International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

A Comprehensive Study on State-Of-Art Learning Algorithms in Emotion Recognition

Kshirod Sarmah[1*], Hem Chandra Das[2], Sisir Kumar Rajbongshi[3], Swapnanil Gogoi[4]

[1*]Department of Computer Science, Pandit Deendayal Upadhyaya Adarsha Mahavidyalaya (A Govt. Model College), Goalpara, 783124, Assam, INDIA, kshirodsarmah@gmail.com
[2]Department of Computer Science and Technology, Bodoland University, Kokrajhar, 783370, Assam, INDIA, hemchandradas78@gmail.com
[3]Department of Computer Science, Pandit Deendayal Upadhyaya Adarsha Mahavidyalaya (A Govt. Model College), Goalpara, 783124, Assam, INDIA, sisirrajbongshi@gmail.com
[4]GUCDOE, Gauhati University, Guwahati-781014, Assam, INDIA, swapnanil@gauhati.ac.in

***Corresponding Author:** Kshirod Sarmah
*Department of Computer Science, Pandit Deendayal Upadhyaya Adarsha Mahavidyalaya (A Govt. Model College), Goalpara, 783124, Assam, INDIA, kshirodsarmah@gmail.com

ABSTRACT:

The potential uses of emotion recognition in domains like human-robot interaction, marketing, emotional gaming, and human-computer interface have made it a prominent research subject. Better user experiences can result from the development of technologies that can accurately interpret and respond to human emotions thanks to a better understanding of emotions. The use of several sensors and computational algorithms is the main emphasis of this paper's thorough analysis of the developments in emotion recognition techniques. Our results show that using more than one modality improves the performance of emotion recognition when a variety of metrics and computational techniques are used. This paper adds to the body of knowledge by thoroughly examining and contrasting several state-of-art computational techniques and measurements for emotion recognition. The study emphasizes how crucial it is to use a variety of modalities along with cutting-edge machine learning algorithms in order to attain more precise and trustworthy emotion assessment. Additionally, we pinpoint prospective avenues for additional investigation and advancement, including the incorporation of multimodal data and the investigation of innovative features and fusion methodologies. This study contributes to the development of technology that can better comprehend and react to human emotions by offering practitioners and academics in the field of emotion recognition insightful advice.

Keywords: Emotions, Sensors, Emotion recognitions techniques, machine learning, deep learning.

1. INTRODUCTION

Expression of emotion is an important aspect of a human's life. Expression portrays perception about the incidents, human interaction, decision-making, and intelligence given their profound influence onour thoughts, behaviors, and exp eriences as a whole, emotions are vital in human interactions ."The struggle to define emotion in scientific terms is as old as the field of psychology," [1] but there is still no agreemen t on a description of emotion in literary works.Emotion is de fined by some researchers as a fleeting conscious experience marked by strong mental activity, while many psychologists view it as a physiologic state that results from various intern al and external stimulus.

Several modalities have been investigated in the field of sensor-based techniques to capture various components of emotional expressions. One extensively researched approach that has proven useful in identifying emotions is facial expression analysis. Conventional methods such as the Facial Action Coding System (FACS) have yielded important insights on the movements of the face muscles and how they correspond with particular emotions [2]. Utilizing 3D facial models and deep learning approaches to extract fine-grained facial information for more accurate emotion recognition is one recent achievement in this field [3] [4].

In addition to facial expressions, researchers have incorporated other physiological signals to complement the understanding of emotions. Electrodermal activity (EDA), which measures the skin's electrical conductivity, and heart rate variability (HRV), reflecting changes in heart rhythm, have been shown to correlate with emotional states [5][6]. The integration of multiple modalities, such as facial expressions, speech analysis, and physiological signals, has demonstrated improved emotion recognition performance [7]. Moreover, the growing availability of textual data from social media platforms has led to the exploration of natural language processing techniques for sentiment analysis and emotion detection [8] [9].

In computer vision, the simplest example is the classification of an image to a specific class, which means the network is built on top of a function or multiple functions that have the purpose of mapping the image data to a specific class. Regarding the computer vision domain, neural networks have been successfully used in image classification and more specifically, face identification, speech emotion recognition and facial emotion recognition applications.

While sensor-based approaches have provided valuable insights, computational methods, particularly machine learning and deep learning algorithms, have revolutionized emotion recognition. Traditional machine learning models, such as Support Vector Machines (SVM) and Random Forests, have been widely used for emotion classification [10]. However, the emergence of deep learning techniques has yielded substantial advancements in this domain. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have shown exceptional performance in capturing spatial and temporal dependencies in emotional data, leading to improved emotion recognition accuracy [11] [12]. Moreover, the integration of attention mechanisms and transfer learning strategies has further enhanced the robustness and generalizability of deep learning models in emotion recognition tasks [13] [14].

Deep neural networks (DNNs) are the most used machine learning solution by facial expression recognition (FER) systems [15]. DNN uses a system of layers of neurons whose weights are dynamic and changing to match incoming information. Deep learning techniques are used in many FER applications due to the results obtained, results that in some cases exceed the results of the best human subjects. The major advantage of DNN over traditional machine learning techniques is the fact that DNN incorporates the feature extraction step of the input elements, whereas this step is usually performed separately by a domain expert in traditional machine learning techniques [16].

Because of their promising performances, DDNs have been used in emotion recognition more and more. The following DNN varieties are very common, particularly in the field of computer vision: The simplest kind of DNN is the multi-layer perceptron (MLP), which consists of several fully connected layers and can be used to get around deep learning architectures' high processing power requirements. In computer vision, convolutional neural networks (CNNs) are mostly used to automatically extract characteristics from input data in order to do certain tasks like picture categorization. CNN models can extract the high-level representation of the input data by utilizing one or more convolutional layers that are composed of filter-based convolutional processes. Recurrent neural network (RNN) models are widely used in language translation, natural language processing (NLP), speech recognition, and picture captioning. RNN models are appropriate for processing sequential data, such as time series or text. The fact that all network levels share the same parameters and that each layer has its own "memory"—where information is recovered from earlier inputs and utilized to affect the current input and output—are some of the traits that set RNNs apart.

While there have been tremendous advances in sensor-based and computational methods, it is important to acknowledge that challenges remain in achieving a perfect and comprehensive understanding of human emotions. Cultural differences and subjectivity in emotional manifestations provide a substantial challenge to the development of universal emotion identification systems [17]. Moreover, protecting confidentiality and upholding moral principles in the collection and handling of emotional data remains a top concern [18].

Thus, this work aims to provide an in-depth examination of the latest advancements in sensors and computational methods associated with emotion identification. Through a study of recent research publications, we hope to present a current overview of the state of the art in this rapidly evolving field. We discuss the advantages, disadvantages, and prospective directions of emotion recognition in the future, with a focus on recent developments and emerging trends as well as major challenges. The information gained from this review will support the responsible and meaningful integration of emotion detection technology into a variety of sectors, as well as its continued development.

2. Different Methods and Sensors of Emotions Recognition System

The collection of emotion evaluation techniques found in the psychology and computer science literature can be divided into two main categories based on the basic metrics used for emotion identification: While objective measurements rely on techniques for processing and evaluating human body signals, subjective measurements mostly rely on self-reporting methods through the use of self-assessment procedures [19] [20]. There is sometimes a division between self-assessment techniques and behavioral modalities within the subjective measurements themselves. The most commonly used physiological measure techniques are the electrocardiogram (ECG) for measuring heart rate [21], the galvanic skin response (GSR) for measuring skin resistance [22], the electroencephalogram (EEG) for measuring cerebral activity [23], the blood volume pulse (BVP) for measuring blood pressure [24], and the electrooculography (EOG) for measuring eye tracker [25][26].

