Wearable-Based Emotion Recognition Using Electrocardiogram and Galvanic Skin Response and Instrumentation Audit: A Systematic Review

Ayse Kosal Bulbul, Muhammad Irfan, Maria K. Jaakkola, Abdulhamit Subasi, and Riku Klén

Abstract—**Simple and unobtrusive wearables, particularly watches and wristbands, now provide continuous monitoring with no inconvenience and are widely used by elderly people. Effective emotion recognition relies not only on advances in models but also on methods for signal acquisition and specification. In response to this imperative, we performed a systematic review of emotion recognition utilizing wearable physiological signals, focusing on two practical modalities: electrocardiography (ECG), which indicates cardiac dynamics and emotional arousal, and galvanic skin response (GSR), a measure of electrodermal activity (EDA) reflecting sympathetic nervous system activation and conveniently recordable at the wrist, which reflects sympathetic nervous system activation and can be conveniently recorded at the wrist. We examined four prominent publishers (ScienceDirect, SpringerLink, Taylor & Francis Online, and IEEE Xplore) and identified 549 papers; after excluding 90 review articles and 75 papers focused solely on electroencephalography (EEG), 384 were subjected to thorough screening. After excluding 20 studies focused solely on stress and 316 papers without ECG/GSR data, 48 studies utilizing machine learning (ML) were identified (2019–2025). Traditional models like support vector machines remain prevalent, though deep learning (DL) techniques, particularly convolutional neural networks (CNNs) and hybrids, are more effective at discerning complex temporal patterns. Simultaneously, we conduct an instrumentation and measurement (I&M) audit on commonly utilized datasets. Protocols and sampling rates are generally documented; however, essential measurement details are frequently absent: electrode type and positioning, skin–electrode impedance, signal quality indices (SQIs), calibration, repeatability, and interdevice comparability. These limitations hinder equitable performance attribution (sensing versus modeling), limit uncertainty assessment, and complicate hardware transfer and cross-study evaluation. Standardizing I&M metadata in public releases, when combined with AI evaluation practices, improves comparability and dependability in healthcare predictions based on wearable physiological information.**

Index Terms—**Electrocardiography (ECG), galvanic skin response (GSR), instrumentation and measurement (I&M), machine learning (ML), wearable emotion recognition.**

I. INTRODUCTION

EMOTION is defined as a psychological reaction generated by both external stimuli and internal cognitive processes, which is accompanied by a variety of physiological changes in the human body. Also, mood is generally characterized as a dominant, conscious emotional state at a specific moment in time [1]. Emotional experiences are often described in terms of two major dimensions: valence, which indicates whether the feeling is positive or negative, and arousal, which reflects how calming or exciting the emotion is [1]. According to Paul Ekman [2], emotions can be classified into six basic categories: happiness, sadness, anger, fear, surprise, and disgust [2]. While there are many approaches to classifying emotions, one widely accepted and useful framework in psychology is the 2-D model of emotions. Despite its simplicity, this model offers a valuable tool for analyzing emotional states and understanding their functional roles in everyday life. Emotions strongly influence both the physiological and psychological aspects of human life. Positive emotions promote improved health and productivity, but negative emotions can lead to a variety of health issues, especially if prolonged [3]. Emotions trigger a wide range of physiological responses that influence various parts of the body, including the brain, heart, skin, muscles, blood circulation, facial expressions, and voice. These changes are often measurable using sensors and can be utilized for emotion recognition, as wearable technology has also progressed from cumbersome clinical instruments to simple wearables such as rings. Both elderly and young individuals are using these nodes for continuous or near-constant monitoring during daily life with minimal user load. This shift from controlled laboratory recordings to free-living wearable data has opened up

Received 14 November 2025; revised 26 March 2026; accepted 25 April 2026. Date of publication 8 June 2026; date of current version 18 June 2026. This work was supported by the AI4HOPE Project under Grant (EU Horizon Project ID 101136769), 2025. The work of Muhammad Irfan and Maria K. Jaakkola was supported in part by the Finnish Cultural Foundation, the Juhani Aho Foundation, and the State Research Funding (VTR) and in part by Shanghai Sci-tech Co-research Program Project under Grant 25HB2703300. The Associate Editor coordinating the review process was Dr. Bruno Andò. *(Corresponding author: Muhammad Irfan.)*

Ayse Kosal Bulbul is with the Institute of Biomedicine, Faculty of Medicine, University of Turku, FI-20014 Turku, Finland (e-mail: aysekosal.a.kosalbulbul@utu.fi).

Muhammad Irfan is with The University of Sydney, Sydney, NSW 2006, Australia, also with Fudan University, Shanghai 200433, China, and also with the Institute of Biomedicine and Turku University Hospital, University of Turku, FI-20520 Turku, Finland (e-mail: muhammad.m.irfan@utu.fi).

Maria K. Jaakkola is with Turku PET Centre, University of Turku, Turku, Finland, and also with Turku University Hospital, FI-20520 Turku, Finland (e-mail: maria.jaakkola@utu.fi).

Abdulhamit Subasi is with the Department of Information Science and Technology, College of Emergency Preparedness, Homeland Security and Cybersecurity, University at Albany—SSUNY, Albany, NY 12222 USA, and also with the Institute of Biomedicine, University of Turku, FI-20014 Turku, Finland (e-mail: abdulhamit.subasi@utu.fi).

Riku Klén is with the University of Turku, FI-20014 Turku, Finland, and also with Turku University Hospital, FI-20520 Turku, Finland (e-mail: riku.klen@utu.fi).

Digital Object Identifier 10.1109/TIM.2026.3701198

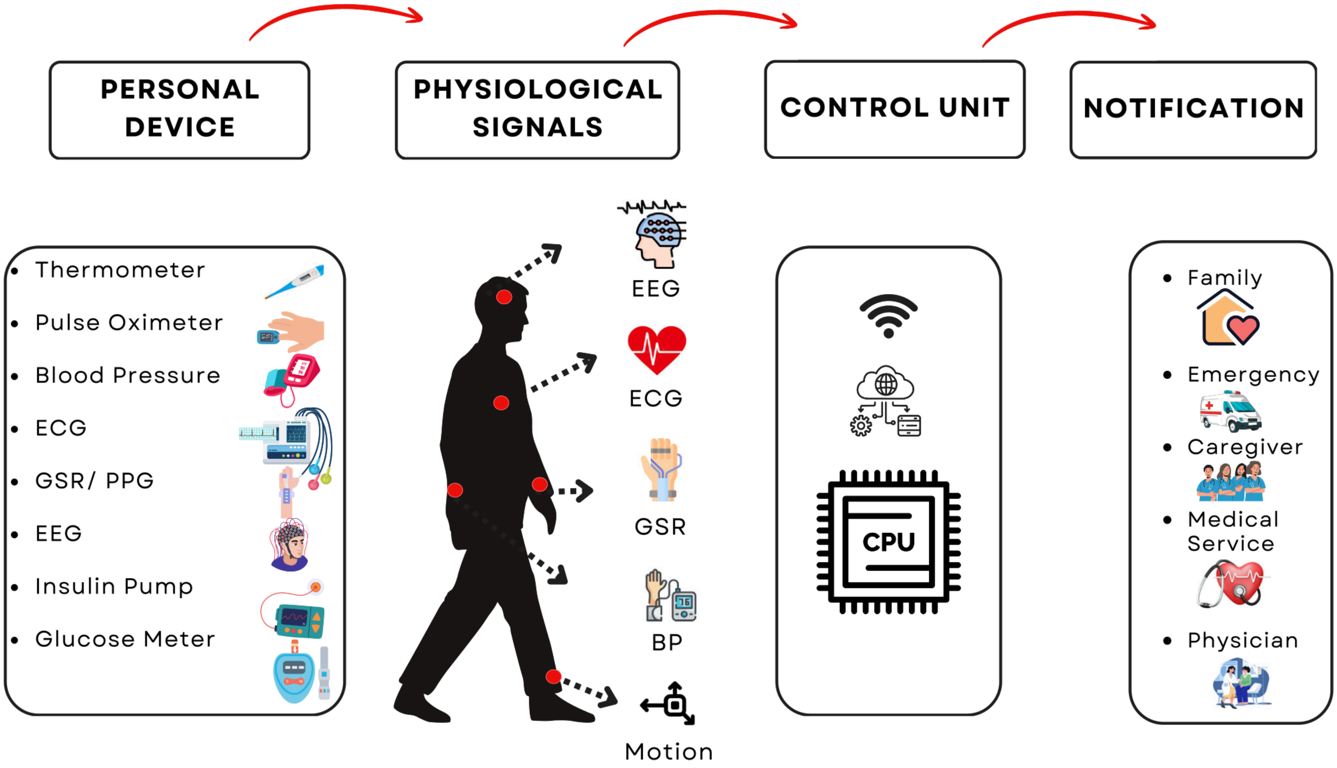


Fig. 1. General diagram for wearable devices.

new emotion-aware applications, but it also introduces signal quality, artifact, and reproducibility issues that affect model performance.

A. AI and Signal-Processing Pipelines and the I&M Chain

Signal preprocessing, feature extraction, and machine learning (ML) model classification or regression are common AI and signal-processing pipelines [4], [5], [6], [7]. Deep learning (DL) methods [8], [9] are becoming more popular because they can learn hierarchical temporal patterns directly from raw or minimally processed signals. These AI-focused advancements have shown encouraging benchmark and cross-subject performance. However, stated metrics sometimes confuse modeling gains with sensor technology, recording techniques, and data quality, factors rarely studied. AI models and signal-processing pipelines for wearable electrocardiography (ECG) or galvanic skin response (GSR) are one primary focus of this article. The measurement chain for wearable emotion recognition includes sensor type and material, electrode placement and contact, skin preparation, analog front-end (AFE) characteristics, sampling rate, filtering, digitization, and on-device preprocessing. Electrode montage, skin–electrode impedance, calibration processes, signal quality indices (SQIs), repeatability, and interdevice comparability affect the extracted ECG and GSR features and the downstream AI model's dependability. However, many public datasets and applied research only partially capture these features, making it impossible to attribute performance to sensing versus modeling, quantify uncertainty, or transfer models across devices, datasets, and clinical or real-world scenarios. A thorough evaluation of instrumentation and measurement (I&M) practices in wearables is our second axis.

B. Applications and Modalities for Emotion Recognition

Emotion recognition has applications, for example, in the care of people with difficulty expressing their emotions to others, mental health support, advertising, robotics, and improved interpersonal communication [10], [11]. Numerous studies have been conducted to identify emotions using various modalities such as facial expressions, body gestures, voice, and physiological signals [12]. However, traditional emotion recognition approaches based on facial expressions or speech features, such as strength, tempo, and tone, often have low accuracy and are influenced by factors such as culture, gender, and age, making them less accessible to all people [13]. The inherent subjectivity and complexity of physiological signals, their susceptibility to motion-induced changes, and the lack of observable emotional signs make it difficult to appropriately label raw data and create trustworthy ground truth [14]. Machine vision-based emotion recognition systems are prone to misrecognizing emotions and can be easily manipulated [15].

