IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 September 2025, accepted 3 November 2025, date of publication 7 November 2025, date of current version 17 November 2025.

*Digital Object Identifier 10.1109/ACCESS.2025.3630563*

SURVEY

# Affective Computing Emotional Body Gesture Recognition: Evolution and the Cream of the Crop

**PEDRO J. VAZ[1,2], JOÃO M. F. RODRIGUES[2], AND PEDRO J. S. CARDOSO[2]**
[1]Faculty of Science and Technology, Universidade do Algarve, 8005-139 Faro, Portugal
[2]NOVA LINCS and ISE, Universidade do Algarve, 8005-139 Faro, Portugal

Corresponding author: Pedro J. Vaz (pjmartins@ualg.pt)

This work was supported by NOVA Laboratory for Computer Science and Informatics (NOVA LINCS) under grant UID/04516/NOVA with the financial support of FCT.IP, and by project AI.EVENT: Monitor Live Audience with AI (ALGARVE-FEDER-01180500, Ref. 17325) co-financed by ALGARVE 2030, Portugal 2030, and the European Union. The work of Pedro J. Vaz was supported by the Portuguese Foundation for Science and Technology (FCT) under PhD grant 2024.04816.BD.

**ABSTRACT** The field of affective computing (AffC) has experienced significant growth, making it challenging to stay up to date with the latest advancements. This surge in interest has likely contributed to a significant rise in the number of systematic reviews or surveys (SRoS) being published across various journals, covering topics like databases, methods, and general perspectives. This paper provides three key contributions: 1) A comprehensive analysis of the evolution of emotion recognition methods from 2002 to 2024, with particular emphasis on emotional body gesture recognition, documenting a clear transition from traditional machine learning to sophisticated deep learning architectures; 2) Identification and detailed analysis of the most impactful papers (the "cream of the crop") that have shaped body-based AffC methods, revealing that modern approaches increasingly use attention mechanisms, graph-based representations for skeletal data, and advanced spatial-temporal modeling techniques; and 3) A systematic categorization and analysis of emotion recognition methods across architectural types (machine learning, deep learning, and hybrid) and modalities (emotional body gesture recognition, facial emotion recognition, multimodal emotion recognition, and speech emotion recognition), demonstrating the field's progression from unimodal to more robust multimodal approaches. Through an analysis of 10 selected SRoS papers published between 2021-2024, referencing 292 papers collectively, this study reveals critical challenges including limited availability of large-scale body-based emotional databases, computational demands of modern architectures, and cross-database generalization issues.

**INDEX TERMS** Affective computing, body-based emotion recognition.

## I. INTRODUCTION

Emotional body gesture recognition (EBGR) has emerged as a fundamental research domain at the intersection of affective computing (AffC) and human-computer interaction, focusing on the automatic detection and analysis/classification of emotions through body posture, movement, and gesture. As outlined in recent surveys and systematic reviews (e.g., [1], [2], [3]), EBGR systems typically follow a structured pipeline that includes human detection, pose estimation, feature extraction, and emotion classification. These systems often employ machine learning (ML) techniques, ranging from traditional ML algorithms to deep learning (DL) models, to analyze extracted features such as joint positions, angles between body parts, and temporal changes in body pose. EBGR applications span diverse domains, from human-robot interaction and healthcare to entertainment and security systems, offering transformative potential for emotionally intelligent human-computer interfaces (e.g., see [2], [4], [5]). However, significant challenges remain,

The associate editor coordinating the review of this manuscript and approving it for publication was Longzhi Yang.

notably the limited availability of large-scale annotated databases and the complex and contextual nature of emotional bodily expression.

The modalities used for EBGR often include visual data from RGB cameras and depth sensors, as well as physiological signals such as electroencephalography (EEG) and Electrocardiography (ECG) [6], [7]. The use of uni- or multimodal approaches has significant implications for accuracy, robustness, and applicability. Unimodal systems, which rely on a single source of data such as body posture, facial expression, or physiological signals, offer simplicity in implementation and lower computational requirements [8], [9]. However, these systems may struggle to capture the full complexity of human emotions, especially in nuanced or ambiguous situations. In contrast, multimodal emotion recognition systems integrate data from multiple sources, providing a more comprehensive and accurate representation of emotional states [7].

Multimodal approaches can compensate for the limitations of individual modalities, making them more resilient to noise and variability in real-world settings. For instance, when certain modalities become unavailable (i.e., occluded facial expressions or absence of voice data), the redundancy inherent in multimodal systems maintains emotion detection capabilities. Additionally, multimodal systems have the potential to capture subtle emotional cues that might be missed by single-modal approaches, leading to more nuanced and context-aware emotion recognition methods [9], [10]. However, a possible drawback of multimodal systems is the increased complexity often required, including increased algorithmic complexity, larger training data requirements, and higher computational demands, which can be challenging for real-time applications or resource-constrained environments [11], [12], [13].