2.1 Empirical Assessments of Emotions

It is true that a variety of physiological signals, in addition to behavioral ones, can express emotions. These physiological signals are reactions that the body has to emotional events, and they can provide important information about the emotional state of the individual. These signals are produced by the physiological systems of the human body operating normally. Specialized tools and methods are used to measure physiological signals. To precisely record the signals, electrodes, sensors, or other monitoring apparatuses are applied to the relevant body parts. After that, the signals are filtered, amplified, and recorded for later examination. The biosensors that are most frequently utilized to gather physiological signals are the electroencephalogram (EEG), electrocardiogram (ECG), galvanic skin response (GSR), blood volume pulse (PVB), and electrooculography (EOG).

The electrical activity produced by the brain's neurons firing is measured by the EEG. The small electrical impulses generated by the brain's neural activity are detected by electrodes applied to the scalp. EEG measures provide information that can be extracted to acquire insights into an individual's emotional state because numerous studies have demonstrated the involvement of the prefrontal cortex, temporal lobe, and anterior cingulate gyrus in the regulation and control of emotions in the human brain. There are numerous works in the literature on intelligence artificial intelligence that focus on creating EEG-based emotion recognition systems, which have a wide range of potential applications. Brain electrical impulses have been categorized into five distinct frequency bands, or brain waves, for the assessment of human emotions [27]: delta (0–4 Hz), theta (4–8 Hz), alpha (812 Hz), beta (12–30 Hz), and gamma (>30 Hz). These frequency ranges, which show various brain activity states, are frequently seen in EEG recordings.

An ECG gadget measures the electrical activity of the heart in various phases and angles depending on the setup and circumstances. Electrodes are applied to the skin to record the ECG signals over time. ECG signals can be used to assess changes in heart activity brought on by emotional events, even though the main function of ECG in medicine is to identify pathological cardiac problems [28][29]. Research has demonstrated the effectiveness of ECG-based emotion identification systems, which are widely utilized [30].

Skin reaction galvanic Among the several electrodermal signals that can detect electrical alterations in a person's skin that are not under conscious control is GSR [31]. These signals are the outcome of modifications to the skin's electrical conductivity brought on by different stimuli or emotional arousal. It is a physiological metric that represents the skin's perspiration gland activity. In fact, sweat glands release more fluid towards the skin's surface when they become active. For example, an individual's sweat glands tend to generate more perspiration when they are anxious or concerned, which causes larger swings in electrical current.

Electrooculography (EOG) is the clinical term for the measuring and recording of electrical impulses produced by eye movement. In order to identify and examine the variations in electrical potentials that arise with eye movement, electrodes are positioned in close proximity to the eyes. In order to evaluate eye movements, track sleep patterns, diagnose specific eye disorders, and investigate different facets of visual perception and cognitive processes, electroencephalography (EOG) is frequently employed in medical and research contexts. The same theory that underpins EMG, which measures and records the electrical potential produced by muscle cells, also underpins the application of EOG to emotion recognition [32]. EOG is a powerful tool for identifying emotions like melancholy, stress, and surprise because it is mostly dependent on the detection of eye blinks. EOG was frequently employed in research to measure focus and weariness (while studying or driving) [33][34].Additionally, RSP and Skin Temperature Measurements (SKT) are related physiological markers.

2.2 Subjective Measures of Emotions

Subjective measures are very basic procedures that rely on conscious responses and have the benefit of being easily collected. The first modality, self-assessment procedures, is unreliable, and it's possible that someone will misinterpret his sentiments or provide false information in response to uncomfortable queries. People may easily manage behavioral modalities such as posture, voice, text expressions, and facial expressions. For instance, when irritated or upset, a person may seem neutral. As a result, it is impossible to guarantee the reliability.

Self-assessment methods typically use a rating scale or pictograms to help users identify their emotions. These tools may be non-verbal or spoken. Psychologists frequently employ verbal instruments, yet nonverbal ones are utilized when the user needs to feel free to evaluate and recognize his feelings. The instruments come in the form of a questionnaire, which may be filled out in a variety of ways [35]. One option is to use a rating scale, such the Likert scale, or a radio button that allows the emotion to be represented in binary. Pictorial tools called Emocards [36] and Self-Assessment Manikin (SAM) [37] are frequently used to gauge dominance, arousal, and pleasure. Other self-report tools, such as the Positive and Negative Affect Schedule (PANAS) [38] and the Product Emotion Measurement (PrEmo) [39], concentrate on specific emotional state.

One of the most common and effective ways to communicate with people on a personal level is to be able to read their emotions from their posture, voice tone, and facial expressions. According to research, nonverbal cues like body language, facial emotions, and gestures convey two thirds of interpersonal communication while verbal cues like speech convey only one third. It follows that the interest in emotion recognition techniques that rely on the examination of gestures, postures, and facial expressions has grown dramatically.

One of the most popular and natural emotion recognition techniques based on visual sensors is FER. A number of sensors, including cameras, eye trackers, electrocardiograms (ECG), electromyograms (EMG), electroencephalographs (EEG), and others, are utilized in several FER systems. However, because of its affordability and ease of use, the camera is the most widely used sensor. There are three steps in the FER process: face detection, facial expression detection, and emotion classification.

The majority of Facial Expression Recognition (FER) systems are designed to recognize the six basic emotional expressions—disgust, anger, surprise, fear, happiness, and sadness—that were first presented by Ekman [40]. Certain FER systems aim to identify more complex emotional states in addition to these fundamental ones. Human facial expression recognition (FER) systems can be broadly divided into two categories: pose-based recognition systems, which concentrate on capturing expressions based on particular facial positions, and spontaneous recognition systems, which examine spontaneous and uncontrolled expressions. [41]. Systems for recognizing spontaneous expressions identify facial expressions that people frequently make on a regular basis, including when they watch movies or preserve their facial hair. On the other hand, pose-based expression recognition systems identify the fake expressions that people make when prompted to.Facial expressions are categorized into two categories by the FER systems: micro-expressions and macro-expressions. Short, uncontrollably made facial expressions that arise naturally in reaction to an emotional cue are known as facial micro-expressions [42]. They are usually quite brief, lasting less than half a second. They are extremely subtle. On the other hand, macro-expressions are longer-lasting, more noticeable expressions that exceed thirty seconds. People can communicate their emotions more clearly when they regulate these expressions, which they do intentionally and willingly [42].

An essential and abundant source of data on speakers is speech as an acoustic signal. Vocal expressions convey emotional information through a variety of means, such as pitch, loudness, rhythm, pace, and tone of voice. But deciphering speech's emotional clues is a difficult task. Vocal manifestations of emotion are culturally distinctive, in contrast to facial emotions, which are standardized by Ekman's Facial Action Coding System (FACS) [42]. The varying expressivity techniques of individuals from diverse cultural backgrounds result in sonic diversity, which has a direct impact on the process of recognizing speech emotions. Acoustic-based and Linguistic-based models are the two main models used by speech emotion recognition systems. In order to deduce emotions, linguistic-based models examine the linguistic features and content of speech. Word choice, phrase structure, intonation patterns, and semantic context are among the characteristics they take into account [43]. In order to identify emotions, the acoustic-based models concentrate on obtaining spectral information from the spoken stream [44].In an effort to enhance speech emotion recognition performance, numerous studies have integrated linguistic and acoustic models [45] [46].

Although there are many ways to express emotions, body language and gestures have been thoroughly researched in the discipline of nonverbal communication to identify the range of human emotions. These nonverbal clues reveal important details about the goals and emotional condition of the person. Research on language and body posture has not only improved our basic understanding of human behavior but has also been crucial to the development of animated conversational agents (ACA) [47]. Numerous research investigated the relationship between emotions and gestures and body posture (Table 1) [48] [49]. Using wearable accelerometers, the authors of [50] created a Body Action and Posture (BAP) coding system to examine 12 emotional actions carried out by professional actors.

