [71]. The process was repeated three times for each of the following emotions; happiness, sadness and neutral. The authors extracted features from the sensor data and walking times labelled by the experimenter, and then built classifiers (personal models) that recognized the emotional state of the user. For the task of emotion recognition using classifiers, Random Forest (RF) [42] and logistic regression with L2 regularization [21], the results suggested that the personal models outperformed personal baselines, with personal models achieving median accuracies higher than 78% for the binary classification of happiness vs sadness. For multiclass classification of happiness vs sadness vs neutral, the accuracies of personal models on average outperformed the baselines.

In the view of using multi-modal methods to recognize emotions, input data are typically segregated into one of two categories: signal and visual. Signal data refers to the data that can be represented sequentially by time. It can be further separated into audio, voice, speech and physiological data. Visual data consists of facial expressions, neuroimages, posture and gesture data. To recognize emotions, models use ML algorithms, including DL techniques such as CNN, LSTM and AutoEncoder. However, to better choose which model(s) we should apply for this purpose, we need to understand the data representation. For example, CNN model is great for images or data that is independent. LSTM is great for signal data. On the other hand, we can use traditional machine learning algorithms such as RF if we need the interpretable model as used in [56]. The other important thing is determining how to combine or extract the effective features from different input data. Using multi-modal data might give more useful findings and the results might be more accurate.

4 CONCLUSIONS

In this review, we summarized a number of studies and theories that have contributed to the debate surrounding the definition of emotion, and whether or not emotion can be empirically measured. We believe that, given the current body of knowledge, there is insufficient evidence to neither prove nor disprove the existence of coherent emotional expression (both within subjects and between subjects). Furthermore, results seem to be heavily influenced by both experimental conditions, as well as by the theoretical assumptions that underpin them. Nevertheless, there does not appear to be substantive evidence to support the hypothesis of a universal one-to-one relationship between specific modalities and discrete emotional states. Nevertheless, we argue that these inconsistencies can, in part, be attributed to methodological discrepancies. These include differences in emotion elicitation techniques, analytical methods, assessment periods, temporal matching between measurement periods and emotional events, and incomplete characterization of autonomic systems across studies that make it difficult to compare results. While some may argue that the search for basic/universal emotional expression/experience is a frivolous endeavour, the ubiquitous role of emotion in our daily lives is undeniable. Therefore, we believe that it is worthwhile to continue to investigate human emotional states. In the future, we will focus on designing and implementing multi-modal approaches to recognize emotions. We

also aim to establish a ground-truth of each emotion by leveraging machine learning, domain knowledge and theories.

REFERENCES

[1] Magda B Arnold. 1960. Emotion and personality. (1960).

[2] Fatemeh Bahari and Amin Janghorbani. 2013. Eeg-based emotion recognition using recurrence plot analysis and k nearest neighbor classifier. In *Biomedical Engineering (ICBME), 2013 20th Iranian Conference on*. IEEE, 228–233. https://doi.org/10.1109/ICBME.2013.6782224

[3] Lisa Feldman Barrett. 2006. Solving the emotion paradox: Categorization and the experience of emotion. *Personality and social psychology review* 10, 1 (2006), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

[4] Julie Beck. 2015. Hard feelings: Science's struggle to define emotions. *The Atlantic* 24 (2015). https://www.theatlantic.com/health/archive/2015/02/hard-feelings-sciences-struggle-to-define-emotions/385711/

[5] Gordon H Bower. 1981. Mood and memory. *American psychologist* 36, 2 (1981), 129.