C. Overview Architecture of Wearable Emotion-Monitoring Systems

Fig. 1 depicts the general architecture of a wearable device system intended for health and emotional monitoring. The relationship between basic emotions and the valence–arousal dimensions is illustrated in Fig. 2. The process commences with an array of personal devices, including thermometers, pulse oximeters, blood pressure (BP) monitors, ECG patches, GSR and photoplethysmogram (PPG) sensors, electroencephalography (EEG) headbands, insulin pumps, and

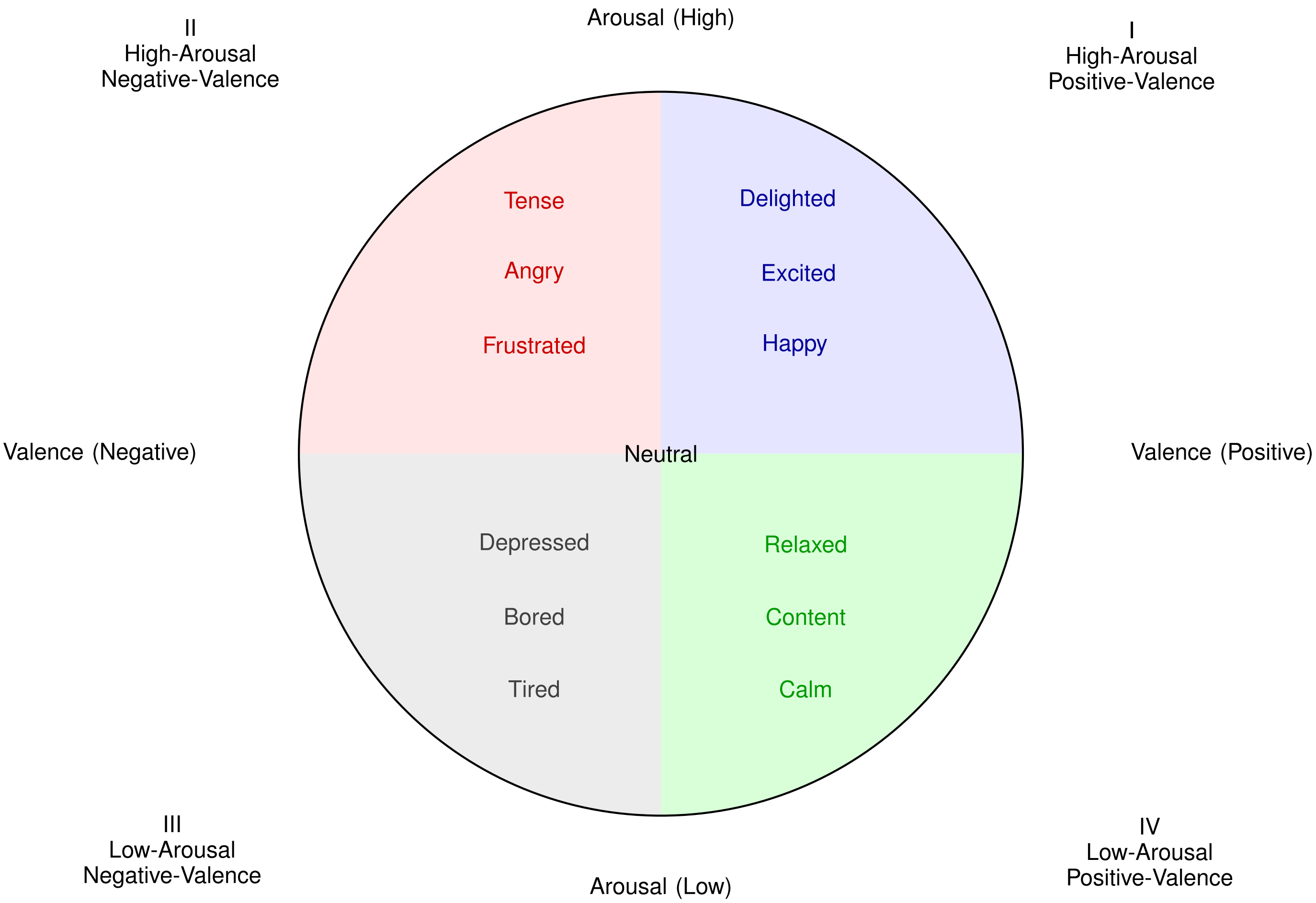


Fig. 2. 2-D valence and arousal model together with several basic emotions.

glucose meters, which gather real-time physiological measurements from the user. The signals encompass, but are not restricted to, ECG, EEG, GSR, BP, and motion data. Upon collection, the data is transmitted to a control device with processing capabilities and wireless connectivity (Wi-Fi, Bluetooth, or cloud-based infrastructure). The control unit analyses incoming signals, whether local or remote, using algorithms to assess the user's physiological and emotional status. In cases of identified anomalies or emergencies, the system can autonomously dispatch notifications to carers, medical practitioners, family members, or emergency services. Under typical circumstances, the system facilitates ongoing assessment of the user's stress, anxiety, and emotional health. This proactive method facilitates early interventions, thereby enhancing user safety and mental well-being.

Importance of ECG and GSR in wearable systems.

1) ECG plays a critical role in monitoring heart activity. In wearable applications, ECG enables continuous cardiac assessment, arrhythmia detection, ECG-derived heart rate variability (HRV) analysis, and emotion recognition, making it vital for both physical and psychological health tracking. GSR
2) GSR measures skin conductance and is highly sensitive to emotional arousal and stress. When used in wearable systems, GSR offers valuable insights into the user's autonomic nervous system activity, making it especially useful for detecting stress and anxiety levels.

Together, ECG and GSR provide complementary insights into both physiological and affective states, reinforcing their importance in developing robust, noninvasive wearable systems for real-time health monitoring and early intervention. While GSR is a marker of sympathetic nervous system activity and is therefore considered to reflect arousal rather than valence, ECG-derived measures may capture aspects related to both arousal and, to some extent, valence. This article examines both how these signals are modeled by AI techniques and how the measurements that make them up are described and shared in public datasets. This connects the sense layer to the performance of emotion recognition further down the line.

D. Motivation

Wearables are now mainstream, with 534.6 M units shipped in 2024, and older adults are already using them (38% of U.S. 50+ own a wearable; 71% use health/wellness apps), making them a practical foundation for emotion-aware care, as reported by IDC [16], [17]. Fig. 1 illustrates the general wearable device system, including the collection of physiological signals from body sensors, cloud-based data transmission, and alerts to family, emergency services, and healthcare providers.

Wearable sensing technologies have also been explored in ergonomics and human-factors research. For example, motion-

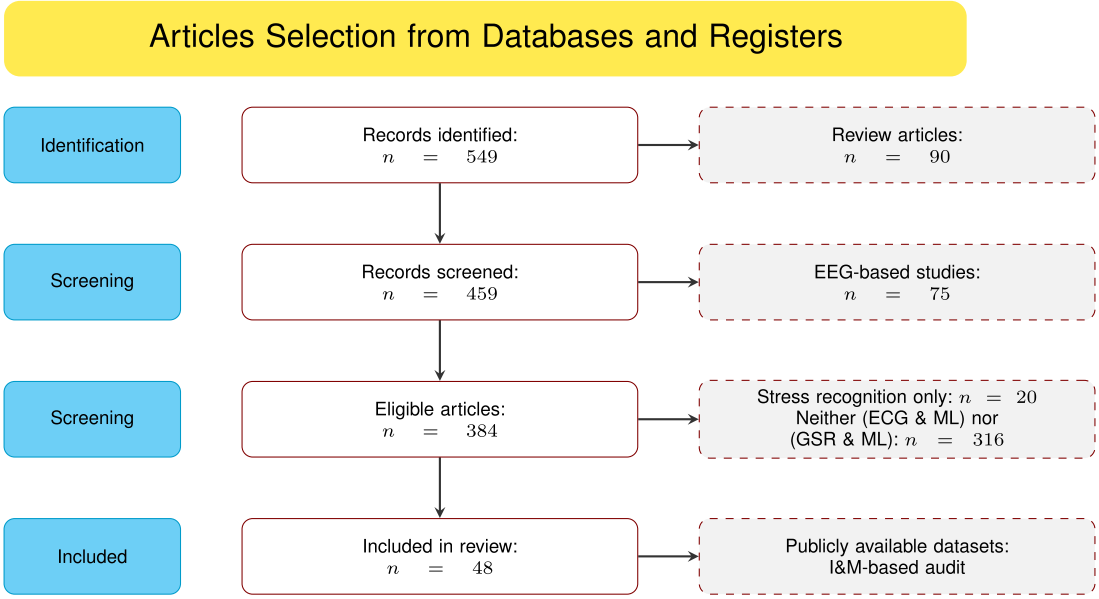


Fig. 3. Screening process and study counts aggregated from four publishers and datasets audits in terms of I&M ML.

based wearable sensors have been used to monitor children's behavior, including ADHD-related studies where hyperactivity behaviors have been identified using wearable sensing systems [18]

Simple signals like ECG/HRV and GSR passively capture autonomic arousal that maps onto stress/emotion, and wearable AI can detect emotions with high accuracy in ML and pooled analyses [16], [19]. However, achieving such performance in practice depends not only on increasingly sophisticated AI models, but also on reliable, well-characterized measurement chains on the devices themselves (sensor type, placement, sampling, and SQIs), which is why our review jointly considers both AI pipelines and I&M aspects.

In dementia, wrist actigraphy is feasible and acceptable, while agitation affects 40%–60% and up to 60% may wander, so detecting rising arousal creates an early, nondrug intervention window that improves safety and preserves quality of life [16], [20]. Recent review studies [21], [22], [23], [24] have highlighted the importance of physiological signals for emotion recognition, noting their reliability compared to speech- or facial-based techniques.

For this reason, many studies have focused on physiological signals, particularly multimodal approaches that combine various physiological data obtained from biosensors such as ECG, GSR, EEG, electromyography (EMG), PPG or blood volume pressure, electrooculography (EOG), electrogastrography (EGG) which has also been explored in autonomic-related contexts [25], and respiratory inductive plethysmography. While multimodal emotion recognition often produces better results, the unimodal technique offers the advantages of faster processing and easier data collection [26]. Several papers studied emotion recognition utilizing a variety of signals. For example, Kessous et al. [12] and D'mello and Kory [27] provided overviews of multimodal emotion identification methods that combine facial expressions, speech, and body motions. Some studies have focused entirely on speech-based emotion recognition [28] or on vision-based systems that use facial expressions [29]. Cosoli et al. [30] and Dai et al. [31] investigated multimodal emotion recognition using wearable biosignals (PPG/EDA), while Du et al. [32] and Li et al. [33] explored noncontact or EEG-based emotion recognition. Li et al. [34] and Apicella et al. [35] studied affective touch and low-channel EEG systems, respectively. Alghoul et al. [36] analyzed PPG-only wearable emotion recognition, and similarly, Irfan et al. [37] studied positive and negative emotions using a single modality (ECG).

Although such approaches exist, current research has revealed an increasing interest in physiological cues. This is because physiological signals are produced involuntarily in reaction to emotional moods and are difficult to manipulate consciously. As a result, they have gained popularity in emotion recognition studies [38]. Yet, very few of these works interrogate how well the underlying wearable measurements are specified, such as sensor materials, electrode placement, calibration, SQIs, or how such I&M details might constrain the generalizability of the AI models that are trained on them, an issue we explicitly address in our study.

E. Gap

While several reviews broadly cover emotion recognition and physiological signals, to date, there has been no comprehensive review in the past five years specifically focusing on emotion recognition using wearable devices' ECG and GSR signals. In addition, no review jointly analyses AI/ML approaches and the I&M practices of the underlying wearable datasets, including sensor characteristics, placement, and signal-quality reporting. This study aims to fill this gap by providing an up-to-date overview of ECG and GSR-based emotion recognition using wearable devices, together with a structured I&M audit of the datasets and measurement chains on which these models rely.

