Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review

Jianhua Zhang Writing - review & editing Writing - original draft Supervision Methodology Investigation [a,*] jianhuaz@oslomet.no, Zhong Yin Writing - original draft [b], Peng Chen Methodology Investigation [c], Stefano Nichele Resources [a]

[a]Dept. of Computer Science, Oslo Metropolitan University, Norway

[b]Dept. of Control Science and Engineering, University of Shanghai for Science and Technology, China

[c]School of Information Science and Engineering, East China University of Science and Technology, China

*Corresponding author.

Abstract

In recent years, the rapid advances in machine learning (ML) and information fusion has made it possible to endow machines/computers with the ability of emotion understanding, recognition, and analysis. Emotion recognition has attracted increasingly intense interest from researchers from diverse fields. Human emotions can be recognized from facial expressions, speech, behavior (gesture/posture) or physiological signals. However, the first three methods can be ineffective since humans may involuntarily or deliberately conceal their real emotions (so-called social masking). The use of physiological signals can lead to more objective and reliable emotion recognition. Compared with peripheral neurophysiological signals, electroencephalogram (EEG) signals respond to fluctuations of affective states more sensitively and in real time and thus can provide useful features of emotional states. Therefore, various EEG-based emotion recognition techniques have been developed recently. In this paper, the emotion recognition methods based on multi-channel EEG signals as well as multi-modal physiological signals are reviewed. According to the standard pipeline for emotion recognition, we review different feature extraction (e.g., wavelet transform and nonlinear dynamics), feature reduction, and ML classifier design methods (e.g., k-nearest neighbor (KNN), naive Bayesian (NB), support vector machine (SVM) and random forest (RF)). Furthermore, the EEG rhythms that are highly correlated with emotions are analyzed and the correlation between different brain areas and

emotions is discussed. Finally, we compare different ML and deep learning algorithms for emotion recognition and suggest several open problems and future research directions in this exciting and fast-growing area of AI.

Keywords: Emotion recognition; Affective computing; Physiological signals; Feature dimensionality reduction; Data fusion; Machine learning; Deep learning

List of main acronyms.

Acronym	Full form	Acronym	Full form
EEG	Electroencephalogram	LDA	Linear Discriminant Analysis
ECG	Electrocardiography	LSTM	Long- Short-Term Memory
EMG	Electromyography	RNN	Recurrent Neural Network
EOG	Electrooculography	SVM	Support Vector Machine
FFT	Fast Fourier Transform	ML	Machine Learning
DBN	Deep Belief Network	ELM	Extreme Learning Machine
CNN	Convolutional Neural Network	HMM	Hidden Markov Model
DBM	Deep Boltzmann Machine	PCA	Principal Component Analysis
RBM	Restricted Boltzmann Machine	ICA	Independent Component Analysis
ECDF	Empirical Cumulative Distribution Function	kNN	K-nearest neighbors
EMD	Empirical Mode Decomposition	DL	Deep Learning
LR	Logistic Regression	RBF	Radial Basis Function

1 Introduction

In recent years, the availability of various electronic products in our lives has made people spend more and more time on social media, online shopping, online video games, etc. However, most of contemporary human-computer interaction (HCI) systems are deficient in interpreting and understanding emotional information and lack emotional intelligence. They are unable to identify human emotional states and use this information for decision-making and action.

Resolving the lack of rapport between humans and machines is critical in advanced intelligent HCI. Any HCI system that ignores human affective states would fail to react to those states properly. To address this problem in HCI, we need to equip machines with the ability to interpret and understand human emotional states. Hence, a prerequisite for implementing intelligent HCI is a reliable, accurate, adaptable and robust emotion recognition system. With the ultimate goal of endowing machine with emotions, more and more researchers in the field of artificial intelligence (AI) have carried out studies of affective computing in general and emotion recognition in particular, making them an emerging and promising research area.

Recently emotion-aware intelligent systems have been used in various areas such as e-health, e-learning, recommender systems, smart home, smart city, and intelligent conversational systems (e.g., chatbot). The use of computer-based automatic emotion recognition has great potential in various intelligent systems, including online gaming, neuromarketing (customers' feedback assessment), and mental health monitoring. Given the importance of mental health in contemporary societies, researchers are now finding ways to accurately recognize human emotions in order to develop intervention schemes for mental health. For example, in a healthcare system with module of emotion recognition module, patients' mental and physical states can be monitored in real time and appropriate therapy can be prescribed accordingly. In the field of HCI, the goal of emotion recognition/detection is to design and implement intelligent systems with optimized HCI, which are adaptable to users' emotional states.

What is emotion? What factors induce emotion? Are emotions computable? Can the computer automatically recognize human emotions? These vague questions were only concern of science fiction about two or three decades ago. Nevertheless, over the past two decades it has become a hot research topic to automatically recognize/detect human emotions and make HCI as natural as human-human interactions.

1.1 Background and motivations

As the interaction/collaboration between man and machines (computers) exists in a variety of environments, more and more researchers in the fields of ergonomics and intelligent systems are trying to improve the efficiency and flexibility of human-computer interaction (HCI). This intelligent HCI system requires the computer to be adaptive, and it is critical to precisely understand the ways of human communications and in turn trigger correct feedback. Human intentions can be expressed through verbal and nonverbal behaviors with different emotions. A key factor of computer adaptability is its ability to understand human emotions and behavior. Most existing HCI systems lack the ability to recognize human emotional states. The computerized automatic recognition of human emotional state is significant to the development of advanced HCI systems. This emerging research field is termed as affective computing.

Emotion is a complex state that combines feelings, thoughts, and behavior and is people's psychophysiological reactions to internal or external stimuli. It plays a vital role in people's decision-making, perception and communication. Affective computing has a wide range of applications. In a HCI system, if the computer can recognize the human operator's emotional state accurately and in real time, the interaction between the machine and the operator can be made more intelligent and user-friendly. The application of emotion recognition in the product design and user experience allows for monitoring in real time the emotional state of

the user when using the product, thereby further improving the user experience. In military and aerospace applications, the high-risk functional state of soldiers and pilots/astronauts can be detected in real time. The emotion recognition can also be applied to public transportation, for example to enhance driving safety by monitoring the emotional state of the driver in real time to prevent dangerous driving under extreme emotional conditions.

Emotion recognition is a key component of affective computing. It is an interdisciplinary field that spans computer science, AI, psychology, and cognitive neuroscience. Human emotions can be identified by facial expression, speech, behavior, or physiological signals [1-4]. However, the first three methods of emotion recognition are somehow subjective. For instance, the subjects under study may deliberately conceal their true feelings, which may be inconsistent with their performance. In contrast, the emotion recognition by means of physiological signals is more reliable and objective [5]. EEG signals are generated by the central nervous system (CNS) and respond more rapidly to emotional changes than other peripheral neural signals. Moreover, EEG signals have been shown to provide important features for emotional recognition [6,7]. This paper is aimed at analyzing the complex correlation between EEG signals and emotional states in humans.

1.2 An overview of emotion recognition approaches

In recent years, the AI field has undergone rapid development and there is an urgent need for intelligent HCI. As an important branch of AI research, affective computing has drawn increasingly intense interest from researchers. Currently the research of emotion recognition is focused on the following topics: (1) the correlation between different types of physiological signals and emotions; (2) stimuli selection methods for inducing the expected emotional states; (3) emotion-characteristic feature extraction algorithms; (4) mechanistic or causal models of emotion generation mechanism; and (5) emotion recognition techniques based on multi-modal information fusion. In the following, we will give a brief overview of earlier and current research developments in those directions.

Picard's Lab at the Massachusetts Institute of Technology (MIT) has conducted a significant amount of research, demonstrating that certain affective states can be recognized by using physiological signals including heart rate, galvanic skin response (GSR), temperature, EMG and respiration rate. For instance, she and her associates used personalized imagery to elicit targeted emotions and collected four channels of physiological signals (EMG, pulse rate, GSR and respiration) to recognize up to eight classes of emotional states [8]. They extracted the time- and frequency-domain features from those physiological signals respectively. The feature selection was performed by forward floating search method, Fisher projection method and a hybrid algorithm. They achieved an overall classification accuracy of 88.3% for 3-class (anger, sadness, and happiness) problem by using kNN classification algorithm and 81% for 8-class problem by using hybrid LDA. Khorrami and his team used visual and auditory cues to induce emotion, and collected four types of physiological signals, namely temperature, galvanic skin response, blood volume fluctuation, and electrocardiogram [9]. The resultant average classification accuracy is 61.8%. Chanel et al. used the international emotional picture system to induce emotions in the subjects, and performed 100 high arousal and low arousal emotion induction on the four subjects, and recorded the EEG, blood pressure, and skin conductance response of the subjects [10] . Heart rate, skin temperature and respiratory signals were extracted, and linear discriminant analysis and naive

Bayes were used for emotion recognition. A classification accuracy of about 55% was reported. Koelstra et al. used music video clips as stimulating material, instructing each of the 32 subjects to watch 40 pieces of music video material, and recorded the self-report (subjective ratings), facial expression, EEG and peripheral physiological signals [11]. A classification accuracy of 0.67.7% was achieved. Schmidt et al. used music to induce four emotions [12]. They found that when using positive musical materials, the EEG activity in the frontal areas of left hemisphere was enhanced, while the EEG activity in the frontal areas of right hemisphere is enhanced when using the negative music materials. The authors conclude that there is a strong correlation between the frontal areas of human brain and the emotion. Wagner has also done a lot of work in this area [13]. They collected four types of physiological signals, including electrocardiogram, galvanic skin response, EOG, and respiration, and extracted the physiological features respectively. Three feature selection methods were tested and compared, namely variance analysis, Fisher projection method and sequence forward drifting selection algorithm. Three classifiers, namely K-nearest neighbor, linear discriminant analysis, and multi-layer perceptron, were used to identify the four emotions of joy, happiness, anger and sadness, and encouraging classification results were achieved.

Lu and his associates carried out research on emotion recognition based on physiological signals [14-17]. In recent years, they have set up EEG-based emotion recognition database accessible to researchers. Other researchers also set up the emotion-related database, consisting of four physiological signals of ECG, galvanic skin response, skin temperature and respiration, and the binary (happy vs. relaxed) classification rate reached 86.7% [18]. Liu and others collected physiological signals of heart rate, respiration, skin conductance response, ECG, EMG and pulse rate from 500 university students, and extracted the features through neural network, random forest, and evolutionary algorithms (such as particle swarm optimization, ant colony optimization, etc.) [19-23]. The average 6-class (happiness, disgust, sadness, fear, anger and surprise) recognition rate was reported to be 60-90%.

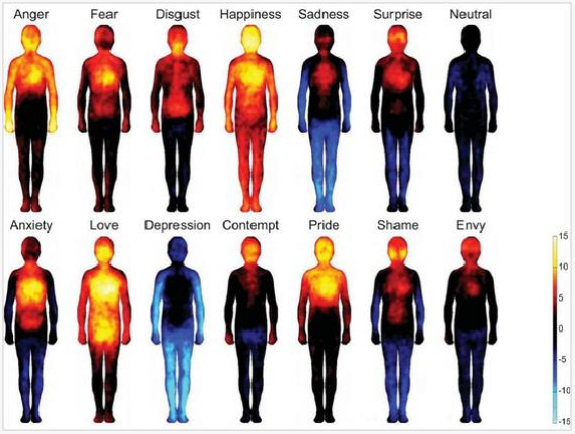
1.3 Affective computing

Affective computing is an emerging cross-disciplinary research field that aims to enable intelligent systems to recognize, infer/predict, and interpret human emotions and spans such domains as computer science, AI, cognitive science, neuroscience, neuropsychology, and social science. The goal of affective computing is to recognize emotional cues during HCI and forming emotional responses. Affective computing is the set of techniques of affect recognition from data in different modalities and granularities. Affective computing research mainly comprises the topics of sentiment analysis and emotion recognition. The former performs coarse-grained affect recognition (usually a task of binary positive vs. negative or 3-class positive, negative, and neutral sentiments classification), whereas the latter involves fine-grained analysis (usually a multiclass classification of big data into a larger set of emotion labels, for example more than 4 classes). Over the past two decades, AI researchers have attempted to endow machines with cognitive capabilities to recognize, interpret and express emotions and sentiments. All such efforts can regarded as affective computing research.

Emotion is a psychophysiological phenomenon. Fig. 1 shows the bodily map of different emotions [24]. Due to its complexity, psychologists have not yet reached a consensus on a unifying definition of emotion or sentiment. There are hundreds of theories related to emotions or sentiments available [25].

alt-text: Fig 1

Fig. 1



The bodily map of human emotions.

The main goal of emotion recognition study is to achieve more natural, harmonious and friendly HCI, and to transform the computer from a logical computing machine into an intuitive perceptron, thus realizing the transformation from the machine- to human-centric design of machines. In order for the machine/computer to achieve such a transformation, it must possess affective computing capacity. Without emotional intelligence, the machine/computer could not achieve true human-level intelligence. In 1997, Picard from the MIT published her seminal book on affective computing [26]. The main contents of her influential book include the following four parts: (1) Extracting human emotion information; (2) Modeling of emotions; (3) perception-based understanding through reasoning/inference; (4) representation of the outcome of the understanding. The general procedure of affective computing based on physiological signals is composed of the following three steps:

Step 1 – Feature extraction: Extract features from heterogeneous physiological signals from different sources (including EEG, EEG, ECG, galvanic skin response, respiration, pulse rate, etc.);

Step 2 – Emotion recognition: Recognition of the emotional state; and

Step 3 – Emotional regulation: Regulation/adjustment of the emotions through psychological measures.

Chen et al. identified the emotions of fear, calm and happiness by extracting the facial expressions and movements of the subjects, and the highest recognition rate reached 86.7%. Their study showed that the human brain's prefrontal cortex (PFC) is responsible for emotional regulation and perception. Through clinical trials it was found that PFC damage can lead to abnormal emotional function [27].

1.4 Categorization of emotions

The number of categories of emotions has always been controversial in psychology. Historically, psychologists have two different methods to model emotions: one is basic emotion theory that label emotions in discrete

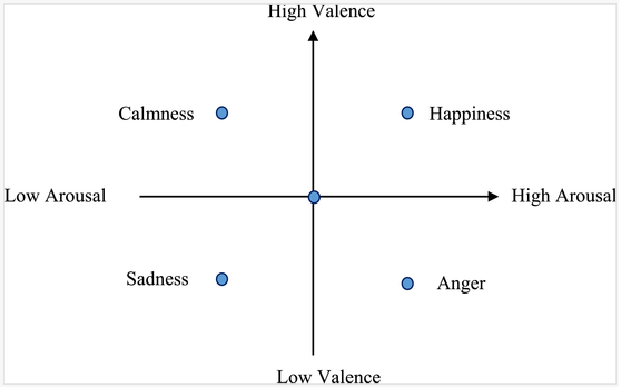
categories and the other is multi-dimensional theory that categorizes emotions on multiple dimensions or scales.

Basic emotion theory holds that there are several basic emotions in humans, e.g., happiness, sadness, fear, anger, disgust and surprise. Other nonbasic emotional states (such as fatigue, anxiety, satisfaction, confusion and frustration) are somehow composed of them. Each category of emotion has unique inner experience, external performance, and physiological patterns. Ekman believes that basic emotions must have the following characteristics: (1) Emotions come from instinct; (2) Different people produce the same emotion under the same situation; (3) Different people express basic emotions in a similar way; (4) When basic emotions are produced, the physiological patterns of different people are consistent. There are six basic emotions: anger, fear, sadness, disgust, surprise, and happiness. From these six basic emotions, other compound (nonbasic) emotions may be derived, such as shyness, guilt, contempt, etc.

The dimensional theory holds that human emotion has a multi-dimensional structure, and each dimension is a characteristic of emotion. Each emotional state can be placed as a point in a multi-dimensional space. Instead of choosing discrete labels or terms, one can indicate his/her impression on several continuous- or discrete-valued scales, e.g., pleasant-unpleasant, attention-rejection, etc. So far, researchers have proposed different multi-dimensional approaches for emotion modeling. Typical examples include (1) Russell's early circumplex model – 2 D emotional model which uses the dimensions of arousal and valence to plot 150 affective labels; (2) Whissell's continuous 2 D space whose dimensions are evaluation and activation; and (3) Schloberg's 3 D emotion model in which he added attention-rejection dimension to the 2 D model. The commonly used approach is Russell's 2 D emotion model, in which the horizontal axis is the arousal dimension (indicating the intensity of the emotional experience, from low to excitement), and the vertical axis is the valence dimension (indicating the degree of joy or cheerfulness, from negative to positive). Fig. 2 shows a schematic of the 2 D emotion model. The 2 D emotion model will be used to determine the target/real/actual emotion-related data labels as the ground truth in order to evaluate the emotion classification accuracy.

alt-text: Fig 2

Fig. 2



The 2-D emotion model.

1.5 Emotion elicitation paradigms

A critical step in emotion recognition based on physiological signals is to induce/trigger/evoke the emotional state of the experimental subject in certain proper ways, that is, the emotional arousal. There are three main ways to induce emotions:

(1) Evoking emotions by presenting the music, pictures, videos and other stimulating materials. This is a commonly used approach to inducing emotions. In order to induce the subjects to generate emotional states and tag/label them in more objective manner, Lang et al. developed the international emotional picture/sound systems [28,29]. They asked subjects of different ages, genders, and ethnicities to score the three dimensions of the valence, arousal, and dominance of each material in the system. This provides a basis for the tagging of the emotional categories induced.

(2) Inducing emotions by constructing simulated scenarios. People often produce some unforgettable emotions at some point in the past. It is also possible to induce emotions by letting the subjects recollect the fragments with different emotional colors in their past experience. However, the disadvantage of this method is that it cannot guarantee that the subject generates the corresponding emotion, and the duration of the corresponding emotion is not measurable.

(3) The subject is required to play a computer/video game or pretend to have certain facial expressions and gestures. The disadvantage of this method is that the participants might conceal their true emotions.

1.6 Main contribution and organization of this paper

This paper reviews the emotion recognition methods based on heterogeneous signals with the following major contributions:

1) The physiological data labeling approaches were surveyed.
2) Different feature extraction methods, such as wavelet transform and nonlinear dynamics, were reviewed.
3) Different feature dimensionality reduction algorithms, including Kernel Spectral Regression (KSR) discriminant analysis, Locality Preserving Projection (LPP), Principal Component Analysis (PCA), mRMR, and ReliefF algorithm, were reviewed.
4) Classification performance of k-Nearest Neighbor (KNN), Naïve Bayes (NB), Support Vector Machine (SVM) and Random Forest (RF) was compared.
5) Guidelines for combining feature extraction, feature reduction, and classification algorithms are suggested.
6) In order to select an optimal number of EEG electrodes, the emotion classification accuracy using the EEG electrodes placed on each brain area was compared to identify the most relevant brain areas.

The rest of the paper is organized as follows: Section 1 describes the background and motivations, an overview of the state-of-the-art emotion recognition techniques, different models of emotions, emotional induction paradigms, as well as application areas of emotion recognition techniques. Section 2 introduces the physiological basis of emotion generation and the role of each area of the brain in the formation of emotions, and analyzes the correlation between emotion and EEG signals. Section 3 considers the emotional databases and provides an overview of EEG signal processing techniques for emotion recognition problem. Section 4 compares several major feature extraction, reduction and classification methods. Section 5 is devoted to analysis of different brain regions that are relevant to emotions, with an aim to find the brain regions that are highly correlated to emotions. Finally, Section 6 summarizes the major findings of the paper and points out some open problems/challenges and future research directions in the field.

2 An overview of emotion and EEG signals

This section provides an overview of the emotion and EEG. Section 2.1 introduces the physiological basis of emotion generation and Papaz's fundamental theory of emotion. Section 2.2 describes the structure and function of the brain. The cerebral cortex is usually divided into four regions, each of which is responsible for different functions. Studies have shown that the prefrontal cortex (PFC) is most closely associated with emotion. Section 2.3 describes the characteristics of EEG signals. The last section describes the major techniques for EEG signal processing and analysis, including time-domain analysis, frequency-domain analysis, time-frequency analysis, and nonlinear dynamics.

2.1 Physiological differentiation of emotions

The accurate measurement of physiological indicators/markers is based on the endocrine changes and vegetative neural activities produced under emotional conditions. Psychophysiological studies have shown that subcortical activity is mainly regulated by the cerebral cortex, and the lower part of the cortex is the main area that is responsible for emotion generation.