Table 1. Emotional Expressive elements of Posture and Gesture [48]

Emotion	Body Posture and Gesture
Joy	Body extended, shoulders up, Head backward, arms raised up or away from the body
Fear	Body muscles tense up, arms are raised forwards, shoulders forwards
Boredom	head backwards and bent sideway, collapsed body posture, Raising the chin
Disgust	Shoulders forwards, head downwards, upper body collapsed, hands close to the body
Surprise	Head backward, chest backward, abdominal twist, Right/left hand going to the head (covering the cheeks/mouth)
Anger	Head backward, arms raised forwards and upwards, shoulders squared

3. State-of-art approaches of Emotion Recognition

Multimodal Approaches:
Researchers are increasingly focusing on integrating information from multiple modalities, such as facial expressions, voice, body language, and physiological signals. Combining these modalities often enhances the robustness and accuracy of emotion recognition systems.

Deep Learning Techniques:
Deep learning methods, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), continue to dominate emotion recognition research. These techniques excel at learning hierarchical features and capturing temporal dependencies, which are crucial for understanding emotional expressions.

Transfer Learning:
Transfer learning, where pre-trained models are adapted to specific emotion recognition tasks, is gaining popularity. This approach helps overcome challenges related to limited labeled data by leveraging knowledge learned from related tasks or domains.

Real-world Applications:
Emotion recognition technologies are being increasingly applied in real-world scenarios, such as human-computer

interaction, virtual reality, customer service, and mental health. Researchers are exploring ways to make these systems more practical and effective in various contexts.

Explainable AI (XAI):
There is a growing emphasis on making emotion recognition models more interpretable and explainable. Understanding how models arrive at specific emotion predictions is crucial for gaining user trust, especially in applications where decisions impact individuals' well-being.

Cross-cultural and Cross-lingual Studies:
Emotion recognition research is expanding to include cross-cultural and cross-lingual studies. Researchers are exploring how cultural and linguistic differences influence emotional expression and how models can be adapted to different cultural contexts.

Edge Computing for Real-time Processing:
With the increasing demand for real-time emotion recognition, there is a trend towards developing lightweight models that can run efficiently on edge devices. This is particularly relevant for applications such as human-computer interaction and wearable devices.

Affective Computing in Human-Robot Interaction:
Emotion recognition is playing a crucial role in human-robot interaction. Researchers are exploring how robots can better understand and respond to human emotions, improving the overall user experience and collaboration.

Privacy and Ethical Considerations:
As emotion recognition technologies become more widespread, there is a growing focus on addressing privacy and ethical concerns. Researchers are exploring ways to ensure responsible use of these technologies and safeguard user privacy.

Continuous and Dynamic Emotion Recognition:
Traditional emotion recognition focused on static snapshots of emotions. Current trends involve capturing the dynamic and continuous nature of emotions, considering how emotional states evolve over time.

4. Emotion Recognition Process

In this section, we focus on the computational methods used in emotion recognition process, which can be classified into two categories as shown in Figure 2.

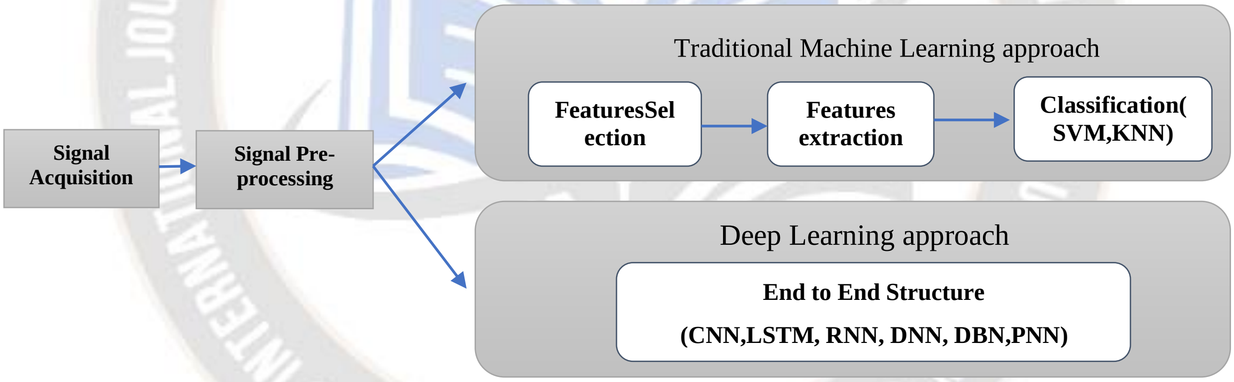

Figure 1. -Emotion Recognition Process

Traditionally, the signal pre-processing stage has been an essential first step in the emotion recognition process. The bulk of data gathered from various sensors is, in fact, quite prone to being noisy, inconsistent, and missing because of their heterogeneous origin, the subjective and complicated character of certain signals, their vulnerability to noise from electromagnetic interference and crosstalk, etc. The actual signal pre-processing method consists of four steps: The term "data cleaning" describes methods for removing duplicate data and outliers, adding missing numbers, and fixing inconsistent data [51].The data integration process entails methods that combine information gathered from several sources into a single, more comprehensive, and consistent picture of the data. Following data integration and cleaning, the data reduction phase starts. To create a condensed representation of the data, unique approaches such principal component analysis are used. The last stage of data pre-processing is data transformation, which changes the data into formats that are suited for data processing. This can entail integrating important variables, defining critical ranges, and organizing unstructured data.

Three steps are often involved in a traditional machine learning analysis of emotions: feature extraction, feature optimization, and classification. The technique of recognizing emotions depends heavily on the initial step, or feature extraction. This attempts to condense the data's original information and turn it into a more digestible set of features. As a result, feature extraction can lessen computational complexity, break free from the dimensionality curse, and enhance models' capacity for generalization [52]. The physiological signals frequently need to have features extracted before being fed into conventional classification models for emotion recognition because of their complexity and non-stationarity. Prosodic characteristics, energy features, frequency spectral features, and frequency spectral coefficients, for instance, are all included in speech signal features [53]. These characteristics include those that provide information and those that convey emotion. Therefore, emotion features should be extracted using features extraction techniques. Generally speaking, the goal of the feature optimization step is to minimize the amount of features in order to lower computational complexity and boost machine learning model effectiveness. Many feature selection methods, such ReliefF ((Relief), Sequential Forward Selection (SFS), Sequential Backward Selection (BS), and Tree-Based Feature Selection (TS), minimize the dimensionality of the features by discarding redundant and irrelevant ones.Emotion recognition systems' main goal is to classify input data in order to identify the emotions that are expressed. In order to identify patterns and connections between the input characteristics and associated emotional states, classification entails either using other classification techniques or constructing a machine learning model. Next, the emotional state of fresh, unseen data is predicted or classified using the learned model. The problem of classifying emotions is a good fit for a number of the classification techniques available in classical machine learning. Among these techniques are random forests, naive Bayes classifiers, logistic regression, decision trees, support vector machines (SVM), k-nearest neighbors (KNN), Gaussian Mixture Modelling (GMM), and Hidden Markov Model (HMM). Conventional techniques work fine with lab-pose datasets, but they struggle with genuinely complex and spontaneous data.

Using deep learning techniques, the model learns by itself, figuring out how the features relate to one another. As a result, an overall phase combines feature extraction, feature optimization, and classification. A number of deep learning techniques, including CNN, have been modified for use in emotion recognition software. Actually, a lot of physiological signals, including EEG, EMG, and ECG, are processed by CNN. CNNs have demonstrated promising results in a variety of applications and have been proved to be useful in extracting relevant features from these signals [54][55]. For emotion recognition tasks, deep learning techniques like Long Short-Term Memory (LSTM) and Deep Belief Networks (DBNs) have been applied in addition to convolutional neural networks (CNNs). Recurrent neural networks (RNNs) of the long short term memory (LSTM) type are excellent in modeling sequential data, such speech signals or time series, and identifying and categorizing emotional states across time. [56]. Generative neural networks with several layers of hidden units are called deep belief networks, or DBNs. DBNs have been effectively used for a number of applications, such as representation and feature learning. DBNs can be used to automatically extract pertinent features from the input data in the context of emotion detection, enabling more accurate representation and categorization of emotions. Additionally, probabilistic neural networks (PNNs) have been investigated for a variety of tasks, including emotion detection and mental state analysis, in the context of EEG data classification [57].