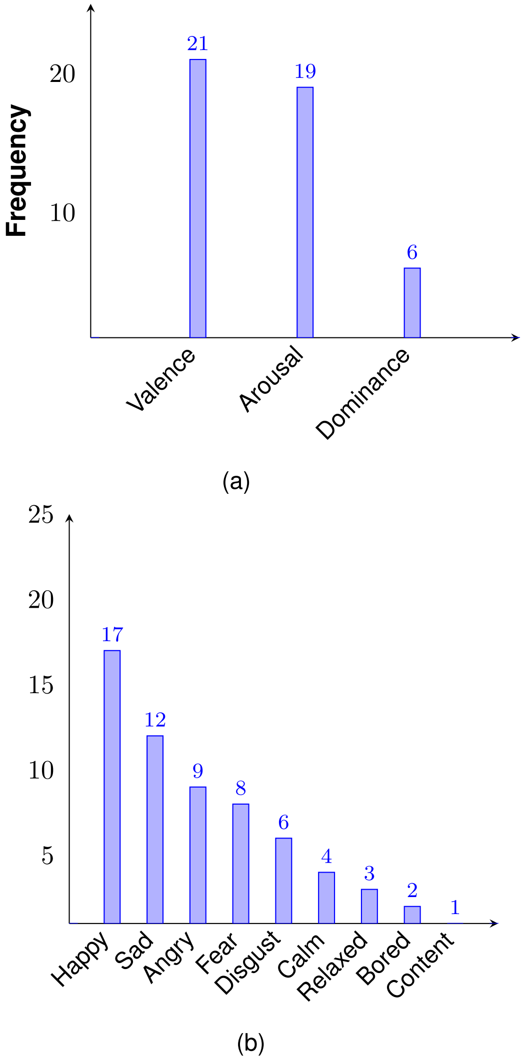
Given advancements in artificial intelligence, particularly in ML algorithms, and in wearable devices, such as smartwatches capable of collecting physiological signals, significant progress has been made in emotion recognition. These developments hold great promise for understanding emotions within the healthcare sector. At the same time, without transparent reporting of how ECG and GSR are acquired, processed, and quality-controlled on wearable platforms, it remains difficult to compare models fairly or to translate them into clinical or home-care settings. In this context, we sought to explore how these innovations can be more effectively utilized for emotion recognition, specifically focusing on the potential benefits they offer to the healthcare industry, not only at the level of AI algorithms but also at the level of I&M standards. To this end, we conducted a review of two primary physiological signals, ECG and GSR, which can be easily obtained from wearable devices and provide valuable insights into emotional states. To ensure the relevance and applicability of our findings, we focused on research published in the past five years and examined how thoroughly these studies and datasets report key I&M metadata needed for robust, transferable emotion-recognition pipelines.

II. METHODS

To identify relevant papers for this review, we conducted a structured literature search across four major academic publisher databases: IEEE Xplore, ScienceDirect, SpringerLink, and Taylor & Francis Online. These databases were chosen for their consistent, advanced search capabilities and their application in engineering, computer science, and health informatics. The search was conducted using terms related to emotion recognition, ML, and wearable sensor technologies. The following Boolean query was used with the syntax modifications appropriate for the database:

"emotion recognition" AND ("ML" OR "DL" OR "neural network (NN)" OR "deep neural network" OR "convolutional neural network" OR CNN OR "recurrent neural network" OR RNN OR "long short-term memory" OR LSTM) AND ("wearable device" OR sensor OR body-worn OR smartphone OR smartwatch OR "face sensor") AND (electrocardiography OR ECG OR "galvanic skin response" OR GSR).

To ensure relevance and quality, we considered peer-reviewed articles written in English and published during the last five years. The term "face sensor" was included in the search string to capture studies involving wearable or multimodal sensing platforms that combine physiological signals with additional sensing modalities. Although the main focus of this review is on ECG- and GSR-based emotion recognition, some wearable systems integrate physiological signals with facial sensing components. Including this term helped ensure that relevant multimodal studies were not unintentionally excluded during the initial search stage. Some widely used emotion datasets like DEAP [39], ASCERTAIN [40], MAHNOB-HCI dataset [41] contain both physiological and facial recordings.The search identified 549 studies combined from the four sources.Although stress and emotion are closely related phenomena, they represent different affective processes. Emotions are typically short-term and momentary responses to internal or external stimuli and can be reflected in physiological signals. In emotion recognition research, these responses are often described using discrete emotional categories such as happiness, sadness, anger, fear, disgust, and surprise, or using the valence–arousal framework. In contrast, stress usually refers to a more prolonged physiological and psychological response to demanding or challenging conditions. For this reason, the present review focuses on emotion recognition rather than stress detection, and studies dealing exclusively with stress recognition were excluded from the analysis. Also, we excluded review articles, EEG-based studies (i.e., without ECG or GSR signals), articles that did not perform emotion recognition, and articles that did not actually utilize GSR or ECG. Finally, 48 studies were left for further analyses (Fig. 3).



(a)

(b)

Fig. 4. Trends in emotion-related research publications (2019–2025). Based on Table I, (a) emotion dimensions and (b) specific emotions.

III. RESULTS

A. *Key Characteristics of the Identified Papers*

Across the reviewed studies, ECG and GSR were generally used as the primary physiological signals for emotion recognition. However, in some studies, these signals were combined with additional modalities such as EEG, PPG,

TABLE I

Summary of the Identified Emotion Recognition Studies. Column "Dataset" Indicates If a Publicly Available Ready Dataset Was Used or If New in-House Data Was Generated. For Readability, Commonly Used Evaluation Metrics Are Abbreviated as: Precision (Prec), Recall (Rec), Accuracy (Acc), $F1$-Score ($F1$), Sensitivity (Sens), Cohen's Kappa (κ), Area Under the ROC Curve (AUC) and Specificity (Spec), Respectively, and Are Defined in [8] and [78]. The Column "# Subjects" Refers to the Number of Individuals Participating in the Study, but Several Samples May Have Been Collected From Each Subject. Studies 1–22 Focus on Arousal/Valence; Studies 23–39 Recognize Distinct Basic Emotions; and Studies 40–47 Focus on Emotion Recognition in Specific Applications

#	Author&Ref.	Dataset	Emotions	Metrics	# Subjects
1	Albraikan *et al.* [64]	MAHNOB, in-house	arousal, valence	acc	43
2	Ali *et al.* [70]	MAHNOB, DEAP, AMIGOS, in-house	arousal, valence	acc	27, 22, 40, 9
3	Althobaiti *et al.* [79]	in-house	arousal, valence	acc, F1	19
4	Fan *et al.* [60]	ASCERTAIN, DREAMER, WESAD	arousal, valence	acc, F1	58, 23, 15
5	Fang *et al.* [48]	AMIGOS, DREAMER, WESAD	arousal, dominance, valence	acc, F1, kappa, prec, rec, spec	40, 23, 15
6	Fiorini *et al.* [58]	in-house	arousal, dominance, valence	acc	15
7	Gahlan *et al.* [49]	AMIGOS, DEAP, DREAMER	arousal, dominance, valence	acc, F1	40, 32, 23
8	Gahlan *et al.* [75]	AMIGOS, DREAMER	arousal, dominance, valence	acc, F1, prec, rec	40, 23
9	Harper *et al.* [80]	DREAMER, AMIGOS	negative, positive, valence	acc	40, 23
10	Ismail *et al.* [55]	DEAP, DREAMER, in-house	arousal, valence	acc, F1	32, 23, 47
11	Kanjo *et al.* [65]	in-house	different degrees of valence	acc, F1, prec, rec	34
12	Li *et al.* [81]	DEAP, WESAD	arousal, valence	acc, F1, prec, rec, spec	32, 15
13	Mokhtari *et al.* [67]	DREAMER, SWELL, WESAD	negative, positive	acc	23, 25, 15
14	Nita *et al.* [61]	DREAMER	arousal, dominance, valence	acc, F1, prec, rec	23
15	Panahi *et al.* [50]	ASCERTAIN	arousal, valence	acc, F1	58
16	Pepa *et al.* [82]	in-house	arousal, valence	acc, F1, prec, rec, spec	19
17	Raheel *et al.* [56]	in-house	arousal, valence	acc, kappa	18
18	Santamaria *et al.* [62]	AMIGOS	arousal, valence	acc, F1	40
19	Sharma *et al.* [42]	DREAMER, AMIGOS	arousal, dominance, valence	acc	23, 40
20	Sweeney *et al.* [83]	ASCERTAIN, DREAMER	arousal, valence	acc, F1, sens, spec	58, 23
21	Veerankia *et al.* [53]	DEAP	arousal, valence	acc, F1, prec, rec, spec	32
22	Woo *et al.* [84]	AMIGOS	arousal, valence	acc	40
23	S. Kumar P and J. Fredo Agastinose Ronickom *et al.* [85]	CASE, WESAD	arousal, valence (CASE); amusement, neutral, stress (WESAD)	acc	30, 15
24	Akbulut *et al.* [63]	in-house	angry, calm, disgust, fear, happy, sad	acc	30
25	Akbulut *et al.* [71]	in-house	angry, calm, disgust, fear, happy, sad	acc, F1, prec, rec	30
26	Alam *et al.* [52]	DEAP	disgust, happy, neutral, relaxed, sad	AUC	32
27	Chen *et al.* [43]	VREED, WESAD	amused, meditative, neutral, stress	acc, F1, prec, rec	34, 15

TABLE I

(Continued.) SUMMARY OF THE IDENTIFIED EMOTION RECOGNITION STUDIES. COLUMN "DATASET" INDICATES IF A PUBLICLY AVAILABLE READY DATASET WAS USED OR IF NEW IN-HOUSE DATA WAS GENERATED. FOR READABILITY, COMMONLY USED EVALUATION METRICS ARE ABBREVIATED AS: PRECISION (PREC), RECALL (REC), ACCURACY (ACC), $F1$-SCORE ($F1$), SENSITIVITY (SENS), COHEN'S KAPPA (κ), AREA UNDER THE ROC CURVE (AUC) AND SPECIFICITY (SPEC), RESPECTIVELY, AND ARE DEFINED IN [8] AND [78]. THE COLUMN "# SUBJECTS" REFERS TO THE NUMBER OF INDIVIDUALS PARTICIPATING IN THE STUDY, BUT SEVERAL SAMPLES MAY HAVE BEEN COLLECTED FROM EACH SUBJECT. STUDIES 1–22 FOCUS ON AROUSAL/VALENCE; STUDIES 23–39 RECOGNIZE DISTINCT BASIC EMOTIONS; AND STUDIES 40–47 FOCUS ON EMOTION RECOGNITION IN SPECIFIC APPLICATIONS

#	Author & Ref.	Dataset	Emotions	Metrics	# Subjects
28	Dissanayake *et al.* [86]	AffectiveROAD, CASE, CLAS, K-EmoCon, PPG-DaLiA, WESAD	happy, sad, angry, cheerful and nervous	acc	10, 30, 62, 21, 15, 15
29	Dominiguez *et al.* [57]	in-house	amused, neutral, sad	acc, AUC	37
30	Francese *et al.* [59]	in-house	sad, contentment, angry, fear	acc	154
31	Houssein *et al.* [87]	in-house	angry, happy, neutral	acc	30
32	Kang *et al.* [88]	IEMOCAP	happy, sad, angry, neutral	acc	11
33	Li *et al.* [89]	WESAD	amused, neutral, stress	acc, F1	15
34	Pradhan *et al.* [44]	WESAD	amused, neutral, stress	acc, F1	15
35	Saha *et al.* [45]	ECSMP	angry, disgust, fear, happy, neutral, sad	acc, F1	52
36	Theerthagiri [46]	WESAD	amused, neutral, stress	acc, F1, prec, rec	15
37	Upadhaya *et al.* [54]	WESAD	amused, meditative, neutral, stress, transient	acc, F1, prec	15
38	Wang *et al.* [47]	in-house	happy, relaxed, calm, sad, fear	acc	15
39	Yang *et al.* [19]	POPANE	amused, disgust, fear, gratitude, sad, tender	acc, F1, prec, rec, spec	616
40	Kumar *et al.* [51]	WESAD, CASE	boring, relaxing, scary, amusement, neutral, stress	acc	15,30
41	Dehzangi *et al.* [90]	in-house	neutral, distracted	acc	15
42	Garg *et al.* [69]	PMEmo, DEAM	angry, fear, anticipation, surprise, joy, sadness, trust, disgust	acc	457
43	Lal *et al.* [72]	in-house	anxiety, neutral, sleepy	acc	unknown
44	Qian *et al.* [91]	in-house	fatigue, frustration and denial	acc	11
45	Sun *et al.* [68]	in-house	angry, calm	acc, F1, prec, rec, spec	32
46	Uluer *et al.* [73]	in-house	neutral, pleasant, unpleasant	acc, cor, F1, prec, rec, spec	19
47	Wang *et al.* [92]	in-house	anxiety, boredom, flow	AUC	27
48	Al Jassmi *et al.* [66]	in-house	happy, neutral, sad	acc	3

or facial features within multimodal emotion-recognition frameworks. Furthermore, different validation strategies were adopted across the reviewed studies. Some studies employed subject-dependent validation, while only a limited number used subject-independent evaluation, and some applied cross-validation to assess model performance.