Emotion is a psychophysiological process triggered by conscious and/or unconscious perception of an object, stimulus, or situation and is often associated with mood, temperament, sentiment, personality and disposition/inclination, and personal motivations. Emotions of humans can be expressed either verbally through emotional words or by nonverbal cues such as intonation of voice, facial expressions and body gestures/postures. Emotion is systematically produced by cognitive processes, subjective feelings, physiological arousal, motivational or action tendencies (also called *stance*), and behavioral reactions. The production of emotion involves the activities in many areas, including the thalamic system, reticular structure, limbic system and the subcutaneous ganglion and thus emotion has a complex central mechanism.

The debate on whether emotions can be discriminated by physiological changes is still under way in the communities of psychology and neurophysiology. A well-known hypothesis was made by James to support the antecedence of physiological specificity in emotional processes, but Cannon rejected the claim. In neurophysiology, the hypothesis can be reduced to the quest for the central circuitry of human emotions, that

is, to find the brain center in the central nervous system (CNS) and the neural center in the peripheral nervous system (PNS) which are both involved in emotional experience.

In 1884 James, the father of American psychology, argued that emotion is a feeling caused by bodily changes. Any category of emotion is accompanied by some changes in the body, such as muscle activity, facial expressions and visceral secretion. Later physiologist Lange proposed a similar view [30], which was called the James-Lange theory of emotions. This theory points out that there is an intrinsic relationship between emotion and physiological activities, but it is also incomplete to simplistically regard emotion as changes in activities of peripheral nervous system, which includes two parts: somatic nervous system and autonomic nervous system (ANS). In 1927 Cannon rejected the James-Lange theory. He believed that the hypothalamus determines the production of emotion. When people feel the stimulus, it is transmitted to the cerebral cortex, which then activates the hypothalamus and thereby produces different emotions. This is called the Cannon-Bade theory [31], in which the thalamus plays a major role in formation of emotions, but the relationship between emotion and PNS is completely disregarded. In 1937, Papez loop, the limbic system mechanism of emotion production, was proposed to link emotions to physiological activities [32]. First, the emotion originates from the hippocampus. After the hippocampus is stimulated, the impulse is transmitted to the hypothalamus, then the anterior nucleus and eventually return to the hippocampus through the cingulates. It is in this process that the emotion is produced. Based on the Papez theory of emotion, Maclean proposed the concept of visceral brain. He believed that the visceral brain is responsible for regulating or modulating all internal organs related to emotion and the hypothalamus is responsible for mediating related bone and visceral responses [33].

Although the definition and generation mechanism of emotion is still controversial, there is no doubt that human emotions are accompanied by physiological changes and are associated with activities of the physiological ANS activity. This provides a neurophysiological basis for recognizing emotions from EEG signals.

2.2 Structure and functionalities of brain

The human brain is divided into three major parts, the cerebrum, cerebellum and the brainstem. The cerebrum consists of the cerebral cortex, the limbic system and the brain nucleus. The cerebral cortex is primarily responsible for the higher-level emotional and cognitive functions. It is located at the outermost layer of the human brain, with a thickness of about 1-4 mm, mainly composed of grey matter, below which is mainly white matter [34]. A central sulcus in the middle of the brain divides it into left and right hemispheres. The brain can be divided into four areas: the Frontal Lobe, the Parietal Lobe, the Occipital Lobe, and the Temporal Lobe. These four areas of the brain have the following different functions:

(1) Frontal lobe: The frontal lobe is located before the central sulcus of the brain. It is responsible for higher cognitive functions. It includes the prefrontal lobe, the primary motion area, and the frontal motion area. The main functions include abstract thinking, inference, judgment, conception, and motion control. They are mainly responsible for planning, thinking, and physiological functions related to individual's emotions and needs.

(2) Parietal lobe: The parietal lobe is located behind the central sulcus and before the occipital

fissure. It is a high-level sensory center. It is mainly responsible for the response to the sensory and spatial information of pain, temperature, pressure, touch, and taste, as well as the integration of somatosensory information. This area is also related to mathematical and logical reasoning.

(3) Occipital lobe: The occipital lobe is located at the back of the hemisphere, behind the occipital sulcus, and is primarily responsible for processing information related to vision. In addition, it is related to individual's memory, behavioral perception, and abstract concepts.

(4) Temporal lobe: The temporal lobe is located under the lateral fissure, with the frontal lobe in front, the parietal lobe above, and the occipital lobe in the back. It is mainly responsible for the processing of auditory information, and is also related to emotion and memory.

2.3 EEG signal processing

2.3.1 EEG signals

Electroencephalogram (EEG), also known as brain wave, is one of the effective tools for monitoring brain activity. In 1929, the Austrian psychiatrist Berger initiated the recording of human EEG and then published the first paper on human EEG [35]. After that, electrophysiologists and neurophysiologists gradually confirmed his research results, making rapid development of EEG research in brain science and clinical medicine. By analyzing the EEG signals, one can understand the changes in emotion. Neuronal potentials can reflect the functional and physiological changes of the central nervous system (CNS). EEG is currently the most sensitive method for monitoring brain function. However, the EEG does not simply reflect the activity of a certain neuron, but reflects the electrical activity of a population of neurons in the brain region where the EEG measurement electrode is placed. Therefore, the EEG signal contains a lot of meaningful and useful psychophysiological information. In medicine, an objective basis for diagnosing certain diseases can be provided through EEG signals classification, computing, and analysis. In neuroengineering, disabled people can control wheelchairs or robotic arms using the EEG signals generated by mind or motion imagery. This is currently a hot research field, called Brain-Computer Interface (BCI). Due to the non-stationarity of EEG signals and the complex environmental factors, the EEG signal processing and analysis is always challenging in brain research.

2.3.2 Characteristics of EEG signals

EEG signal is one of the most important physiological signals. It is a direct reflection of brain activity, and plays an important role in the study of physiological phenomena of human brain. It possesses the following main characteristics:

(1) Noisy: The EEG recordings are usually noisy and susceptible to environmental interferences. The EEG signal usually has a low amplitude (generally around 50 uV with the maximum of about 100 uV). The EEG signals are usually mixed with a number of other signals (such as EOG, EMG, and ECG), noises, interferences, or artifacts.

(2) Nonlinear: EEG signals can be divided into spontaneous and evoked signals. Spontaneous EEG

or evoked potentials are inevitably affected by other peripheral physiological signals during the signal acquisition. Physiological regulation or adaptation of human tissues makes EEG signals highly nonlinear.

(3) Nonstationary: The change of EEG signals is unstable, sensitive to the external environment factors, and exhibits strong non-stationarity property. Many studies use statistical analysis techniques to detect and identify features of EEG signals.

(4) Frequency-domain characteristics: The frequency range of EEG signals is generally 0.5-100 Hz, but the frequency band most relevant to cognition is the low frequency range of 0.5-30 Hz. Usually the researchers divide it into five frequency sub-bands, each corresponding to different cognitive function.

EEG signals are usually classified into two types: spontaneous and evoked. Spontaneous EEG is a rhythmic potential fluctuation produced by the nervous system without any external stimuli. The external stimulation/excitation of the human sensory organs leads to a detectable potential change in the cerebral cortex, called evoked potentials. Brain cells are rhythmically discharged, so the EEG signals are rhythmic. The EEG signal can be divided into five frequency bands, as illustrated in Fig. 3 [26,35].

(1) Delta wave (1-4 Hz): The signal amplitude is generally 20-200 μV, which usually occurs in the frontal cortex. When one is in a state of sleep, lack of oxygen, or being anesthetized, the delta wave is usually detected. When an adult is in a normal awake state, the wave would disappear.

(2) Theta wave (4-8 Hz): The signal amplitude is generally 100-150 μV. It usually appears in the temporal lobe and parietal lobe when one is relaxed, indicating that the CNS is in a state of inhibition. It is usually associated with working memory load and can help people with long-term and deep memory.

(3) Alpha wave (8-13 Hz): The signal amplitude is generally 20-100 μV. It mainly occurs in the parietal lobe and occipital lobe. When one is in a resting state with eyes closed, the alpha wave can be detected. Alpha waves can be significantly reduced or even disappeared under some external stimuli, such as auditory or visual stimuli, or when people are carrying out some mental activity. Alpha waves are usually thought to be involved in preparatory activities of brain.

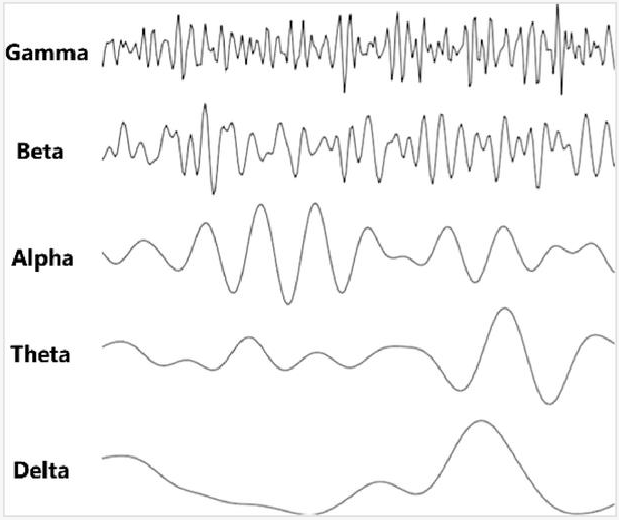
(4) Beta wave (13-30 Hz): The signal amplitude is generally 5-20 μV. When one is in a resting state with eyes closed, the beta wave is usually only detected in the frontal lobe; but when one is thinking, the beta wave appears in a wide range of areas. When the human body is in a relaxed state, alpha waves dominate the cerebral cortex, and this rhythm gradually disappears as the emotional activity becomes stronger. Conversely, when the CNS is in a state of tension/strain/stress, the amplitude of the Alpha wave is reduced while its frequency becomes higher, and the alpha wave gradually turns a Beta wave. Its appearance usually implies that the cerebral cortex is in an excited state.

(5) Gamma wave (>30 Hz): This is the high frequency component of the brain wave. The

amplitude is usually lower than 2 μV. Gamma waves play an important role in the cognitive activities of the brain. They are also related to high-level functions, such as reception, transmission, processing, integration, and feedback of information in the brainstem, and activities that require intense attention (concentration). The gamma wave is often found in the process of multi-modal sensory processing. Some studies have shown that the gamma waves can directly reflect the activity of the brain.

alt-text: Fig 3

Fig. 3



The waveforms of five typical EEG rhythms.

2.3.3 EEG signal processing methods

The procedure of emotion recognition based on EEG signals can be divided into the following steps: (1) emotional induction; (2) acquisition of EEG signals; (3) preprocessing of EEG signals; (4) extraction of EEG features; (5) EEG feature dimensionality reduction; (6) emotional pattern learning and classification.

EEG signals contain a wealth of information about psychophysiological activities, which is a direct indication of neuronal activity. However, the EEG signals are weak with small amplitude and highly noisy. An important issue is how to extract useful information from complex EEG signals. In 1932, Dietch first proposed the analysis of EEG signals by Fourier transform. After that, various modern signal processing methods have been used in the analysis of EEG signals [36]. Since 1980s more sophisticated EEG signal analysis methods have been proposed, opening up new opportunities for in-depth research on EEG signals. The EEG feature extraction methods can be roughly divided into classical and modern analysis methods. Classical analysis methods include time-domain analysis, frequency-domain analysis, and bi-spectral analysis. Modern analysis methods include time-frequency multi-resolution analysis, matching and tracking methods, and nonlinear dynamics (such as chaos, fractal, complexity, and entropy theory). In the following, we will briefly introduce those analysis methods:

A. Time-domain analysis

It is an early analysis method by exploiting the statistics of EEG signals (such as mean, variance, amplitude, skewness and kurtosis). The commonly-used time-domain analysis methods include zero-crossing analysis, analysis of variance, correlation, histogram, waveform recognition, etc. Since the time-domain waveform contains all the information of EEG without loss of information, the time-domain analysis methods was developed first. Because it is intuitive and easy to interpret, many researchers are still using the time-domain analysis. However, due to the complexity of the EEG signals, there is no particularly effective time-domain waveform analysis method.

B. Frequency-domain analysis

The frequency-domain analysis method assumes that the EEG signal is stationary, and only considers the frequency-domain information of the EEG signal while ignoring the time information. One of the primary frequency-domain analysis is power spectrum estimation. It is a spectrogram that changes the amplitude of the EEG over time into the power of the EEG as a function of frequency, so that the EEG rhythms can be observed intuitively.

Power spectrum estimation methods can usually be divided into classical and modern methods. Classical spectral estimation method estimates the power spectrum of an EEG signal by using Fourier transform and time window. The periodogram method is the simplest method. In modern analytical methods, the most widely used spectral estimation method is based on parametric model. Firstly, according to some a priori information or some assumptions about EEG signals, a stochastic parametric model is determined, and then the autocorrelation delayed sequence or sampling sequence is used to estimate the parameters of the model. Finally, the model identified is used to calculate the power spectrum of the EEG signal. The AR, MA and ARMA models are commonly used in modern spectral estimation. However, the spectral estimation requires the assumption that the EEG signal is stationary. However, the measured EEG signal is typically non-stationary in nature, so the spectral analysis must be based on the partition of the EEG signal into several quasi-stationary segments. The common frequency-domain features include power spectrum, power spectral density, and energy.

C. Time-frequency analysis

Much of the information in EEG signals is contained in their time-domain waveforms. However, it is difficult to extract the useful features directly using waveform analysis. In many situations, it is easier to examine EEG signals in the frequency domain. However, the local property of the EEG signal cannot be obtained only from its frequency-domain characteristics. Therefore, for EEG signals, some features cannot be extracted by simple time- or frequency-domain analysis. The time-frequency analysis method has better localized analysis and other important properties in both the time and frequency domain. The commonly used methods for time-frequency analysis are short-time Fourier transform and wavelet transform.

In short-time Fourier transform, the composition and phases of the local sine waves of the time-varying signal are obtained by moving the window function. The resolution of the short-time Fourier transform depends on

the window function selected to use. For a quasi-stationary or a piecewise stationary time series, the short-time Fourier transform usually leads to satisfactory analysis results. Nevertheless, for the non-stationary EEG signal, high temporal resolution is required. The short-time Fourier transform cannot balance frequency and time resolution and wavelet transform is more suitable for non-stationary EEG signals.

In 1984 Grossman and Morlet formulated Wavelet Transform (WT). In 1987 Mallet introduced the idea of multi-scale analysis in wavelet analysis, including the construction of wavelet functions and the decomposition and reconstruction of signals by WT. After selecting the appropriate base wavelet, the orthogonal function generates wavelets through binary translation and scaling. Wavelet analysis actually uses a fast-decaying, finite-length mother wavelet to represent the signal, which is translated and scaled to match the input signal. WT is divided into two categories: Discrete Wavelet Transform (DWT) and Continuous Wavelet Transform (CWT). Discrete transform uses a specific subset of all scaling and translation values, while continuous transformation operates on all possible scaling and translation transformations. The mathematical principles of wavelet analysis will be described in more detail in Sect. 3.

D. Nonlinear dynamics

EEG signals are a mixture of neuronal activities in the brain, characterized by complexity and irregularities, so it is difficult to analyze EEG signals using only traditional analytical methods. In recent years, some research has shown that the human brain is a nonlinear dynamic system, and EEG signals can be considered as the output of such a system. Therefore, researchers are trying to analyze EEG signals using nonlinear dynamics. In general, the methods of nonlinear dynamics can be divided into two categories: one is chaos theory and the other is information theory. The methods based on chaos theory include correlation dimension, Lorenz scatter plot, Lyapunov exponent, and Hurst exponent. Information theory based methods include Approximate Entropy (APoEn), Sample Entropy (SampEn), Permutation Entropy (PeEn), and complexity. The nonlinear dynamic analysis of EEG signals can provide information that cannot be obtained by conventional EEG analysis methods, which is reproducible and insensitive to the impact of outliers or artifacts in the EEG time series.

3 EEG-based emotion recognition

This section introduces the emotion datasets, EEG feature extraction, feature dimensionality reduction, and classification methods. Section 3.1 introduces the benchmark DEAP database used in various emotion recognition studies and describes the methods for EEG data preprocessing and the determination of the target/actual emotion classes. Section 3.2 introduces two EEG feature extraction methods: wavelet transform and nonlinear dynamics. Section 3.3 introduces several EEG feature reduction/selection algorithms, including Kernel Regression Discriminant Analysis (KSR), Local Preserving Projection (LPP), ReliefF, and minimal-Redundancy-Maximal-Relevance (mRMR) algorithms. Finally, we briefly describe two ML-based emotion classifiers: support vector machine (SVM) and random forest (RF).

3.1 Datasets and data preprocessing

3.1.1 Available physiological datasets

In this section, we describe widely-used datasets for multimodal emotion recognition, especially emotion recognition from physiological signals.

Recent advances in emotion recognition have motivated the creation of datasets containing emotional expressions in different modalities. Most datasets contain speech (acoustic), visual data, or audio-visual data (e.g., [37-41]). The audio modality covers genuine or posed emotional speech in different languages. The visual modality includes facial expressions and/or body gestures (or postures).

Healey [42,43] recorded one of the first affective physiological datasets. She recorded 24 participants driving around Boston area and annotated the dataset by the drivers' stress level. Responses of 17 out of the 24 participants are publicly available (http://www.phsyionet.org/pn3/drivedb/). Her recordings include ECG, galvanic skin response (GSR), EMG, and respiration patterns.

A publicly available multimodal emotional datasets which include both physiological responses and facial expressions are the enterface 2005 emotional database and MAHNOB HCI [44,45]. The database recorded by Savran et al. [45] includes two sets. The first one has EEG, peripheral physiological signals, functional near infrared spectroscopy (fNIRS) and facial videos from five male participants. The 2nd dataset only has fNIRS and facial videos from 16 participants of both genders. Both datasets recorded spontaneous emotional responses to images from the International Affective Picture System (IAPS) [46].The MAHNOB HCI dataset [44] consists of two experiments. The emotional responses including EEG, physiological signals, eye gaze, audio, and facial expressions of 30 people were measured.

There has been a large number of publications in the area of emotion recognition from physiological signals [42,47-51]. There are also various studies on music emotion characterization from acoustic features [52-54]. In the study [55], EEG and physiological signals of six participants were recorded as each watched 20 music videos. The participants rated arousal and valence dimensions and the EEG and physiological signals for each video were classified into low/high A/V classes. Some datasets used in recent literature are summarized in Table 1, in which the emotions in all papers were *induced*, except for Ref. [60] (*natural* emotions).

alt-text: Table 1

Table 1

The presentation of Tables and the formatting of text in the online proof do not match the final output, though the data is the same. To preview the actual presentation, view the Proof.

Modality of datasets for emotion recognition used in recent literature.

Data modality	Ref.
CP + *C* + PP	Koelstra et al. (2012) [11]
A + PP	Kim (2007) [56]
V + PP	Bailenson et al. (2008) [57]

CP + PP	Khalali and Moradi (2009) [58]
A + PP	Kim and Lingenfelser (2010) [59]
CP + PP	Chanel et al. (2011) [60]
A + PP	Walter et al. (2011) [61]
V + PP	Hussain et al. (2012) [62]
V + PP	Monkaresi et al. (2012) [63]
CP + Gaze	Soleymani et al. (2012) [64]
CP + *C*	Wang et al. (2014) [65]

Legenda: *V* = Video, *A* = Audio, *C* = Content/Context, CP = Central Physiology, PP = Peripheral Physiology; Gaze = eye Gaze.

3.1.2 Benchmark DEAP emotion dataset

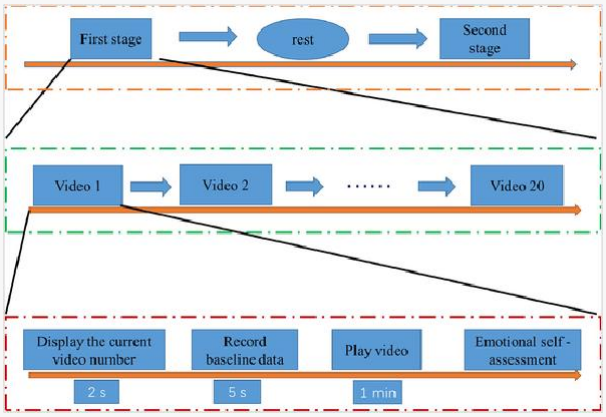
In this section, the DEAP emotion database will be described. Based on the 2 D model of emotions, Koelstra, Muhl and Soleymani et al. [63] used 40 music video clips as the visual stimuli to elicit different emotions, and 32 subjects (half of them were male, half were female; aged between 19 and 37 with an average age of 26.9 y/o) watched 40 1 m highlight of music videos while their EEG and peripheral physiological signals and facial expressions were recorded. The advantages of using music video as emotional stimuli include: (i) It can express emotions consistently and continuously; and (ii) It is easier to induce the emotional state of the subjects, without requiring the subjects to induce emotions by watching the slowly developed plots in movie clips.