While considerations of modal approaches are essential, a fundamental challenge in AffC lies in clearly defining what is being analyzed. In this regard, emotion and sentiment analysis have emerged as fundamental research areas within AffC [12], [14] and, while inherently related, emotions and sentiments represent distinct psychological phenomena [15]. Emotions are characterized as intense, relatively brief affective states triggered by specific events or stimuli [16], often manifesting through complex physiological responses and distinct facial expressions [17], [18]. For instance, receiving unexpected good news might trigger immediate joy, accompanied by increased heart rate, smiling, or even tears of happiness [19]. In contrast, sentiments typically represent long-term affective states [15], generally reflecting established opinions or attitudes toward particular subjects [20]. A consumer's long-term satisfaction with a product [21], or evolving opinion about a concert or album [22] exemplify such sentiment states. This distinction is particularly evident in computational analysis, where sentiment is commonly categorized as positive, negative, or neutral, while emotion recognition aims to identify and analyze emotional states [15], [23]. For example, a product review might express overall positive sentiment [21], while containing expressions of various emotions such as initial frustration during initial setup, followed by satisfaction with the final result [23].

Theoretical models for representing emotions and sentiments can be broadly divided into discrete (categorical) and continuous (dimensional) approaches, each offering distinct advantages for affective computing applications [1], [24], [25]. Categorical models, exemplified by Ekman's six basic emotions (happiness, sadness, anger, fear, disgust, and surprise) and Plutchik's wheel of emotions (which, for its primary emotions, extends Ekman's model with trust and anticipation), define distinct emotion classes [26], [27]. These discrete models have gained widespread adoption in emotion recognition systems due to their interpretability and straightforward implementation. However, they may oversimplify the complexity of human emotional experiences [28], [29].

Dimensional models offer an alternative perspective by representing emotions in continuous space. Russell's circumplex model characterizes emotions along two primary dimensions: valence (positive to negative) and arousal (low to high activation) [30]. The Pleasure-Arousal-Dominance (PAD) model extends this framework by incorporating a third dimension of dominance, enabling the representation of more subtle emotional variations [31]. When using continuous models, each emotional state can be mapped along valence (the pleasantness of the experience, from negative to positive), arousal (the intensity of activation, from calm to excited), and dominance (the degree of perceived control, from submissive to dominant) [32]. For instance, joy combines high valence, moderate to high arousal, and high dominance, while fear typically involves low valence, high arousal, and low dominance [31], [33]. Similarly, relaxation can be characterized by high valence, low arousal, and moderate dominance, demonstrating how these dimensions capture the nuanced nature of emotional experiences [34].

Additionally, beyond the general binary (positive, negative), or ternary (positive, negative, neutral) classification of sentiments [35], [36], different categorical and continuous models of sentiment intensity representation exist. For example, ordinal scales, such as 5-star ratings common in review systems, or 5-point (-2 to +2) and 7-point (-3 to +3) Likert scales, enable more nuanced sentiment intensity distinctions [37], [38]. Continuous approaches have also been explored, particularly in dimensional models where sentiment is represented as real-valued scores, often normalized between -1 and +1, or 0 to 1 [39], [40], [41]. The use of discrete or continuous scales often reflects a trade-off between annotation reliability and granularity, since continuous scales can capture subtle variations in sentiment intensity, however, they typically show lower inter-annotator agreement compared to discrete categorical schemes [23]. This variety in sentiment representations has led to the development of specialized evaluation metrics and annotation guidelines for different granularity levels [42].

The rapid growth of the AffC field has led to the proliferation of methods and models, as well as a high number of systematic review or survey (SRoS) papers, making it challenging for researchers to navigate the vast landscape of available approaches. In this context, this study examines 10 selected SRoS papers published between 2021 and September 2024 that focus on AffC emotion recognition methods (ERMs). These surveys collectively reference 292 papers, covering not only EBGR, but also facial emotion recognition (FER), speech emotion recognition (SER), and multimodal emotion recognition (MMER), as well as databases.

While this study emphasizes EBGR analysis, systematic data from the selected SRoS regarding FER, SER, and MMER is also analyzed, though with lesser detail since these modalities, while complementary to EBGR, are not the main focus of this work. Since the ERMs presented in this study were identified through other SRoS, it was considered that they represent the cream of the crop in the field of AffC, especially in the EBGR modality, since they are the ones with the higher visibility. To be clear, the ERM analysis provided on this work is based on the systematic information collected on other SRoS authors, possibly disregarding other papers that might be relevant but were not included in the SRoS analyzed, as this was the adopted methodology. An empirical assessment revealed heterogeneity in methodological rigor across included SRoS papers, particularly regarding protocol registration, eligibility criteria, and quality appraisal.

The main contributions of this paper are: i) the evolution of emotion recognition methods over time, with special emphasis on EBGR, and how they are documented in the selected literature; ii) identification and analysis of the most significant papers (the cream of the crop) that have shaped the development of body-based AffC methods; and iii) a comprehensive analysis including a) classification by architecture (ML, DL, other) of the ERMs presented in the SRoS papers, b) classification by modality (EBGR, FER, MMER, SER) of the ERMs presented in the SRoS papers, c) detailed analysis and discussion of the most prevalent body-based AffC methods and models, and d) discussion of the trends, potential and future direction of the EBGR field, within the general scope of AffC. The need for this study lies in its ability to consolidate fragmented insights from recent surveys, offering a structured and comparative analysis that is crucial for guiding future research in emotional body gesture recognition and multimodal affective computing.