[6] Walter B Cannon. 1927. The James-Lange theory of emotions: A critical examination and an alternative theory. *The American journal of psychology* 39, 1/4 (1927), 106–124. https://doi.org/10.2307/1415404

[7] Alessia Celeghin, Matteo Diano, Arianna Bagnis, Marco Viola, and Marco Tamietto. 2017. Basic emotions in human neuroscience: neuroimaging and beyond. *Frontiers in Psychology* 8 (2017), 1432. https://doi.org/10.3389/fpsyg.2017.01432

[8] Ling Cen, Fei Wu, Zhu Liang Yu, and Fengye Hu. 2016. A Real-Time Speech Emotion Recognition System and its Application in Online Learning. In *Emotions, Technology, Design, and Learning*. Elsevier, 27–46.

[9] Guillaume Chanel, Julien Kronegg, Didier Grandjean, and Thierry Pun. 2006. Emotion assessment: Arousal evaluation using EEG's and peripheral physiological signals. In *International workshop on multimedia content representation, classification and security*. Springer, 530–537. https://doi.org/10.1007/11848035_70

[10] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* 2, 3 (2011), 27. https://doi.org/10.1145/1961189.1961199

[11] Vasileios Charisis, Stelios Hadjidimitriou, Leontios Hadjileontiadis, Deniz Uğurca, and Erdal Yilmaz. 2015. EmoActivity-An EEG-based gamified emotion HCI for augmented artistic expression: The i-Treasures paradigm. In *International Conference on Universal Access in Human-Computer Interaction*. Springer, 29–40.

[12] Darwin Charles, Ekman Paul, and Prodger Phillip. 1872/2005. The expression of the emotions in man and animals. *Electronic Text Center, University of Virginia Library* (1872/2005). https://doi.org/10.1037/10001-000

[13] Mo Chen, Junwei Han, Lei Guo, Jiahui Wang, and Ioannis Patras. 2015. Identifying valence and arousal levels via connectivity between EEG channels. In *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*. IEEE, 63–69. https://doi.org/10.1109/ACII.2015.7344552

[14] Seong Youb Chung and Hyun Joong Yoon. 2012. Affective classification using Bayesian classifier and supervised learning. (2012), 1768–1771.

[15] Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G Taylor. 2001. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine* 18, 1 (2001), 32–80.

[16] Paul Ekman. 1971. Universals and cultural differences in facial expressions of emotion.. In *Nebraska symposium on motivation*. University of Nebraska Press.

[17] Paul Ekman and Erika L Rosenberg. 1997. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA.

[18] P Elkman. 1999. Emotional and conversational nonverbal signals. *Gesture, speech, and sign* (1999), 44–55.

[19] Phoebe C Ellsworth and Klaus R Scherer. 2003. Appraisal processes in emotion. *Handbook of affective sciences* 572 (2003), V595.

[20] Hans J Eysenck. 1994. Systematic reviews: Meta-analysis and its problems. *Bmj* 309, 6957 (1994), 789–792. https://doi.org/10.1136/bmj.309.6957.789

[21] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *Journal of machine learning research* 9, Aug (2008), 1871–1874.

[22] Haytham M Fayek, Margaret Lech, and Lawrence Cavedon. 2017. Evaluating deep learning architectures for Speech Emotion Recognition. *Neural Networks* 92 (2017), 60–68. https://doi.org/10.1016/j.neunet.2017.02.013

[23] Nico H Frijda. 1986. *The emotions*. Cambridge University Press.

[24] Maria Gendron and Lisa Feldman Barrett. 2018. Emotion perception as conceptual synchrony. *Emotion Review* 10, 2 (2018), 101–110.

[25] Vinícius P Gonçalves, Eduardo P Costa, Alan Valejo, PR Geraldo Filho, Thienne M Johnson, Gustavo Pessin, and Jó Ueyama. 2017. Enhancing intelligence in multimodal emotion assessments. *Applied Intelligence* 46, 2 (2017), 470–486. https://doi.org/10.1007/s10489-016-0842-7

[26] Kun Han, Dong Yu, and Ivan Tashev. 2014. Speech emotion recognition using deep neural network and extreme learning machine. In *Fifteenth annual conference of the international speech communication association*.