There were 40 channels of physiological signals, including 32-channel EEG signals and 8-channel peripheral physiological signals (such as galvanic skin response, respiratory signal, skin temperature, ECG, blood volume, EMG, and EOG). More specifically, for 22 participants, frontal face video was recorded. Also included is the subject ratings from the initial online subjective annotation and the list of 120 videos used. EEG was recorded at a sampling rate of 512 Hz using 32 electrodes. Thirteen peripheral physiological signals were also recorded.

The experiment started with a 2 m baseline recording, during which a fixation cross was displayed to the participant who was asked to relax during this baseline period. Then the 40 music videos were presented in 40 trials. After 20 trials (videos), the participants took a short break. Then the experimenter checked the signal quality and the electrodes placement and asked the participants to continue the 2nd stage of the experiment. The experimental procedure is shown in Fig. 4, where the top panel represents that each experiment consisted of two stages with a short break in between, the middle panel represents that during each stage the subject was asked to watch 20 music video clips in order to elicit his/her emotional states, and the bottom panel represents that during watching each music video the video number was displayed on the monitor for 2 s, a 5 s baseline physiological data was recorded, the video clip was then played for 1 min for the subject to watch, finally the subject was asked to complete self-assessment of his/her emotional states while watching the 1 min video clip. This challenging benchmark database can be used to compare different affect recognition methods.

alt-text: Fig 4

Fig. 4



The procedure of emotion elicitation experiment.

3.1.3 Data preprocessing

EEG signals are generated by the CNS and suited for emotion recognition. Therefore, EEG signals have been extensively used for emotion classification and analysis. The EEG signals are usually down-sampled during data preprocessing step. Then the EOG artifact is removed by using a band-pass filter.

The pre-processed EEG data includes emotion-related and baseline (emotionless) EEG data. In addition, there are significant individual differences (i.e., subject-to-subject variability) in the physiological signals. Even for the same subject and the same stimulus material, different emotions may be triggered at different times and/or under different environments. Therefore, in order to minimize the influence of the previous stimulus material on the subsequent emotional state and the influence of individual differences in physiological signals, the baseline EEG features (prior to the emotional material stimulation) were subtracted from the EEG features after the emotional material stimulation. Finally, the resultant residual features are normalized into the unit interval [0, 1].

A major problem in emotion recognition research is that there are individual differences in subjective emotional experiences for the same stimulus material. Therefore, the number of emotion classes in most studies is usually small. In the study of DEAP emotion recognition, many studies focus on the binary (positive vs. negative or high vs. low arousal) classification problems [66-69], and the target emotional labels are usually obtained by simple, and subjective hard threshold of subjective rating/scoring data of the subjects.

3.1.4 Clustering and validation

K-means clustering of the subjective scoring data can be used to obtain the target classes of emotion. When k-means clustering algorithm is used, the initial clusters must be set mainly based on two considerations: (1) whether the target classes obtained by data clustering can be reasonably explained by the two-dimensional emotion model, i.e., the clusters obtained can be found on the 2 D emotion plane; (2) two indices are used to

evaluate the clustering performance. Since the true labels are unknown, the Silhouette coefficient and Calinski-Harabasz index are used.

The silhouette coefficient applies when the actual clusters are unknown. For a sample, the Silhouette coefficient is defined by [70]:

$$S = \frac{b - a}{\max(a, b)} \tag{1}$$

Where represents the average distance of the sample from other samples in the same cluster, represents the average distance of the sample from all samples in the closest (different) cluster, S denotes a measure of the clustering quality. Generally, the larger the S, the higher the clustering quality.

The Calinski-Harabasz index can also be used in the case that the true clusters are unknown [71]. The Calinski-Harabasz index is calculated by:

$$s_K = \frac{N - K}{K - 1} \cdot \frac{Tr(B_k)}{Tr(W_k)} \tag{2}$$

$$W_K = \sum_{q=1}^{K} \sum_{x \in \mathbf{c}_q} (\mathbf{x} - \mathbf{c}_q)(\mathbf{x} - \mathbf{c}_q)^T \tag{3}$$

$$B_K = \sum_{q} N_q (\mathbf{c}_q - \mathbf{c})(\mathbf{c}_q - \mathbf{c})^T \tag{4}$$

where$Tr(\cdot)$ denotes the trace of a matrix,B_k represents covariance matrix between different clusters, W_k represents covariance matrix within the same cluster, N is the total number of samples, K is the number of clusters, $\mathbf{c}_q$ is the set of samples in the q-th cluster, N_q represents the number of samples in the q-th cluster.

The above formulae show that the larger the covariance between the clusters, the smaller the covariance within the cluster, and the larger the Calinski-Harabasz index, the higher the clustering quality.

3.2 EEG feature extraction

The main purpose of feature extraction is to extract the information from EEG signals that can significantly reflect the emotional state, which can be further used for the emotion recognition/classification algorithms. The features extracted determine, to a large extent, the accuracy of emotion recognition. Therefore, it is crucial to extract the salient EEG features of emotional state. In this section, two methods of EEG feature extraction

are reviewed: wavelet transform (time-frequency analysis) and approximate entropy and sample entropy (nonlinear dynamics).

3.2.1 Wavelet transform

Wavelet decomposition is a typical and practicable time-frequency analysis method. It is a localized analysis method based on time window and frequency window. The EEG signal is non-stationary and is characterized by slow change of the lower-frequency components and fast variability of the higher-frequency components, so wavelet transform is ideally suited to its signal analysis. The multi-scale analysis of EEG signals using wavelet transform allows for the EEG signal to exhibit both details and approximations at different wavelet scales. By wavelet decomposition of EEG signals, a series of wavelet coefficients can be obtained at different scales. These coefficients can completely describe the characteristics of the signal and thus can be used as a feature set of the signal. In general, the wavelet function $\psi(\cdot)$ is defined by:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{5}$$

Where b and a represent the time-shift and scale factor, respectively.

The wavelet transform has the following characteristics: (1) multi-resolution; (2) constant relative bandwidth; (3) by selecting appropriate base wavelets, the local characteristics of the target signals can be represented in both the time and frequency domains. The wavelet function is like a band-pass filter. The original EEG signal is decomposed into different scales by high- and low-pass filters through dilation and contraction transformations. The wavelet coefficients obtained after high-pass filtering are details (D), while the wavelet coefficient obtained after low-pass filtering is the approximations (A). The approximate component is then further decomposed into a detail component and an approximate component.

For a given signal $x(t)$, its wavelet decomposition can be expressed by:

$$x(t) = \sum_{k=-\infty}^{\infty} C_{N,k}\phi\left(2^{-N}t-k\right) + \sum_{j=1}^{N}\sum_{k=-\infty}^{\infty} D_{j,k}2^{-j/2}\psi\left(2^{-j}t-k\right) \tag{6}$$

Where $C_{N,k}$ denotes the k-th approximate component of the N-th level of wavelet decomposition, $D_{j,k}(j=1,2,\cdots,N)$ represents the k-th detail component of the j-th level of the wavelet decomposition, $\psi(t)$represents wavelet function, and $\phi(t)$represents the dilation/contraction coefficient.

Compared with the short-time Fourier transform, there are many types of wavelet base functions, which have different properties and scopes of applicability, in the wavelet transform. Commonly used wavelet bases include Daubechies wavelets, Meyer wavelet, Morlet mother wavelet and Haar mother wavelet. However,

because different wavelets have different properties in terms of symmetry, smoothness, orthogonality and compact support, it is difficult to construct a wavelet function that possesses all the four properties.

Considering the orthogonality and compact support of the Daubechies wavelet and the nearly optimal localization of the Db4 wavelet base [72], this function is used as the wavelet basis function to decompose the EEG signal into five levels. Thereby the frequency components of the EEG signal in five frequency bands can be extracted.

From Table 2, it can be seen that the frequency bands of the EEG signal obtained by five-level wavelet decomposition is in good agreement with the known rhythms of the EEG signal. Fig. 5 illustrates the five-level wavelet decomposition of EEG signal into approximate signals (A1-5 in higher frequency bands) and details (D1-5 in lower frequency bands). For example, A2 and D2 result from the further decomposition of the approximate signal A1. The specific frequency range of each signal (A1-5, D1-5) is given in Table 2.

alt-text: Table 2

Table 2

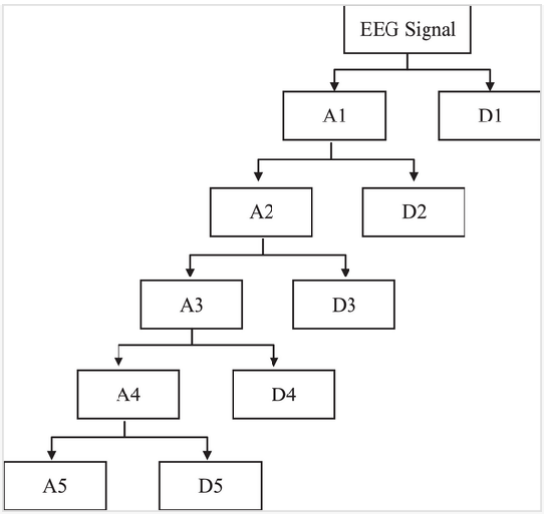
The presentation of Tables and the formatting of text in the online proof do not match the final output, though the data is the same. To preview the actual presentation, view the Proof.

Wavelet decomposition of EEG

Decomposition level	Frequency range (Hz)	EEG rhythm	Frequency range (Hz)
D1	32-64	Gamma	30-50
D2	16-32	Beta	13-30
D3	8-16	Alpha	8-13
D4	4-8	Theta	4-8
A5, D5	0-4	Delta	0.5-4

alt-text: Fig 5

Fig. 5



Schematic of five-level wavelet decomposition of EEG signal.

For EEG signal from each channel, three features are derived from the wavelet coefficients of each sub-band, including wavelet energy, wavelet energy ratio (the ratio of each sub-band energy in the total energy of all sub-bands), and wavelet entropy which are defined as follows:

(1) wavelet energy

$$E(i) = \sum_{j=1}^{n_i} D_{i,j}^2 \quad (7)$$

(2) wavelet energy ratio

$$R(i) = E(i) / \sum_{j=1}^{n} E(j) \quad (8)$$

(3) wavelet entropy

$$W_e = \sum_{i=1}^{n} R(i) \ln R(i) \quad (9)$$

where $D_{i,j}$ represents wavelet coefficients of the corresponding decomposition levels, as introduced in (6).

3.2.2 Nonlinear dynamics

EEG signals are highly complex and nonlinear. In recent years, nonlinear analysis (e.g., entropy and other complexity measures) has been widely used in the analysis of EEG signals [73-75]. Among them, two nonlinear dynamic methods, approximate entropy and sample entropy, are important tools for quantifying the complexity of time series. This section mainly studies the effectiveness of the two nonlinear dynamic methods in extracting EEG features.

A. Approximate entropy

Approximate entropy (ApEn) was proposed by Pincus [76] and can be used to describe the irregularity of complex systems. If the time series is more irregular, the corresponding ApEn is larger. The ApEn indicates the probability of the generation of new pattern when the dimension of a time series increases from m to $m+1$. The larger the ApEn, the greater the probability of generating a new pattern, implying the more complex the time series. Usually only a short data segment is required to obtain a robust estimation of ApEn. What the ApEn actually reflects is the degree of self-similarity of the patterns corresponding to a time series, that is, the probability of generating a new pattern when the dimension of the time series varies. Compared with other nonlinear dynamic measures, such as the Lyapunov exponent and correlation dimension, ApEn has the following characteristics:

(1) A robust estimate can be obtained by using a relatively shorter data segment. It can be estimated by taking 100–5000 data points when applied to biosignals.

(2) Strong resistance to transient interference.

(3) It can be applied to many types of signals, such as stochastic signals, deterministic signals, or mixed signals comprising both.

The specific algorithmic steps for estimating ApEn are as follows:

Step 1: Given time series of length n, $\{x(i), i = 1, 2, \cdots, N\}$, construct in turn the m-dimensional vector as:

$$\mathbf{x}_m(i) = [x(i), x(i+1), \cdots, x(i+m-1)], 1 \le i \le N-m+1 \tag{10}$$

Step 2: Compute the distance between vectors $\mathbf{x}_m(i)$ and $\mathbf{x}_m(j)$ and define the maximum distance between each component as the distance of maximum contribution:

$$d\left[\mathbf{x}_m(i), \mathbf{x}_m(j)\right] = \max\left\{\left|\mathbf{x}_m(i+k) - \mathbf{x}_m(j+k)\right|\right\} \tag{11}$$

where $1 \le k \le m-1; 1 \le i, j \le N-m+1, i \ne j$.

Step 3: Given a positive threshold $r > 0$ and the embedding dimensionality m, the regularity probability of the time series X can be obtained by:

$$C_i^m(r) = N^m(i) / (N - m + 1) \tag{12}$$

where $N^m(i)$ denotes the number of $d[\mathbf{x}_m(i), \mathbf{x}_m(j)] \le r$.

Step 4: For each $C_i^m(r)$, calculate its logarithmic average:

$$\phi^m(r) = \left(\sum_{i=1}^{N-m+1} \ln C_i^m(r) \right) / (N - m + 1) \tag{13}$$

Step 5: For each embedding dimensionality $m + 1$, repeat Step 1–4 to obtain $\phi^{m+1}(r)$. Finally, the ApEn can be defined by:

$$ApEn(m, r, N) = \phi^m(r) - \phi^{m+1}(r) \tag{14}$$

The values of *M*, *r*, *N* determine the value of the approximate entropy. We usually take ~~*m*~~*m* = 2 and ~~*r*~~*r* = 0.1–0.2 * SD_X , where SD_X represents the standard deviation of the original data, to have better statistical properties of ApEn.

B. Sample entropy

Sample Entropy (SampEn) is a complexity measure of time series based on improved ApEn, which was proposed by Richman et al. [77]. The ApEn has the following disadvantages: (1) It involves the comparison of its own data segments in the calculation, so it is biased; (2) The consistency of the ApEn results is poor.

The calculation of SampEn does not need to perform its own matching, thus in principle it is more accurate than the ApEn. Intuitively, when the time series is more complex, the SampEn is larger, and vice versa. SampEn has the following characteristics:

(1) It has good consistency.

(2) It requires less data points and does not require explicit granulation of the original data.

(3) It can be used to analyze mixed signal comprising deterministic and stochastic signals.

Because of the above characteristics, the SampEn is more suitable for EEG signal processing. The specific algorithmic steps for calculating the SampEn are as follows:

Step 1: Given the original data sequence $\{x(i), i = 1, 2, \cdots, N\}$, construct in turn the *m*-dimensional vector:

$$\mathbf{x}_m(i) = [x(i), x(i+1), \cdots, x(i+m-1)], 1 \le i \le N-m+1 \tag{15}$$

where m is the window length, also called embedding dimensionality.

Step 2: Define the distance between vectors $\mathbf{x}_m(i)$ and $\mathbf{x}_m(j)$ as:

$$d\left[\mathbf{x}_m(i), \mathbf{x}(j)\right] = \max\left\{\left|\mathbf{x}_m(i+k) - \mathbf{x}_m(j+k)\right|\right\} \tag{16}$$

where $1 \le k \le m-1; 1 \le i, j \le N-m+1, i \ne j$.

Step 3: Given a positive threshold r and embedding dimensionality m, calculate the probability of regularity of the time series $C_i^m(r)$ by:

$$C_i^m(r) = N^m(i) / (N-m+1) \tag{17}$$

where $i \le N-m$. Then calculate the average of all $C_i^m(r)$ by:

$$B^m(r) = \frac{1}{N-m}\sum_{i=1}^{N-m} C_i^m(r) \tag{18}$$

Step 4: Set the embedding dimensionality as $m+1$, repeat Steps 1-3, the SampEn can be computed by:

$$SampEn(m, r) = \lim_{N\to\infty}\left[-\ln\frac{B^{m+1}(r)}{B^m(r)}\right] \tag{19}$$

In practice N is taken a limited number, thus we have:

$$SampEn(m, r, N) = -\ln\frac{B^{m+1}(r)}{B^m(r)} \tag{20}$$

Likewise, the value of SampEn also depends on the selection of the values of m, r, and N. According to Pincus [76], to make the calculated SampEn have good statistical properties, the embedding dimension m can

be selected as 1 or 2 and *r* is generally chosen between 0.1**SD* and 0.25**SD* (*SD* is the standard deviation of the original time series).

3.3 EEG feature reduction and feature selection

Dimensionality reduction of EEG features is an important step in EEG-based emotion recognition. Selecting an effective feature reduction and selection algorithm can improve not only the efficiency of model training, but also the accuracy of model prediction. Feature reduction and selection is usually required to: (1) help with data visualization and understanding; (2) reduce the training time of the model; (3) overcome the curse of dimensionality, thereby improving the model prediction performance (or generalizability).

In this section we will introduce three dimensionality reduction algorithms (KSR, LPP, and PCA) and two feature selection algorithms (mRMR, Relief) on EEG features. PCA is a commonly used dimensionality reduction algorithm, which is generally used as a baseline for comparison with advanced dimensionality reduction and feature selection algorithms.

3.3.1 Graph embedding framework for dimensionality reduction

Suppose that we have dataset $\{\mathbf{x}_i\}_{i=1}^{m} \subset R^n$ containing *m* samples. The goal of a dimensionality reduction algorithm is to find a lower-dimensional representation of the dataset. Given a graph *G* with *m* vertices, each representing a data point. The data points are connected by a weighted edge, which can be represented by a *m* × *m* symmetric matrix *W*. Graph embedding actually represents the vertices in graph *G* by lower-dimensional vector, and describes the similarity between any two samples by the weight of the edge [78].

The one-dimensional mapping of $\mathbf{x} = [x_1, x_2, \cdots, x_m]$ is $\mathbf{y} = [y_1, y_2, \cdots, y_m]$, then the optimal $\mathbf{y}$ under certain constraints would minimize

$$\sum_{i,j} w_{ij} (y_i - y_j)^2 = 2\mathbf{y}^T (D - W)\mathbf{y} = 2\mathbf{y}^T L\mathbf{y} \tag{21}$$

where $L = D - W$ and D is a diagonal matrix with the elements on the main diagonal $d_{ii} = \sum_j w_{ji}$.

Then the above minimization problem can be rewritten as:

$$\mathbf{y}^* = \arg\left(\min_{\mathbf{y}^T D\mathbf{y}} \mathbf{y}^T L\mathbf{y}\right) = \arg\left(\min \frac{\mathbf{y}^T L\mathbf{y}}{\mathbf{y}^T D\mathbf{y}}\right) \tag{22}$$

Since $L = D - W$, the problem becomes equivalent to:

$$\mathbf{y}^* = \arg\left(\max_{\mathbf{y}^T D\mathbf{y}} \mathbf{y}^T W\mathbf{y}\right) = \arg\left(\max \frac{\mathbf{y}^T W\mathbf{y}}{\mathbf{y}^T D\mathbf{y}}\right) \tag{23}$$

Now the solution of the above optimization problem can be obtained by solving the following maximum eigenvector problem:

$$W\mathbf{y} = \lambda D\mathbf{y} \tag{24}$$

In order to obtain the mapping of all training and test samples, we select the linear function $y_i = f(\mathbf{x}_i) = \mathbf{a}^T\mathbf{x}_i$, i.e. $\mathbf{y} = X^T\mathbf{a}$, then the above formula can be rewritten as:

$$\mathbf{a}^* = \arg\left[\max_{\mathbf{a}} \frac{\mathbf{a}^T XWX^T\mathbf{a}}{\mathbf{a}^T XDX^T\mathbf{a}}\right] \tag{25}$$

The optimal **a** can be obtained by solving the following maximum eigenvalue problem:

$$XWX^T\mathbf{a} = \lambda XDX^T\mathbf{a} \tag{26}$$

and

$$KWK^T\alpha = \lambda KDK^T\alpha \tag{27}$$

This method is called Linear extension of Graph Embedding (LGE). Through selecting different *W*, different subspace learning algorithms result, such as linear discriminant analysis (LDA) and locality preserving projection (LPP). Nevertheless, a common drawback of these algorithms is that the eigendecomposition of the redundant matrix is computationally expensive. In LDA algorithm, assume that we have *c* classes, the *s*-class contains m_s samples, thus the total number of samples is $m = \sum_{s=1}^{c} m_s$. Then we define:

$$w_{ij} = \begin{cases} 1/m_s, & \text{if} \mathbf{x}_i \text{and} \mathbf{x}_j \text{belong to the same} s-\text{th class} \\ 0, & \text{otherwise} \end{cases} \tag{28}$$

Then it is easy to check $d_{ii} = \sum_j w_{ji} = \frac{1}{m_s} \cdot m_s = 1$ and $D = \mathbf{I}_m$ (a *m*-dimensional identity matrix).