5. Literature Review on Emotion Recognition

The current works typical for each emotion recognition measuring method are shown in the following tables, along with pertinent details about the number of participants, the emotions tested, the classification scheme, and the recognition rate. The following are the primary keywords that were utilized in the literature search: Emotion Recognition; Emotion Classification; Physiological Sensor for Emotion Recognition; Computational Models of Emotion Recognition. Excluded were any works that contained erroneous information on the tools, characteristics, or precision of the calculations. This analysis also omitted studies that examined participants' ability to recognize emotions when they had physical or mental health issues. We begin our review with studies that used a single physiological indicator to identify emotions in science (Table 2).

Table 2. - Review of previous researches focused on emotion recognition used single physiological signal

Signal	Year	Emotion	Classification Method	Recognition Accuracy	Ref
ECG	2012	Valence, Arousal	LDA	Arousal: Bipartition 76.19% C.36% Valence: from 52% to 89%	[58]
ECG	2014	Valence, Arousal	SVM	Valence: 79.15% Arousal: 83.55%	[59]
ECG	2016	Negative, Positive	SVM	71.40%	[60]
GSR	2016	Level of arousal	KNN	At 1 Hz and 10 Hz: 62.5% to 63.34% At 100 Hz: 71.67%	[61]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

EOG	2017	Positive, Neutral and Negative	SVM and Naïve Bayes	Horizontal Eye Movement: - Positive:78.43%, - Negative: 74.61% - Neutral: 76.34%. Vertical Eye Movement: -Positive :77.11% -Negative: 74.03% -Neutral: 75.84%	[62]
EEG	2018	Engagement, enjoyment, boredom, frustration, workload	KNN	95.00%	[63]
HRV	2019	High/low valence and arousal	CNN	Valence: 75.3% Arousal: 76.2%	[64]
EEG	2020	-Negative, positive, and neutral. -Amusement, excitement, happiness, calmness, anger, disgust, fear, sadness, and surprise	CNN Dynamical graph	90.40%	[65]
EEG	2021	-Disgust, sadness, surprise and anger - Positive, negative, and neutral	LSTM	4 class: 94.12% 3 class: 92.66%	[66]
EEG	2021	Valence, Arousal	Deep forest	Valence 97.69% Arousal 97.53 %	[67]

In contrast to the previous example, a number of studies have sought to fuse and integrate signal components before feeding them into an emotion classifier (Table 3).

Table 3. -Review of previous researches focused on emotion recognition used a combination of physiological signals

Signal	**Year**	**Emotion**	**Classification Method**	**Recognition Accuracy**	**Ref**
GSR, HR, RSP, SKT	2015	valence, arousal (HVHA, HVLA, LVHA, LVLA)	LDA, QDA	HVHA: 98%, HVLA: 96%, LVHA: 93%, LVLA: 97%	[68]
ECG, EMG, RSP, SC	2016	joy, anger, sadness, pleasure	Decision Tree	92%	[69]
EMG, RSP	2016	valence, arousal	SVM	Valence :74%, Arousal: 74% Liking: 76%	[70]
GSR, RSP, BVP EEG, EMG, EOG	2016	valence, arousal	SVM	3 level valence: 88.33% 3 level arousal: 90.56%	[71]
EEG, ECG	2017	amusement, fear, sadness, joy, anger, and disgust	BN	98.06%	[72]
ECG, GSR, HR, GSR, SKT	2019	Anger, Happy, Sad, Joy	ANN	75.38 %	[73]
RB, PPG, and fingertip	2020	Valence, Arousal	RF, SVM, LR	Arousal: 69.86 % - 73.08 % Valence:	[74]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 10 September 2023 **Revised:** 20 October 2023 **Accepted:** 30 October 2023*

temperature (FTT)				69.53% - 72.18%	

Emotion recognition techniques that analyze facial expressions, body postures, and gestures operate on the same assumption as physiological signals, which is that body postures and gestures are also implicated in emotional responses. Generally, the databases of facial expressions, postures, and gestures can come in various forms such as static images, videos, 3D models, or real-world recordings, depending on the purpose and techniques used to create them. Researchers in the field of facial expression recognition have access to a wide range of databases that can be utilized for their research. Most of these databases are annotated with the six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) in addition to a neutral expression. The field of emotion recognition is emergent, and it needs large databases, obtained especially in the wild where the conditions are very dynamic. The performance of FER systems is highly dependent on the training databases which must be diverse because facial expressions have slight variations from person to person, may mix different emotional states at the same time, or people may not even express emotions.Table 4 provide an overview of the most widely used FER databases.

Table 4. –Some important databases that used for facial expression recognition

Database	**Samples**	**Types**	**Emotions**	**Ref**
Affect Net	450,000 images	Posed and Spontaneous	Anger, disgust, fear, happiness, sadness, surprise, and neutral	[75]
Emotio Net	1,000,000 images	Posed and Spontaneous	26 emotions: 6 basic emotion anger, disgust, fear, happiness, sadness, surprise) and 20 compound emotions (e.g., happy-surprise, sad-disgust, etc.).	[76]
MMI	2900 videos and 740 images	Posed	Anger, disgust, fear, happiness, sadness, surprise, and neutral	[77]
FER-2013	35,887 images	Posed and Spontaneous	Anger, disgust, fear, happiness, sadness, surprise, and neutral	[78]
CK+	593 images	Posed	Anger, disgust, fear, happiness, sadness, surprise, neutral and contempt.	[79]
JAFEE	213 images	Posed	Anger, disgust, fear, happiness, sadness, surprise, and neutral	[80]
RaFD	8040 images	Posed	Anger, disgust, fear, happiness, sadness, surprise, neutral and contempt.	[81]
KDEF	4900 images	Posed	Anger, disgust, fear, happiness, sadness, surprise, and neutral	[82]

While databases for facial expressions are most frequently utilized, gesture and body posture databases are also extensively utilized for the purpose of identifying human emotions. There are numerous databases with information on the body language and gestures connected to particular feelings. The process of gathering data for these databases usually entails employing motion capture technologies to record the motions and postures of human participants. Some remarkable databases are given below:

- Chalearn Gesture Dataset: A large-scale dataset of RGB-D videos of people performing 249 gestures in front of a Kinect camera.
- NTU RGB+D Dataset: A large-scale dataset of RGB-D videos of 60 action classes, including various human actions and interactions.
- UT-Kinect Gesture Dataset: A dataset of 10 different hand gestures performed by 10 different subjects in front of a Kinect camera. These include gestures such as swipe left, swipe right, wave, circle, and others.
- MPII Human Pose Dataset: A dataset of human pose estimation with more than 25,000 images of people in various poses. This dataset has been widely used for human pose estimation, and various models and algorithms have been developed and evaluated using this dataset [82][83].

A summary of researches focused on emotions recognition using facial expressions, body posture and gestures with classification methods and recognition accuracy is provided in Table 5.