The remainder of this paper is organized as follows. Section II presents the review methodology and related works, Section III presents a summary of the emotion recognition methods present in the selected SRoS, and Section IV presents the discussion. Finally, Section V presents the conclusions and future work.

## II. REVIEW METHODOLOGY AND RELATED WORKS

To the best of the authors' knowledge, this is the first time that an in-depth analysis of SRoS papers on AffC ERMs is presented. Although the authors previous work [43] followed a similar methodology to the one presented in this paper, the focus there was on analyzing AffC databases, covering a general scope across modalities, namely: visual, speech, gesture, body, and physiological. The present study focuses on the evolution of emotion recognition methods, with particular emphasis on body-based emotion recognition, and how they are documented in the selected literature. Regarding the SRoS analyzed in this work, all of them present a general overview of the field but do not go into detail when comparing the information they provide against each other. The remainder of this section presents the selection criteria and a brief analysis of each of the selected 10 SRoS papers.

### A. SELECTION CRITERIA

The papers selected for analysis in this study met the following criteria: i) they must be a systematic review or survey paper, ii) they must have been published in a journal, iii) they must be written in English, iv) they must be accessible online, v) they must provide systematic information on affective computing methods, and vi) they must have been published between 2021 and September 2024.

The search was conducted across major academic databases, including IEEE Xplore, ACM Digital Library, ScienceDirect, Scopus, Google Scholar, and ResearchGate. The following search keywords were used: "affective computing", "emotion recognition", "body emotion recognition", "sentiment analysis", "affective computing methods", and "affective computing surveys". The systematic search, as defined above, was conducted in September 2024, resulting in the identification of 31 SRoS papers on AffC. After applying the selection criteria, 10 SRoS papers were chosen for further analysis [1], [3], [4], [10], [13], [24], [44], [45], [46], [47].

### B. SELECTED REVIEW PAPERS SUMMARY

As already mentioned, the 10 selected SRoS papers ( [1], [3], [4], [10], [13], [24], [44], [45], [46], [47]) were published between 2021 and September 2024, providing an overview of the state-of-the-art in emotion recognition methods.

From a statistical point of view, Fig. 1 illustrates the annual distribution of the 292 ERM papers that were referenced across the selected SRoS. A first observation is that the SRoS papers reference ERMs with publication dates spanning from 2002 to 2024. Naturally, given the SRoS publication dates, a larger number of references are observed in the most recent years, reflecting the growing interest in emotion recognition research and the increasing adoption of artificial intelligence (AI) techniques in the field.

In general, the SRoS authors provide extensive reviews that typically categorize approaches based on the type of method (e.g., ML-based, DL-based) and the modalities used, documenting their application across different emotional databases. They also discuss implementation approaches,

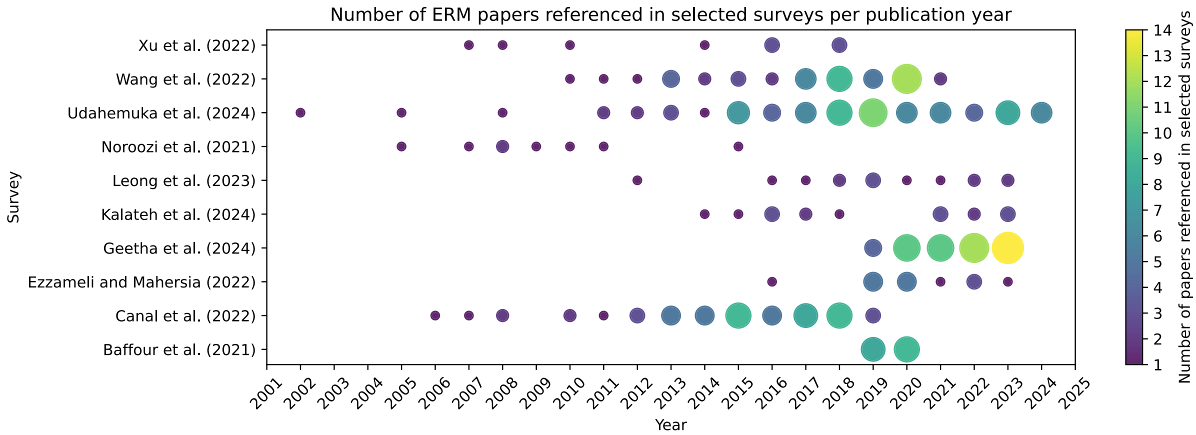


**FIGURE 1. Number of ERM papers referenced across the selected SRoS papers per year.**

architectural decisions, and recent technological advances in the field. A brief analysis of each of the 10 selected SRoS papers is now presented.

Noroozi et al. [1] present a comprehensive survey examining the state-of-the-art in EBGR, demonstrating that EBGR is a relatively unexplored domain within AffC in comparison with FER. The authors present a detailed analysis of body language as a form of non-verbal communication, discussing its cultural and gender dependencies, as well as its importance in transmitting emotions and thoughts during interactions. The survey covers key technical aspects including human detection, static and dynamic body pose estimation in both RGB and 3D modalities, and feature extraction methods for emotion recognition. While pre-processing technologies like human detection and pose estimation have reached maturity for robust large-scale analysis, the authors identify significant challenges in the field, notably the scarcity of labeled emotional gesture data and inconsistency in emotion taxonomies across different databases.