In LPP algorithm, $N_k(\mathbf{x}_i)$represents the k-neighborhood set of$\mathbf{x}_i$. We define:

$$w_{ij}=\begin{cases} e^{-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|^2}{2\sigma^2}}, & \text{if} \mathbf{x}_i \in N_k(\mathbf{x}_j) \text{ or} \mathbf{x}_j \in N_k(\mathbf{x}_i) \\ 0, & \text{otherwise} \end{cases} \tag{29}$$

3.3.2 Spectral regression kernel discriminant analysis (SRKDA)

When dealing with massive data, the spectral regression algorithm is very effective.

Theorem 1 [79]: Let $y$and λbe the eigen-vector and eigen-value, respectively of the problem (24). If $X^T\mathbf{a}=y$, the eigen-vector and eigen-value of the problem (26) are **a** and λrespectively. If $K\alpha=y$, the eigen-vector and eigen-value of the problem (27) are α and λrespectively.

The above theorem gives the solution to problem (26). The linear embedding function can be obtained by the following two steps:

Step 1: Solve the problem (24) to get the eigen-vector **y**.

Step 2: Find **a**such that $X^T\mathbf{a}=\mathbf{y}$. However, **a** may not exist, hence a feasible way is to find **a** that satisfies the least squares formula:

$$\mathbf{a}^*=\arg\left[\min_{\mathbf{a}}\sum_{i=1}^{m}(\mathbf{a}^T\bar{\mathbf{x}}_i-\bar{y}_i)\right] \tag{30}$$

wherey_i is the i th component of **y**.

The advantages of the above two-step procedure are:

(1) Since D is positive definite, stable solution exists for the problem (24). In addition, L and D are both sparse matrices.

(2) There exists standard method to solve the LS problem [80]. In the case that there are more features than samples, (30) is ill-defined, thus $X^T\mathbf{a}=\mathbf{y}$ has an infinite number of solutions. To overcome this problem, a common method is to introduce regularization as follows:

$$\mathbf{a}^*=\arg\left[\min_{\mathbf{a}}\sum_{i=1}^{m}(\mathbf{a}^T\bar{\mathbf{x}}_i-\bar{y}_i)^2+\alpha\|\mathbf{a}\|^2\right] \tag{31}$$

where α is called scale parameter. The regularized LS is also known as ridge regression [81].

(3) The regression model allows the regularization technique to be well integrated. Even with many features, a stable and meaningful solution can be obtained. If we replace linear regression with spectral regression, an embedding function can be obtained in the reproducing kernel Hilbert space (RKHS). This algorithm performs data regression after the graphical spectral analysis, so it is called spectral regression (SR).

Given the labeled samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_l$ and unlabeled samples $\mathbf{x}_{l+1}, \mathbf{x}_{l+2}, \cdots, \mathbf{x}_m$ in real space $\mathbb{R}^n$. Assume that there are c classes of samples and l_k represents the number of samples in the k-th class ($\sum_{k=1}^{c} l_k = l$). SR algorithm consists of the following computational steps:

(1) Create adjacency graph

Let G be a graph with M vertices and the i th point $\mathbf{x}_i$. G can be created by the following two steps:

Step1: If $\mathbf{x}_i$ is the p-nearest neighbor of $\mathbf{x}_j$, the two points are connected.

Step2: If $\mathbf{x}_i$ and$\mathbf{x}_j$ have the same label, the two points are connected; Otherwise, their connection is removed.

(2) *Select the weight function W* that is a $m \times \times m$ sparse matrix with w_{ij} representing the connection weight between $\mathbf{x}_i$ and$\mathbf{x}_j$. If there is no edge between them, the weight is set zero; otherwise, we have:

$$w_{ij} = \begin{cases} 1/l_k, if\, \mathbf{x}_i and \mathbf{x}_j belong to the same lek - th class \\ \delta \cdot s(i,j), \text{otherwise} \end{cases} \tag{32}$$

where $0 \le \delta \le 1$ is a parameter that evaluates the neighborhood information and $s(i, j)$ is a preselected function used to describe the degree of similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$.

(3) Feature decomposition

Let the eigenvectors corresponding to the c maximum eigenvalues of problem (24) be $\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_{c-1}$.

(4) Regularized LS

Find (c-1) vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_{c-1} \in \mathbb{R}^n$ as the solution to the following regularization problem:

$$\mathbf{a}_k = \arg\min_{\mathbf{a}} \left[\sum_{i=1}^{l} \left(\mathbf{a}^T \mathbf{x}_i - y_i^k \right)^2 + \sum_{i=l+1}^{m} \left(\gamma \mathbf{a}^T \mathbf{x}_i - y_i^k \right)^2 + \alpha \|\mathbf{a}\|^2 \right] \tag{33}$$

where y_i^k is the i-th component of $\mathbf{y}_k$ and $\lambda \leq 1$is a parameter that is used to adjust the weights of unlabeled samples. Let $\bar{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_l, \gamma\mathbf{x}_{l+1}, \cdots, \gamma\mathbf{x}_m]$, then (33) can be rewritten as:

$$\mathbf{a}_k = \arg\min_{\mathbf{a}} \left[\left\| \bar{X}^T \mathbf{a}^T - \mathbf{y}_k \right\|^2 + \alpha \|\mathbf{a}\|^2 \right] \tag{34}$$

$\mathbf{a}_k$ is the solution to the following system of linear equations:

$$\left(\bar{X}\bar{X}^T + \alpha \mathbf{I} \right) \mathbf{a}_k = \bar{X}\mathbf{y}_k \tag{35}$$

(5) Regression discriminant analysis embedding

Let the $n \times (c-1)$transformation matrix $A = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_{c-1}]$, the samples can be embedded in the (c-1)-dimensional subspace where

$$\mathbf{x} \to \mathbf{z} = A^T \mathbf{x} \tag{36}$$

If the embedding function is obtained in RKHS, **Step4** can be modified to the following form.

(6) Regularized kernel LS

Find (c-1) vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_{c-1} \in \mathbb{R}^n$ as the solution to the system of equations as follows:

$$(K + \alpha I)\, \boldsymbol{\alpha}_k = \mathbf{y}_k \tag{37}$$

where $K$is$m \times m$Cramer matrix. It is easy to prove that the function $f(\mathbf{x}) = \sum_{i=1}^m \alpha_i^k K(\mathbf{x}, \mathbf{x}_i)$ where $\alpha_i^k$is the i th component of $\boldsymbol{\alpha}_k$, is the solution to the following optimization problem:

$$\min_{f \in H_k} \sum_{i=1}^m \left(f(\mathbf{x}_i) - y_i^k \right)^2 + \alpha \|f\|_K^2 \tag{38}$$

(7) Spectral regression discriminant analysis embedding

Let $m \times (c-1)$ transformation matrix $\Theta = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{c-1}]$, the samples can be embedded in the $(c-1)$-dimensional subspace in the following form:

$$\mathbf{x} \to \mathbf{z} = \Theta^T K(:, \mathbf{x}) \tag{39}$$

where $K(:, \mathbf{x}) = [K(\mathbf{x}_1, \mathbf{x}), \cdots, K(\mathbf{x}_m, \mathbf{x})]^T$.

3.3.3 RELIEF-F algorithm for feature selection

RELIEF is a multivariable feature selection algorithm, which is used to compute the weights of features based on sample learning [82]. The algorithm is mainly used for feature selection in two-class problem. The basic idea is to determine the importance of features according to their ability of distinguishing instances that are near each other. The more important feature should make samples in the same class closer and those in different classes farther apart. In other words, the RELIEF algorithm mainly measures the difference of features between within-class samples and cross-class samples, and then measures the distinguishing ability of the features. If a feature has a small difference between samples in the same class and a large difference in samples from different classes, it has better discriminant ability. For any sample in the dataset, the RELIEF algorithm first selects its two nearest neighbors: one from the same class (called *nearest hit*) and the other from different class (called *nearest miss*).

In RELIEF, the degree of importance of the feature f_i is estimated as the difference of two probabilities: the probability that the nearest instance from the same class takes different value of feature f_i minus the probability that the nearest instance from different class takes different value of f_i.

In 1994 Kononenko extended the RELIEF algorithm and proposed the RELIEF-F algorithm [83]. The algorithm extends the applicability of RELIEF from only two-class to multi-class problem by converting the latter into multiple one-to-many problems. For the multi-class problem, instead of finding one nearest neighbor from different class, the algorithm finds one nearest neighbor for each different class, and then evaluates the quality of a feature through its ability of differentiating the nearest neighbors from any two classes.

The pseudocode implementation of the RELIEF-F algorithm is as follows:

RELIEF-F algorithm
Inputs: Instance set S and the number of classes C
Output: Weight vector $\mathbf{w}$
Step 1: For any feature $f_a, a = 1, 2, \cdots, d$, set the initial weight $w(a) = 0$.

Step 2: for $i = 1$ to m do
Randomly select $\mathbf{x}_i$ from S;
Select the k-nearest neighbors $\mathbf{h}_j$ from the same class of $\mathbf{x}$;
Select the k-nearest neighbors $\mathbf{m}_j(c)$ from different class from $\mathbf{x}$.
for $a = 1$ to d do
Update the weight by:
$(a) \quad (a) - \sum_{j=1}^{k} \frac{dist(a,\mathbf{x}_i,\mathbf{h}_j)}{m \cdot k} + \sum_{c \neq class(\mathbf{x}_i)} \left\{ \frac{P(c)}{1-P[class(\mathbf{x}_i)]} \times \sum_{j=1}^{k} dist \right. \quad k)$
End
End

where $dist(a, \mathbf{x}, \mathbf{y})$ is the distance between instances $\mathbf{x}$ and $\mathbf{y}$ under the feature a, $P(c)$ denotes the probability of the c-th class which can be obtained as the ratio between the size of the c-th class and the total number of instances, $\mathbf{m}_j(c)$ denotes the j-th sample from the c-th class, m is the number of iterations, and k is the number of nearest neighbors.

3.3.4 minimal-Redundancy-Maximal-Relevance (mRMR) algorithm for feature selection

mRMR is a typical information-based feature selection algorithm [84]. The core idea of the algorithm is to find the m features in the feature space of the given samples, which have maximum relevance to the target class but minimal redundancy with other features. Mutual information is used to measure the relevance between features and target classes or other features in the feature space:

$$D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c) \tag{40}$$

$$R = \frac{1}{|S|} \sum_{x_i, x_j \in S} I(x_i, x_j) \tag{41}$$

where S is the feature set, c is the target class, $I(x_i; c)$ denotes mutual information between feature i and target class c, $I(x_i, x_j)$ denotes mutual information between feature i and j.

The relevance between feature subset S and target class c can be maximized by maximizing (40). Conversely, by minimizing (40) we can make the cross-relevance between features in S minimal.

Assume that we have two random variables x and y with p.d.f. and joint p.d.f. $p(x)$, $p(y)$, and $p(x, y)$, then their mutual information can be defined as:

$$I(x,y) = \int\int p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy \tag{42}$$

The mutual information between S_m and the target class c can be defined as:

$$I(S_m;c) = \int\int p(S_m,c) \log \frac{p(S_m,c)}{p(S_m)p(c)} dS_m dc \tag{43}$$

By combining (40) and (41), we can get the feature selection criterion for mRMR algorithm:

$$\max (D-R) \tag{44}$$

Based on this criterion, we can find the optimum feature subset by sequential forward search. Firstly, we obtain the feature that is most relevant to the target class and then add it to the set S_m . Other features are computed and added in the analogous manner. Suppose that there are already m features in the feature subset S_m , other features can be selected in the remaining sample set $\{S-S_m\}$ by:

$$\max_{x_i \in S-S_m} \left[I(x_i;c) - \frac{1}{m} \sum_{x_j \in S_m} I(x_i,x_j) \right] \tag{45}$$

3.4 Machine learning classifiers

The process of constructing the emotion recognition model includes data collection, emotion-related feature extraction, feature reduction, and classifier model building. After the first three steps are completed, the final task is to design an effective emotion classifier model based on certain classification performance criteria. The accuracy of the emotion recognition depends largely on the classifier developed [85]. In order to obtain accurate recognition and to validate the effectiveness of the feature extraction and dimensionality reduction algorithms, we will briefly describe in the following two types of ML-based classifiers, namely SVM and RF. SVM and RF outperform other types of ML classifiers in terms of 4-class emotion classification accuracy based on our recent empirical studies and performance comparison of different ML-based classifiers [36,86].

3.4.1 Support vector machine (SVM)

Given a labeled dataset, the task is to find a linear classification hyperplane $\mathbf{w}^T\mathbf{x}+b$ that satisfies:

$$y_i\left(\mathbf{w}^T\mathbf{x}_i+b\right)\geq 1, i=1,2,\cdots,N \tag{46}$$

where $\mathbf{x}_i$ denotes the feature vector and y_i denotes the corresponding actual label.

If the sample cannot be correctly classified with the linear classifier, the problem can be described by:

$$y_i\left(\mathbf{w}^T\mathbf{x}_i+b\right)\geq 1-\xi_i, i=1,2,\cdots,N \tag{47}$$

where the relaxation variable ξ_i indicates the allowed degree of deviation from the ideal linear separability condition.

Finding the optimal hyperplane among all subsets of linear classification hyperplane is equivalent to minimizing the following cost function:

$$\frac{1}{2}\mathbf{w}^T\mathbf{w}+C\sum_{i=1}^{N}\xi_i \tag{48}$$

where the constant $C > 0$ is a penalty factor that controls the degree of penalty to misclassification of samples.

The constraints are:

$$\begin{cases} y_i\left(\mathbf{w}^T\mathbf{x}_i+b\right)\geq 1-\xi_i, i=1,2,\cdots,N \\ \xi_i\geq 0 \end{cases} \tag{49}$$

Under the above constraints, the minimization of the cost function (42) is equivalent to maximizing the classification boundary. The Lagrange multiplier approach can be used to solve the optimization problem: Given a training set $\{(x_i,y_i)\}, i=1,2,\cdots,N$, find the Lagrange multiplier α_i to maximize the objective function:

$$Q\left(\alpha_1,\alpha_2,...,\alpha_N\right)=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\mathbf{x}_i^T\mathbf{x}_i \tag{50}$$

such that

$$\sum_{i=1}^{N}\alpha_i y_i\geq 0, \text{and } 0\leq\alpha_i\leq C, i=1,2,\cdots,N$$

(51)

Finally, the new sample can be classified by:

$$f\left(\mathbf{x}_{new}\right)=sign\left(\sum_{i=1}^{L}y_i\alpha_i\mathbf{x}_{new}^{T}\mathbf{x}_i+b\right) \tag{52}$$

where L is the number of support vectors.

In practice, before performing linear classification the data is firstly mapped to higher-dimensional feature space through nonlinear transformation. Then the following decision function is applied:

$$f\left(\mathbf{x}_{new}\right)=sign\left\{\sum_{i=1}^{L}y_i\alpha_i\left[\varphi\left(\mathbf{x}_{new}\right)\right]^{T}\varphi\left(\mathbf{x}_i\right)+b\right\} \tag{53}$$

where $\varphi(x)$ denotes the nonlinear mapping that maps the data to higher-dimensional space.

Thus the inner product in the objective function ${x_i}^{T}x_i$ is changed to $[\varphi(x_i)]^{T}\varphi(x_i)$. However, in order to avoid the operation of dot product in higher-dimensional feature space, we use kernel function to calculate the inner product: $K\left(\mathbf{x}_{new},\mathbf{x}_i\right)=\varphi\left(\mathbf{x}_{new}\right)\cdot\varphi\left(\mathbf{x}_i\right)$, then the decision function is modified to:

$$f\left(\mathbf{x}_{new}\right)=sign\left[\sum_{i=1}^{L}y_i\alpha_iK\left(\mathbf{x}_{new},\mathbf{x}_i\right)+b\right] \tag{54}$$

For multi-class classification problem, we can use the “one-to-one” or “one-to-many” method. “One-to-one” refers to the use of $k(k-1)/2$ binary classifiers and when there are new samples, the outputs of all those classifiers are aggregated by majority voting approach. “One-to-many” means that the solution of multiple classification hyperplane parameters is formulated as an optimization problem with modified objective function. Finally, the multi-class classification is directly realized by solving the optimization problem, but this method is computationally expensive and not easy to implement. In addition, the SVM-based classification performance depends on: (1) the parameter optimization; and (2) the selection of appropriate kernel function.

3.4.2 Random forest (RF)

RF is a classifier formed by combining decision trees. It is a kind of ensemble learning algorithm based on the idea of Bagging algorithm. The final output of RF is determined by voting of all decision trees [87].

RF is an ensemble classifier comprising K decision trees $\{h(X,\theta_k), k=1,2,\cdots,K\}$, each used as a base classifier, where $\{\theta_k, k=1,2,\cdots,K\}$ is a sequence of random variables determined by the two ideas of randomization of RF:

1) Bagging: From the original sample set X, K training sets of the same size as X, $\{T_k, k=1,2,\cdots,K\}$, are randomly selected (bootstrapping), and a decision tree is constructed using each training set T_k.
2) Feature subspace: When splitting each node of the decision tree, we randomly extract a subset of attributes from all attributes with the same probability, and then select an optimal attribute from this subset to split the nodes. We grow each decision tree using the following algorithm:

Step 1: Give the training set with the number of samples N and the number of features M.

Step 2: Randomly select $m < M$ features in the feature set.

Step 3: Extract n samples from the given sample set to form a new set of training samples for growing decision tree.

Step 4: For the splitting of a node, calculate its best attribute using the previously selected m features.

Step 5: Fully grown each decision tree without pruning.

3.4.3 Classification performance metrics

In order to evaluate the performance of the emotion classifier, the classification confusion matrix (see the binary classification case in Table 3) is used, based on which four classification performance indices, namely Precision, Sensitivity, Specificity, and F-score, are usually calculated. Consider binary classification problem for the sake of simplicity, then the four metrics, namely precision, sensitivity, specificity, and F score, are defined respectively as:

$$\begin{cases} Prec = \frac{TP}{TP+FP} \\ Sens = \frac{TP}{TP+FN} \\ Spec = \frac{TN}{TN+FP} \\ F_{\text{score}} = \frac{2 \cdot Prec \cdot Sens}{Prec+Sens} \end{cases} \tag{55}$$

alt-text: Table 3

Table 3

The presentation of Tables and the formatting of text in the online proof do not match the final output, though the data is the same. To preview the actual presentation, view the Proof.

Binary classification confusion matrix.

		Predicted Class	
		Positive (+)	Negative (-)
Target Class	Positive (+)	*True Positive (TP)*	*False Negative (FN)*
	Negative (-)	*False Positive (FP)*	*True Negative (TN)*

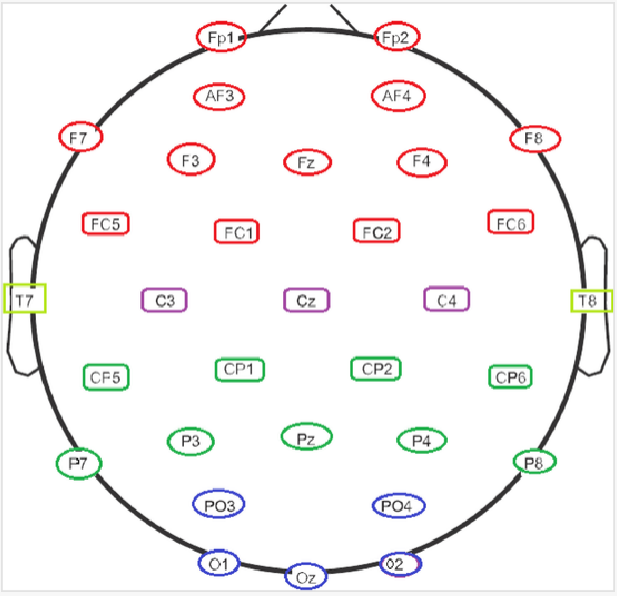
3.5 Emotion-relevant brain regions

This section discusses the brain regions most correlated to emotions with an aim to use fewer electrodes for satisfactory emotion classification performance. In subsection 2.2, we introduced different areas of the brain and their distinctive functions. Physiological studies have shown that the cerebral cortex is primarily responsible for the higher emotional cognitive functions in humans.