The number and type of identified emotions varied across the reviewed studies (Fig. 4), with arousal and valence as the most commonly recognized emotions. The study-specific emotions, used datasets, evaluation metrics, and number of subjects are summarized in Table I. In the bar chart, the "happy" category includes the emotions "amused" and "joy."

1) Methodology: Among the reviewed studies, some focus on preprocessing the data and extracting features using various methods, including wavelet-based techniques (see Fig. 5). In contrast, others emphasize optimizing DL models, fine-tuning them specifically for emotion classification tasks (see Fig. 6). A number of studies (see [42], [43], [44], [45], [46], [47]) emphasize preprocessing of physiological signals, particularly ECG signals [48], [49], [50], [51], to enhance the quality of input data. Notably, Pradhan et al. employed separate preprocessing techniques tailored for each physiological signal type [44]. Fig. 5 shows the detailed pipeline of the models, including preprocessing and feature extraction steps.

In addition to preprocessing, feature extraction is another common focus area in these studies. Panahi et al. [50], for instance, utilized the fractional Fourier transform for feature extraction, while Alam et al. [52] leveraged deep CNNs for automated feature extraction. Also, Wang et al. applied a CNN model and used a feature fusion method for feature extraction,

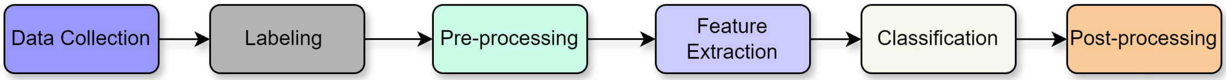


Fig. 5. Flow diagram for models performing preprocessing and feature extraction.

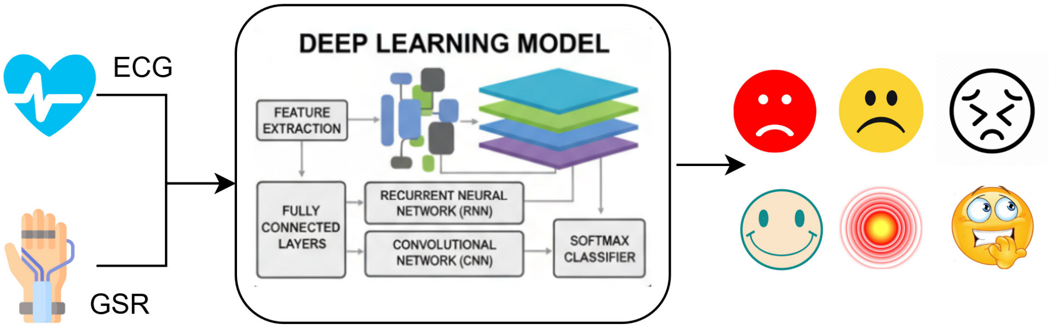


Fig. 6. General diagram of DL approaches implemented across several studies reviewed in this article.

TABLE II

KEY DETAILS OF DATA WHILE CONSIDERING ASPECTS OF MEASUREMENT TECHNIQUES.NR = NOT REPORTED; FS = SAMPLING FREQUENCY; DR/NOISE (NR IN ANY OF THE LISTED STUDIES); AFE = ANALOG FRONT-END; ADC = ANALOG-TO-DIGITAL CONVERTER RESOLUTION; MET. ACC. = MEASUREMENT ACCURACY IN THE METROLOGICAL SENSE (NR IN ANY OF THE LISTED STUDIES); Z_{se} = SKIN–ELECTRODE IMPEDANCE; SQI = SQIS; INTER-DEV = INTERDEVICE COMPARISON

Ref	Device	Elec type	Material	Site	Fs (Hz)	Leads	EDA sep	Calibration	Reprod	Ambient	Motion	Inter-dev	Additional missing measurements
[76]	RespiBAN Professional	ECG: wet Ag/AgCl; GSR: dry	gel (ECG)	chest	ECG: 700; GSR: 4	3-lead ECG	Yes	NR	Partial	lab	multiple tasks incl. stress, rest, meditation	No	Z_{se}, AFE, ADC, BW, DR/Noise, SQI, Inter-dev
[76]	Empatica E4	PPG: optical; GSR: dry	Optical	wrist	PPG: 64; GSR: 4	NR	Yes	Factory	Partial	lab	simultaneous use with RespiBAN	No	Z_{se}, AFE, ADC, BW, DR/Noise, SQI, Inter-dev
[39]	BioSemi ActiveTwo (+ peripheral)	GSR: NR	NR	left-hand fingers (middle & index; distal phalanges)	512→256 (paper) / 128 (preproc)	10-20	No	NR	NR	lab (controlled illumination)	seated viewing (40 × 1 min)	No	Z_{se}, AFE, ADC (bits), BW, DR/Noise, SQI, calibration/traceability, reproducibility, explicit electrode/material details
[77]	Emotiv EPOC Neuroheadset	NR	NR	EEG scalp (10–20); ECG/GSR: –	EEG 128; 14 bit resolution	EEG scalp (10–20)	–	NR	NR	Lab	Seated viewing; individual & group	No	Z_{se}, AFE, BW, DR/noise, SQIs, calibration/traceability, repeatability/reproducibility
[77]	Shimmer 2R (ECG, GSR)	ECG: NR; GSR: NR	NR	ECG: R/L arm crooks + left ankle (ref); GSR: left-hand middle & index phalanges	ECG 256; GSR 128, 12 bit resolution	3-electrode ECG	NR	NR	NR	Lab	Seated viewing; individual & group	No	Z_{se}, AFE, BW, DR/noise, SQIs, calibration/traceability, repeatability/reproducibility
[74]	Emotiv EPOC	NR	NR	EEG scalp (10–20)	EEG 128	International 10-20 system (14 ch used)	–	NR	NR	Lab	Seated viewing (18 clips)	No	Z_{se}, AFE/ADC, BW, DR/noise, SQI, calibration/traceability, reproducibility
[74]	Shimmer (wireless ECG sensor)	NR	NR	NR	ECG 256	1-lead (single-channel)	–	NR	NR	Lab (controlled illumination)	Audio and visual stimuli in the form of film clips (18 film clips)	No	Z_{se}, AFE/ADC, BW, DR/noise, SQI, calibration/traceability, ECG electrode type/material/site, reproducibility
[40]	EEG: NeuroSky (single dry); ECG: 3-ch (device NR); GSR: NR	EEG: dry; ECG/GSR: NR	NR	NR	NR	ECG: 3-ch (lead cfg NR)	No (band-power features only)	NR	NR	Lab (movie clips)	Seated viewing	No	Z_{se}, AFE/ADC, BW, DR/noise, electrode material/site (ECG/GSR), SQIs, calibration/traceability, repeatability/reproducibility

combining both CNN and LSTM-MLP architectures [47]. Veerankia et al. [53] utilized a transition network to carry out effective feature extraction, and Upadhaya et al. [54] implemented a sequential feature extractor. In addition, several studies emphasized feature extraction techniques applied directly to raw signal data to enhance robustness [42], [44], [45], [55]. Gahlan and Sethia [49] and Raheel et al. [56], on the other hand, extracted statistical, time-domain, and frequency-domain features to comprehensively represent the physiological data. While most studies focus on feature extraction, Domínguez-Jiméenez et al. [57] applied the random forest recursive feature elimination method for feature selection. Table II focuses on the measurement techniques, while Table III summarizes the physiological signals and representative extracted features reported in the reviewed studies. The most commonly used signals include ECG, EDA/GSR, and PPG. From ECG signals, studies typically extracted heart rate (HR), HRV, RR intervals, and frequency-domain features. EDA/GSR signals were mainly used to obtain statistical measures, skin conductance level (SCL), skin conductance responses (SCR), and peak-related features. For PPG signals, features such as interbeat intervals and HR/HRV-related measures were reported.

A wide range of ML and DL methods were employed for the actual emotion recognition across the reviewed studies. Fig. 7 presents a summary of the methodological selections identified in our review (2019–2025), indicating that DL is the most common, followed by classical ML, with a limited number of

TABLE III
PHYSIOLOGICAL SIGNALS AND EXTRACTED FEATURES USED IN THE REVIEWED STUDIES

Ref	Signals	Extracted Features
[50]	ECG, GSR	HR, HRV, IBI, LF/HF, statistical and derivative features
[52]	ECG, EDA	HRV, pNN50, PSD (LF, HF), Poincaré features, SCL, SCR
[47]	EEG, ECG	HRmean, RRmean, SDNN, RMSSD, pNN50, LF, HF
[53]	EDA	tonic/phasic decomposition, transition network features
[54]	ECG, EDA, EMG, Resp.	Feature selection on multimodal sensor channels (temperature, respiration, acceleration)
[44]	EDA, ECG, EMG, Resp.	Deep features extracted from signal images using CNN
[45]	EEG, ECG, PPG	Time-, frequency-, and time–frequency-domain features
[55]	ECG, PPG	PQRST peaks, HRV, spectral and multiscale entropy (PPG)
[42]	EEG, ECG	Entropy features (IP, CEC)
[49]	EEG, ECG, GSR, Resp.	Statistical, time-domain, and frequency-domain features
[56]	EEG, GSR, PPG	EEG time- and frequency-domain features, GSR statistical features, PPG HR/HRV
[57]	PPG, GSR	HRV features, spectral features, EMD-based GSR features

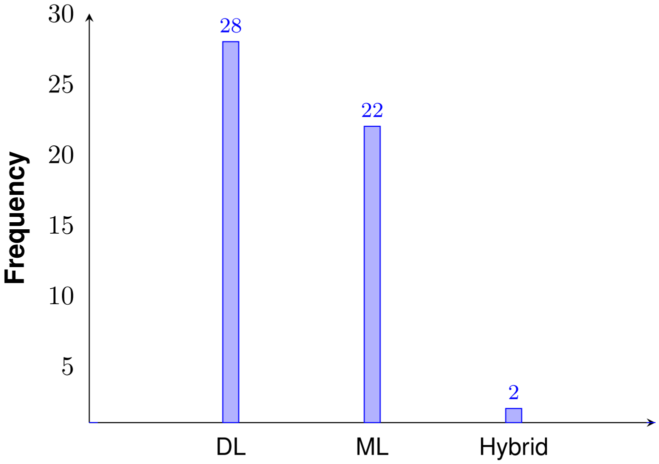


Fig. 7. Distribution of studies by methodology: DL, traditional ML, and hybrid approaches (2019–2025).