The survey reveals that most current approaches rely on basic geometrical features (such as joint coordinates, distances, and orientations) and simple motion cues, with limited exploration of DL approaches compared to other domains of AffC. A key insight from the authors is that unlike FER, which has achieved consensus around basic emotion models and action units, emotional body gesture analysis lacks such standardization in how emotions are labeled and categorized. The authors also provide a comprehensive review of publicly available databases for training automatic recognition systems and discuss various applications of EBGR technology. The work concludes by highlighting the need for larger, high-quality databases and consistent emotion taxonomies to advance the field toward more sophisticated representations of affective body language.

Baffour et al. [44] provide a systematic survey of DL algorithms for FER, focusing on the current state-of-the-art architectures and their applications in AffC. The authors highlight the importance of FER in human-computer communication, noting that facial expressions transmit 55% of emotional and mental states in face-to-face interactions. The survey examines key technical aspects including DL architectures for FER, with particular focus on CNNs, which have shown dominance over other architectures like RNNs and SVMs. The survey reviews major public databases available for FER training and development, preprocessing techniques, and various model architectures, comparing their contributions, performance, and limitations.

A key finding is that while CNNs have achieved high accuracy rates, challenges remain around computation power requirements and the availability of large facial emotion databases. The authors highlight an emerging trend toward multimodal approaches that combine facial expressions with other modalities like speech to improve overall emotion recognition accuracy. The work concludes by identifying open research challenges and future directions, particularly emphasizing the need for larger databases and more sophisticated approaches to handle continuous, spontaneous facial expressions in real-world scenarios.

Canal et al. [45] presents a survey on FER techniques, analyzing traditional and DL approaches in AffC. The authors conduct a systematic review of 51 papers published between 2006 and 2019, containing 94 distinct methods, categorizing them into classical approaches (based on traditional computer vision and pattern recognition) and neural network-based (NNB) methods. The authors find that while classical approaches achieve marginally better recognition precision in controlled settings, NNB methods offer greater generalization capability when dealing with larger and more varied databases.

The study also examines the complete FER pipeline, including preprocessing techniques, feature extraction methods, and classification algorithms, with SVMs emerging as the predominant classical classifier and CNNs dominating the NNB category. The authors also evaluate popular FER databases, noting limitations in existing collections regarding resolution, size, and demographic diversity. While highlighting significant progress in the field, the survey indicates that FER remains an unsolved challenge, particularly in unconstrained real-world environments, and suggests that future research should focus on developing more comprehensive databases and use the generalization capabilities of NNB approaches.

Wang et al. [24] survey provides a broad review of AffC, examining emotion models, databases, and recent methodological advances in the field. The authors examine both unimodal affect recognition approaches across textual, audio, visual, and physiological modalities, as well as multimodal affective analysis that combines multiple data types, and categorizing methods into traditional ML-based and modern DL-based approaches. For the visual modalities, the authors include both FER and EBGR. While FER remains the most studied approach, the survey emphasizes how body gestures can provide emotional cues that are often more reliably perceived than subtle facial changes, with EBGR research analyzing both full-body movements and upper-body gestures, including hand movements, head positioning, and general posture. For EBGR specifically, traditional approaches rely on hand-crafted features like statistical analysis of body movement and gesture expressivity metrics, while modern DL methods, particularly CNN-LSTM (long short-term memory) architectures, can automatically learn spatio-temporal features from body motion sequences.

Through analysis of over 380 research papers, the work evaluates architectural designs and performance characteristics, examining different fusion strategies for multimodal systems including feature-level, decision-level, and hybrid approaches. Key challenges identified include the limited availability of large-scale naturalistic body gesture databases, as most existing databases contain acted expressions captured in controlled laboratory settings. The authors highlight the need for more research on spontaneous (non-acted) body expressions and their correlation with genuine emotional states, alongside the development of more comprehensive benchmark databases and better integration of multiple modalities.

Xu et al. [46] survey examines the emerging field of emotion recognition through gait analysis, emphasizing its potential as a non-invasive and hard-to-imitate method compared to traditional biometric approaches, like facial expression or speech analysis. The authors review how different emotional states manifest in distinctive gait patterns and walking characteristics, presenting evidence that gait can serve as a viable biometric modality for emotion detection. The survey details various approaches to gait-based emotion recognition, including data collection methodologies, preprocessing techniques, and classification methods. Of particular interest are the body-based emotion recognition approaches discussed, such as the effort-shape method which evaluates qualitative movement factors including space, time, energy, and flow, along with body shape parameters.

The research demonstrates that specific emotions correlate with distinct movement patterns. For instance, sad gaits typically exhibit contracted torso shape and slower movements, while angry gaits show expanded limb shape and more forceful energy. The authors also explore how various emotions affect spatio-temporal gait parameters such as stride length, walking speed, and arm swing amplitude. The survey identifies several advantages of gait-based emotion recognition over traditional methods like facial or speech analysis, including the ability to perform remote observation and reduced susceptibility to conscious manipulation. Looking forward, the authors suggest future research directions including implementation of advanced DL techniques, creation of large-scale databases, and development of real-time emotion prediction systems.