Table 5. -Review of previous studies focused on emotions recognition using facial expressions, body posture and gestures

Year	**Features**	**Emotion**	**Methods**	**Recognition Accuracy**	**Ref**
2010	speech, body gestures, and	Anger, Despair, Interest, Pleasure,	Bayesian classifier	78.3%	[84]

	facial expressions	Sadness, Irritation, Joy and Pride			
2019	Skeletal Movement	neutral state, sadness, surprise, fear, anger, disgust and happiness	CNN RNN RNN-LSTM	Case of 4 emotions CNN: 63.6% RNN:80.8% RNN-LSM:82.7% Case of 6 emotions: CNN: 54.2% RNN:59.2% RNN-LSM:72%	[85]
2019	Fusion of Facial expression and Body Gesture	disgust, anger, fear, neutral, surprise, sad, and happy	Multi SVM	Anger: 95% Fear:91% Happy:95% Disgust:96% Neutral:95% Sad:94% Surprise:93%	[86]
2020	Facial expressions	For JAFEE Dataset: Happy, Neutral, Fear, Sad, Disgust, Surprise and Anger For CK+ Dataset: Contempt, Anger, Disgust, Fear, Sadness, Surprise and Happiness For YALE FACE Dataset: Happy, Center-light, Left-light, W/no glasses, W/glasses, Sad, Right-light, Normal, Sleep, Surprised and Wink	LBP for features extraction and SVM for classification CNN	LBP: CK+: 96.66% JAFEE: 76.23% YALEFACE:74% CNN: CK+: 97.32% JAFEE:77.27% YALFACE:31.82%	[87]
2021	Facial expression	For CK+ dataset: anger, contempt, fear, disgust, happiness, surprise, and sadness For JAFEE Dataset: happiness, sadness, surprise, anger, disgust, and fear and neutral FER2013: angry, disgust, fear, happy, sad, surprise, including neutral	A feedforward learning model	JAFFE:96.8% CK:86.5% FER2013:62.5%	[88]
2021	Body Movements	Happiness, sadness, fear, anger, and neutral.	Two-layer feature-selection process	During walking: 90% During setting: 96% Action independent-scenario: 86.66%	[89]
2022	Facial expression	For CK+ dataset: anger, contempt, fear, disgust, happiness, surprise, and sadness For JAFEE Dataset: happiness, sadness, surprise, anger, disgust, and fear and neutral For RAF dataset: anger, happiness, fear,	MLP SVM KNN LR	MLP: JAFFE:90% CK+:94% RAF:67% SVM: JAFFE:88% CK+:94% RAF:67% KNN:	[90]

		surprise, disgust, sadness and neutral		JAFFE:95% CK+:97% RAF:63% LR: JAFFE:86% CK+:87% RAF:66%	

6. Speech Emotion Recognition

There are several methods for recognizing emotion in speech. Many methods make use of different sound characteristics, such as pitch, accent, musicality, timbre, and many others. Phonemes are the names for these parameters. Expression of emotions differs amongst cultures, but it can also vary from person to person within the same lifestyle. Depending on the context of the discourse, different emotions can be expressed in the same speech. Therefore, one of the most important steps in speech analysis is recovering clean speech from signal characteristics. Speech recognition software uses the nuances and shifts in audio signals to identify emotions in speech. Speech emotion detection is widely utilized in speech recognition, call centers, customer service, and other fields nowadays. It has several uses. Feature extraction and classification are the two main phases involved in speech emotion recognition. A two-stage technique that employs convolutional neural networks (CNNs) is proposed in [91]. Initially, unlabeled samples are used to compute the local invariant features (LIFs) using a sparse auto-encoder (SAE) with reconstruction segregation. The feature extractor in the second stage then receives LIF as an input. In [92], an alternative approach is put forth that employs contractive CNN for feature extraction after semi-CNNs in the first step. A technique for better outcomes than MFCC is presented in [93] and it makes use of Fourier Parameters and MFCC. The Multilayer Perceptron (MLP) and Support Vector Machines (SVMs) are trained using Gaussian Mixture Models (GMMs) in the front and back end systems of the system. In order toaddress issues with skewed data, anchor models are presented in [94]. These models are mostly based on Euclidean or Cosine distance. Gaussian Mixture Models are employed in [95] for the category-conditional distribution of speech features, and the EM technique (Expectation Maximization algorithm) is then utilized to estimate the parameters. Another approach, covered in [96], tests the Gaussian Mixture Models and CHMM on datasets in the German and English languages.

A different approach [96] that makes use of a fuzzy interface system and tests all the prosodic aspects of speech—that is, pitch, length, and energy features—on a self-generated dataset that consists of a user conversing with a call center computer has produced noteworthy findings. A different approach that was covered in [96] uses features from the Linear Predictive Cepstral Coefficient (LPCC) to identify dysfluencies in audio sources. Both k-NN and linear discriminant analysis (LDA) are used in classification. Teager Energy Operator (TEO) attributes are one of the greatest features for voice recognition according to a method described in [98]. The method is evaluated on the SUSAS database and employs the Bayesian hypothesis.

Another technique covered in [96] use DWT to extract low-frequency speech components. It is evaluated on a self-generated Malayalam language database using a three-layer MLP classifier. A novel approach to feature selection for lowering the dimensionality of the features has been presented by the procedure outlined in [97]. A biogeography-based approach with Particle Swarm Optimization support is used to run various simulations on various datasetsof speech with an accuracy of up to 99.5%.

Table 6. -Review of previous studies focused on emotions recognition techniques using speech signal

Year	Specifications	Method	Database	Recognition Accuracy
2004	Uses Pitch and energy as features	GMM-CHMM	Self-generated	GMM-86.8% CHMM-77.8%
2009	Discrete Wavelet Transform (DWT), Artificial neural network.	Artificial Neural Networks tested on Gender-Dependent Databases	Self-generated	Male:72.05% Female:65.5%
2013	Anchor Models based on cosine distances	Anchor Models	FAU-AIBO	44.2%
2014	LIF, SAE, SDFA	CNN	SAVEE DES MES	71.8% 60.4% 57.8%
2014	Contractive CNN to learn	Semi-CNN	SAVEE	73.6%

	candidate features, a novel function for classification		DES MES	79.9% 78.3%
2015	Restricted Boltzmann Machine(RBM) based unsupervised learning, and DNN-HMMs with discriminative learning	Hybrid Deep Neural Network Hidden Markov Model (DNN-HMM)	Berlin Emotional Database	77.92%
2017	PSO and BBO based algorithms	PSO and BBO	BES Database	99.47%

7. Discussion

This study conducted a thorough examination of a large number of recent articles and research papers that addressed sentiment analysis and emotion recognition in the field. Indeed, there is a great need for emotion recognition systems, and they have many uses in the Internet of Things (IoT) and artificial intelligence (AI) domains. Emotions recognition systems are in demand for a variety of reasons, of which the following are just a few:

- ✓ Human- Computer Interaction: Emotion recognition systems enhance human-computer interaction by enabling computers and machines to understand and respond to human emotions.
- ✓ Personalized Experiences: Emotion recognition can help tailor experiences based on individual emotions. For instance, it can be used in entertainment and gaming industries to adjust the content or difficulty level based on the user's emotional state, providing a more personalized and engaging experience.
- ✓ Healthcare and Well-being: Emotion recognition systems have potential applications in healthcare and mental well-being. They can be used for monitoring and diagnosing conditions like depression, anxiety, and stress, enabling timely interventions and personalized treatments.
- ✓ Market Research and Advertising: Emotion recognition can be utilized in market research and advertising to measure consumer emotional responses to products, advertisements, or user interfaces. This information can help companies to optimize their marketing strategies and improve user satisfaction.
- ✓ Security and Surveillance: Emotion recognition systems can enhance security and surveillance systems by identifying suspicious or threatening behaviors based on facial expressions or voice analysis. They can be employed in public spaces, airports, or high-security areas to improve safety measures.
- ✓ Education and Learning: Emotion recognition technology can assist in educational settings by gauging student engagement and emotional states during learning activities. This information can help educators customize teaching methods, provide additional support, and create a more effective learning environment.

Owing to the increasing need for emotion detection systems, scientists have been working hard to increase these systems' accuracy by concentrating on a number of important areas, including multimodal approaches. Multiple modalities, including speech analysis, body language, physiological data (such heart rate variability), and facial expressions, are sometimes included into more reliable and accurate emotion identification systems.

Partially true is the claim that physiological signals are the most dependable signals in this situation since they cannot be controlled by conscious thought and are therefore extremely dependable for identifying emotions. Electroencephalography (EEG), skin conductance, and heart rate are examples of physiological signals that can provide important information about an individual's emotional state. These mostly involuntary signals may be an indication of underlying physiological processes connected to emotions. Physiological cues are less susceptible to intentional manipulation than verbal or facial clues, which can be deliberately controlled or disguised. It is crucial to remember that physiological signals are not impervious to outside influences. Even while physiological reactions are not always under an individual's direct control, a number of circumstances can nevertheless affect these signals. For instance, a number of variables, including stress, physical activity, drugs, and specific medical disorders, might influence physiological arousal. Furthermore, individual variations in baseline physiological responses may potentially affect how these signals are interpreted.