Hybrid techniques. Several studies, particularly those focusing on feature processing, tested multiple basic ML approaches at the emotion classification stage [53], [54], [55], [58], [59]. Different NN-based approaches, particularly CNN structures, were the most popular emotion classifiers among the reviewed studies [43], [46], [47], [60], [61], [62], [63]. Some studies also introduced hybrid models [51], also combining NN with other approaches [44], [63], [64], [65] or compared deep and shallow networks [63]. Besides NN, several lighter ML methods have been successfully applied in the literature. For example, Al Jassmi et al. [66] and Mokhtari et al. [67] utilized random forest models for emotion classification. Also, SVM has been used in a couple of studies: Panahi et al. [50] used binary SVMs for classification, and Sun et al. [68] used it to identify the emotional stage during driving. One study applied the k-nearest neighbor (k-NN) algorithm to detect emotions from multimedia content after extracting time–frequency-domain features [56]. Finally, different regression models have been used for emotion classification after careful feature processing [47], [48]

2) Other Signals Than ECG and GSR: In addition to ECG and GSR, which were used as search criteria in this review, the identified studies also used other signals. In particular, EEG has been used in several studies [52], [56], [57], [69]. PPG is another frequently used signal, featured in multiple works introduced here [45], [55], [56], [57], [70]. Additional features such as HR and body temperature were applied in studies including those by Domínguez-Jiméenez et al. [57], Akbulut et al. [63], [71], and Al Jassmi et al. [66]. Less commonly used signals include accelerometer data [57], blood volume pulse [66], while derived parameters such as oxygen saturation (SpO_2) [71], and respiration-related measures, respiration rate, and general respiratory data [64], [66] were also considered. In addition to such physiological signals and features, a few studies utilized facial information [70], and both facial and audio modalities were used in some studies [64], and in some studies, audio data were used [69], [72], [73]. These diverse modalities highlight the broad spectrum of physiological and multimodal features explored in recent emotion recognition systems.

3) Obtained Accuracies: While comparing accuracy results across studies can be challenging due to methodological differences, it is still informative to present performance variations on the same public dataset. However, it is important to note that class definitions, label discretization, and data-splitting strategies are not always identical. These methodological differences may influence the reported accuracy values and limit the direct comparability of results. The DREAMER [74] dataset, in particular, has been widely used, and various studies have reported differing accuracy scores. Fang et al. [48] achieved high performance on DREAMER with 0.95 for valence, 0.98 for arousal, and 0.90 for dominance. Gahlan and Sethia [75] reported an accuracy of 0.84 on DREAMER while also achieving 0.88 on the AMIGOS dataset.

Similarly, Nita et al. [61] reported 0.95 for valence, 0.85 for arousal, and 0.77 for dominance. Fan et al. [60] experimented with different CNN architectures on DREAMER, obtaining 0.83 for both valence and arousal with a base CNN and slightly improved performance with CNN + CBAM (0.84 for valence and 0.83 for arousal). They also achieved 0.87 accuracy for valence classification using a three-class division. Ismail et al. [55] observed that ECG performed better for arousal (accuracy of 0.68), while PPG was more effective for valence (0.64) and emotion dimension classification (0.37). Sharma and Bhattacharyya [42] achieved a high dominance classification accuracy of 0.92 using the DREAMER dataset. In another study, Gahlan and Sethia [49] reported an accuracy of 0.79 on DREAMER. These varying results demonstrate the influence of the preprocessing steps, feature extraction approaches, model architectures, and classification strategies on performance outcomes, even when using the same dataset. However, the reported accuracies are not directly comparable despite using the same dataset, as the split between training and testing varies across studies, and the number of classes to predict also influences accuracy. While traditional ML models like SVM and RF are still commonly used, DL models, particularly CNN and hybrid architectures, often outperform them in terms of accuracy, especially when working with time-dependent physiological signals, such as ECG or GSR, and ECG showed higher classification accuracy for detecting arousal levels than GSR in several studies.

4) Training and Testing: Notably, all of the studies presented in this article use cross-validation or simple data splits into training, testing, and validation sets to evaluate their proposed methods. Ideally, a completely independent validation dataset would be used; otherwise, it remains unclear how well the method generalizes to data with minor differences from the training data, such as when a different device is used to measure the input signals. This is a clear gap in the literature.

B. Commonly Used Public Datasets

In this section, we discuss several public datasets used in multiple studies reviewed. The DEAP dataset [39] was used in six emotion recognition studies selected in our review paper. The dataset comprises EEG and physiological signals, including GSR and ECG, from 32 participants while they viewed 40 short music video clips. For a subset of participants, face videos were also recorded. All participants rated the arousal, valence, like/dislike, dominance, and familiarity for each video. The dataset is publicly available, but accessing it requires filling out the end-user license agreement.

Dataset WESAD [76] contains physiological signals recorded using a chest-worn device and a wrist-worn device during three emotional states (stressed, neutral, amused) from 15 subjects. The stress condition was induced by a speech in front of a judgmental panel, combined with a mentally demanding arithmetic task, whereas the amusement was stimulated by watching funny videos. The participants also completed self-reports as objective ground truth. The dataset is publicly available for scientific, noncommercial purposes.

Also, the AMIGOS dataset [77] contains physiological signals, such as EEG, ECG, and GSR, while participants watch emotion-provoking videos. The 40 participants watched 16 short video clips alone and 4 longer ones either alone or in a group. Several emotional aspects, namely valence, arousal, control, familiarity, liking, and basic emotions, during video watching were self-reported. In addition, participants were filmed during the experiment, and an external panel of evaluators ranked their arousal and valence from the recorded facial videos. Like the DEAP dataset, AMIGOS is publicly available, but accessing the data requires completing the end-user license agreement.

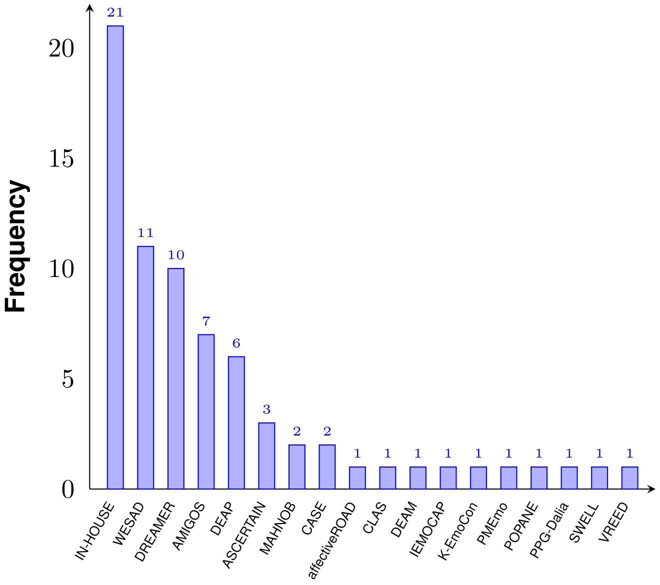


Fig. 8. Counts of datasets used in emotion-related studies (2019–2025) based on this review.

DREAMER [74] provides features derived from EEG and ECG signals from 23 young, healthy participants while they watched 18 video clips that stimulated different emotions. The participants self-assessed their affective state regarding arousal, dominance, and valence after the different video clips. The dataset, available on Zenodo, must be requested via a simple form.

Similar to DEAP, ASCERTAIN [40] contains EEG, ECG, GSR, and facial landmarks during the 58 participants watching 36 video clips. The physiological signals were combined with self-reported arousal, valence, liking, engagement, and familiarity. The distinguishing feature of this dataset is the provided personality trait scores, which can be investigated alongside emotional responses. Access to ASCERTAIN data requires the end-user license agreement. As a high-level overview of adoption within our corpus, the datasets most frequently used are illustrated in Fig. 8.

1) Dataset Details From an I&M Viewpoint: In physiological ML, model accuracy is reliant upon the reliability of the underlying measurements. High input impedance, precisely defined AFE, and explicitly stated bandwidths mitigate morphology loss and bias in derived features such as HRV, while converter resolution and effective number of bits constrain quantization noise and dynamic range (DR), thereby directly influencing signal quality and downstream separability. The documentation of skin–electrode impedance and the materials/sites of electrodes elucidates the limitations of contact quality and motion sensitivity; standardized signal quality metrics formalize data inclusion and exclusion, rendering cross-study standards justifiable. Ultimately, calibration, traceability, and interdevice concordance serve as the connection between "functions on my device" and portable, uncertainty-aware artificial intelligence. In the absence of this metadata, enhancements attributed to algorithms may only indicate sensor or front-end characteristics; with it, we can accurately attribute improvements, transfer models across hardware, and present error bars that comply with measurement-science norms [33], [93], [94], [95], [96].

2) Survey of Measurement Techniques in Public Physiological Datasets for Emotion Recognition: In the field of I&M, the accurate acquisition, processing, and analysis of physiological signals, such as electrocardiogram (ECG), electroencephalogram (EEG), and GSR, are paramount for advancing measurement science in biomedical applications. This survey examines key public datasets (WESAD, DEAP, AMIGOS, DREAMER, and ASCERTAIN) used in emotion recognition, focusing on their measurement methodologies, instrumentation specifications, and gaps in metrological reporting. By evaluating electrode types, AFEs, sampling frequencies (F_s), DR/noise, skin–electrode impedance (Z_{se}), bandwidth (BW), SQIs, calibration/traceability, and interdevice comparability, we highlight how incomplete documentation undermines measurement traceability and uncertainty quantification. Addressing these gaps through standardized metadata enhances the reproducibility of measurement protocols, supports the design of robust instrumentation systems, and facilitates uncertainty-aware applications in ML, driven healthcare predictions, such as stress detection via HRV analysis [22].

Table II summarizes the measurement techniques across datasets, revealing consistent omissions in critical I&M parameters that affect signal fidelity and cross-study validation. From an instrumentation perspective, WESAD utilizes the RespiBAN Professional for chest-mounted ECG/GSR measurements and the Empatica E4 for wrist-based PPG/GSR, following a structured protocol (baseline, stress, amusement, meditation). It delineates the per-channel sampling frequencies F_s, such as ECG 700 Hz, BVP 64 Hz, EDA 4 Hz, and synchronization resources [76]. Nonetheless, deficiencies in metrology, such as electrode materials (wet Ag/AgCl inferred from images), ECG lead configuration, Z_{se}, AFE/ADC specifications, bandwidth, DR/noise, signal quality index, calibration/traceability, reproducibility, and interdevice agreement impede uncertainty estimation and comparability, as highlighted in recent I&M studies regarding wearable signal fidelity [97], [98].

DEAP employs the BioSemi ActiveTwo system for 32-channel EEG and peripheral signals (EOG, EMG, GSR at distal phalanges), recorded at 512 Hz (downsampled to 128 Hz for public access) during the viewing of music videos accompanied by affect evaluations [39]. This configuration facilitates comprehensive measurement methods but is deficient in electrode materials, AFE/ADC specifications, bandwidth, DR/noise, Z_{se}, calibration, and SQI, hindering the assessment of measurement quality in emotion benchmarks [39]. Explicit inclusion would conform to I&M standards for traceable physiological monitoring [93], [99].

AMIGOS documentation Shimmer 2R specifications for ECG (256 Hz, 12-bit, arm/ankle locations) and GSR (128 Hz, phalangeal sites), synchronized with Emotiv EPOC EEG in laboratory observation procedures [77]. These facilitate repeatable segmentation; nevertheless, the lack of metrology (Z_{se}, AFE/BW, noise/DR, SQI, calibration, repeatability) constrains error budgeting.

DREAMER incorporates Emotiv EPOC EEG (128 Hz, 14 channels, 10–20 system) and Shimmer ECG (256 Hz, RA-LL lead) within a dimly lit laboratory, utilizing clip viewing and SAM ratings [74]. It enables HRV extraction and band analysis but excludes details on electrode/Z_{se}, AFE/BW/ADC/ENOB, noise, SQI, calibration, and interdevice testing, limiting morphological studies and generalizability [74]. Future improvements may encompass gain/passband specifications, as well as Z_{se} ranges for justifiable prefilters.

ASCERTAIN records synchronized single-channel dry EEG (NeuroSky), three-lead ECG, and GSR with quality annotations during clip observation involving 58 patients [40]. Strengths comprise covariates for variability; weaknesses involve under-reported metrology (materials, Z_{se}, AFE/ADC/BW/noise/DR, calibration, repeatability), which constrain SNR and benchmarking [40].

Our review shows that although most commonly used public datasets report sensor types, acquisition setups, and sampling frequencies, the measurement accuracy of wearable devices is rarely reported.