Ezzameli and Mahersia [47] survey examines emotion recognition across various modalities, including facial expressions, body gestures, speech, text, and physiological signals, reviewing both traditional ML methods and modern DL architectures. Body gestures are presented as valuable for emotion detection since gross body motions are typically unconscious and therefore not subject to social editing, unlike some other modalities. The authors analyze how body movements, postures, and gestures contribute to emotional expression, while reviewing the evolution of methodologies in this domain from traditional approaches to DL methods. The survey discusses how body gesture-based emotion recognition generally involves four stages: detecting humans in the input data, estimating their poses, extracting relevant features, and recognizing the emotions conveyed by the gestures. The authors examine various DL architectures that can capture the characteristics of body movements for emotion recognition.

In multimodal systems, the authors demonstrate that combining different modalities can improve overall recognition accuracy. Body gesture databases like FABO, GEMEP, and HUMAINE, which provide valuable resources for developing and evaluating emotion recognition systems are also reviewed. The authors discuss the development of emotion recognition techniques in this domain, from traditional ML approaches using KNN, decision trees and SVM to more recent DL methods. Current methodological trends show increasing adoption of DL approaches and attention mechanisms for processing body movement data, with particular focus on capturing temporal relationships in gesture sequences. The work identifies future research opportunities, particularly in developing more robust body-based features and improving fusion with other modalities, while acknowledging the challenges of collecting and annotating body gesture data for emotion recognition.

Leong et al. [3] review examines recent advancements in FER and EBGR within AffC, analyzing 30 selected papers published between 2012 and 2022. The study investigates the relationship between visual features and emotion models, finding that facial expressions can identify more distinct emotions compared to body gestures. The research reveals that the seven basic emotions model is predominantly used for FER, while four-emotion classification is common for EBGR. Interestingly, the survey does not explicitly specify which four emotions comprise this model. However, analysis of the cited papers that use the four emotion model ([48], [49], [50], [51]) suggests that it consists of happy, sad, angry, and neutral emotions.

The review highlights that body gesture recognition typically involves a four-stage process: human detection, pose estimation, feature extraction, and emotion classification. For body gesture analysis, graph convolutional network (GCN) and temporal-based approaches have emerged as particularly effective, especially when dealing with skeletal data sequences. The analysis shows that despite the growing sophistication of FER, EBGR remains a less explored but promising field, with significant potential for improvement in detecting a broader range of emotions. The review identifies a notable research gap in body gesture studies, as only 8 of the 30 analyzed papers focused on EBGR, suggesting an important area for future research while acknowledging the complementary role of FER in comprehensive emotion recognition systems.

Geetha et al. [4] survey examines recent advancements in Multimodal Emotion Recognition using DL, providing a systematic analysis of the field's theoretical foundations, methodologies, and applications. The authors present an in-depth review of emotion theories, DL architectures, and fusion strategies used in MMER systems, categorizing approaches into simple, sequential, contextual, and graph-based methodologies. The survey extensively covers pre-processing techniques, feature extraction methods, and fusion mechanisms across different modalities including facial, audio, textual, and physiological data. It analyzes various benchmark databases and evaluation metrics while highlighting key challenges such as heterogeneous modality integration, limited annotated data availability, and real-time processing constraints. The review also addresses important considerations for responsible AI development in MMER systems, including fairness, privacy, and transparency. Furthermore, it explores diverse applications across healthcare, security, education, and human-computer interaction domains. The authors identify as a future direction the need for continued advancement in multimodal approaches to enhance the accuracy and robustness of emotion recognition systems.

Udahemuka et al. [10] survey examines MMER systems, using DL, that include body gesture and movement-based recognition alongside visual/facial, vocal, and physiological signals. The review acknowledges how EBGR, which analyzes postural changes, gestures, and full-body motion patterns, offers unique advantages in detecting emotional states from distance and multiple viewing angles. While the survey primarily focuses on facial and multimodal approaches, it also presents EBGR methods, including an implementation by Santhoshkumar et al. [51] that reported a 95.4% accuracy on the GEMEP database, using CNNs on video data.

The authors note that while facial expressions have traditionally dominated emotion recognition research, body movement analysis provides complementary information that is especially valuable in scenarios where facial data may be unclear or unavailable. The survey examines the integration of EBGR with other modalities, including the challenges in synchronizing body movement data with facial expressions, voice, and physiological signals. The authors conclude that while multimodal approaches show promise, several challenges remain, including the need for larger databases of naturalistic body expressions, better handling of occlusions and viewpoint variations in body movement analysis, and more robust fusion strategies for combining body-based features with other modalities, while maintaining robustness across diverse contexts and cultural variations.

Kalateh et al. [13] review analyzes MMER between 2014 and 2024, with body gestures and movements representing one of several modalities alongside facial expressions, physiological signals, and speech. Regarding EBGR specifically, the review discusses how emotions can be conveyed through body language, including posture, gestures, and movements. The review identifies that body movement analysis may require specialized equipment and expertise in motion capture and biomechanics, presenting both challenges and opportunities for emotion detection. While maintaining comprehensive coverage of other modalities, the review highlights how body gesture recognition systems face particular challenges in capturing movements and complex gestures in real-time applications. Further, the authors note integration challenges when combining body-based approaches with other modalities, acknowledging technical challenges in areas like feature extraction, fusion techniques, and real-world implementation of body gesture analysis systems.