In general, psychological techniques for identifying human emotions are thought to be less complex than physiological techniques. In order to infer emotional states, psychological approaches use subjective self-report measures along with the observation and interpretation of observable behavioral clues. These techniques recognize and classify emotions by analyzing facial expressions, body language, vocal tones, and speech content. Since psychological techniques don't call for sophisticated data processing or specific equipment, they can be applied with relative ease. They frequently entail employing standardized questionnaires or rating scales to have trained observers or the participants themselves report their emotional experiences. These techniques are frequently applied in many different sectors and can offer insightful information on emotions. It's crucial to remember that psychological approaches have their limitations. They rely on outside cues and self-report, which are subject to individual variances, cultural differences, and social desirability, among other things. Furthermore, proficiency and training are necessary for correctly understanding behavioral signals and self-report measures in order to reduce bias and guarantee accurate results. Conversely,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 10 September 2023 **Revised:** 20 October 2023 **Accepted:** 30 October 2023*

physiological approaches—which were previously discussed—involve the measurement of physiological signals linked to emotions, such as skin conductance, heart rate, or brain activity. Physiological approaches offer an extra layer of objective information and can catch emotional responses that would not be seen through behavioral signals alone, although being more complicated in terms of equipment and data analysis. As a result, the decision between the two approaches is based on the particular objectives, the situation, and the resources available for emotion recognition.

As stated in [98] and [99], recent research on the subject of human emotion recognition shows that there isn't a single approach that is perfect for recognizing emotions in every circumstance. Instead, combining psychological and physiological approaches can lead to a more thorough understanding of emotions. Additionally, combining different approaches can offer chances for verification and cross-validation. Research and development in the field of multimodal emotion recognition are still ongoing, despite variations in the accessibility and availability of these services. It is anticipated that the number of businesses and applications that take advantage of multimodal emotion identification systems will rise as the technology develops and spreads.

8. Challenges of Emotion Recognition

Several challenges persisted in emotion recognition research. Keep in mind that the field is dynamic, and new challenges may have emerged since then. Here are some issues that were relevant at that time:

Ambiguity and Subjectivity:
Emotions can be subjective and context-dependent, making it challenging to create a universal model that accurately recognizes emotions across diverse individuals and cultures.

Multimodal Emotion Recognition:
Integrating information from multiple modalities (such as facial expressions, voice, and physiological signals) poses challenges in terms of feature fusion and model complexity.

Real-world Conditions:
Emotion recognition systems often struggle when faced with real-world conditions such as noisy environments, varying lighting, and different camera angles. Adapting models to handle these conditions is an ongoing challenge.

Limited Labeled Data:
Acquiring large, labeled datasets for training emotion recognition models is often difficult. Limited data can lead to overfitting and may hinder the generalization of models to real-world scenarios.

Cultural and Individual Variability:
Emotion expression varies across cultures and individuals. Developing models that account for this variability is essential for accurate and inclusive emotion recognition.

Dynamic Nature of Emotions:
Emotions are dynamic, and their expression evolves over time. Capturing the temporal dynamics of emotions and understanding how they change is a persistent challenge.

Ethical and Privacy Concerns:
Implementing emotion recognition in various contexts raises ethical concerns related to privacy, consent, and potential misuse of the technology. Striking a balance between technological advancement and ethical considerations is crucial.

Explainability and Interpretability:
Emotion recognition models often lack transparency, making it difficult to understand how they arrive at a particular emotion classification. Ensuring interpretability and explainability is important, especially in applications where decisions impact individuals' lives.

Cross-cultural Validity:
Cultural differences in the interpretation and expression of emotions pose challenges for models trained on data from one cultural context and applied to another. Developing models with cross-cultural validity is important for global applications.

Emotion Elicitation:
Standardizing emotion elicitation techniques for research studies is challenging. The methods used to induce specific emotions in study participants can impact the data collected and, consequently, the performance of emotion recognition models.

9. Conclusion

This review paper has examined the techniques and sensors used in emotion recognition, emphasizing various modalities and computational approaches. The talk demonstrated the advances in the field by focusing on the analysis and classification of emotions using both deep learning and conventional machine learning approaches. Furthermore, the strengths and limits of the various emotion recognition modalities—such as facial expressions, voice cues, physiological signals, textual data,voice and brain activity—were investigated.

It's crucial to remember that, despite advancements, there is still no ideal technique for identifying emotions. Every strategy or method offers a unique set of benefits and drawbacks. Sensitivity to emotion is a persistent challenge due to individual differences, cultural variances, and contextual effects, among other factors.In order to overcome these obstacles, the paper emphasizes the necessity of ongoing study and advancement in emotion identification. Future developments can entail improving on already used computational techniques, investigating cutting-edge sensor technology, and taking into account multimodal approaches, which integrate several modalities to provide a more thorough knowledge of emotions. In the end, a comprehensive and reliable method for recognizing emotions is still a work in progress. We should expect more developments in emotion recognition as technology progresses and interdisciplinary cooperation thrive, setting the way.

REFERENCE

[1] S. Menon, *Brain, self and consciousness: Explaining the conspiracy of experience*. New Delhi, India: Springer, 2016.

[2] N. M. Szajnberg, "What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS) what the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS). Edited by Erika L. rosenberg and Paul Ekman. 3rd ed. New York: Oxford university press, 2020, xxii + 627 pp., 80.00 hardcover, 29.99 paperback," *J. Am. Psychoanal. Assoc.*, vol. 70, no. 3, pp. 591–595, 2022

[3] Elham S. Salama, Reda A. El-Khoribi, Mahmoud E. Shoman, Mohamed A. Wahby Shalaby, "A 3D-convolutional neural network framework with ensemble learning techniques for multi-modal emotion recognition," *Egyptian Informatics Journal*, vol. 2, no. 2, pp. 167–176, 2021.

[4] M.-Z. Poh, N. C. Swenson, and R. W. Picard, "A wearable sensor for unobtrusive, long-term assessment of electrodermal activity," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 5, pp. 1243–1252, 2010.

[5] X. Liu, Z. Xu, and K. Huang, "Multimodal emotion recognition based on cascaded multichannel and hierarchical fusion," *Comput. Intell. Neurosci.*, vol. 2023, p. 9645611, 2023.

[6] G. Valenza, L. Citi, A. Lanatà, E. P. Scilingo, and R. Barbieri, "A nonlinear heartbeat dynamics model approach for personalized emotion recognition," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2013, pp. 2579–2582, 2013.

[7] J.Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010.

[8] H. Wang, X. Li, Z. Ren, M. Wang, and C. Ma, "Multimodal sentiment analysis representations learning via contrastive learning with condense attention fusion," *Sensors (Basel)*, vol. 23, no. 5, 2023.

[9] R. W. Picard, *Affective Computing*. London, England: MIT Press, 2000.

[10] D. Liu, Z. Wang, L. Wang, and L. Chen, "Multi-modal fusion emotion recognition method of speech expression based on deep learning," *Front. Neurorobot.*, vol. 15, p. 697634, 2021.

[11] R. S. Gaikwad and S. C. Gandage, "MCNN: Visual Sentiment Analysis using Various Deep Learning Framework with Deep CNN," *InternationalJournalof INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING*, vol. 11, no. 2, pp. 265–278, 2023.

[12] K. Feng and T. Chaspari, "A review of generalizable transfer learning in automatic emotion recognition," *Front. Comput. Sci.*, vol. 2, 2020.

[13] S. Liu, P. Gao, Y. Li, W. Fu, and W. Ding, "Multi-modal fusion network with complementarity and importance for emotion recognition," *Inf. Sci. (Ny)*, vol. 619, pp. 679–694, 2023.

[14] K. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Commun.*, vol. 40, no. 1–2, pp. 227–256, 2003.

[15] F.Ma, Y. Li, S.Ni, S.L. Huang and L.Zhang, Data Augmentation for Audio-Visual Emotion Recognition with an Efficient Multimodal Conditional GAN. Appl. Sci. 2022, 12, 527.