Publicly incorporating established metrology standardizes measurement techniques, facilitating traceable preprocessing, quantified uncertainty budgets, and interinstrument comparability, fundamental to I&M theory and practice. This reduces domain shifts such as lab-to-wearable for physiological monitoring, enhancing ECG/GSR accuracy in emotion detection [22]. In AI healthcare, it facilitates uncertainty-aware ML such as SQI-gated predictions, improving repeatability in cross-dataset benchmarks and expediting clinical translation with FDA-traceable equipment [100]. This survey enhances I&M by pinpointing deficiencies and suggesting metadata frameworks, and by promoting innovative measurement techniques for multimodal physiological systems.

C. *Summary of the Literature*

1) Arousal and Valence: Eight of the studies detecting arousal and valence used in-house datasets. Albraikan et al. [64] proposed a system combining base models for emotion classification and enhanced with a meta-learning ensemble approach. The system was evaluated using their dataset and the MAHNOB-HCI dataset [41]. The in-house dataset included peripheral physiological signals, whereas MAHNOB-HCI provided multimodal data, including facial videos, audio, eye gaze, and physiological signals. Also, Ali and Hughes [70] and Ismail et al. [55] combined their datasets with publicly available datasets for validation. Ali and Hughes combined physiological signals with facial information, while Ismail and colleagues focused on feature extraction combined with several basic ML classifiers. The remaining five studies with in-house data did not use any public datasets to support validation. Althobaiti et al. [79] employed supervised classification to predict valence and arousal during interaction with horses, while Pepa et al. [82] evaluated several classical ML models on data collected from people with Parkinson's disease. Also, Fiorini et al. [58] evaluated several methods, comparing unsupervised with supervised learning. Kanjo et al. [65] and Raheel et al. [56] focused on feature processing; the first one introduced a model that processes raw sensor data directly, bypassing manual feature extraction, and the second one extracted them in time and frequency domains for classification using k-NN with feature-level fusion.

Two studies using public datasets compared different classical ML algorithms for the recognition of arousal and valence. Santamaria-Granados et al. [62] applied a deep CNN to ECG and GSR data from the AMIGOS dataset and compared the results with classical ML algorithms. Veeranki et al. [53] used a transition network for feature extraction, followed by classification using logistic regression, MLP, random forest, and SVM. The studies by Sharma and Bhattacharyya [42] and Panahi et al. used SVM to classify emotions based on carefully extracted and preprocessed features. Sharma and Bhattacharyya used sliding-mode singular spectrum analysis for preprocessing and extracted features such as information potential and centered correntropy. Meanwhile, Panahi and colleagues used the fractional Fourier transformation for feature extraction. Besides SVM, ridge regression has been used for emotion classification, as in Fang et al. [48], who applied it to ECG signals processed with random convolutional kernels.

Different NN approaches were well represented in the actual classification step. For example, Fan et al. [60] introduced a deep CNN with residual structures and a convolutional block attention module, whereas Li et al. [81] focused on feature processing using a cross-modal transformer and low-rank fusion, followed by a self-attention transformer, while employing a simple Conv1D as the final classification model. Sweeney-Fanelli and Imtiaz [83] introduce a temporal CNN for real-time emotion detection, and Woo et al. [84] developed a deep multimodal architecture with a modality-aware attention mechanism to classify levels of arousal and valence. Harper and Southern [80] presented an end-to-end DL approach incorporating a Bayesian framework to model uncertainty in predictions.

Besides methodology related to feature extraction/processing and the actual classification, other topics have been the focus of the revised studies. For example, two of them utilized data augmentation. Mokhtari et al. [67] used ECG signals from augmented datasets to classify emotions with a random forest classifier. Nita et al. [61], on the other hand, applied a seven-layer CNN with data augmentation techniques such as randomizing, concatenating, and resampling. Privacy is often an overlooked factor in emotion recognition, but Gahlan and Sethia [49], [75] published two studies simultaneously that address the recognition of arousal, valence, and dominance from physiological signals using a privacy-sensitive federated learning approach.

2) Multiple Basic Emotions: Several studies that classify basic emotions build and evaluate their approaches on the public WESAD dataset. Upadhaya et al. [54] and Pradhan et al. [44] focused on preprocessing the available signals using a framework based on a sequential feature extractor and a multistep pipeline including preprocessing, signal-to-image conversion, and feature extraction/selection, respectively. Li et al. studied the same dataset with a focus on general and personalized models [89], while Theerthagiri [46] proposed a deep belief network and transfer learning method using CNNs and discrepancy reduction techniques. Studies by Chen et al. [43], and Dissanayake et al. [86] also use datasets other than WESAD. Dissanayake et al. [86] used up to four datasets to demonstrate the performance of their proposed method, with special attention to robustness against losses in the input signals. Chen et al. [43] evaluated random forest models and regression analyses for emotion recognition from WESAD combined with the VREED dataset. VREED contains behavioral and physiological signals and affective states during virtual reality stimuli [101].

The emotion classification task becomes harder as the number of emotions increases. Our search identified five studies that classified at least six basic emotions. Akbulut and colleagues have identified six basic emotions in two studies. During the first one, they collected an in-house dataset and performed emotion classification using a hierarchical approach combined with feature extraction via autoregressive hidden Markov models [63]. Later, Akbulut et al. [71] proposed a hybrid model combining a CNN and a random forest to improve their classification accuracy. A fairly similar set of emotions, only calmness replaced with neutral, was classified by Saha et al. [45]. They developed a 1-D DL model for feature extraction and fed the extracted features to multiple basic ML tools to classify emotions from the publicly available ECSMP dataset [102]. Yang et al. [19] converted trained artificial neural networks into spiking feed-forward NNs using weight normalization techniques to classify basic emotions from a very large POPANE dataset [103].

Among the remaining six studies identifying at most five basic emotions from other datasets than WESAD, most used in-house data. The two exceptions were the study by Alam et al. [52], which focused on feature extraction with a deep CNN for emotion recognition using the previously introduced DEAP dataset. Kang et al. [88] utilized the EMOCAP dataset for training and an in-house dataset for model evaluation in emotion classification tasks. The emotions were "happy, sad, angry, and neutral." Domínguez-Jiméenez et al. [57] used video clips to induce emotions in 37 participants while collecting physiological signals. They classified the emotions using SVM combined with random forest recursive feature elimination for feature selection. Francese et al. [59], on the other hand, used multiple classifiers, among which the decision tree classifier achieved the best performance with accuracies of 0.84 and 0.91 in two experiments. In this study, images were used to stimulate different emotions. Wang et al. [47] classified five basic emotions of 15 participants who listened to different music tracks to provoke different emotions. Emotion classification was performed using a long short-term memory multilayer perceptron and a CNN architecture for feature fusion and classification. Exceptionally, Houssein et al. [87] did not trigger different emotions, but the participants simply self-reported them as they naturally occurred. Their study focuses on the explainability of the results.

3) Special Applications: In this section, we introduce studies focusing on emotion recognition in specific situations (the last block in Table I). The most common application was driving a vehicle. Dehzangi et al. [90] aimed to identify distracted drivers from focused ones using GSR as an input. They obtained their best accuracy of 0.935 using an embedded random forest feature selection combined with an ensemble bagged classifier. Instead of distraction, Sun et al. [68] recognized drivers' anger using SVM on ECG and behavioral signals, such as speed and steering wheel angle. The third driving-related study is relatively more complex, as Lal et al. combined emotion recognition with multiple approaches. Their proposed scheme, based on an SVM with a Radial Basis Function kernel, achieves a high accuracy of 0.95, outperforming other methods such as the mood classification approach (0.55), driver mood analysis (0.65), and the music recommender system (0.83) [72].

Besides driving, other practical conditions have been studied. Al Jassmi et al. [66] monitored construction workers' physiological signals and mood over four days for emotion classification and to analyze the connection between emotions and productivity. The study by Wang et al. [92] identifies anxiety, boredom, and flow during online learning experiences. Among the evaluated emotions, flow was the easiest to identify, as a multilayer perception based on the fusion of ECG and posture information achieved an accuracy of 0.8089 for it [92]. Another special application is the consideration of emotional factors for successful rehabilitation after a stroke, as studied by Qian et al. [91]. Using an NN on ECG and other signals, they aimed to identify different negative emotions, as these may negatively impact the rehabilitation of stroke patients [91]. Uluer et al. [73], on the other hand, investigated how children with hearing disabilities react to the same task done conventionally, on a tablet, or with a social humanoid robot. Besides GSR, facial videos were used as physiological signals for the emotion classification. Finally, Garg et al. [69] focused on predicting music mood from audio signals and mapping it to human emotions using EEG, GSR, ECG, and pulse data in real-time.

IV. DISCUSSION AND CONCLUSION

A. Discussion

In this systematic review, we explored how recent studies have used physiological signals and ML to recognize emotions through wearable devices. These studies show a wide range of approaches across the types of signals used, data processing, datasets chosen, model architectures, and feature extraction and evaluation techniques.

In our review, we focused on papers that used ECG and GSR signals for emotion classification, as these are among the most commonly used physiological markers in this field. ECG was frequently reported to be more effective for detecting arousal; for example, according to Ismail et al. [55], GSR also showed high correlations with arousal-related changes. Also, multimodal signals outperformed unimodal ones [51]. In addition to ECG and GSR, several studies found that additional signals, such as PPG, body temperature, and EEG, can boost model performance by capturing complementary features of emotional states. For example, in several circumstances, PPG outperformed other methods in valence categorization. A few studies also analyzed less commonly used signals, such as respiration rate (Albraikan et al. [64]), SpO_2 (oxygen level in the blood) derived from PPG signal (Akbulut et al. [71]), and facial or postural data (Wang et al. [47]), demonstrating an increasing interest in multimodal emotion recognition methods.

The most frequently used datasets in emotion recognition research were DREAMER, WESAD, AMIGOS, and DEAP. As noted in the Results section, methodological differences limit the direct comparability of performances, even when the same datasets are used. For example, studies using the DREAMER dataset reported a wide range of accuracy scores, from 0.68 to 0.98. These differences were mostly due to variations in data preprocessing steps, the choice of ML models, and the emotion dimensions analyzed (such as valence or arousal). Similar findings can be seen in the studies by Fang et al. [48], Gahlan and Sethia [49], and Sharma et al. [42].

One of the most significant findings is the methodological variability among studies. Some concentrated on advanced preprocessing of raw physiological signals, particularly ECG, to eliminate noise and improve signal quality (Fang et al. [48], Pradhan et al. [44]). Others focused on feature extraction, either by employing traditional statistical and frequency-domain techniques (Gahlan and Sethia [49], Raheel et al. [56]) or by utilizing DL for automated feature learning.

Also, in studies on emotion recognition, a variety of ML algorithms have been explored. Traditional techniques, such as SVM, KNN, random forest, and decision trees, were widely used because of their simplicity, interpretability, and performance on small datasets. DL models, specifically CNNs, LSTMs, and hybrid models, performed better at learning complex, time-dependent patterns in physiological inputs. CNN–LSTM models (Kanjo et al. [65]) and attention-based architectures (Fan et al. [60]) obtained higher accuracies, demonstrating the benefit of combining spatial and temporal learning. Valence and arousal were commonly used to classify emotions, following the dimensional approach, which includes multiple emotions such as happiness, sadness, disgust, fear, surprise, and anger. However, the classification prediction results often varied. In several studies, ECG signals were found to be more predictive of arousal than valence, whereas others, including PPG, showed improved performance in valence classification. These findings suggest that distinct physiological signals are better suited to detecting distinct emotional dimensions.

Beyond this, our study also highlights significant diversity and recurrent shortcomings in the I&M-related reporting of research publications concerning emotion recognition utilizing wearable technology and AI. Numerous publicly available datasets provide inadequate information on sensor type, electrode material, electrode positioning, skin preparation, sampling rate settings, filtering, or the use of SQIs. Specifications for calibration processes, repeatability, and interdevice comparability are infrequently provided. This complicates the determination of whether reported improvements stem from enhanced algorithms, superior sensor hardware and protocols, or more rigorous data cleansing. It further complicates device transfer, cross-dataset evaluation, and any efforts to develop uncertainty-aware or SQI-aware emotion identification pipelines.