Taking into account the previous summaries, the following partial conclusions can be made regarding the selected SRoS papers. The reviewed literature shows a clear evolution from traditional ML to DL architectures, with an increasing trend toward multimodal approaches. EBGR, while less explored than FER, is gaining attention for its ability to provide robust emotional cues. A key challenge remains the limited availability of large-scale naturalistic databases, particularly for body-based approaches, which impacts the development of new DL methodologies. In the next section, the ERMs presented in the selected SRoS papers are analyzed and discussed.

## III. EMOTION RECOGNITION METHODS

For the analysis of ERMs, the categorization into ML-based, DL-based, and hybrid (ML+DL) approaches was primarily extracted from the systematic data presented in the analyzed SRoS papers. Where categorization was not explicit, it was inferred from the architectural descriptions in the SRoS

**TABLE 1.** Summary of selected SRoS ERM papers per model architecture.

| Year | DL-based | ML+DL-based | ML-based |
|---|---|---|---|
| 2002 | [52] | | |
| 2005 | [53] | | [54] |
| 2006 | | | [55] |
| 2007 | | | [56]–[58] |
| 2008 | [59], [60] | | [7], [61]–[63] |
| 2009 | | | [64] |
| 2010 | | | [65]–[70] |
| 2011 | | | [71]–[74] |
| 2012 | | | [75]–[81] |
| 2013 | [82]–[84] | | [85]–[91] |
| 2014 | [92] | | [93]–[101] |
| 2015 | [102]–[105] | | [102], [106]–[122] |
| 2016 | [123]–[127] | [128], [129] | [50], [130]–[139] |
| 2017 | [140]–[151] | | [150], [152]–[160] |
| 2018 | [161]–[180] | [181], [182] | [166], [175], [183]–[192] |
| 2019 | [51], [193]–[219] | [220] | [219], [221]–[228] |
| 2020 | [5], [48], [229]–[257] | | [233], [253], [258]–[263] |
| 2021 | [49], [264]–[280] | [8], [281]–[283] | [284]–[286] |
| 2022 | [287]–[303] | | |
| 2023 | [304]–[325] | [326] | [319], [327]–[330] |
| 2024 | [331]–[336] | | |

or consulted the original referenced papers. Papers using traditional ML algorithms (e.g., SVM or Random Forests) were classified as *ML-based*, those using neural networks and DL architectures as *DL-based*, and approaches combining both as *hybrid*. For papers referenced across multiple SRoS, categorization consistency was verified, resolving any discrepancies through examination of the original papers. As a note, from the universe of 292 ERM papers, only 17 were referenced in more than one SRoS, with 15 referenced only in two different SRoS, and the remaining 2 referenced in three SRoS.

The analysis of emotion recognition methods across the selected SRoS reveals distinct trends in methodological evolution and modality distribution over time. As summarized in Table 1 and depicted in Fig. 2 (top left), there is a clear three-phase progression in methodological approaches: i) an ML-dominated phase from 2002 to 2014 with 84.3% of papers using traditional ML methods; ii) a transition phase from 2015 to 2017 marked by the emergence of hybrid approaches and increasing DL adoption, where DL methods grew from 15.7% (between 2002 and 2014) to 33.9% of publications; and iii) a DL-dominated phase from 2018 onwards where DL-based methods constitute 75.9% of all approaches. During this evolution, DL architectures progressed from basic CNNs to sophisticated temporal models incorporating attention mechanisms, with DL implementations increasing from 3 papers in 2013 to 31 papers in 2020, while traditional ML approaches maintained relevance in specific applications where computational efficiency is a key factor.

The distribution across modalities, see Fig. 2 (bottom left) and Table 2, shows distinct temporal patterns. FER dominated early research from 2002 to 2018, representing 64.1% of all papers. MMER approaches gained significant traction from 2019 onwards, growing from 6.3% of publications in 2018 to 27.0% by 2019. This evolution reflects both technological advancement and increasing recognition of the benefits of multimodal approaches. From the 292 ERM papers referenced in the selected SRoS, 50.3% are about FER, 28.2% about MMER, 15.1% about SER, and only 6.4% of the publications focus on EBGR. While EBGR shows steady but slower growth from 2016 onwards, SER maintains a consistent presence throughout the period, with a peak of 21.9% annual publications in 2018.

Regarding the databases used across different methodological approaches, Fig. 2 (top right and bottom right) presents the top 20 most referenced databases from a total of 127 databases identified in the analyzed SRoS papers. CK+ emerges as the most utilized database with 39 references, followed by JAFFE with 33 references, both of which are typically used in FER approaches. Also of notice is the high usage of proprietary databases, with a total of 38 references. The distribution of methods across databases (Fig. 2, top right) reveals that newer databases tend to be predominantly associated with DL-based approaches, while established databases show a more balanced distribution of methodologies. Table 2 further details this distribution across modalities.

The temporal evolution of the top 20 most utilized databases as seen in Fig. 3 shows distinct patterns, with some databases maintaining consistent usage throughout the analyzed period while others show more concentrated temporal adoption. Notable trends include the sustained use of established databases like CK+ and JAFFE, contrasting with the emergence of newer databases commonly associated with DL approaches, such as FER2013 and RaFD, all of which are typically used in FER context.