[27] Pavol Harar, Jesus B Alonso-Hernandezy, Jiri Mekyska, Zoltan Galaz, Radim Burget, and Zdenek Smekal. 2017. Voice pathology detection using deep learning: a preliminary study. In *Bioinspired Intelligence (IWOBI), 2017 International Conference and Workshop on*. IEEE, 1–4. https://doi.org/10.1109/IWOBI.2017.7985525

[28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[29] Alicia Heraz, Ryad Razaki, and Claude Frasson. 2007. Using machine learning to predict learner emotional state from brainwaves. In *Advanced Learning Technologies, 2007. ICALT 2007. Seventh IEEE International Conference on*. IEEE, 853–857. https://doi.org/10.1109/ICALT.2007.277

[30] Xin Hu, Jianwen Yu, Mengdi Song, Chun Yu, Fei Wang, Pei Sun, Daifa Wang, and Dan Zhang. 2017. EEG correlates of ten positive emotions. *Frontiers in human neuroscience* 11 (2017), 26. https://doi.org/10.3389/fnhum.2017.00026

[31] Che-Wei Huang and Shrikanth S Narayanan. 2016. Attention Assisted Discovery of Sub-Utterance Structure in Speech Emotion Recognition.. In *INTERSPEECH*. 1387–1391. https://doi.org/10.21437/Interspeech.2016-448

[32] Keisuke Ishino and Masafumi Hagiwara. 2003. A feeling estimation system using a simple electroencephalograph. In *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, Vol. 5. IEEE, 4204–4209. https://doi.org/10.1109/ICSMC.2003.1245645

[33] Carroll E Izard. 1977/2013. *Human emotions*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4899-2209-0

[34] WILLIAM JAMES. 1884. II.—WHAT IS AN EMOTION ? *Mind* os-IX, 34 (1884), 188–205. https://doi.org/10.1093/mind/os-IX.34.188

[35] Lauren AJ Kirby and Jennifer L Robinson. 2015. Affective mapping: An activation likelihood estimation (ALE) meta-analysis. *Brain and cognition* (2015). https://doi.org/10.1016/j.bandc.2015.04.006

[36] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2012. Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing* 3, 1 (2012), 18–31. https://doi.org/10.1109/T-AFFC.2011.15

[37] Sylvia D Kreibig. 2010. Autonomic nervous system activity in emotion: A review. *Biological psychology* 84, 3 (2010), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

[38] James D Laird and Katherine Lacasse. 2014. Bodily influences on emotional feelings: Accumulating evidence and extensions of William James's theory of emotion. *Emotion Review* 6, 1 (2014), 27–34. https://doi.org/10.1177/1754073913494899

[39] Charlyn Pushpa Latha and Mohana Priya. 2016. A Review on Deep Learning Algorithms for Speech and Facial Emotion Recognition. *APTIKOM Journal on Computer Science and Information Technologies* 1, 3 (2016), 88–104.

[40] Richard S Lazarus. 1991. Progress on a cognitive-motivational-relational theory of emotion. *American psychologist* 46, 8 (1991), 819. https://doi.org/10.1037/0003-066X.46.8.819

[41] Jinkyu Lee and Ivan Tashev. 2015. High-level feature representation using recurrent neural network for speech emotion recognition. (2015). https://www.microsoft.com/en-us/research/publication/high-level-feature-representation-using-recurrent-neural-network-for-speech-\emotion-recognition/

[42] Andy Liaw, Matthew Wiener, et al. 2002. Classification and regression by randomForest. *R news* 2, 3 (2002), 18–22.

[43] Yuan-Pin Lin, Chi-Hong Wang, Tzyy-Ping Jung, Tien-Lin Wu, Shyh-Kang Jeng, Jeng-Ren Duann, and Jyh-Horng Chen. 2010. EEG-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering* 57, 7 (2010), 1798–1806. https://doi.org/10.1109/TBME.2010.2048568.