B. Conclusion and Future Work

This review examined recent advances in emotion recognition using physiological signals and ML, with particular emphasis on ECG and GSR data. Our data indicate that ECG is frequently more successful for arousal classification and may also provide information related to emotional valence, although GSR primarily reflects changes in arousal rather than valence. Traditional ML methods, such as SVM, RF, and k-NN, remain popular, but DL approaches, such as CNNs and hybrid models, outperform them, particularly for complex, time-dependent physiological data. The use of multimodal signals, such as ECG, GSR, PPG, and body temperature, frequently enhances accuracy by capturing many components of emotional states. While many studies report relatively high accuracy, challenges such as limited dataset sizes, variations in emotion labeling, and a lack of standardization across studies continue to affect the generalizability of results.

From an I&M perspective, our audit of the most commonly used public datasets (WESAD, DEAP, AMIGOS, DREAMER, and ASCERTAIN) reveals a consistent pattern of underreporting. While sensor types and sampling frequencies are generally documented, critical metrological parameters are almost universally absent: Z_{se}, AFE specifications, ADC resolution and effective number of bits, input-referred noise, DR, bandwidth, materials, and SQIs. None of the audited datasets reports calibration against a reference instrument or traceability to a metrological standard, and interdevice comparability assessments were not performed. These omissions make it nearly impossible to construct a measurement uncertainty budget for the extracted physiological features and, consequently, to determine whether reported accuracy improvements originate from superior modeling or from differences in sensor hardware and recording conditions. For instance, the wide accuracy range observed on the DREAMER dataset may partly reflect undocumented variability in electrode contact quality and front-end characteristics rather than solely algorithmic differences. Establishing minimum I&M reporting standards, analogous to those in clinical measurement science, is therefore essential for the maturation of wearable emotion recognition as a reproducible discipline.

Future wearable studies and dataset releases should: 1) report sensor hardware (manufacturer, AFE gain and passband, and ADC resolution), electrode configuration (type, material, and anatomical site), and acquisition settings (sampling rate, filtering, and synchronization); 2) Z_{se} ranges and skin preparation procedures; 3) provide SQI and segment inclusion/exclusion criteria; 4) include calibration data with traceability to reference instruments, together with evidence of repeatability (same device) and reproducibility (different operator/device); and 5) release this I&M metadata alongside the physiological recordings to enable fair cross-study comparison, device-transfer experiments, and uncertainty-aware AI pipelines. Such a reporting framework would help bridge the gap between sensing and modeling, allowing performance gains to be attributed more reliably and supporting emotion-recognition systems whose reliability can be quantified for clinical and home-care use.

CONFLICT OF INTEREST

The authors claim no conflict of interest.

REFERENCES

[1] E. A. Kensinger, "Remembering emotional experiences: The contribution of valence and arousal," *Rev. Neurosciences*, vol. 15, no. 4, pp. 241–252, Jan. 2004.

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

[3] E. Diener, S. Thapa, and L. Tay, "Positive emotions at work," *Annu. Rev. Organizational Psychol. Organizational Behav.*, vol. 7, no. 1, pp. 451–477, 2020.

[4] M. Irfan et al., "Smart IoT-based solutions for neonatal sleep stratification: Single-dual channel EEG, AdaptiSelect, multiview fusion, and rotational ensemble stacking," *IEEE Internet Things J.*, vol. 12, no. 22, pp. 46018–46037, Nov. 2025.

[5] M. Irfan et al., "A novel NICU sleep state stratification: Multiperspective features, adaptive feature selection and ensemble model," *IEEE Trans. Biomed. Eng.*, vol. 72, no. 9, pp. 2684–2697, Sep. 2025.

[6] M. Irfan et al., "Multidomain selective feature fusion and stacking based ensemble framework for EEG-based neonatal sleep stratification," *IEEE J. Biomed. Health Informat.*, vol. 30, no. 4, pp. 2828–2838, Apr. 2026.

[7] G. Bossavi, R. Yan, and M. Irfan, "A novel convolutional neural network deep learning implementation for cuffless heart rate and blood pressure estimation," *Appl. Sci.*, vol. 13, no. 22, p. 12403, Nov. 2023.

[8] M. Irfan, A. Subasi, N. Mustafa, T. Westerlund, and W. Chen, "An evaluation of pretrained convolutional neural networks for stroke classification from brain ct images," in *Applications of Artificial Intelligence in Healthcare and Biomedicine*. Amsterdam, The Netherlands: Elsevier, 2024, pp. 111–135.

[9] A. Nawaz, M. Irfan, and T. Westerlund, "Optical character recognition using optimized convolutional networks," in *Proc. 8th Int. Conf. Fog Mobile Edge Comput. (FMEC)*, Sep. 2023, pp. 107–114.

[10] Z. Ahmad and N. Khan, "A survey on physiological signal-based emotion recognition," *Bioengineering*, vol. 9, no. 11, p. 688, Nov. 2022.

[11] K. Kamble and J. Sengupta, "A comprehensive survey on emotion recognition based on electroencephalograph (EEG) signals," *Multimedia Tools Appl.*, vol. 82, no. 18, pp. 27269–27304, Jul. 2023.

[12] 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 Interface*, vol. 3, nos. 1–2, pp. 33–48, Mar. 2010.

[13] C. Park, J. Ryu, J. Sohn, and H. Cho, "An emotion expression system for the emotional robot," in *Proc. IEEE Int. Symp. Consum. Electron.*, Jun. 2007, pp. 1–6.

[14] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Med. Biol. Eng. Comput.*, vol. 42, no. 3, pp. 419–427, May 2004.

[15] ALI Ghali and Mohamad-Bassam Kurdy, "Emotion recognition using facial expression analysis," *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 18, pp. 6117–6129, 2018.

[16] International Data Corporation. (2025). *Worldwide Quarterly Wearable Device Tracker*. Accessed: Sep.9, 2025. [Online]. Available: https://www.idc.com/promo/wearablevendor/

[17] AARP. (2025). *Aarp Research Insights on Technology*. Accessed: Sep.9, 2025. [Online]. Available: https://www.aarp.org/pri/topics/technology/internet-media-devices/aarp-research-insights-technology/

[18] A. Kulkarni and J. R. Prasad, "Design considerations for adaptive sensor-based systems for pediatric behavioral monitoring," in *Proc. IEEE 5th Int. Conf. ICT Bus. Ind. Government (ICTBIG)*, Dec. 2025, pp. 1–9.

[19] X. Yang et al., "Emotion recognition based on multimodal physiological signals using spiking feed-forward neural networks," *Biomed. Signal Process. Control*, vol. 91, May 2024, Art. no. 105921.

[20] B. Wong, P. Wu, Z. Ismail, J. Watt, and Z. Goodarzi, "Detecting agitation and aggression in persons living with dementia: A systematic review of diagnostic accuracy," *BMC Geriatrics*, vol. 24, no. 1, p. 559, Jun. 2024.

[21] E. Gkintoni, A. Aroutzidis, H. Antonopoulou, and C. Halkiopoulos, "From neural networks to emotional networks: A systematic review of EEG-based emotion recognition in cognitive neuroscience and real-world applications," *Brain Sci.*, vol. 15, no. 3, p. 220, Feb. 2025.

[22] S. Zhang, Y. Yang, C. Chen, X. Zhang, Q. Leng, and X. Zhao, "Deep learning-based multimodal emotion recognition from audio, visual, and text modalities: A systematic review of recent advancements and future prospects," *Expert Syst. Appl.*, vol. 237, Mar. 2024, Art. no. 121692.

[23] H. Liu et al., "EEG-based multimodal emotion recognition: A machine learning perspective," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–29, 2024.

[24] R. Guo, H. Guo, L. Wang, M. Chen, D. Yang, and B. Li, "Development and application of emotion recognition technology—A systematic literature review," *BMC Psychol.*, vol. 12, no. 1, p. 95, Feb. 2024.

[25] T. Gruden et al., "Electrogastrography in autonomous vehicles—An objective method for assessment of motion sickness in simulated driving environments," *Sensors*, vol. 21, no. 2, p. 550, Jan. 2021.

[26] M. Pantic, G. Caridakis, E. André, J. Kim, K. Karpouzis, and S. Kollias, "Multimodal emotion recognition from low-level cues," in *Emotion-Oriented Systems: The Humaine Handbook*. Cham, Switzerland: Springer, 2010, pp. 115–132.

[27] S. K. D'mello and J. Kory, "A review and meta-analysis of multimodal affect detection systems," *ACM Comput. Surv.*, vol. 47, no. 3, pp. 1–36, Apr. 2015.

[28] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognit.*, vol. 44, no. 3, pp. 572–587, Mar. 2011.

[29] D. Canedo and A. J. R. Neves, "Facial expression recognition using computer vision: A systematic review," *Appl. Sci.*, vol. 9, no. 21, p. 4678, Nov. 2019.

[30] G. Cosoli, A. Poli, L. Scalise, and S. Spinsante, "Heart rate variability analysis with wearable devices: Influence of artifact correction method on classification accuracy for emotion recognition," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2021, pp. 1–6.

[31] Y. Dai, X. Wang, X. Li, and P. Zhang, "Reputation-driven multimodal emotion recognition in wearable biosensor network," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2015, pp. 1747–1752.

[32] G. Du, Q. Tan, C. Li, X. Wang, S. Teng, and P. X. Liu, "A noncontact emotion recognition method based on complexion and heart rate," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.

[33] Y. Li et al., "A tale of single-channel electroencephalography: Devices, datasets, signal processing, applications, and future directions," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–20, 2025.

[34] Y.-K. Li, Q.-H. Meng, T.-H. Yang, Y.-X. Wang, and H.-R. Hou, "Touch gesture and emotion recognition using decomposed spatiotemporal convolutions," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–9, 2022.

[35] A. Apicella, P. Arpaia, F. Isgro, G. Mastrati, and N. Moccaldi, "A survey on EEG-based solutions for emotion recognition with a low number of channels," *IEEE Access*, vol. 10, pp. 117411–117428, 2022.

[36] K. Alghoul, H. Al Osman, and A. El Saddik, "Enhancing generalization in PPG-based emotion measurement with a CNN-TCN-LSTM model," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2025, pp. 1–6.

[37] M. Irfan et al., "Emotion recognition with minimal wearable sensing: Multi-domain feature, hybrid feature selection, and personalized vs. generalized ensemble model analysis," 2025, *arXiv:2510.22498*.

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

[39] S. Koelstra et al., "DEAP: A database for emotion analysis;Using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012.

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

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

[42] L. D. Sharma and A. Bhattacharyya, "A computerized approach for automatic human emotion recognition using sliding mode singular spectrum analysis," *IEEE Sensors J.*, vol. 21, no. 23, pp. 26931–26940, Dec. 2021.

[43] J. Chen, Y. Hu, L. Garg, T. R. Gadekallu, G. Srivastava, and W. Wang, "Graph-enhanced low-resource ECG representation learning for emotion recognition based on wearable Internet of Things," *IEEE Internet Things J.*, vol. 11, no. 24, pp. 39056–39068, Dec. 2024.

[44] A. Pradhan and S. Srivastava, "Hybrid densenet with long short-term memory model for multi-modal emotion recognition from physiological signals," *Multimedia Tools Appl.*, vol. 83, no. 12, pp. 35221–35251, Sep. 2023.

[45] P. Saha et al., "Novel multimodal emotion detection method using electroencephalogram and electrocardiogram signals," *Biomed. Signal Process. Control*, vol. 92, Jun. 2024, Art. no. 106002.

[46] P. Theerthagiri, "Stress emotion recognition with discrepancy reduction using transfer learning," *Multimedia Tools Appl.*, vol. 82, no. 4, pp. 5949–5963, Feb. 2023.