In the specific context of EBGR, the detailed analysis presented in Table 3 reveals significant patterns in the evolution of body-based emotion recognition methods. Early works from 2005 to 2018 primarily utilized traditional ML methods such as Decision Trees, $k$-Nearest Neighbors (kNN), and SVM approaches with handcrafted features. A notable transition period emerges around 2018, where initial DL implementations appear, though often still combined with traditional feature extraction methods. From 2019 onwards, there is a clear dominance of pure DL approaches, with increasing architectural complexity.

The architectural evolution of these methods follows a distinct progression pattern. Early implementations focused on ML and basic classifiers, primarily processing geometric features and motion parameters. The transition period after 2018 saw the introduction of basic neural networks and early CNN implementations. Modern approaches from 2019 onward use sophisticated architectures including specialized CNNs, GCNs, and transformer-based approaches, with increasing emphasis on temporal modeling capabilities.

Database usage patterns in EBGR research, as illustrated in Fig. 4, show notable trends, with early research predominantly relying on proprietary databases, reflecting the limited availability of standardized body-gesture databases at the time. From 2018 onwards, there is a shift towards specialized

**TABLE 2.** Summary of selected SRoS ERM papers per modality.

| Year | SER | FER | MMER | EBGR |
|---|---|---|---|---|
| 2002 | | [52] | | |
| 2005 | | [53] | [54] | |
| 2006 | | [55] | | |
| 2007 | | [57] | [56] | [58] |
| 2008 | [60] | [62], [63] | [7], [61] | [59] |
| 2009 | | | [64] | |
| 2010 | [65] | [68], [69] | [66], [67] | [70] |
| 2011 | | [72]–[74] | [71] | |
| 2012 | | [75]–[80] | [81] | |
| 2013 | [85] | [83], [84], [86]–[90] | [82], [91] | |
| 2014 | [92], [93], [100] | [94]–[98], [101] | | [99] |
| 2015 | [115], [116] | [102]–[108], [111]–[114], [117]–[122] | [109], [110] | |
| 2016 | [123], [124], [128], [129] | [125]–[127], [131]–[134], [136]–[139] | [130] | [50], [132], [135] |
| 2017 | [140], [141], [144], [157] | [142], [143], [146]–[156], [158]–[160] | [144], [145] | |
| 2018 | [161]–[164], [173], [179], [192] | [165]–[172], [175]–[178], [181], [183]–[187], [191] | [180], [182] | [174], [188]–[190] |
| 2019 | [210], [211], [218], [219] | [193], [194], [197]–[204], [208], [209], [212]–[217], [221]–[223], [228] | [195], [196], [205]–[207], [220], [224]–[227] | [51] |
| 2020 | [253], [257], [258], [263] | [229]–[231], [237]–[245], [250]–[252], [254]–[256], [259], [262] | [5], [232]–[236], [246]–[249], [253], [260], [261] | [48], [250] |
| 2021 | [273]–[276], [279] | [264]–[266], [278], [285], [286] | [8], [267]–[272], [280]–[284] | [49], [277] |
| 2022 | [300], [301], [303] | [297], [298] | [287]–[296] | [299], [302] |
| 2023 | [317], [318], [322], [323] | [315], [316], [320], [321], [329], [330] | [304]–[314], [317], [319], [324]–[328] | [319] |
| 2024 | [332], [333] | [331] | [334]–[336] | |

databases like GEMEP and FABO, designed specifically for emotional body and gesture movement analysis. Recent studies from 2021 to 2023 demonstrate increased use of multi-purpose databases like IEMOCAP and CREMA-D, particularly in the context of multimodal approaches, suggesting a trend toward more sophisticated emotion recognition systems. To be clear, the databases shown in Fig. 4 are extensive, and represent all EBGR databases, including database aggregations (like YORK; GEMEP) that were referenced from the ERM papers extracted from the original SRoS papers.

In summary, recent implementations demonstrate significant methodological sophistication through the integration of attention mechanisms for improved feature selection, adoption of graph-based representations for skeletal data, and implementation of advanced spatial-temporal modeling techniques. There is also an increased focus on real-time processing capabilities, reflecting the growing emphasis on practical applications of EBGR systems.

## IV. DISCUSSION

The systematic analysis that was conducted reveals several fundamental implications about the field's trajectory and future challenges in emotion recognition methodologies. The temporal evolution documented in Table 3, combined with the shifting modality preferences illustrated in Fig. 2, demonstrates a fundamental transformation in how researchers conceptualize and approach emotion recognition. This progression reflects not merely technological advancement but a deeper understanding of emotion as a complex spatial-temporal phenomenon requiring sophisticated modeling approaches.