It would be desirable to find the brain areas that are closely relevant to emotion through EEG-based emotion recognition [88]. The electrodes are first grouped according to the respective cerebral cortex where they are distributed. The area division of the electrodes is shown in Fig. 6, where red electrodes are distributed in the frontal cortex, green electrodes in the parietal cortex, blue ones in the occipital cortex, yellow ones in the temporal cortex, and squares represents the electrodes distributed in the central area. EEG features are extracted from each group of electrodes and then emotion classification are performed. All the electrodes are sorted/ranked according to their degree of importance by the feature selection algorithm, and then the importance of the electrodes is visualized through the brain topographic map, thereby facilitating the identification of the brain areas in which the higher ranked electrodes are distributed.

alt-text: Fig 6

Fig. 6



Grouping of EEG measurement electrodes according to different lobes.

4 Emotion recognition techniques: recent advances

4.1 Emotion recognition using multi-modal signals

Sensors can measure different types of signals and can be embedded in daily devices such as smartphones, smart watches, and other wearable medical devices. Based on the type of the measurement sensor, human emotion recognition/detection can be roughly classified into audio-visual based and smartphone and other wearable approaches.

So far a significant amount of work on emotion recognition has been carried out using audio (speech and voice), visual (facial expressions), and movement (body posture and gesture) data. audio-visual approaches deploy speech/video sensors that capture speech and facial expressions to recognize emotions.

The growing popularity of sensors, low-power integrated circuits and wireless networks has led to the development of affordable and wearable devices that can measure and transmit data for a long time. Wearable devices are worn by the users for non-intrusive (un-obtrusive) monitoring of physiological signals. In recent years, many wearable devices are equipped with a range of sensors which can continuously monitor HR, movement/motion, and location data. This led to the emergence of big data in a variety of areas such as health-care, smart home, and smart city.

With the advent of low-cost wearable sensors (e.g., smartphone, smartwatch and wristband), there is also an emerging interest in using multi-modal physiological signals (e.g., EEG, heart rate, galvanic skin response, etc.) for emotion recognition. Smartphone has become a ubiquitous personal device with a rich set of sensors embedded, such as accelerometer, GPS, gyroscope, and microphone, for health monitoring, pedestrian localization and navigation. The use of smartphone and wireless network signals for human activity recognition (including emotion detection) has also gained considerable attention.

4.1.1 Physiological signals

Most efforts in emotion assessment used audio-visual data, that is, facial expressions, emotional speech and body gestures. Physiological signals can also be used for emotion assessment, but they have received less attention [89]. The main difficulties in using physiological signals for emotion recognition include:

1) It is hard to map uniquely physiological patterns onto specific emotional states because physiological patterns may largely vary across subjects and experimental sessions.
2) Traditionally, recording of physiological signals requires the user to be connected with biosensors and the signal measurement is sensitive to motion artifacts. Furthermore, analysis of multi-channel homogenous or heterogeneous (multi-source) physiological signals is a complex multivariate time series analysis problem, which requires knowledge about and insight into neuro-psychological processes and functions.
3) It is hard to obtain the ground truth of physiological data (i.e., the target/actual class label of each data point) since we cannot perceive emotions directly from the real-time physiological

signals. This results in difficulties in physiological data labeling/annotation.

However, the following major benefits can be gained by using physiological signals for emotion recognition:

1) We can continuously monitor affective states. Consider the cases that people resort to 'poker face' or simply do not say anything when they get angry. In those cases, the emotional states of the user cannot be detected by traditional audio-visual data driven method.
2) Since the activation of autonomic nervous system (ANS) is largely involuntary and cannot be easily triggered by any conscious or intentional (deliberate) control, using physiological signals from ANS for emotion recognition would be robust against artifacts caused by human *social masking*. For instance, sometimes people smile in negative emotional experiences. Such smile results from social masking, where one regulates emotions for a good interpersonal relationship, and does not express his/her actual emotional state [44].
3) Experimental results revealed significant cross-cultural consistency (or invariance) in the ANS physiological patterns among different emotions [90].

Atkinson and Campos [91] combined the mutual information based EEG feature selection approach and support vector machine to estimate emotions. Two levels on arousal and valence dimensions were classified with the accuracy of 60.7% and 62.4%, respectively. Moreover, Chen et al. [92] improved the EEG-based emotion classifier system by employing a set of ontological models to represent EEG feature sets. Their results showed the ontological approach achieved the binary classification accuracy of 69.9% and 67.9%, respectively. Li et al. [93] proposed the individual-specific models to improve the machine learning based emotion classifier. In particular, four types of emotions, neutral, sadness, fear, and pleasure, were recognized. Three physiological indicators were used as the classifier inputs, i.e., electrocardiogram (ECG), galvanic skin response (GSR), and photo plethysmography (PPG). Verma and Tiwary [94] proposed a multi-resolution approach for emotion estimation using physiological signals. The clustering algorithm was applied to determine valence, arousal and dominance thresholds of multiple emotions. The features were extracted with the discrete wavelet transform (DWT) method. Yoon and Chung [95] proposed a Bayesian weighted-log-posterior function method to identify the optimal weights of an artificial neural network (ANN) for classifying emotions. The EEG features were used as the classifier input. For the three-class emotion classification problem, an average accuracy of 55.4% and 55.2%, respectively, has been achieved.

Petrantonakis and Hadjileontiadis [96] defined an Asymmetry Index (AsI) to represent the difference of the activities in left and right hemispheres of the frontal cortex. They used three-channel EEG measurements to classify six type of emotions. The SVM classifier was used to model the EEG data. Wang et al. [97] proposed a Fourier parametric model using the 1 st- and 2 nd-order differences for speech-based emotion recognition. Wang et al. [98] employed functional near infrared spectroscopy (fNIRS) to estimate the variation of the patients' emotions in healthcare context. They found a decrement of 22.2% in the classification accuracy between two experimental sessions with a 3-week interval. A feature selection method was developed to find the stable indicators of emotion to achieve comparable recognition accuracy. Nakisa et al. [99] utilized long- and short-term memory (LSTM) networks to process the time courses of the EEG and blood volume pulse signals for emotion recognition. They found that the generalizability of the LSTM model can be improved by

tuning its hyper-parameters via differential evolution algorithm. Yin et al. [22,27,66] investigated the problem of how to derive the salient EEG features that is closely correlated to the emotion variations.

4.1.2 Audiovisual data and text data

Several representative recent work on emotion recognition using audio, visual and text data is summarized in Table 4.

A. Audio data

alt-text: Table 4

Table 4

The presentation of Tables and the formatting of text in the online proof do not match the final output, though the data is the same. To preview the actual presentation, view the Proof.

Recent work on emotion recognition using audio-visual data and text data.

Dataset	Features	Model	Ref.
404 YouTube vloggers (194 M, 210 F)/ YouTube personality dataset	A-V, lexical, POS psycholinguistic, emotional and traits	SVM	Alam et al. (2014) [100]
404 YouTube vloggers/ YouTube personality dataset	A-V, text, demographic and sentiment	LR with ridge estimator	Sarkar et al. (2014) [101]
47 (27 M 20 F)/YouTube dataset	Softwares using CLM-Z and GAVAM, openEAR and using CNN	MKL	Poria et al. (2015) [102]
42/ ISEAR, CK + +, eNTERFACE dataset	66 FCP using Luxand software, JAudio software, BOC, Sentic features and Negation	SVM	Poria et al. (2015) [103]
230 videos/ Rallying a Crowd (RAC) dataset	Softwares using CAFFEE and features (prosody, MFCC, or spectrogram) and using SATSVM and DCM	RBF-SVM and LR	Siddiquie et al. (2015) [104]
47 (27 M and 20 F)/YouTube dataset SenticNet	Softwares using Luxand FSDK 1.7 and GAVAM, openEAR and Concept-gram and SenticNet-based features	ELM	Poria et al. (2016) [105]

Legenda: *A* = Audio; *V* = Video; *T* = Text; ML = Machine Learning; LR = Logistic Regression; MKL = Multiple Kernel Learning; ELM = Extreme Learning Machine; CNN = Convolutional Neural Network; RBF = Radial Basis Function.

Recently Zhalehpour et al. [106] built a database, named BAUM-1, of audio-visual information for validating the effectiveness of the emotion recognition algorithms. Eight classes of emotion were recognized, including happiness, anger, sadness, disgust, fear, surprise, boredom and contempt. Eyben et al. [107] proposed a

standard acoustic parameters set for emotion recognition via automatic voice analysis. The features were selected based on their usefulness to indicate affective changes in voice, the automatedness of extraction procedure, and the theoretical significance. Wu et al. [108] classified emotions via affective speech that combines acoustic-prosodic information (AP) and semantic labels (SLs) with multiple classifiers. The Gaussian mixture model (GMM), support vector machine (SVM), and the multi-layer perceptron (MLP) were integrated to extract spectra, formant, and pitch related features. Yoon and Park [109] proposed a speech-based emotion recognition framework for consumer electronic applications. They extended binary (neutral vs. anger) emotion classifier to a hierarchical structure exploiting emotional characteristics and gender difference. Wu et al. [110] proposed an eigenface conversion based approach to filtering the facial expressions. The GMM model and the decision three were used to recognize emotions. Four types of emotions defined on the VA plane were classified based on the articulatory attribute (AA) of speech segments. Schuller et al. [111] studied the generalizability of the emotion recognition systems on multiple speech datasets. They performed a cross-corpora evaluation across six benchmark databases and found a serious performance degradation when compared with the intra-corpus testing. Abdelwahab et al. [91] introduced the principle of the transfer learning in speech-based emotion recognition since it is plagued by the mismatch of the data distributions between the training and testing sets. They learned a common representation of features based on the predicted emotional-attribute descriptors of arousal, valence, or dominance. Bisio et al. [112] proposed subject-specific emotion recognition systems by using the audio signal registrations. In particular, they improved the classification accuracy by combining the emotion recognition with the gender recognition. Park et al. [113] studied the speech emotion recognition in service robot interacting with human users. They selected discriminative vectors among overlapped features to design the binary-class (negative vs. positive) classifier. Chen et al. [114] evaluated four component tying methods, namely single group tying, quadrant-wise tying, hierarchical tying, and random tying, for personalized emotion recognition. They adapted acoustic emotion GMM models to individual users.

B. Visual data

Guo et al. [115] performed fine-grained analysis of emotions by using images with a compound facial emotion. In particular, they built the dataset, termed iCV-MEFED which includes 50 labeled classes of compound emotions. They designed the deep convolutional neural network (CNN) as the emotion classifier. Jing et al. [116] employed local binary pattern (LBP) and k-nearest neighbors (kNN) to classify facial expressions into different emotions in the smart home with eldercare robots. Yan et al. [117] proposed a bimodal emotion classification method that combines the facial expression and speech. They also developed sparse kernel reduced-rank regression (SKRRR) fusion method to integrate the bimodal features. In particular, the scale-invariant feature transform was used to extract emotion-related indicators in facial expressions. Petrantonakis and Hadjileontiadis [118] used the higher-order crossings (HOC) analysis to find the representative features from EEG signals. They combined HOC feature extraction method with the quadratic discriminant analysis (QDA), KNN, Mahalanobis distance model, and SVMs to recognize six types of basic emotions. Shojaeilangari et al. [119] proposed a robust emotion recognizer based on extreme sparse learning. They combined the extreme learning machine with the sparse representation in order to cope with the noisy images recorded in natural settings. Chakraborty et al. [120] developed a fuzzy relation based emotion classifier using facial expressions. The Mamdani-type relational model was adopted to control the transition of

emotion dynamics towards a desired state. Ferreira et al. [121] employed the deep neural network to classify facial expressions into different emotion states. They proposed an end-to-end NN architecture along with a well-designed loss function based on the prior knowledge that facial expressions are the result of the motions of some facial muscles. Zhang et al. [122] employed biorthogonal wavelet entropy method to extract the multiscale features for facial expression recognition. A stratified cross-validation accuracy of 97% was achieved for recognizing seven types of emotions (happy, sadness, surprise, anger, disgust, fear, and neutral).

C. Text data

Li et al. [123] considered the issues in affective lexicon. They applied the support vector regression to automatically estimate affective representation of words from the word embedding. The extended lexicons are publicly accessible. Albornoz and Milone [124] developed a language-independent emotion recognition system. The concept of universality of emotions is employed to map and predict emotions in unforeseen languages. In particular, they developed an ensemble classifier using the emotion profiles technique in order to map features from diverse languages in a more tractable space. Xu et al. [125] proposed an integrative framework for transferring knowledge from heterogeneous sources of image and text to facilitate three related tasks in video emotion analysis and understanding: emotion recognition, emotion attribution, and emotion-oriented summarization.

4.1.3 Time series signals

Quan and Ren [126] identified compound emotions in the text using weighted high-order hidden Markov models (HMMs). They encoded emotion of the text by a sequence of spectral vectors under its temporal structure. Karyotis et al. [127] represented an AV-AT emotion model with a genetically optimized adaptive fuzzy logic framework and built a hybrid cloud intelligence architecture to examine the user sentiments and affects. A personalized learning system was applied to validate the performance of the fuzzy model. He and Zhang [128] employed an assisted learning strategy to improve the training performance of the CNN model for the problem of image based emotion recognition. Jain et al. [129] employed a hybrid convolution-recurrent neural network (RNN) for facial expression recognition problem. Architecturally, the network consists of several convolution layers linked to an RNN, in order to reveal the relations between facial image sequences.

Lin et al. [130] investigated emotion-related EEG dynamics during music listening. They obtained an average classification accuracy of 82.29% for 26 subjects. Yang et al. [131] developed a hierarchical network structure, consisting of several subnetworks, to improve the emotion classification performance. Three subnetworks were designed to model three types of human emotions (positive, neutral, and negative). Katsigiannis and Ramzan [132] released a neurophysiological database for emotion classification. The ongoing EEG and ECG signals were measured under audio-visual stimuli. All neurophysiological signals were recorded by using portable, wearable, and wireless devices. Soleymani et al. [133] combined LSTM, RNN, and continuous conditional random fields (CCRF) to estimate continuous emotions.

4.2 Emotion recognition based on multi-modal information fusion

Most previous studies on emotion recognition focused on use of single sensor modality, features and classifiers, which are ineffective to discriminate complex emotion classes. Fusion of multiple modalities aims at improving classification accuracy by exploiting the complementarity of different modalities. For effective emotion recognition, data, features and classifiers may need to be fused with appropriate strategies as follows.

Data-level fusion: involves integration of multi-source data to increase reliability, robustness and generalizability of the emotion recognition system.

Feature-level fusion: Features are combined using ML algorithms such as SVM, decision tree, and HMM to discriminate the data into higher-level abstraction. Furthermore, automatic feature representation using DL has been proposed to solve the issues of temporal and spatial dependencies. DL automatically extracts translational invariant and robust features from data to minimize application-specificity and simplify feature extraction and selection steps.

Classifier-level fusion: involves combination of multiple weak base classifiers to reduce uncertainty by fusion of outputs of different classifiers to achieve higher accuracy that are unlikely when the classifiers are used in isolation. Data manipulation, feature manipulation, model diversification and random initializations are often used to build multiple classifier systems.

In feature-level fusion (or called early aggregation), the features extracted from signals of different modalities are concatenated to form a composite feature vector and then input to an emotion classifier. In decision- or classifier-level fusion (or called late aggregation), each modality is processed separately by the corresponding classifier and the outputs of the individual classifiers are aggregated to yield the final output.

Either approach has its own strengths. For example, implementing a feature-fusion-based system is straightforward, while a decision-fusion-based system can be constructed from existing unimodal classifier systems. Moreover, feature fusion can consider synchrony of the multiple modalities, whereas decision fusion can flexibly model their asynchronous characteristics.

An interesting advantage of decision fusion over feature fusion is that we can easily employ an optimal weighting scheme to adjust the degree of the contribution of each modality to the final decision result based on the reliability of individual modalities.

A common weighting scheme can be formulated as follows. For a given test data X, the decision output of the fusion system is:

$$c^{*} = \arg\max_{i} \left\{ \prod_{m=1}^{M} P_i(X|\lambda_m)^{\alpha_m} \right\}, \tag{56}$$

where M is the number of modalities, λ_m is the classifier for the m-th modality, and $P_i(X|\lambda_m)$ is its output for the i th class. The weights α_m, which satisfy $0 \le \alpha_m \le 1 \text{and} \sum_{m=1}^{M} \alpha_m = 1$ represents the modality's reliability which determines its degree of contribution to the final decision.

A simple method for determining the optimal weights is based on the training data. The optimal weights are estimated by exhaustive search of the grid space, where each weight is increased from 0 to 1 with a step size of 0.01 and the weights producing the best classification of the training data are selected.

For example, Subramanian et al. [134] developed a multimodal database for emotion recognition by using commercial physiological sensors. The database, termed as ASCERTAIN, includes the personality traits and affective behaviors. The 58 participants' data of subjective ratings, EEG, ECG, GSR, and facial activity revealed the personality difference issue for emotion classification. In 2014 Zacharatos [135] reviewed emerging techniques for emotion recognition, the application areas of emotion classification, and the techniques for body movement segmentation. Wang et al. [136] developed a music emotion recognition system for multi-label classification. In their work, hierarchical Dirichlet process mixture model (HPDMM) was used to share model components in different emotions. Kim and André [137] comprehensively investigated the measures of EMG, ECG, GSR, and respiration activity for emotion recognition. They proposed a multi-level dichotomous classifier and achieved the subject-dependent (or -independent) classification accuracy of 95% (or 70%). Mariooryad and Busso [138] investigated the cross-modal properties of an emotion classifier system based on facial expression and speech information. Under a mutual information evaluation framework, the authors found that the facial and acoustic features are indicative of similar affective behavior. Zheng et al. [139] modified the least square regression (LSR) into an incomplete sparse LSR to model the correlation between the speech features and the corresponding emotion labels. In particular, both the labeled and unlabeled speech data were used for semi-supervised learning. Wagner et al. [140] considered the missing data issue when applying the ensemble learning method to facial, vocal, and gestural data. In order to compare in a fair manner different facial expression based emotion recognition algorithms, Valstar et al. [141] proposed protocols for the emotion database, classification paradigms and the performance evaluation metrics.

4.3 Emotion recognition using deep learning techniques

4.3.1 Comparison between traditional ML and DL

Researchers have investigated the relationship between the physiological data and human emotions. A majority of earlier work employed traditional statistical analysis techniques.

Recently, numerous studies on using machine learning and deep learning (DL) approaches for automatic emotion recognition have been reported, although they are relatively new when comparing with the long history of emotion research in psychophysiology.

Feature extraction and dimensionality reduction identify a smaller feature set to increase classification accuracy and reduce computational time. Feature extraction can be divided into the extraction of shallow and deep features. Shallow features refer to hand-crafted features in different analysis domains, such as time-domain, frequency-domain, and Hilbert-Huang transform. The higher-dimensional features are reduced using principal component analysis (PCA) or empirical cumulative distribution functions. Unfortunately, shallow features rely heavily on heuristics and require a large number of labeled data that may be hard to collect in real-world application scenarios.

The extraction and selection of hand-crafted features (usually statistical variables such as mean, variance, kurtosis, entropy, and multi-scale entropy) are usually laborious and time-consuming, but have a decisive impact on the performance of ML models. The hand-crafted shallow features are often domain-specific and hard to be reused in similar problems.

Traditional feature engineering and ML algorithms may not be efficient to elicit the complex and nonlinear patterns in multivariate time series datasets. Additionally, selecting the salient features in a large feature set is critical and will require dimensionality reduction techniques. Furthermore, feature extraction and selection are computationally expensive. For instance, the computational cost of feature selection may increase exponentially with the increase of feature dimensionality. In general, search algorithms may not be able to converge to optimal feature set for a given ML model.

In order to overcome the difficulties in obtaining effective and robust features from time series data, many researchers have paid attention to DL approaches. DL relieves the burden of extracting hand-crafted features for ML models. Instead, it can learn a hierarchical feature representation automatically. This eliminates the need for data preprocessing and feature space reconstruction in standard ML pipeline. DL techniques, such as autoencoder, convolutional neural network, and recurrent neural network, have made significant impact (with almost human-level performance) on computer vision, speech recognition, object recognition, natural language processing (NLP), and machine translation. Since DL can realize high-level abstraction of data, it has been used to develop reconfigurable architectures for emotion recognition in recent years.