[16] K. Ter Burg and H. Kaya, Comparing Approaches for Explaining DNN-Based Facial Expression Classifications. Algorithms 2022, 15, 367.

[17] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affect. Comput.*, vol. 1, no. 1, pp. 18–37, 2010.

[18] J. M. Salsman *et al.*, "Emotion assessment using the NIH Toolbox," *Neurology*, vol. 80, no. 11 Suppl 3, pp. S76-86, 2013.

[19] T.-M. Bynion and M. T. Feldner, "Self-Assessment Manikin," in *Encyclopedia of Personality and Individual Differences*, Cham: Springer International Publishing, 2020, pp. 4654–4656.

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

[21] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, "Physiological signals based human emotion Recognition: a review," in *2011 IEEE 7th International Colloquium on Signal Processing and its Applications*, 2011.

[22] C. Qing, R. Qiao, X. Xu, and Y. Cheng, "Interpretable emotion recognition using EEG signals," *IEEE Access*, vol. 7, pp. 94160–94170, 2019.

[23] A. Kushki, J. Fairley, S. Merja, G. King, and T. Chau, "Comparison of blood volume pulse and skin conductance responses to mental and affective stimuli at different anatomical sites," *Physiol. Meas.*, vol. 32, no. 10, pp. 1529–1539, 2011.

[24] J. Z. Lim, J. Mountstephens, and J. Teo, "Emotion recognition using eye-tracking: Taxonomy, review and

current challenges," *Sensors (Basel)*, vol. 20, no. 8, p. 2384, 2020.

[25] W.-L. Zheng, B.-N. Dong, and B.-L. Lu, "Multimodal emotion recognition using EEG and eye tracking data," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2014, pp. 5040–5043, 2014.

[26] F. Z. Canal *et al.*, "A survey on facial emotion recognition techniques: A state-of-the-art literature review," *Inf. Sci. (Ny)*, vol. 582, pp. 593–617, 2022.

[27] P. Ekman, R. W. Levenson, and W. V. Friesen, "Real-Time Recognition of the Affective User State with Physiological Signals," in *Proc. Doctoral Consortium Conf. Affective Computing and Intelligent Interaction*, 2007, pp. 1–8.

[28] J. Anttonen and V. Surakka, "Emotions and heart rate while sitting on a chair," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2005.

[29] K. Rattanyu, Graduate School of Functional Control Systems Engineering, Shibaura Institute of Technology, 3-7-5 Toyosu, Koto-ku, Tokyo 135-8548, Japan, M. Mizukawa, and Department of Electrical Engineering, Shibaura Institute of Technology, 3-7-5 Toyosu, Koto-ku, Tokyo 135-8548, Japan, "Emotion recognition based on ECG signals for service robots in the intelligent space during daily life," *J. Adv. Comput. Intell. Intell. Inform.*, vol. 15, no. 5, pp. 582–591, 2011

[30] G. Udovičić, J. Đerek, M. Russo, and M. Sikora, "Wearable emotion recognition system based on GSR and PPG signals," in *Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care*, 2017.

[31] C. Zong and M. Chetouani, "Hilbert-Huang transform based physiological signals analysis for emotion recognition," in *2009 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2009.

[32] B. Tag, A. Vargo, A. Gupta, G. Chernyshov, K. Kunze, and T. Dingler, "Continuous Alertness Assessments: Using EOG Glasses to Unobtrusively Monitor Fatigue Levels In-The-Wild," in *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–12.

[33] M. Kolodziej *et al.*, "Fatigue detection caused by office work with the use of EOG signal," *IEEE Sens. J.*, vol. 20, no. 24, pp. 15213–15223, 2020.

[34] P. Mishra, A. Biancolillo, J. M. Roger, F. Marini, and D. N. Rutledge, "New data preprocessing trends based on ensemble of multiple preprocessing techniques," *Trends Analyt. Chem.*, vol. 132, no. 116045, p. 116045, 2020.

[35] M. M. Bradley and P. J. Lang, "Measuring emotion: the Self-Assessment Manikin and the Semantic Differential," *J. Behav. Ther. Exp. Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[36] P. Desmet, *Designing Emotions*. The Netherlands, 2002.

[37] P. Desmet, K. Overbeeke, and S. Tax, "Designing products with added emotional value: Development and appllcation of an approach for research through design," *Des. J.*, vol. 4, no. 1, pp. 32–47, 2001.

[38] P. Ekman, *Emotion in the human face: Guide-lines for research and an integration of findings*. Pergamon, 1972.

[39] D.Watson, L.A. Clark and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the PANAS scales," *J. Pers. Soc. Psychol*, vol. 54, no. 6, pp. 1063–1070, 1988.

[40] Q. Gan, C. Wu, S. Wang, and Q. Ji, "Posed and spontaneous facial expression differentiation using deep Boltzmann machines," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2015.

[41] J. A. Harrigan, R. Rosenthal, and K. R. Scherer, *The new handbook of methods in nonverbal behavior research*. London, England: Oxford University Press, 2005.

[42] R. Rajoo and C. C. Aun, "Influences of languages in speech emotion recognition: A comparative study using Malay, English and Mandarin languages," in *2016 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 2016.

[43] S. Bhosale, R. Chakraborty, and S. K. Kopparapu, "Deep encoded linguistic and acoustic cues for attention-based end to end speech emotion recognition," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.

[44] M. Sakurai and T. Kosaka, "Emotion recognition combining acoustic and linguistic features based on speech recognition results," in *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*, 2021.

[45] E. Kim and J. W. Shin, "DNN-based emotion recognition based on bottleneck acoustic features and lexical features," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.

[46] G. Ball and J, Breese, E*motion and personality in a conversational agent*. Cambridge: USA: Embodied conversational agents, 2000.

[47] S. Lee, M. Bae, W. Lee, and H. Kim, "CEPP: Perceiving the emotional state of the user based on body posture," *Appl. Sci. (Basel)*, vol. 7, no. 10, p. 978, 2017.

[48] R. Calvo, S. D'Mello, J. Gratch, A. Kappas, M. Lhommet, and S. C. Marsella, "Expressing emotion through posture and gesture," in *The Oxford Handbook of Affective Computing*, Oxford University Press, 2015.

[49] N. Dael, M. Mortillaro, and K. R. Scherer, "Emotion expression in body action and posture," *Emotion*, vol. 12, no. 5, pp. 1085–1101, 2012.

[50] B. Kaur, D. Singh, and P. P. Roy, "EEG based emotion classification mechanism in BCI," *Procedia Comput. Sci.*, vol. 132, pp. 752–758, 2018.

[51] Y. Cai, X. Li, and J. Li, "Emotion recognition using different sensors, emotion models, methods and datasets: A comprehensive review," *Sensors (Basel)*, vol. 23, no. 5, 2023.

[52] M. Swain, A. Routray, and P. Kabisatpathy, "Databases, features and classifiers for speech emotion recognition: a review," *Int. J. Speech Technol.*, vol. 21, no. 1, pp. 93–120, 2018.

[53] R. Qiao, C. Qing, T. Zhang, X. Xing, and X. Xu, "A novel deep-learning based framework for multi-subjectemotion recognition," in *Proceedings of the International Conference on Information, Cybernetics and Computational Social Systems*, Dalian, 2017, pp. 181–185.

[54] J. Huang, X. Xu, and T. Zhang, "Emotion classification using deep neural networks and emotional patches," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017.

[55] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[56] J. Zhang, M. Chen, S. Hu, Y. Cao, and R. Kozma, "PNN for EEG-based emotion recognition," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016.

[57] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 532–541, 2020

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

[59] G. Valenza, L. Citi, A. Lanatá, E. P. Scilingo, and R. Barbieri, "Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics," *Sci. Rep.*, vol. 4, no. 1, p. 4998, 2014

[60] H.-W. Guo, Y.-S. Huang, C.-H. Lin, J.-C. Chien, K. Haraikawa, and J.-S. Shieh, "Heart rate variability signal features for emotion recognition by using principal component analysis and support vectors machine," in *2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*, 2016.