[44] Kristen A Lindquist, Tor D Wager, Hedy Kober, Eliza Bliss-Moreau, and Lisa Feldman Barrett. 2012. The brain basis of emotion: a meta-analytic review. *Behavioral and brain sciences* 35, 3 (2012), 121–143. https://doi.org/10.1017/S0140525X11000446

[45] David Matsumoto and Hyi Sung Hwang. 2011. Reading facial expressions of emotion. *Psychological Science Agenda* 25, 5 (2011). http://www.apa.org/science/about/psa/2011/05/facial-expressions.aspx

[46] Iris B Mauss and Michael D Robinson. 2009. Measures of emotion: A review. *Cognition and emotion* 23, 2 (2009), 209–237. https://doi.org/10.1080/02699930802204677

[47] Angeliki Metallinou, Martin Wollmer, Athanasios Katsamanis, Florian Eyben, Bjorn Schuller, and Shrikanth Narayanan. 2012. Context-sensitive learning for enhanced audiovisual emotion classification. *IEEE Transactions on Affective Computing* 3, 2 (2012), 184–198. https://doi.org/10.1109/T-AFFC.2011.40

[48] Johannes Michalak, Nikolaus F Troje, Julia Fischer, Patrick Vollmar, Thomas Heidenreich, and Dietmar Schulte. 2009. Embodiment of sadness and depression-gait patterns associated with dysphoric mood. *Psychosomatic medicine* 71, 5 (2009), 580–587.

[49] Seyedmahdad Mirsamadi, Emad Barsoum, and Cha Zhang. 2017. Automatic speech emotion recognition using recurrent neural networks with local attention. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2227–2231. https://doi.org/10.1109/ICASSP.2017.7952552

[50] Agnes Moors. 2009. Theories of emotion causation: A review. *Cognition and emotion* 23, 4 (2009), 625–662.

[51] Fionnuala C Murphy, IAN Nimmo-Smith, and Andrew D Lawrence. 2003. Functional neuroanatomy of emotions: a meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience* 3, 3 (2003), 207–233. https://doi.org/10.3758/CABN.3.3.207

[52] Daimi Syed Naser and Goutam Saha. 2013. Recognition of emotions induced by music videos using DT-CWPT. In *Medical Informatics and Telemedicine (ICMIT), 2013 Indian Conference on*. IEEE, 53–57. https://doi.org/10.1109/IndianCMIT.2013.6529408

[53] Paula M Niedenthal, Silvia Krauth-Gruber, and Francois Ric. 2006. Psychology of emotion: Interpersonal, experimental and cognitive approaches. , 432 pages.

[54] Jaak Panksepp. 1998/2004. *Affective neuroscience: The foundations of human and animal emotions*. Oxford university press.

[55] K Luan Phan, Tor Wager, Stephan F Taylor, and Israel Liberzon. 2002. Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage* 16, 2 (2002), 331–348. https://doi.org/10.1006/nimg.2002.1087

[56] Juan C Quiroz, Elena Geangu, and Min Hooi Yong. 2018. Emotion-Recognition Using Smart Watch Sensor Data: Mixed-Design Study. *arXiv preprint arXiv:1806.08518* (2018).

[57] Rainer Reisenzein. 1983. The Schachter theory of emotion: Two decades later. *Psychological bulletin* 94, 2 (1983), 239. https://explorable.com/schachter-singer-theory-of-emotion

[58] Fabien Ringeval, Andreas Sonderegger, Juergen Sauer, and Denis Lalanne. 2013. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 1–8. https://doi.org/10.1109/FG.2013.6553805

[59] James A Russell and Lisa Feldman Barrett. 1999. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology* 76, 5 (1999), 805. https://doi.org/10.1037/0022-3514.76.5.805

[60] Stanley Schachter and Jerome Singer. 1962. Cognitive, social, and physiological determinants of emotional state. *Psychological review* 69, 5 (1962), 379. https://doi.org/10.1037/h0046234

[61] Klaus R Scherer. 2005. What are emotions? And how can they be measured? *Social science information* 44, 4 (2005), 695–729.