DL can be dated back to the work in 1980s. The neocognitron [142] was arguably the first artificial neural network (ANN) that possessed the “deep” property and took into account neurophysiological insights. In 2006, Hinton and Salakhutdinov [143] initiated a breakthrough in feature extraction, which was followed up in later years [144-147]. Various studies [143,145,148-150] showed that multilayer NNs with iterative or non-iterative methods can be used for feature representation/learning.

DL methods apply high-level data representation to extract salient features with deep neural network. An interesting advantage of DL techniques is that they can work directly with raw data and automate the feature extraction and selection processes. Time series samples are fed into the network, and after each nonlinear transformation, a hidden representation of inputs from the previous layer is generated to form a hierarchical structure of data representation. In other words, each layer in a deep network model combines outputs from the previous layer and transforms them to a new feature set via a nonlinear mapping. The typical DL method, convolutional neural networks (CNNs), has been applied to EEG data in different application scenarios [151, 152].

4.3.2 Recent work on emotion recognition using DL approaches

Recently the DL methods have become prevalent for analysis of physiological signals for emotion recognition. In most state-of-the-art emotion recognition studies, DL was used due to its effectiveness for deep feature extraction/representation.

DL-based learning methods have penetrated into the field of EEG-based emotion recognition. For example, DL has been applied to EEG-based emotion recognition [153-156]. Sheng et al. [17] trained a deep belief

network (DBN) with DE features and achieved a classification accuracy of 87.62%. Tsiaris et al. [157] applied the deep convolutional neural network (CNN) to extract the intermediate speech features for emotion recognition. In addition, a 50-layer deep residual network was constructed for visual feature representations. Finally they implemented a long short-term memory networks to improve the testing performance of deep model and to handle the outliers in the dataset. Li and Deng [158] developed a deep locality-preserving convolutional neural network (DLP-CNN) that exploits the advantages of deep and manifold learning approaches. The basic idea is to improve the discriminative power of high-level features by measuring the locality closeness. In particular, the authors built a new database, called RAF-DB, with 30 k images for facial expression based emotion recognition. Attabi and Dumouchel [159] studied anchor classification models. They used an unbalanced dataset of children speech under various emotions. Their results indicated that the Euclidean or cosine distances are best suitable in a GMM and SVM based anchor classifier. Zhang et al. [160] adopted CNN deep model to extract features from audio-visual data and integrated them via deep belief networks (DBNs). Then a linear SVM was used to classify emotions. Deng et al. [161] investigated the domain adaptability of an adaptive denoising autoencoder for the speech based emotion recognition. Xia and Liu [162] attempted to combine the arousal and valence dimensions of the emotion model into a hidden variable. They applied deep belief networks (DBN) as a regression model to find such variable and to output the predicted emotion class. Tariq et al. [163] used a set of features of images to perform the subject-independent emotion classification. Hierarchical Gaussianization markers, scale-invariant features and coarse motion features were input into a SVM classifier. A classification accuracy of 66% was reported for the person-independent emotion recognition. Chen et al. [164] employed the deep sparse autoencoder network (DSAN) to find the hidden information in facial features and recognize different emotions via a Softmax regression model. They found that the training complexity of the deep classifier model (and therefore computational burden of the model training) is relatively lower than CNNs.

A DL framework was proposed on the basis of a sparse auto-encoder, which combines emotion-related features as inputs of two logistic regression models for arousal and valence classification [165]. A DL-based information fusion system was developed using the speech and the video data [166]. The speech signal was processed to extract Mel-spectrogram features and then combined with the representative frames of videos. The extracted features were fed to CNN to estimate different emotional states.

The unsupervised deep belief network (DBN) was used for fusion of the features from Electro-Dermal Activity (EDA), zygomaticus ElectroMyoGraphy (zEMG), and PhotoPlethysmoGram (PPG) signals [167]. Subsequently the DBN-elicited features are combined with statistical features of EDA, PPG and zEMG signals to constitute a feature vector, which is then used as inputs of the Fine Gaussian Support Vector Machine (FGSVM) with radial basis function kernel to classify five basic types of emotions (i.e., Happy, Relaxed, Disgust, Sad, and Neutral).

The recurrent neural network (RNN) based DL model was applied to cope with the complexity of the emotions in texts [168]. The deep RNN adopted dropout layers and the weighted loss function with the regularization term. The frame-based formulation of minimal speech processing technique was used as features of the RNN architecture to model intra-utterance dynamics for emotion recognition [169]. The deep CNN model was also suitable for recognizing emotions based on facial expressions. A deep CNN was applied to emotion

recognition problem, with an architecture of 6 convolutional layers, 3 max pooling layers, 2 residua layers, and 2 fully-connected layers [170]. The deep CNN based emotion recognition system was also applied to process body movement features [171].

A hybrid deep feature extraction model, based on LDA and PCA, was used for speech emotion recognition [172]. In particular, the authors developed an improved emotion learning system by using genetic algorithm for parameter optimization. The deep CNN was also used for image emotion recognition [173]. A LSTM-based CNN was applied for understanding emotions in text [174]. They combined semantic and sentiment feature representations to classify three classes of emotions (i.e., happy, sad and angry) in the texts. In [170] Jain et al. utilized an extended deep convolutional neural network for facial emotion classification. The deep network architecture contains several convolution layers and deep residual blocks. The proposed model can learn the subtle features that discriminate the seven facial emotions (i.e., sad, happy, surprise, angry, neutral, disgust, and fear) from facial images.

Finally, the performance comparison of several recent DL methods for emotion recognition problem is given in Table 5.

alt-text: Table 5

Table 5

The presentation of Tables and the formatting of text in the online proof do not match the final output, though the data is the same. To preview the actual presentation, view the Proof.

Performance comparison of DL methods for emotion recognition based on audio-visual data.

Dataset	DL method	Accuracy [%]	Ref.
Enterface	mel-spectrogram, face images, cnn for audio, 3 d cnn for video	85.97	Zhang et al. (2017) [160]
Audio-visual big data of emotion	2 d cnn for speech, 3 d cnn for video, elm-based fusion	99.9	Hossain et al. (2019) [166]
eNTERFACE	Audio features, facial features, triple stream DBN	66.54	Jiang et al. (2011) [175]
IEMOCAP; facial markers	Feature selection and DBN	70.46–73.78	Kim et al. (2013) [176]
EmotiW 2014	CNN for video, DBN for audio, 'bag of mouth' model, and auto-encoder	47.67	Kahou et al. (2016) [177]
eNTERFACE	Multidirectional regression, SVM	84	Hossain et al. (2016) [178]
enterface	mdr, ridgelet transform, elm	83.06	Hossain et al. (2016)

			[179]
Audio-visual big data of emotion	lbp features for speech, idp features for face images, svm classifier	99.8	Hossain et al. (2018) [180]
MAHNOB-HCI	CDBN	58.5	Ranganathan et al. (2016) [181]
EmotiW 2015; CK+	Audio features, dense features, CNN extracted features	54.55; 98.47	Kaya et al. (2017) [182]

Legenda: DBN = Deep Belief Network; CDBN = Convolutional DBN; CNN = Convolutional Neural Network; SVM = Support Vector Machine; ELM = Extreme Learning Machine; MDR = Multi-Directional Regression; LBP = Local Binary Pattern; IDP = Interlaced Derivative Pattern.

4.4 Our work related to emotion recognition

From 2005 Zhang and his co-workers started to investigate the problems of quantification of human mental stress/workload and human operator functional state in safety-critical human-machine interaction environment with an ultimate goal of realizing adaptive function/task allocation (a.k.a. adaptive automation or adjustable autonomy of human-automation systems) between humanistic and machine agents of intelligent human-machine systems. The work is intrinsically related to the problem of detecting low valence state (or high-risk or vulnerable functional state) in humans. The multimodal physiological signals, including EEG, ECG and ECG, were measured and pre-processed. The extracted features were fed to the supervised learning classifiers. To estimate/predict the quantitative level of mental stress, in 2016 Zhang et al. employed the Gaussian mixture model to elicit soft clusters [183]. Then an ensemble of SVM classifiers was adopted to recognize the variations in mental stress over time. In particular, the diversity of the member classifiers was enhanced by using different hyper-parameters in the objective function for the SVM training. In [184,185], Zhang et al. showed the effectiveness of these methods and their variants, based on NARX-LSSVM and adaptive SVM, for the pattern recognition of mental stress.

In recent years, we have focused on developing the improved DL architectures, models and algorithms for human emotion and mental stress recognition. In our recent work [22,27,186,187], ensemble SAE DL architectures have been proposed to classify binary arousal and valence levels. Both the CNS and PNS features were used as the input of each weak SAE model for abstract high-level neurophysiological representations. A transfer dynamical SAE was developed for cross-subject mental stress classification with an aim to handle the non-stationarity of the physiological features [188]. Similar architectures have been applied in extreme learning machine with multi-layer network [34]. In order to reduce the dimensionality of the input features and improve the emotion classification accuracy, we also developed several EEG feature selection approaches based on feature recursive elimination [22,27,66,186]. Although these DL methods lead to acceptable emotion recognition performance, their online version still needs to be developed and validated in the future work. On the other hand, we found that the emotion recognition performance of the DL models depends severely upon the size of the feature set as well as the size of the training dataset.

5 Summary and future outlook

In this paper, we have surveyed more than 220 papers, not only discussing the state of the art emotion recognition techniques proposed in recent years (up to 2019), but also considering the available datasets and illustrating principal constituents of a data-driven emotion recognition pipeline. In this section, we summarize some of our major findings drawn from this survey.

5.1 Summary of major findings

With the advancement of automation and human-machine systems technologies, emotion recognition has become a hot topic in the field of human-computer interaction. This paper reviews multi-modal psychophysiological data driven emotion recognition techniques. EEG responds in real time to changes in emotions. The EEG features can be extracted, reduced and then used for emotion classification. The general procedure of EEG-based emotion recognition consists of the following computational steps: Data acquisition, data preprocessing, feature extraction, feature dimensionality reduction, and design of optimal classifier models. This paper focuses on different techniques for feature extraction, feature dimensionality reduction, classifier model optimization, and selection of brain regions that are most correlated to emotions. The main findings of this expository paper are summarized as follows:

(1) In many literature [68,69,189], only binary classification of each dimension of emotion was considered.

(2) Traditionally the labeling of the actual/true/target emotion classes is based on a preselected threshold of the subjective rating data. Unfortunately the proper threshold is hard to select. A new idea is to look at the valence and arousal dimensions simultaneously and use data clustering algorithms to obtain the target classes of emotion.

(3) The incorporation of baseline EEG data: In many emotion recognition studies (e.g., [4,5,68,69]), the researchers only used the EEG data under different emotional conditions while ignoring the baseline EEG data. In [18,19,36], the features were extracted from both the baseline/spontaneous EEG (when the subjects were not aroused emotionally) and evoked potentials (or event-related potentials). The difference between the two features is then used as the input features to the classifier. The comparative results showed that the emotion classification accuracy can be significantly improved by taking into account the baseline EEG features.

(4) Different feature extraction methods are reviewed, including wavelet transform and nonlinear dynamics. Through extensive comparisons, we found that the gamma sub-band features lead to the highest classification accuracy, indicating that the gamma frequency band is most sensitive to the emotional changes. We also found that the classification accuracy can be improved when the ApEn and SampEn features are used jointly.

(5) Different dimensionality reduction algorithms are reviewed. We found that the best feature dimensionality reduction algorithm varies with different feature extraction methods.

(6) Different types of machine learning based classifiers are reviewed, including kNN, NB, SVM and RF. It is found that SVM and RF perform better than kNN and NB for EEG-based emotion recognition task.

(7) The emotion classification accuracy based on the EEG signals from different brain regions is discussed. We found that using only a dozen of EEG electrodes placed on the frontal lobe can achieve a classification accuracy of over 90%, which may provide a basis for online EEG-based emotion recognition.

5.2 Open problems and future research vista

As this survey paper has demonstrated, in addition to rich opportunities there still exist significant research challenges in this multi-disciplinary field. Although EEG-based emotion recognition systems have been developed recently, the use of EEG headcap for signal measurement is inconvenient and it usually takes a long time to measure experimentally the multi-channel signals. The users' acceptability can be also an issue for EEG-based emotion recognition. For real-world applications of emotion recognition technology, we can use wearable and wireless devices (with active dry electrodes) to measure EEG signals and advanced ML and DL algorithms for EEG signal analysis. However, some open problems still remain. For instance, the existing approaches to emotion recognition using physiological signals achieved average classification accuracy of over 80%, which seems acceptable for practical applications, but the recognition accuracy are application- specific and strongly dependent upon the datasets under consideration.

Several open problems and promising research directions in the field of emotion recognition are outlined as follows.

(8) In addition to EEG signals, other types of physiological signals, such as ECG, EOG, and EMG signals, can also indicate change of emotions. In future studies, we need to use mobile, wearable (portable) sensors to collect facial expression, EEG, ECG, and other multi-modal physiological signals and fuse them in an appropriate framework to further improve the accuracy of online, real-time feature extraction and emotional state estimation [135,190-199].

A challenge in the future will be the studies of multi-modal emotion recognition. Humans use several modalities jointly to interpret emotional states, since emotion affects almost all modes – audiovisual (facial expression, voice, gesture, posture, etc.), physiological (EEG, RSP, skin temperature, etc.), and contextual (goal, preference, environment, social situation, etc.) states in human communications.

Recently emotion recognition by combining multiple modalities have been reported in literature, mostly by fusing features extracted from audiovisual modalities (such as facial expressions and speech). However, it should be noted that combining multiple modalities by equally weighting them does not always guarantee improved classification accuracy. A crucial issue is how to properly aggregate the multiple modalities. An essential step towards human-like fine-grained recognition of emotions would therefore be to find the innate priority/preference among the modalities for each emotional state.

It was found that for arousal there are negative correlations between the subjective ratings and the EEG theta, alpha, and gamma bands. The central alpha power decrease for higher arousal [200] and an inverse relationship exists between alpha power and the arousal level [201,202].

Valence showed the strongest correlation with EEG signals and correlation with the subjective ratings were found in all EEG frequency bands. In the lower-frequency bands, an increase of valence led to an increase of theta and alpha power. A central decrease and an occipital and right temporal increase of beta power were found. Increased beta power over right temporal sites was associated with positive emotional self-induction and external stimulation [203]. Similarly, Onton and Makeig [204] found a positive correlation of valence and high-frequency (i.e., beta and gamma bands) power emanating from anterior temporal cerebral sources. In addition, a significant increase of left and especially right temporal gamma power was also observed. However, it should be noted that the EMG activity is also prominent in the high frequencies, especially from anterior and temporal electrodes [205].

Most engineering approaches to emotion recognition provides evidence that the accuracy of arousal classification is usually higher than that of valence differentiation. The reason may be that the change of the arousal level directly correlated to the ANS activities (e.g., blood pressure and skin conductivity) which are easy to measure, whereas valence level discrimination requires a factor analysis of cross-correlated ANS responses. Therefore, we need to develop an emotion-specific classification scheme and to extract a wide range of valence-relevant features from EEG and other peripheral physiological signals in various analysis domains (e.g., time, frequency, sub-band spectra, time-frequency, space- and spectrum-time, entropy, and multi-scale entropy).

A viable solution might be to decompose an emotion recognition problem into several refining processes using additional modalities, for example, arousal (dimension) recognition using peripheral physiological signals, valence (dimension) recognition using audiovisual and EEG signals, and then resolution of subtle overlapping between adjacent emotion classes. In this connection, the physiological signals should be considered as a *baseline channel* in developing a multi-modal emotion recognition system.

(9) Development of advanced ML techniques: In fact, the human emotion generation is a complex and subjective process. Emotions reflect the cognitive processes associated with biological understanding and psychophysiological phenomena. Thus it is difficult to propose a recognition method which is solely based on traditional ML methods. Domain-independent adaptive, deep and transferrable ML techniques need to be developed for speech, text, and physiological data based affective computing. In particular, the EEE is essentially space- and spectrum-time data (SSTD), specialized deep ML approaches can be used to analyze the EEG signals and to extract a richer set of emotion-related features from them. Also we need to apply DL techniques for feature- or classifier-level information fusion in order to further improve the emotion classification accuracy [88,166,206-210]. On the other hand, traditional time series analysis methods need to be combined with ML techniques for continuous-time monitoring of the temporal variations in emotion [211-217].

According to recent studies, the thalamus, basal ganglia, insular cortex, amygdala, and frontal cortex are all involved in emotion recognition [218]. Furthermore, more and more biological evidence [219, 220] indicates that neuron activity in a mammal's prefrontal cortex is heterogeneous, random, and

disordered. The combined features extracted from mixed selectivity neurons may be central to complex cognition.

(10) The existing research mainly considers the subjective-specific/dependent emotion recognition problem, that is, for each subject we need to design an individualized/personalized classifier. In real-world scenarios, a subject-independent (or called generic) emotion recognition model, which fits a group of subjects, would be of significant importance. However, the subject-independent classifier model needs to be combined with transfer learning technique to obtain emotion recognition accuracy that is stable across subjects.

(11) Higher-dimensional models of emotion need to be built. Currently the 2 D emotion model is often used. Higher-dimensional emotion models need to be constructed for multi-class emotion recognition. For example, we can then predict the '*stance'* dimension in a 3 D emotion model (i.e., arousal, valence and stance) by cumulative analysis of the subject's context/situational information. Researchers have subsumed the associated action tendencies under the term *stance*. For example, fear is associated with the action pattern of 'flight', while anger is associated with the urge to 'fight'. Nevertheless, it is still not clear what action patterns are associated with positive emotions (such as happiness, contentment, and amusement). Such positive emotions seem to lack autonomic activation. This may be why so far there has been less research progress on positive emotions than that on negative emotions. Fredricson and Levenson [221] reported the *undoing* effect of positive emotions, which supported the idea of a symmetric process under the emotion system (i.e., positive emotions speed up return to homeostasis, while negative ones help human being escape from it).

(12) More work needs to be done in order to develop more accurate method of labeling massive data in the emotion-related database. There is also uncertainty about the physiological data labeling due to individual self-reports and the situational variables in ANS activity [222]. In this connection, subjective ratings data can be used to obtain the target classes, but it is also possible to obtain them based on the particular content of emotion-elicitation stimuli. The physiological datasets used in most existing work were measured by using visual elicitation materials under lab settings. The emotional state of the subjects prior to the experiments was not taken into account in most previous work. Such individual differences can induce inconsistency in the datasets. On the other hand, the practical effectiveness and reliability of subjective data clustering and threshold methods need to be further validated.

CRediT authorship contribution statement

Jianhua Zhang: Investigation, Methodology, Supervision, Writing - original draft, Writing - review & editing. **Zhong Yin:** Writing - original draft. **Peng Cheng:** Investigation, Writing - initial draft preparation. **Stefano Nichele:** Resources.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported in part by OsloMet Faculty TKD Lighthouse Project [grant no. 201369-100]. Z. Yin's work was funded by the National Natural Science Foundation of China [grant no. 61703277] and the Shanghai Sailing Program[grant no. 17YF1427000]. We gratefully acknowledge the support from Dr. Salvador García, University of Granada, Spain. We would also like to thank the anonymous reviewers for their insightful and constructive comments and suggestions, which helped to improve this paper.

References

[1] Petrushin V.A., Emotion in speech: recognition and application to call centers, Proceeding of the 1999 Conference on Artificial Neural Networks in Engineering, 1999, pp. 7–10.

[2] Anderson K., McOwan P.W., A real-time automated system for the recognition of human facial expressions, IEEE Trans. Syst. Man Cybern. Pt B (Cybern.) 36 (1) (2006) 96–105.

[3] Pantic M., Rothkrantz L.J.M., Automatic analysis of facial expressions: the state of the art, IEEE Trans. Pattern Anal. Mach. Intell. 22 (12) (2000) 1424–1445.

[4] Zhong Y., Zhao M., Wang Y., et al., Recognition of emotions using multimodal physiological signals and an ensemble deep learning model, Comput. Methods Programs Biomed. 140 (2017) 93–110 C.

[5] Wang X.W., Nie D., Lu B.L., Emotional state classification from EEG data using machine learning approach, Neurocomputing 129 (4) (2014) 94–106.

[6] Petrantonakis P.C., Hadjileontiadis L.J., A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition, IEEE Trans. Inf. Technol. Biomed. 15 (5) (2011) 737–746.

[7] Li X., Hu B., Zhu T., et al., towards affective learning with an EEG feedback approach,, Proc. of the 1st Acm International Workshop on Multimedia Technologies for Distance Learning, ACM, 2009, pp. 33–38.