[61] A. Greco, A. Lanata, L. Citi, N. Vanello, G. Valenza, and E. Scilingo, "Skin admittance measurement for emotion recognition: A study over frequency sweep," *Electronics (Basel)*, vol. 5, no. 4, p. 46, 2016

[62] S. Paul, A. Banerjee, and D. N. Tibarewala, "Emotional eye movement analysis using electrooculography signal," *Int. J. Biomed. Eng. Technol.*, vol. 23, no. 1, p. 59, 2017

[63] J. Fan, J. W. Wade, A. P. Key, Z. E. Warren, and N. Sarkar, "EEG-based affect and workload recognition in a virtual driving environment for ASD intervention," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 43–51, 2018.

[64] M. S. Lee, Y. K. Lee, D. S. Pae, M. T. Lim, D. W. Kim, and T. K. Kang, "Fast emotion recognition based on single pulse PPG signal with convolutional neural network," *Appl. Sci. (Basel)*, vol. 9, no. 16, p. 3355, 2019

[65] J. Cheng *et al.*, "Emotion recognition from multi-channel EEG via deep forest," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 2, pp. 453–464, 2021.

[66] A. Sakalle, P. Tomar, H. Bhardwaj, D. Acharya, and A. Bhardwaj, "A LSTM based deep learning network for recognizing emotions using wireless brainwave driven system," *Expert Syst. Appl.*, vol. 173, no. 114516, p. 114516, 2021.

[67] K. K. Talluri, M.-A. Fiedler, and A. Al-Hamadi, "Deep 3D convolutional neural network for facial micro-expression analysis from video images," *Appl. Sci. (Basel)*, vol. 12, no. 21, p. 11078, 2022.

[68] S. Basu *et al.*, "Emotion recognition based on physiological signals using valence-arousal model," in *2015 Third International Conference on Image Information Processing (ICIIP)*, 2015.

[69] P. Gong, H. T. Ma, and Y. Wang, "Emotion recognition based on the multiple physiological signals," in *2016 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, 2016.

[70] L. M. M. Sadeghi, A. Yazdani, and J. Vesin, "Using cardio-respiratory signals to recognize emotions elicited by watching music video clips," in *Proceedings of the 2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP)*, Montreal: QC Canada, 2016, pp. 1–5.

[71] H. F. Garcia, M. A. Alvarez, and A. A. Orozco, "Gaussian process dynamical models for multimodal affect recognition," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2016, pp. 850–853, 2016.

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

[73] S. Tiwari, S. Agarwal, M. Syafrullah, and K. Adiyarta, "Classification of physiological signals for emotion recognition using IoT," in *2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2019.

[74] D. Ayata, Y. Yaslan, and M. E. Kamasak, "Emotion recognition from multimodal physiological signals for emotion aware healthcare systems," *J. Med. Biol. Eng.*, vol. 40, no. 2, pp. 149–157, 2020.

[75] M. Mahoor, "AffectNet.". [Online] Available: http://mohammadmahoor.com/affectnet/.

[76] Ohio-state.edu, "EmotioNet Database.". [Online] Available: http://cbcsl.ece.ohio state.edu/dbform_emotionet.html.

[77] M. Eu, "MMI facial expression database.". [Online] Available: https://mmifacedb.eu. [Accessed: 08-Jun-2023].

[78] "Facial Expression Recognition 2013 Database.". [Online] Available: https://www.kaggle.com/c/challenges-inrepresentation-learning-facial-expressionrecognition-challenge/data.

[79] "The Japanese Female Facial Expression Database.". [Online] Available: http://www.consortium.ri.cmu.edu/ckagree/.

[80] "The Extended Cohn–Kanade.". Available: http://www.kasrl.org/jaffe.html

[81] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. Van Knippenberg, "Presentation and validation of the Radboud faces database, "Cogn," *Cogn. Emot*, vol. 24, no. 8, pp. 1377–1388, 2010.

[82] I. Petrov, V. Shakhuro, and A. Konushin, "Deep probabilistic human pose estimation," *IET Comput. Vis.*, vol. 12, no. 5, pp. 578–585, 2018.

[83] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019

[84] L.Kessous, G. Castellano, and G. Caridakis, "Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis," *J. Multimodal User Interfaces*, vol. 3, no. 1–2, pp. 33–48, 2010.

[85] T. Sapiński, D. Kamińska, A. Pelikant, and G. Anbarjafari, "Emotion recognition from skeletal movements," *Entropy (Basel)*, vol. 21, no. 7, p. 646, 2019.

[86] T. Keshari and S. Palaniswamy, "Emotion recognition using feature-level fusion of facial expressions and body gestures," in *2019 International Conference on Communication and Electronics Systems (ICCES)*, 2019.

[87] R. Ravi, S. V. Yadhukrishna, and R. Prithviraj, "A Face Expression Recognition Using CNN & LBP," in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, 2020.

[88] B. Yusra, J. Afshan, N. Nudrat, M. Haroon, Y. Serestina, and V. Sergio, "Facial Expression Recognition of Instructor Using Deep Features and Extreme Learning Machine," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–17, 2021.

[89] F. Ahmed, A. S. M. H. Bari, and M. L. Gavrilova, "Emotion recognition from body movement," *IEEE Access*, vol. 8, pp. 11761–11781, 2020.

[90] A. I. Siam, N. F. Soliman, A. D. Algarni, F. E. Abd El-Samie, and A. Sedik, "Deploying Machine Learning techniques for human emotion detection," *Comput. Intell. Neurosci.*, vol. 2022, p. 8032673, 2022.

[91] Q. Mao, M. Dong, Z. Huang and Y. Zhan, "Learning Salient Features for Speech Emotion Recognition Using Convutional Neural Networks", *IEEE Transactions on Multimedia*. pp. 2203-2213. 10.1109/TMM.2014.2360798, 16,2014.

[92] Huang, Zhengwei Dong, Ming Mao, Qirong Zhan, Yongzhao, "Speech Emotion Recognition Using CNN". MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia. pp.801-804. 10.1145/2647868.2654984,2014.

[93] E. H. Kim, K. H. Hyun, S. H. Kim and Y. K. Kwak, "Speech Emotion Recognition Using Eigen-FFT in Clean and Noisy Environments", RO-MAN 2007 - The 16th *IEEE International Symposium on Robot and Human Interactive Communication,* Jeju. (2007) .pp. 689-694. 10.1109/ROMAN.2007.4415174,2007.

[94] Y. Attabi and P. Dumouchel, " Anchor Models for Emotion Recognition from Speech*", IEEE Transactions on Affective Computing*. pp.280-290 . 10.1109/T-AFFC.2013.17,2013.

[95] J. Liu, C. Chen, J. Bu, M. You and J. Tao, "Speech Emotion Recognition using an Enhanced Co-Training Algorithm," *IEEE International Conference on Multimedia and Expo,*. pp. 999-1002 . 10.1109/ICME.2007.4284821. Beijing, 2007.

[96] M.Yashaswi Alva, M. Nachamai and J. Paulose, "A comprehensive survey on features and methods for speech emotion detection", *IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT),* Coimbatore.(2015).1-6. 10.1109/ICECCT.2015.7226047,2015.

[97] Yogesh C.K., M. Hariharan, RuzelitaNgadiran, Abdul Hamid Adom, Sazali Yaacob, Chawki Berkai, Kemal Polat, "A new hybrid PSO assisted biogeographybased optimization for emotion and stress recognition from speech signal," *Expert Systems with Alications,* 69(2017). pp.149-158, 2017.

[98] K. Takahashi, "Remarks on emotion recognition from multi-modal bio-potential signals," in *2004 IEEE International Conference on Industrial Technology, 2004. IEEE ICIT '04*, 2005.

[99] Lin, Lin, Wang, and Wu, "Multiple Convolutional Neural Networks fusion using Improved Fuzzy Integral for facial emotion recognition," *Appl. Sci. (Basel)*, vol. 9, no. 13, p. 2593, 2019