[62] Erika H Siegel, Molly K Sands, Wim Van den Noortgate, Paul Condon, Yale Chang, Jennifer Dy, Karen S Quigley, and Lisa Feldman Barrett. 2018. Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychological bulletin* 144, 4 (2018), 343. https://doi.org/10.1037/bul0000128

[63] Gerhard Stemmler, Tatjana Aue, and Jan Wacker. 2007. Anger and fear: Separable effects of emotion and motivational direction on somatovisceral responses. *International Journal of Psychophysiology* 66, 2 (2007), 141–153. https://doi.org/10.1016/j.ijpsycho.2007.03.019

[64] Mark Stokes. 2015. What does fMRI measure? Retrieved July 10, 2018 from https://www.nature.com/scitable/blog/brain-metrics/what_does_fmri_measure

[65] Kazuhiko Takahashi et al. 2004. Remarks on emotion recognition from biopotential signals. In *2nd International conference on Autonomous Robots and Agents*, Vol. 3. 1148–1153.

[66] Silvan S Tomkins. 1984. Affect theory. *Approaches to emotion* 163, 163–195 (1984).

[67] Cristian A Torres-Valencia, Hernan F Garcia-Arias, Mauricio A Alvarez Lopez, and Alvaro A Orozco-Gutiérrez. 2014. Comparative analysis of physiological signals and electroencephalogram (EEG) for multimodal emotion recognition using generative models. In *Image, Signal Processing and Artificial Vision (STSIVA), 2014 XIX Symposium on*. IEEE, 1–5. https://doi.org/10.1109/STSIVA.2014.7010181

[68] George Trigeorgis, Fabien Ringeval, Raymond Brueckner, Erik Marchi, Mihalis A Nicolaou, Björn Schuller, and Stefanos Zafeiriou. 2016. Adieu features? end-to-end speech emotion recognition using a deep convolutional recurrent network. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 5200–5204. https://doi.org/10.1109/ICASSP.2016.7472669

[69] Efthymios Tzinis and Alexandras Potamianos. 2017. Segment-based speech emotion recognition using recurrent neural networks. In *Affective Computing and Intelligent Interaction (ACII), 2017 Seventh International Conference on*. IEEE, 190–195. https://doi.org/10.1109/ACII.2017.8273599

[70] Katherine Vytal and Stephan Hamann. 2010. Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *Journal of cognitive neuroscience* 22, 12 (2010), 2864–2885. https://doi.org/10.1162/jocn.2009.21366

[71] David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* 54, 6 (1988), 1063.

[72] Zhong Yin, Mengyuan Zhao, Yongxiong Wang, Jingdong Yang, and Jianhua Zhang. 2017. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Computer methods and programs in biomedicine* 140 (2017), 93–110. https://doi.org/10.1016/j.cmpb.2016.12.005

[73] Jianhua Zhang, Zhong Yin, and Rubin Wang. 2017. Pattern classification of instantaneous cognitive task-load through GMM clustering, laplacian eigenmap, and ensemble SVMs. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 14, 4 (2017), 947–965. https://doi.org/10.1109/TCBB.2016.2561927

[74] Qing Zhang and Minho Lee. 2009. Analysis of positive and negative emotions in natural scene using brain activity and GIST. *Neurocomputing* 72, 4-6 (2009), 1302–1306. https://doi.org/10.1016/j.neucom.2008.11.007

[75] Xiaodan Zhuang, Viktor Rozgic, and Michael Crystal. 2014. Compact unsupervised eeg response representation for emotion recognition. In *Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on*. IEEE, 736–739. https://doi.org/10.1109/BHI.2014.6864469

View publication stats