[8] Picard R.W., Vyzas E., Healey J., Toward machine emotional intelligence: analysis of affective physiological state, IEEE Trans. Pattern Anal. Mach. Intell. 23 (10) (2001) 1175–1191 Oct.

[9] Brady K., Gwon Y., Khorrami P., et al., Multi-modal audio, video and physiological sensor learning for continuous emotion prediction, Proc. of the 6th International Workshop on Audio/Visual Emotion Challenge, ACM, 2016, pp. 97–104.

[10] Chanel G., Kronegg J., Grandjean D., et al., Emotion Assessment: Arousal Evaluation Using EEG's and Peripheral Physiological Signals, Springer, Berlin, Heidelberg, 2006, pp. 530–537.

[11] Koelstra S., Muhl C., Soleymani M., Lee J.-S., Yazdani A., Ebrahimi T., Pun T., Nijholt A., Patras I., DEAP: a database for emotion analysis using physiological signals, IEEE Trans. Affect. Comput. 3 (1) (2012) 18–31.

[12] Schmidt L.A., Trainor L.J., Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions, Cogn. Emot. 15 (4) (2001) 487–500.

[13] Wagner J., Kim J., Andre E., From physiological signals to emotions: implementing and comparing selected methods for feature extraction and classification, Proceeding of IEEE International Conference on Multimedia and Expo, 2005, pp. 940–943.

[14] Zheng W.L., Lu B.L., Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks, IEEE Trans. Auton. Ment. Dev. 7 (3) (2015) 162–175.

[15] Zheng W.L., Zhu J.Y., Lu B.L., Identifying stable patterns over time for emotion recognition from EEG, IEEE Trans. Affect. Comput. (2018) early access, doi:10.1109/TAFFC.2017.2712143.

[16] Zheng W.L., Guo H.T., Lu B.L., Revealing critical channels and frequency bands for emotion recognition from EEG with deep belief network, Proceeding of the 7th IEEE/EMBS International Conference on Neural Engineering (NER), 2015, pp. 154–157.

[17] Zheng W.-L., Zhu J.-Y., Peng Y., Lu B.-L., EEG-based emotion classification using deep belief networks,, Jul. 2014, pp. 1–6.

[18] Chen P., Zhang J., Performance comparison of machine learning algorithms for EEG-signal-based emotion recognition, in: Lintas A., Rovetta S., Verschure P., Villa A. (Eds.), ICANN2017, Part I, LNCS, 10613, Springer Int. Publ. AG, 2017, pp. 208–216.

[19] Chen P., Zhang J.H., Wen Z., Xia J., Li J., Emotion recognition of EEG based on kernel spectral regression and random forest algorithms, J. East China Univ. Sci. Technol. 44 (5) (2018) 744–751.

[20] Yang Z., Emotion recognition based on nonlinear features of skin conductance response, J. Inf. Comput. Sci. 10 (12) (2013) 3877–3887.

[21] Cheng J., Liu G., Yang Z., Construction of human-computer affective interaction assistant, Adv. Inf. Sci. Serv. Sci. 4 (17) (2012) 83–90.

[22] Yin Z., Liu L., Liu L., Zhang J.H., Wang Y.G., Dynamical recursive feature elimination technique for neurophysiological signal-based emotion recognition,, Cogn. Technol. Work 19 (2017) 667–685.

[23] Yan F., The research on material selection algorithm design with improved OWA in affective regulation system based on human-computer interaction, J. Inf. Comput. Sci. 10 (14) (2014) 4477–4486.

[24] Nummenmaa L., Glerean E., Hari R., Hietanen J.K., Bodily maps of emotions, Proc. Natl. Acad. Sci. 111 (2) (2014) 646–651.

[25] Strongman K.T., The Psychology of Emotion: From Everyday Life to Theory, John Wiley & Sons, 2003 5th ed.

[26] Picard R.W., Affective Computing, MIT Press, 1997.

[27] Yin Z., Wang Y., Liu L., Zhang W., Zhang J.H., Cross-subject EEG feature selection for emotion recognition using transfer recursive feature elimination, Front. Neurorobot 11 (2017) 1–16.

[28] Frantzidis C.A., Bratsas C., Papadelis C.L., Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli, IEEE Trans. Inf. Technol. Biomed. 14 (3) (2010) 589–597.

[29] Lang P.J., Bradley M.M., Cuthbert B.N., *International A*ffective *P*icture *S*ystem *(IAPS): technical M*anual *and A*ffective *R*atings, NIMH Center for the Study of Emotion and Attention, 1997, pp. 39–58.

[30] Lange C.G., The emotions: a psychophysiological study (I.A. Haupt, Trans. from the authorized German translation of H. Kurella; original work published 1885), in: Dunlap K. (Ed.), The Emotions, Williams & Wilkins Company, Baltimore, 1922, pp. 33–90 (vol. I).

[31] Cannon W., The James-Lange theory of emotions: a critical examination and an alternative theory, Am. J. Psychol. 39 (1927) 106–124.

[32] Dalgleish T., The emotional brain, Nat. Rev. Neurosci. 5 (7) (2004) 582–589.

[33] Maclean P.D., Psychosomatic disease and the visceral brain: recent developments bearing on the Papez theory of emotion, Psychosom. Med. 11 (6) (1949) 338–353.

[34] Yin Z., Zhang J.H., Task-generic mental fatigue recognition based on neurophysiological signals and dynamical deep extreme learning machine, Neurocomputing 283 (2018) 266–281.

[35] Berger H., Uber das elektroenkephalogramm des menchen, Eur. Arch. Psychiatry Clin. Neurosci. 87 (1929) 527–570.

[36]

Chen P., Zhang J.H., Performance comparison of machine learning algorithms for EEG-signal-based emotion recognition, Proc. of 26th International Conference on Artificial Neural Networks (ICANN17), 2017, pp. 11–15 Sep.Alghero, Sardinia, Italy.

[37] Pantic M., Valstar M., Rademaker R., Maat L., Web-based database for facial expression analysis, Proceeding Int'l Conference Multimedia and Expo, 2005, pp. 317–321.

[38] Douglas-Cowie E., Cowie R., Schro¨der M., A new emotion database: considerations, sources and scope, Proc. Int'l Symp. Computer Architecture, 2000, pp. 39–44.

[39] Gunes H., Piccardi M., A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior, Proc. 18th Int'l Conf. Pattern Recognit. 1 (2006) 1148–1153.

[40] Fanelli G., Gall J., Romsdorfer H., Weise T., Van Gool L., A 3-D audio-visual corpus of affective communication, IEEE Trans. Multimedia 12 (6) (2010) 591–598 Oct.

[41] Grimm M., Kroschel K., Narayanan S., The Vera am Mittag German audio-visual emotional speech database, Proc. Int'l Conf. Multimedia and Expo (2008) 865–868.

[42] Healey J.A., Wearable and Automotive Systems For Affect Recognition From Physiology, MIT, 2000 phd dissertation.

[43] Healey J.A., Picard R.W., Detecting stress during real-world driving tasks using physiological sensors, IEEE Trans. Intell. Transport. Syst. 6 (2) (2005) 156–166 June.

[44] Ekman P., The argument and evidence about universals in facial expressions of emotion, Handbook of Social Psychophysiol., John Wiley & Sons, 1989, pp. 143–164.

[45] Savran A., Ciftci K., Chanel G., Mota J.C., Viet L.H., Sankur B., Akarun L., Caplier A., Rombaut M., Emotion detection in the loop from brain signals and facial images, Proc. Enterface (2006) July.

[46] Lang P., Bradley M., Cuthbert B., Technical Report A-8, Univ. of Florida, 2008.

[47] Lang P., Greenwald M., Bradely M., Hamm A., Looking at pictures – Affective, facial, visceral, and behavioral reactions, Psychophysiology 30 (3) (May 1993) 261–273.

[48] Kim J., Andre´ E., Emotion recognition based on physiological changes in music listening, IEEE Trans. Pattern Anal. Mach. Intell. 30 (12) (2008) 2067–2083 Dec.

[49] Wang J., Gong Y., Recognition of multiple drivers' emotional state, Proc. Int'l Conf. Pattern Recognit. (2008) 1–4.

[50] Lisetti C.L., Nasoz F., Using noninvasive wearable computers to recognize human emotions from physiological signals, EURASIP J. Appl. Process. (1) (2004) 1672–1687 (2004)Jan.

[51] Chanel G., Kierkels J., Soleymani M., Pun T., Short-term emotion assessment in a recall paradigm, Int'l J. Hum.-Comput. Stud. 67 (8) (2009) 607–627 Aug.

[52] Liu D., Automatic mood detection from acoustic music data, Proc. Int'l Conf. Music Information Retrieval, 2003, pp. 13–17.

[53] Lu L., Liu D., Zhang H.-J., Automatic mood detection and tracking of music audio signals, IEEE Trans. Audio Speech Language Process. 14 (1) (2006) 5–18 Jan.

[54] Yang Y.-H., Chen H.H., Music emotion ranking,, Proc. Int'l Conf. Acoustics, Speech, and Signal Processing, 2009, pp. 1657–1660.

[55] Koelstra S., Yazdani A., Soleymani M., Mu¨ hl C., Lee J.-S., Nijholt A., Pun T., Ebrahimi T., Patras I., Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos, Proc. Brain Informatics, 2010, pp. 89–100.

[56] Kim J., Bimodal emotion recognition using speech and physiological changes, INTECH, Open Access Publisher, 2007.

[57] Bailenson J.N., Pontikakis E.D., Mauss I.B., Gross J.J., Jabon M.E., Hutcherson C.A., Nass C., John O., Real-time classification of evoked emotions using facial feature tracking and physiological responses, Int. J. Hum.-Comput. Stud. 66 (5) (2008) 303–317.

[58] Khalili Z., Moradi M.H., Emotion recognition system using brain and peripheral signals: using correlation dimension to improve the results of EEG, Proc. of 2009 IEEE Int'l Joint Conf. on Neural Networks, 2009, pp. 1571–1575.

[59] Kim J., Lingenfelser F., Ensemble approaches to parametric decision fusion for bimodal emotion recognition, Biosignals (2010) 460–463.

[60] Chanel G., Rebetez C., Bétrancourt M., Pun T., Emotion assessment from physiological signals for adaptation of game difficulty, IEEE Trans. Syst. Man Cybern. Part A 41 (6) (2011) 1052–1063.

[61] Walter S., Scherer S., Schels M., Glodek M., Hrabal D., Schmidt M., Böck R., Limbrecht K., Traue H.C., Schwenker F., Multimodal emotion classification in naturalistic user behavior, in: human-Computer interaction, Towards Mobile and Intelligent Interaction Environments, Springer, 2011, pp. 603–611.

[62] Hussain M., Monkaresi H., Calvo R.A., Combining classifiers in multimodal affect detection,, Proc. of the 10th Australasian Data Mining Conf.-Volume 134, Australian Computer Society, Inc, 2012, pp. 103–108.

[63] Monkaresi H., Hussain M.S., Calvo R.A., Classification of affects using head movement, skin color features and physiological signals,, Proc. of 2012 IEEE Int'l Conf. on Systems, Man, and Cybernetics (SMC), 2012, pp. 2664–2669.

[64] Soleymani M., Pantic M., Pun T., Multimodal emotion recognition in response to videos, IEEE Trans. Affect. Comput 3 (2) (2012) 211–223.

[65] Wang S., Zhu Y., Wu G., Ji Q., Hybrid video emotional tagging using users' EEG and video content, Multimed. Tools Appl. 72 (2) (2014) 1257–1283.

[66] Yin Z., Zhao M., Wang Y., et al., Recognition of emotions using multimodal physiological signals and an ensemble deep learning model, Comput. Methods Programs Biomed. 140 (2017) 93–110.

[67] Petrantonakis P.C., Hadjileontiadis L.J., A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition, IEEE Trans. Inf. Technol. Biomed. 15 (5) (2011) 737–746.

[68] Daimi S.N., Saha G., Classification of emotions induced by music videos and correlation with participants' rating, Expert Syst. Appl. 41 (13) (2014) 6057–6065.

[69] Yoon H.J., Chung S.Y., EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm, Comput. Biol. Med. 43 (12) (2013) 2230–2237.

[70] Rousseeuw P.J., Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Computat. Appl. Math. 20 (20) (1987) 53–65.

[71] Caliński T., Harabasz J., A dendrite method for cluster analysis, Communicat. Stat. 3 (1) (1974) 1–27.

[72] Subasi A., EEG signal classification using wavelet feature extraction and a mixture of expert model, Expert Syst. Appl. 32 (4) (2007) 1084–1093.

[73] Zhang C., Wang H., Fu R., Automated detection of driver fatigue based on entropy and complexity measures, IEEE Trans. Intell. Transport. Syst. 15 (1) (2014) 168–177.

[74] Vijith V.S., Jacob J.E., Iype T., et al., Epileptic seizure detection using nonlinear analysis of EEG, Proc. of Int. Conf. on Inventive Computation Technologies, 2016, pp. 1–6.

[75] Guido R.C., A tutorial review on entropy-based handcrafted feature extraction for information fusion, Inf. Fusion 41 (2018) 161–175.

[76] Pincus S.M., Approximate entropy as a measure of system complexity, Proc. Nat. Acad. Sci. 88 (6) (1991) 2297–2301.

[77] Richman J.S., Moorman J.R., Physiological time-series analysis using approximate entropy and sample entropy, Am. J. Physiol.-Heart Circ. Physiol. 278 (6) (2000) H2039-H2049.

[78] Yan S., Xu D., Zhang B., Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Trans. Pattern Anal. Mach. Intell. 29 (1) (2007) 40–51.

[79] Cai D., He X., Han J., Speed up kernel discriminant analysis, VLDB J. 20 (2011) 21–33.

[80] Golub G.H., Van Loan C.F., Matrix Computations, JHU Press, 2012.

[81] Zitnik M., Nguyen F., Wang B., et al., Machine learning for integrating data in biology and medicine: principles, practice, and opportunities, Inf. Fusion 50 (2019) 71–91.

[82] Kira K., Rendell L.A., A practical approach to feature selection, Proc. of the 9th Int. Workshop on Machine Learning, Aberdeen, Scotland, UK, 1992, pp. 249–256 July 1-3.

[83] Kononenko I., Estimating attributes: analysis and extensions of relief, Proc. of European Conf. On Machine Learning, Berlin, Heidelberg, Springer, 1994, pp. 171–182.

[84] Peng H., Long F., Ding C., Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.

[85] Kim K.H., Bang S.W., Kim S.R., Emotion recognition system using short-term monitoring of physiological signals, Med. Biol. Eng. Comput. 42 (3) (2004) 419–427.

[86] Zhang J., Chen P., Nichele S., Yazidi A., Emotion recognition using time-frequency analysis of EEG signals and machine learning, Proc of 2019 IEEE Symposium Series on Computational Intelligence (IEEE SSCI 2019), Xiamen, China, 2019 Dec. 6-9(accepted).

[87] Breiman L., Random forests, Mach. Learn. 45 (2001) 5–32.

[88] Zheng W.L., Lu B.L., Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks, IEEE Trans. Auton. Ment. Dev. 7 (3) (2015) 162–175.

[89] Cowie R., Douglas-Cowie E., Tsapatsoulis N., Votsis G., Kollias S., Fellenz W., Taylor J.G., Emotion recognition in human-computer interaction, IEEE Signal Process. Mag. 18 (2001) 32–80.

[90] Levenson R.W., Ekman P., Heider P., Friesen W.V., Emotion and autonomic nervous system activity in the Minangkabau of West Sumatra, J. Personal. Soc. Psychol. 62 (1992) 972–988.

[91] Abdelwahab M., Busso C., Domain adversarial for acoustic emotion recognition, IEEE/ACM Trans. Audio Speech Language Process. 26 (2018) 2423–2435.

[92] Chen J., Hu B., Moore P., Zhang X., Ma X., Electroencephalogram-based emotion assessment system using ontology and data mining techniques, Appl. Soft Comput. 30 (2015) 663–674.

[93] Li C., Xu C., Feng Z., Analysis of physiological for emotion recognition with the IRS model, Neurocomputing 178 (2016) 103–111.

[94]

Verma G., Tiwary U., Multimodal fusion framework: a multiresolution approach for emotion classification and recognition from physiological signals, Neuroimage 102 (2014) 162–172.

[95] Yoon H.-J., Chung S.-Y., EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm, Comput. Biol. Med. 43 (2013) 2230–2237.

[96] Petrantonakis P.C., Hadjileontiadis L.J., A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition, IEEE Trans. Inf. Technol. Biomed. 15 (5) (2011) 737–746.

[97] Wang K., An N., Li B.N., Zhang Y., Li L., Speech emotion recognition using Fourier parameters, IEEE Trans. Affect. Comput. 6 (2015) 69–75.

[98] Wang F., Mao M., Duan L., Huang Y., Li Z., Zhu C., Intersession instability in fNIRS-based emotion recognition, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (2018) 1324–1333.

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

[100] Alam F., Riccardi G., Predicting personality traits using multimodal information, Proc. of the 2014 ACM Multi Media on Workshop on Computational Personality Recognition, ACM, 2014, pp. 15–18.

[101] Sarkar C., Bhatia S., Agarwal A., Li J., Feature analysis for computational personality recognition using youtube personality data set, Proc. of the 2014 ACM Multi Media On Workshop On Computational Personality Recognition, ACM, 2014, pp. 11–14.

[102] Poria S., Cambria E., Gelbukh A., Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis, Proc. EMNLP (2015) 2539–2544.

[103] Poria S., Cambria E., Hussain A., Huang G.-B., Towards an intelligent framework for multimodal affective data analysis, Neural Netw. 63 (2015) 104–116.

[104] Siddiquie B., Chisholm D., Divakaran A., Exploiting multimodal affect and semantics to identify politically persuasive web videos, Proc. of the 2015 ACM On International Conference On Multimodal Interaction, ACM, 2015, pp. 203–210.

[105] Poria S., Cambria E., Howard N., Huang G.-B., Hussain A., Fusing audio, visual and textual clues for sentiment analysis from multimodal content, Neurocomputing 174 (2016) 50–59.

[106] Zhalehpour S., Onder O., Akhtar Z., Erdem C.E., BAUM-1: a spontaneous audio-visual face database of affective and mental states, IEEE Trans. Affect. Comput. 8 (2017) 300–313.

[107]

Eyben F., et al., The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing, IEEE Trans. Affect. Comput. 7 (2016) 190–202.

[108] Wu C., Liang W., Emotion recognition of affective speech based on multiple classifiers using acoustic-prosodic information and semantic labels, IEEE Trans. Affect. Comput. 2 (2011) 10–21.

[109] Yoon W., Park K., Building robust emotion recognition system on heterogeneous speech databases, IEEE Trans. Consum. Electron. 57 (2011) 747–750.

[110] Wu]C, Wei W., Lin J., Lee W., Speaking effect removal on emotion recognition from facial expressions based on eigenface conversion, IEEE Trans. Multimed. 15 (2013) 1732–1744.

[111] Schuller B., Cross-corpus acoustic emotion recognition: variances and strategies, IEEE Trans. Affect. Comput. 1 (2010) 119–131.

[112] Bisio I., Delfino A., Lavagetto F., Marchese M., Sciarrone A., Gender-driven emotion recognition through speech signals for ambient intelligence applications, IEEE Trans. Emerg. Top. Comput. 2 (2013) 244–257.

[113] Park J., Kim J., Oh Y., Feature vector classification based speech emotion recognition for service robots, IEEE Trans. Consum. Electron. 55 (2009) 1590–1596.

[114] Chen Y., Wang J., Yang Y., Chen H.H., Component tying for mixture model adaptation in personalization of music emotion recognition, IEEE/ACM Trans. Audio Speech Language Process. 25 (2017) 1409–1420.

[115] Guo J., et al., Dominant and complementary emotion recognition from still images of faces, IEEE Access 6 (2018) 26391–26403.

[116] Jing H., Xie L., Dan L., He Z., Wang Z., Cognitive emotion model for eldercare robot in smart home, China Commun. 12 (2015) 32–41.

[117] Yan J., Zheng W., Xu Q., Lu G., Li H., Wang B., Sparse kernel reduced-rank regression for bimodal emotion recognition from facial expression and speech, IEEE Trans. Multimed. 18 (2016) 1319–1329.

[118] Petrantonakis P.C., Hadjileontiadis L.J., Emotion recognition from EEG using higher order crossings, IEEE Trans. Inf. Technol. Biomed. 14 (2010) 186–197.

[119] Shojaeilangari S., Yau W., Nandakumar K., Li J., Teoh E.K., Robust representation and recognition of facial emotions using extreme sparse learning, IEEE Trans. Image Process. 24 (2015) 2140–2152.