[47] X. Wang, J. Zhang, C. He, H. Wu, and L. Cheng, "A novel emotion recognition method based on the feature fusion of single-lead EEG and ECG signals," *IEEE Internet Things J.*, vol. 11, no. 5, pp. 8746–8756, Mar. 2024.

[48] A. Fang, F. Pan, W. Yu, L. Yang, and P. He, "ECG-based emotion recognition using random convolutional kernel method," *Biomed. Signal Process. Control*, vol. 91, May 2024, Art. no. 105907.

[49] N. Gahlan and D. Sethia, "Federated learning inspired privacy sensitive emotion recognition based on multi-modal physiological sensors," *Cluster Comput.*, vol. 27, no. 3, pp. 3179–3201, Jun. 2024.

[50] F. Panahi, S. Rashidi, and A. Sheikhani, "Application of fractional Fourier transform in feature extraction from ELECTROCARDIOGRAM and GALVANIC SKIN RESPONSE for emotion recognition," *Biomed. Signal Process. Control*, vol. 69, Aug. 2021, Art. no. 102863.

[51] P. S. Kumar, P. K. Govarthan, A. A. S. Gadda, N. Ganapathy, and J. F. A. Ronickom, "Deep learning-based automated emotion recognition using multimodal physiological signals and time-frequency methods," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–12, 2024.

[52] Md. G. R. Alam, S. F. Abedin, S. I. Moon, A. Talukder, and C. S. Hong, "Healthcare IoT-based affective state mining using a deep convolutional neural network," *IEEE Access*, vol. 7, pp. 75189–75202, 2019.

[53] Y. R. Veeranki, H. F. Posada-Quintero, and R. Swaminathan, "Transition network-based analysis of electrodermal activity signals for emotion recognition," *IRBM*, vol. 45, no. 4, Aug. 2024, Art. no. 100849.

[54] S. Upadhaya, B. Brahma, H. K. S, R. Panigrahi, and A. K. Bhoi, "Decoding emotions and unveiling stress: A non-invasive approach through sequential feature extraction and multiclass classifiers," *Health Technol.*, vol. 14, no. 6, pp. 1149–1160, Nov. 2024.

[55] S. N. M. S. Ismail, N. A. A. Aziz, and S. Z. Ibrahim, "A comparison of emotion recognition system using electrocardiogram (ECG) and photoplethysmogram (PPG)," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3539–3558, Jun. 2022.

[56] A. Raheel, M. Majid, and S. M. Anwar, "DEAR-MULSEMEDIA: Dataset for emotion analysis and recognition in response to multiple sensorial media," *Inf. Fusion*, vol. 65, pp. 37–49, Jan. 2021.

[57] J. A. Domínguez-Jiménez, K. C. Campo-Landines, J. C. Martínez-Santos, E. J. Delahoz, and S. H. Contreras-Ortiz, "A machine learning model for emotion recognition from physiological signals," *Biomed. Signal Process. Control*, vol. 55, Jan. 2020, Art. no. 101646.

[58] L. Fiorini, G. Mancioppi, F. Semeraro, H. Fujita, and F. Cavallo, "Unsupervised emotional state classification through physiological parameters for social robotics applications," *Knowl.-Based Syst.*, vol. 190, Feb. 2020, Art. no. 105217.

[59] R. Francese, M. Risi, and G. Tortora, "A user-centered approach for detecting emotions with low-cost sensors," *Multimedia Tools Appl.*, vol. 79, nos. 47–48, pp. 35885–35907, Dec. 2020.

[60] T. Fan et al., "A new deep convolutional neural network incorporating attentional mechanisms for ECG emotion recognition," *Comput. Biol. Med.*, vol. 159, Jun. 2023, Art. no. 106938.

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

[62] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-González, E. Abdulhay, and N. Arunkumar, "Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS)," *IEEE Access*, vol. 7, pp. 57–67, 2019.

[63] F. Patlar Akbulut, H. G. Perros, and M. Shahzad, "Bimodal affect recognition based on autoregressive hidden Markov models from physiological signals," *Comput. Methods Programs Biomed.*, vol. 195, Oct. 2020, Art. no. 105571.

[64] A. Albraikan, D. P. Tobón, and A. El Saddik, "Toward user-independent emotion recognition using physiological signals," *IEEE Sensors J.*, vol. 19, no. 19, pp. 8402–8412, Oct. 2019.

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

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

[67] B. Mokhtari, A. Merizig, and H. Zerdoumi, "Enhanced emotion recognition in an IoMT platform: Leveraging data augmentation and the random forest algorithm for ECG-based E-health," *Int. J. Inf. Technol.*, vol. 18, no. 1, pp. 599–609, Jan. 2026.

[68] W. Sun, Y. Liu, S. Li, J. Tian, F. Wang, and D. Liu, "Research on driver's anger recognition method based on multimodal data fusion," *Traffic Injury Prevention*, vol. 25, no. 3, pp. 354–363, Apr. 2024.

[69] A. Garg, V. Chaturvedi, A. B. Kaur, V. Varshney, and A. Parashar, "Machine learning model for mapping of music mood and human emotion based on physiological signals," *Multimedia Tools Appl.*, vol. 81, no. 4, pp. 5137–5177, Feb. 2022.

[70] K. Ali and C. E. Hughes, "A unified biosensor–vision multi-modal transformer network for emotion recognition," *Biomed. Signal Process. Control*, vol. 102, Apr. 2025, Art. no. 107232.

[71] F. Patlar Akbulut, "Hybrid deep convolutional model-based emotion recognition using multiple physiological signals," *Comput. Methods Biomechanics Biomed. Eng.*, vol. 25, no. 15, pp. 1678–1690, Nov. 2022.

[72] K. N. Lal and Lekhraj, "Automatic multimedia classification based on mood recognition of drivers in Internet-of-Vehicle using fog computing," *Wireless Netw.*, vol. 31, no. 3, pp. 2027–2038, Mar. 2025.

[73] P. Uluer, H. Kose, E. Gumuslu, and D. E. Barkana, "Experience with an affective robot assistant for children with hearing disabilities," *Int. J. Social Robot.*, vol. 15, no. 4, pp. 643–660, Apr. 2023.

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

[75] N. Gahlan and D. Sethia, "AFLEMP: Attention-based federated learning for emotion recognition using multi-modal physiological data," *Biomed. Signal Process. Control*, vol. 94, Aug. 2024, Art. no. 106353.

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

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

[78] B. B. Felix, B. Vincent, M. Irfan, and D. Amine, "An embedded photodiode-based neural network sensor for solar irradiance estimation with BLE beacon," *TechRxiv*, pp. 1–12, Feb. 2026, doi: 10.36227/techrxiv.177162453.33511645/v1.

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

[80] R. Harper and J. Southern, "A Bayesian deep learning framework for end-to-end prediction of emotion from heartbeat," *IEEE Trans. Affect. Comput.*, vol. 13, no. 2, pp. 985–991, Apr. 2022.

[81] J. Li, N. Chen, H. Zhu, G. Li, Z. Xu, and D. Chen, "Incongruity-aware multimodal physiology signals fusion for emotion recognition," *Inf. Fusion*, vol. 105, May 2024, Art. no. 102220.

[82] L. Pepa, L. Spalazzi, M. G. Ceravolo, and M. Capecci, "Supervised learning for automatic emotion recognition in Parkinson's disease through smartwatch signals," *Expert Syst. Appl.*, vol. 249, Sep. 2024, Art. no. 123474.

[83] T. C. Sweeney-Fanelli and M. H. Imtiaz, "ECG-based automated emotion recognition using temporal convolution neural networks," *IEEE Sensors J.*, vol. 24, no. 18, pp. 29039–29046, Sep. 2024.

[84] S. Woo, M. Zubair, S. Lim, and D. Kim, "Deep multimodal emotion recognition using modality-aware attention and proxy-based multimodal loss," *Internet Things*, vol. 31, May 2025, Art. no. 101562.

[85] S. Kumar P and J. Fredo Agastinose Ronickom, "Emotion classification through optimal segments of EDA and texture analysis of time-encoded images with artificial intelligence," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–15, 2025.

[86] V. Dissanayake, S. Seneviratne, R. Rana, E. Wen, T. Kaluarachchi, and S. Nanayakkara, "SigRep: Toward robust wearable emotion recognition with contrastive representation learning," *IEEE Access*, vol. 10, pp. 18105–18120, 2022.

[87] E. H. Houssein, S. Mohsen, M. M. Emam, N. A. Samee, R. I. Alkanhel, and E. M. G. Younis, "Leveraging explainable artificial intelligence for emotional label prediction through health sensor monitoring," *Cluster Comput.*, vol. 28, no. 2, p. 86, Apr. 2025.

[88] D. Kang et al., "Beyond superficial emotion recognition: Modality-adaptive emotion recognition system," *Expert Syst. Appl.*, vol. 235, Jan. 2024, Art. no. 121097.

[89] J. Li and P. Washington, "A comparison of personalized and generalized approaches to emotion recognition using consumer wearable devices: Machine learning study," *JMIR AI*, vol. 3, May 2024, Art. no. e52171.

[90] O. Dehzangi, V. Sahu, V. Rajendra, and M. Taherisadr, "GSR-based distracted driving identification using discrete & continuous decomposition and wavelet packet transform," *Smart Health*, vol. 14, Dec. 2019, Art. no. 100085.

[91] Z. Qian, D. Lv, Y. Lv, and Z. Bi, "Modeling and quantification of impact of psychological factors on rehabilitation of stroke patients," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 683–692, Mar. 2019.

[92] Y. Wang, B. Chen, H. Liu, and Z. Hu, "Understanding flow experience in video learning by multimodal data," *Int. J. Hum.–Comput. Interact.*, vol. 40, no. 12, pp. 3144–3158, Jun. 2024.

[93] F. Yan and J. Liu, "Analog front-end input-impedance boosting techniques for bio-potential monitoring—A review," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–14, 2024.

[94] M. F. Wagdy and S. S. Awad, "Determining ADC effective number of bits via histogram testing," *IEEE Trans. Instrum. Meas.*, vol. 40, no. 4, pp. 770–772, Aug. 1991.

[95] F. Zhang et al., "Wearable fiber Bragg grating sensors for physiological signal and body motion monitoring: A review," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–24, 2025.

[96] C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Vallance, and L. Tarassenko, "Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 3, pp. 832–838, May 2015.

[97] C. Wang et al., "Motion artifact suppression method based on dynamic background field adaptive filtering for wearable OPM-MEG," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–16, 2025.

[98] Z. Sun, A. Mu, Y. Qian, Z. Lv, and Y. Wang, "Robust heart rate monitoring from wrist pulse signals corrupted by motion artifacts using dual microphones and adaptive filtering," *IEEE Sensors J.*, vol. 25, no. 15, pp. 29336–29346, Aug. 2025.

[99] R. S. Gamad and D. K. Mishra, "Computation of error in estimation of nonlinearity in ADC using histogram technique," *Engineering*, vol. 3, no. 6, pp. 583–587, 2011.

[100] E. K. Naeini, I. Azimi, A. M. Rahmani, P. Liljeberg, and N. Dutt, "A real-time PPG quality assessment approach for healthcare Internet-of-Things," *Proc. Comput. Sci.*, vol. 151, pp. 551–558, Jan. 2019.

[101] L. Tabbaa et al., "VREED: Virtual reality emotion recognition dataset using eye tracking & physiological measures," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 4, pp. 1–20, Dec. 2021.

[102] Z. Gao, X. Cui, W. Wan, W. Zheng, and Z. Gu, "ECSMP: A dataset on emotion, cognition, sleep, and multi-model physiological signals," *Data Brief*, vol. 39, Dec. 2021, Art. no. 107660.

[103] M. Behnke, M. Buchwald, A. Bykowski, S. Kupiński, and L. D. Kaczmarek, "Psychophysiology of positive and negative emotions, dataset of 1157 cases and 8 biosignals," *Sci. Data*, vol. 9, no. 1, p. 10, Jan. 2022.