[120] Chakraborty A., Konar A., Chakraborty U.K., Chatterjee A., Emotion recognition from facial expressions and its control using fuzzy logic, IEEE Trans. Syst. Man Cybernet.—Part A 39 (2009) 726–743.

[121] Ferreira P.M., Marques F., Cardoso J.S., Rebelo A., Physiological inspired deep neural networks for emotion recognition, IEEE Access 6 (2018) 53930–53943.

[122] Zhang Y., et al., Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation, IEEE Access 4 (2016) 8375–8385.

[123] Li M., Lu Q., Long Y., Gui L., Inferring affective meanings of words from word embedding, IEEE Trans. Affect. Comput. 8 (2017) 443–456.

[124] Albornoz E.M., Milone D.H., Emotion recognition in never-seen languages using a novel ensemble method with emotion profiles, IEEE Trans. Affect. Comput. 8 (2017) 43–53.

[125] Xu B., Fu Y., Jiang Y., Li B., Sigal L., Heterogeneous knowledge transfer in video emotion recognition, attribution and summarization, IEEE Trans. Affect. Comput. 9 (2018) 255–270.

[126] Quan C., Ren F., Weighted high-order hidden Markov models for compound emotions recognition in text, Inf. Sci. (Ny) 329 (2016) 581–596.

[127] Karyotis C., Doctor F., Iqbal R., James A., Chang V., A fuzzy computational model of emotion for cloud based sentiment analysis, Inf. Sci. (Ny) 433–434 (2018) 448–463.

[128] He X., Zhang W., Emotion recognition by assisted learning with convolutional neural networks, Neurocomputing 291 (2018) 187–194.

[129] Jain N., Kumar S., Kumar A., Shamsolmoali P., Zareapoor M., Hybrid deep neural networks for face emotion recognition, Pattern Recognit. Lett. 115 (2018) 101–106.

[130] Lin Y., et al., EEG-based emotion recognition in music listening, IEEE Trans. Biomed. Eng. 57 (2010) 1798–1806.

[131] Yang Y., Wu Q.M.J., Zheng W., Lu B., EEG-based emotion recognition using hierarchical network with subnetwork nodes, IEEE Trans. Cogn. Dev. Syst. 10 (2018) 408–419.

[132] Katsigiannis S., Ramzan N., DREAMER: a database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices, IEEE J. Biomed. Health Inf. 22 (2018) 98–107.

[133] Soleymani M., Asghari-Esfeden S., Fu Y., Pantic M., Analysis of EEG signals and facial expressions for continuous emotion detection, IEEE Trans. Affect. Comput. 7 (2016) 17–28.

[134] Subramanian R., Wache J., Abadi M.K., Vieriu R.L., Winkler S., Sebe N., ASCERTAIN: emotion and personality recognition using commercial sensors, IEEE Trans. Affect. Comput. 9 (2018) 147–160.

[135] Zacharatos H., Gatzoulis C., Chrysanthou Y.L., Automatic emotion recognition based on body movement analysis: a survey, IEEE Comput. Graph. Appl. 34 (2014) 35–45.

[136] Wang J., Lee Y., Chin Y., Chen Y., Hsieh W., Hierarchical Dirichlet process mixture model for music emotion recognition, IEEE Trans. Affect. Comput. 6 (2015) 261–271.

[137] Kim J., André E., Emotion recognition based on physiological changes in music listening, IEEE Trans. Pattern Anal. Mach. Intell. 30 (12) (2008) 2067–2083.

[138] Mariooryad S., Busso C., Exploring cross-modality affective reactions for audiovisual emotion recognition, IEEE Trans. Affect. Comput. 4 (2013) 183–196.

[139] Zheng W., Xin M., Wang X., Wang B., A novel speech emotion recognition method via incomplete sparse least square regression, IEEE Signal Process. Lett. 21 (2014) 569–572.

[140] Wagner J., Andre E., Lingenfelser F., Kim J., Exploring fusion methods for multimodal emotion recognition with missing data, IEEE Trans. Affect. Comput. 2 (2011) 206–218.

[141] Valstar M.F., Mehu M., Jiang B., Pantic M., Scherer K., Meta-analysis of the first facial expression recognition challenge, IEEE Trans. Syst. Man Cybern. Part B (Cybernetics) 42 (2012) 966–979.

[142] Fukushima K., Neocognitron: a self-organizing neural network for a mechanism of pattern recognition unaffected by shift in position, Biol. Cybern. 36 (4) (1980) 193–202.

[143] Hinton G.E., Salakhutdinov R.R., Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507 July.

[144] Bengio Y., Lamblin P., Popovici D., Larochelle H., Greedy layerwise training of deep networks, Proc. Adv. Neural Inf. Process. Syst., Vancouver, BC, Canada, 2007, pp. 153–160.

[145] Vincent P., Larochelle H., Lajoie I., Bengio Y., Manzagol P.-A., Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion, J. Mach. Learn. Res. 11 (2010) 3371–3408 Dec.

[146] Kasun L.L.C., Zhou H., Huang G.-B., Vong C.-M., Representational learning with extreme learning machine for big data, IEEE Intell. Syst. 28 (6) (2013) 31–34 Nov.

[147] Chen M., Weinberger K., Xu Z., Sha F., Marginalizing stacked autoencoders, J. Mach. Learn. Res. 22 (2) (2015) 191–194.

[148] Cao J., et al., Landmark recognition with sparse representation classification and extreme learning machine, J. Frankl. Inst. 352 (10) (2015) 4528–4545 Oct.

[149] Yang Y., Wu Q.M.J., Multilayer extreme learning machine with subnetwork nodes for representation learning, IEEE Trans. Cybern. 46 (11) (2016) 2570–2583 Nov.

[150] Cao J., Zhang K., Luo M., Yin C., Lai X., Extreme learning machine and adaptive sparse representation for image classification, Neural Netw. 81 (2016) 91–102 Sep.

[151] Zhang J., Wu Y., Automatic sleep stage classification of single-channel EEG by using complex-valued convolutional neural network, Biomed. Eng./Biomed. Tech. 63 (2) (2017) 177–190.

[152] Sarkar S., Reddy K., Dorgan A., Fidopiastis C., Giering M., Wearable EEG-based activity recognition in PHM-related service environment via deep learning, Int. J. Prognost. Health Manag. 7 (2016) 1–10 021ISSN2153-2648.

[153] Y. Gao, H.J. Lee, R.M. Mehmood, Deep learning of EEG signals for emotion recognition, in Proc. of 2015 IEEE Int. Conf. on Multimedia & Expo Workshops (ICMEW), pp. 1-5.

[154] Tripathi S., Acharya S., Sharma R.D., Mittal S., Bhattacharya S., Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset, Proc. of 29th AAAI Conf. on Innovative Applications (IAAI-17), 2017, pp. 4746–4752.

[155] Alhagry S., Fahmy A.A., El-Khoribi R.A., Emotion recognition based on EEG using LSTM recurrent neural network, Emotion 8 (10) (2017) 355–358.

[156] Xu T., Zhou Y., Wang Z., Peng Y., Learning emotions EEG-based recognition and brain activity: a survey study on bci for intelligent tutoring system, Procedia Comput. Sci. 130 (2018) 376–382.

[157] Tzirakis P., Trigeorgis G., Nicolaou A.M., Schuller B.W., Zafeiriou S., End-to-end multimodal emotion recognition using deep neural networks, IEEE J. Sel. Topics Signal Process. 11 (2017) 1301–1309.

[158] Li S., Deng W., Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition, IEEE Trans. Image Process. 28 (2019) 356–370.

[159] Attabi Y., Dumouchel P., Anchor models for emotion recognition from speech, IEEE Trans. Affect. Comput. 4 (2013) 280–290.

[160] Zhang S., Zhang S., Huang T., Gao W., Tian Q., Learning affective features with a hybrid deep model for audio–visual emotion recognition, IEEE Trans. Circuits Syst. Video Technol. 28 (2018) 3030–3043.

[161] Deng J., Zhang Z., Eyben F., Schuller B., Autoencoder-based unsupervised domain adaptation for speech emotion recognition, IEEE Signal Process. Lett. 21 (2014) 1068–1072.

[162] Xia R., Liu Y., A multi-task learning framework for emotion recognition using 2 D continuous space, IEEE Trans. Affect. Comput. 8 (2017) 3–14.

[163] Tariq U., et al., Recognizing emotions from an ensemble of features, IEEE Trans. Syst. Man Cybern. 42 (2012) 1017–1026 Part B (Cybernetics).

[164] Chen L., Zhou M., Su W., Wu M., She J., Hirota K., Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction, Inf. Sci. (Ny) 428 (2018) 49–61.

[165] Zhang Q., Chen X., Zhan Q., Yang T., Xia S., Respiration-based emotion recognition with deep learning, Comput. Ind. 92–93 (2017) 84–90.

[166] Hossain M.S., Muhammad G., Emotion recognition using deep learning approach from audio-visual emotional big data, Inf. Fusion 49 (2019) 69–78.

[167] Hassan M.M., Alam M.G.R., Uddin M.Z., Huda S., Almogren A., Fortino G., Human emotion recognition using deep belief network architecture, Inf. Fusion 51 (2019) 10–18.

[168] Kratzwald B., Ilić S., Kraus M., Feuerriegel S., Prendinger H., Deep learning for affective computing: text-based emotion recognition in decision support, Decis. Support Syst. 115 (2018) 24–35.

[169] Fayek H.M., Lech M., Cavedon L., Evaluating deep learning architectures for speech emotion recognition, Neural Netw. 92 (2017) 60–68.

[170] Jain D.K., Shamsolmoali P., Sehdev P., Extended deep neural network for facial emotion recognition, Pattern Recognit. Lett. 120 (2019) 69–74.

[171] Santhoshkumar R., Geetha M.K., Deep learning approach for emotion recognition from human body movements with feedforward deep convolutional neural networks, Procedia Comput. Sci. 152 (2019) 158–165.

[172] Liu Z.-T., Xie Q., Wu M., Cao W.-H., Mei Y., Mao J.-W., Speech emotion recognition based on an improved brain emotion learning model, Neurocomputing 309 (2018) 145–156.

[173] He X., Zhang W., Emotion recognition by assisted learning with convolutional neural networks, Neurocomputing 291 (2018) 187–194.

[174] Chatterjee A., Gupta U., Chinnakotla M.K., Srikanth R., Galley M., Agrawal P., Understanding emotions in text using deep learning and big data, Comput. Hum. Behav. 93 (2019) 309–317.

[175] D. Jiang, Y. Cui, X. Zhang, P. Fan, I. Ganzalez, H. Sahli, Audio visual emotion recognition based on triple-stream dynamic Bayesian network models, in: D'Mello (Ed.), ACII,Part I, LNCS6974, 2011, pp. 609-618.

[176] Kim Y., Lee H., Provost E.M., Deep learning for robust feature generation in audiovisual emotion recognition, Proc. of the IEEE Int'l Conf. on Acoustics, Speech and Signal Processing, Vancouver, BC, 2013, pp. 3687–3691 2013.

[177] Kahou S.E., Bouthillier X., Lamblin P., et al., EmoNets: multimodal deep learning ap- proaches for emotion recognition in video, J. Multimodal. User Interf. 10 (2) (2016) 99–111 June.

[178] Hossain M.S., Muhammad G., Alhamid M.F., Song B., Al-Mutib K., Audio-visual emotion recognition using big data towards 5 G, Mobile Netw. Appl. 221 (5) (2016) 753–763 October.

[179] Hossain M.S., Muhammad G., Audio-visual emotion recognition using multi-directional regression and ridgelet transform, J. Multimodal. User Interf. 10 (4) (2016) 325–333.

[180] Hossain M.S., Muhammad G., Emotion-aware connected healthcare big data towards 5 G, IEEE Internet Things J. 5 (4) (2018) 2399–2406 August, doi:10.1109/JIOT.2017.2772959.

[181] Ranganathan H., Chakraborty S., Panchanathan S., Multimodal emotion recognition using deep learning architectures,, Proc. of the IEEE Winter Conf. on Applications of Computer Vision (WACV), Lake Placid, NY, 2016, pp. 1–9 2016.

[182] Kaya H., Gürpinar F., Salah A.A., Video-based emotion recognition in the wild using deep transfer learning and score fusion, Image Vis. Comput. 65 (2017) 66–75.

[183] Zhang J., Yin Z., Wang R., Pattern classification of instantaneous cognitive task-load through GMM clustering, Laplacian eigenmap and ensemble SVMs, IEEE/ACM Trans. Comput. Biol. Bioinformat. 14 (4) (2016) 947–965.

[184] Zhang J., Yin Z., Wang R., Recognition of mental workload levels under complex human-machine collaboration by using physiological features and adaptive support vector machines, IEEE Trans. Hum.-Mach. Syst. 45 (2) (2015) 200–214.

[185] Zhang J., Yin Z., Wang R., Nonlinear dynamic classification of momentary mental workload using physiological features and NARX-model-based least-squares support vector machines, IEEE Trans. Hum.-Mach. Syst. 47 (4) (2017) 536–549.

[186] Yin Z., Zhang J., Cross-session classification of mental workload levels using EEG and an adaptive deep learning model, Biomed. Signal Process. Control 33 (2017) 30–47.

[187] Yang S., Yin Z., Wang Y., Zhang W., Wang Y., Zhang J., Assessing cognitive mental workload via EEG signals and an ensemble deep learning classifier based on denoising autoencoders, Comput. Biol. Med. 109 (2019) 159–170.

[188] Yin Z., Zhao M., Zhang W., Wang Y., Wang Y., Zhang J., Physiological-signal-based mental workload estimation via transfer dynamical autoencoders in a deep learning framework, Neurocomputing 347 (2019) 212–229.

[189] Zhuang X., Rozgic V., Crystal M., Compact unsupervised EEG response representation for emotion recognition, Proc. of IEEE-EMBS Int. Conf. on Biomedical and Health Informatics, 2014, pp. 736–739.

[190] Poria S., Cambria E., Bajpai R., Hussain A., A review of affective computing: from unimodal analysis to multimodal fusion, Inf. Fusion 37 (2017) 98–125.

[191] Kumar S., Yadava M., Roy P.P., Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction, Inf. Fusion 52 (2019) 41–52.

[192] Anderson K., Mcowan P.W., A real-time automated system for the recognition of human facial expressions, IEEE Trans. Syst. Man Cybern. Part B Cybern 36 (1) (2006) 96–105.

[193] Gravina R., Alinia P., Ghasemzadeh H., Fortino G., Multi-sensor fusion in body sensor networks: state-of-the-art and research challenges, Inf. Fusion 35 (2017) 68–80.

[194] Khaleghi B., Khamis A., Karray F.O., Razavi S.N., Multisensor data fusion: a review of the state-of-the-art, Inf. Fusion 14 (2013) 28–44.

[195] Nweke H.F., Wah T.Y., Mujtaba G., Al-garadi M.A., Data fusion and multiple classifier systems for human activity detection and health monitoring: review and open research directions, Inf. Fusion 46 (2019) 147–170.

[196] Wang X.W., Nie D., Lu B.L., Emotional state classification from EEG data using machine learning approach, Neurocomputing 129 (2014) 94–106.

[197] Zitnik M., Nguyen F., Wang B., Leskovec J., Goldenberg A., Hoffman M.M., Machine learning for integrating data in biology and medicine: principles, practice, and opportunities, Inf. Fusion 50 (2019) 71–91.

[198] Chen M., Zhang Y., Li Y., Hassan M.M., Alamri A., AIWAC: affective interaction through wearable computing and cloud technology, IEEE Wirel. Commun. Mag. (Feb. 2015) 20–27.

[199] Soleymani M., Lichtenauer J., Pun T., Pantic M., A multi-modal database for affect recognition and implicit tagging, IEEE Trans. Affect. Comput. 3 (1) (2012) 42–55 Jan.-Mar.

[200] Koelstra S., Yazdani A., Soleymani M., Mu¨ hl C., Lee J.-S., Nijholt A., Pun T., Ebrahimi T., Patras I., Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos, Proc. Brain Inf. (2010) 89–100.

[201] Barry R.J., Clarke A.R., Johnstone S.J., Magee C.A., Rushby J.A., EEG differences between eyes-closed and eyes-open resting conditions, Clin. Neurophysiol. 118 (12) (2007) 2765–2773 Dec.

[201] Barry R.J., Clarke A.R., Johnstone S.J., Brown C.R., EEG differences in children between eyes-closed and eyes-open resting conditions, Clin. Neurophysiol. 120 (10) (2009) 1806–1811 Oct.

[203] Cole H., Ray W.J., EEG correlates of emotional tasks related to attentional demands, Int'l J. Psychophysiol. 3 (1) (1985) 33–41 July.

[204] Onton J., Makeig S., High-frequency broadband modulations of electroencephalographic spectra, Front. Hum. Neurosci. 3 (2009) 61 Dec, doi:10.3389/neuro.09.061.2009.

[205] Goncharova I., McFarland D.J., Vaughan J.R., Wolpaw J.R., EMG contamination of EEG: spectral and topographical characteristics, Clin. Neurophysiol. 114 (9) (Sept. 2003) 1580–1593.

[206] Kanjo E., Younis E.M.G., Ang C.S., Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection, Inf. Fusion 49 (2019) 46–56.

[207] Rouast P.V., Adam M., Chiong R., Deep learning for human affect recognition: insights and new development, IEEE Trans. Affect. Comput. 1 (2019) IEEE Early Access article (publication date: Jan, doi:10.1109/TAFFC.2018.2890471.

[208] Hassan M.M., Alam M.G.R., Uddin M.Z., Huda S., Almogren A., Fortino G., Human emotion recognition using deep belief network architecture, Inf. Fusion 51 (2019) 10–18.

[209] Uddin M.Z., Hassan M.M., Activity recognition for cognitive assistance using body sensors data and deep convolutional neural network, IEEE Sens. J. (2018), doi:10.1109/JSEN.2018.2871203.

[210] Jain N., Kumar S., Kumar A., Shamsolmoali P., Zareapoor M., Hybrid deep neural networks for face emotion recognition, Pattern Recognit. Lett. 115 (2018) 101–106.

[211] Hadjidimitriou S.K., Hadjileontiadis L.J., Toward an EEG-based recognition of music liking using time-frequency analysis, IEEE Trans. Biomed. Eng. 59 (12) (2012) 3498–3510.

[212] Atkinson J., Campos D., Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers, Expert Syst. Appl. 7 (2016) 35–41.

[213] Moreira M.W.L., Rodrigues J.J.P.C., Kumar N., Saleem K., Illin I.V., Postpartum depression prediction through pregnancy data analysis for emotion-aware smart systems, Inf. Fusion 47 (2019) 23–31.

[214] Zheng W., Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis, IEEE Trans. Cogn. Dev. Syst. 9 (3) (2017) 281–290.

[215] Jenke R., Peer A., Buss M., Feature extraction and selection for emotion recognition from EEG, IEEE Trans. Affect. Comput. 5 (3) (2014) 327–339, doi:10.1109/TAFFC.2014.2339834.

[216] Maaten L., Hinton G., Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605.

[217] Chatterjee R., Maitra T., Islam S.H., Hassan M.M., Alamri A., Fortino G., A novel machine learning based feature selection for motor imagery EEG signal classification in internet of medical things environment, Future Gener. Comput. Syst. 98 (2019) 419–434 Mar.

[218] Khosrowabadi R., Quek C., Ang K.K., Wahab A., ERNN: a biologically inspired feedforward neural network to discriminate emotion from EEG signal, IEEE Trans. Neural Netw. Learn. Syst. 25 (3) (2014) 609–620 Mar.

[219] Fusi S., Miller E.K., Rigotti M., Why neurons mix: high dimensionality for higher cognition, Curr. Opin. Neurobiol. 37 (2016) 66–74 Apr.

[220]

Rigotti M., et al., The importance of mixed selectivity in complex cognitive tasks, Nature 497 (2013) 585–590 May.

[221] Fredricson B.L., Levenson R.W., Positive emotions speed recovery from the cardiovascular sequelae of negative emotions, Cogn. Emot. 12 (2) (1998) 191–220.

[222] Stemmler G., The autonomic differentiation of emotions revisited: convergent and discriminant validation, Psychophysiology 26 (1989) 617–632.

Highlights

- Several major EEG feature extraction methods are introduced.
- Several major EEG feature reduction methods are introduced.
- Different types of machine learning classifiers for emotion recognition are reviewed.
- Several open problems and future research directions in the area of emotion recognition are identified.