The analysis uncovers three interrelated challenges that characterize the current state of the field: i) the computational demands of modern architectures present practical constraints, particularly as the field moves toward transformer-based implementations, as documented in Table 3. Further, Hazmoune and Bougamouza [337] review highlights transformers' growing dominance in emotion recognition, with innovations like skeleton-based emotion transformers demonstrating their effectiveness [338]. While these architectures excel at temporal modeling and multimodal integration, their substantial computational requirements may impede deployment in resource-constrained environments; ii) despite the apparent variety of databases shown in Fig. 2, persistent data challenges remain, as previously discussed in Vaz et al. [43]. The temporal patterns revealed in Fig. 3 indicate significant gaps in comprehensive database availability, particularly critical for skeleton-based approaches requiring specialized motion data; and iii) cross-database generalization emerges as a fundamental concern, with current methodological distributions across modalities suggesting

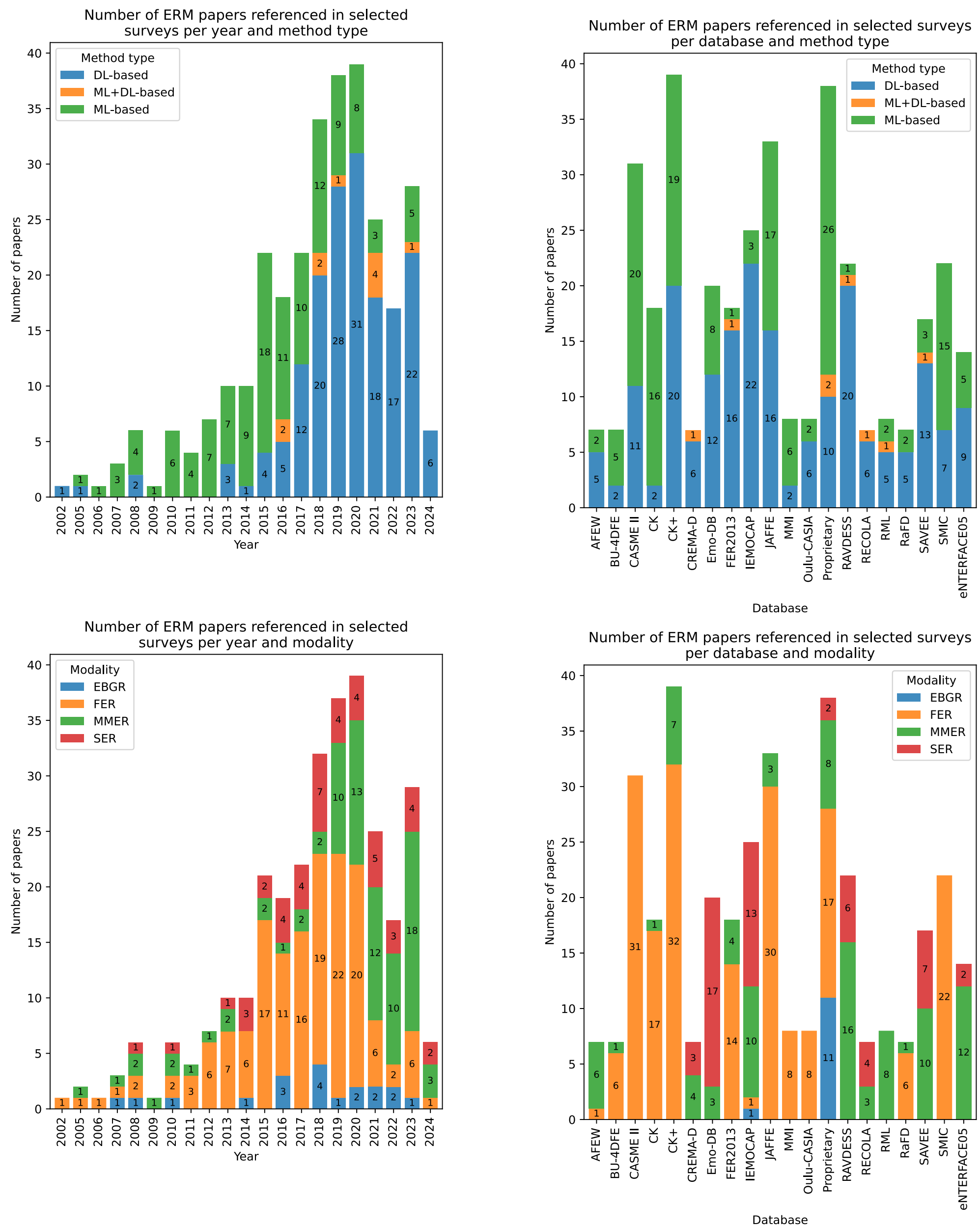


**FIGURE 2.** **Number of ERM papers referenced in selected surveys: per year and method type (top left), per database and method type (top right), per modality and year (bottom left), and per database and modality (bottom right).**

excessive dependence on database-specific characteristics rather than universal aspects of emotional expression.

The increasing adoption of multimodal approaches presents both opportunities and challenges for the field's development. While multimodal integration offers potential for more robust recognition, as evidenced by the growing proportion of MMER studies (see Fig. 2), it introduces complexities in temporal alignment and computational optimization. Additionally, the methodological progression, documented in Table 1, suggests an emerging recognition of the need to balance sophisticated modeling capabilities with practical implementation constraints. This balance may be achieved through hybrid approaches that combine the computational efficiency of traditional methods with the modeling power of modern DL architectures.

Furthermore, the implications go beyond technical aspects, reaching broader application domains. As emotion recognition systems are employed in sensitive fields such as healthcare and human-computer interaction, the demand for interpretable and reliable systems becomes essential. The methodological evolution documented in this analysis highlights that future developments must prioritize both performance and practical applicability. Our findings suggest that the next phase of development will likely focus on

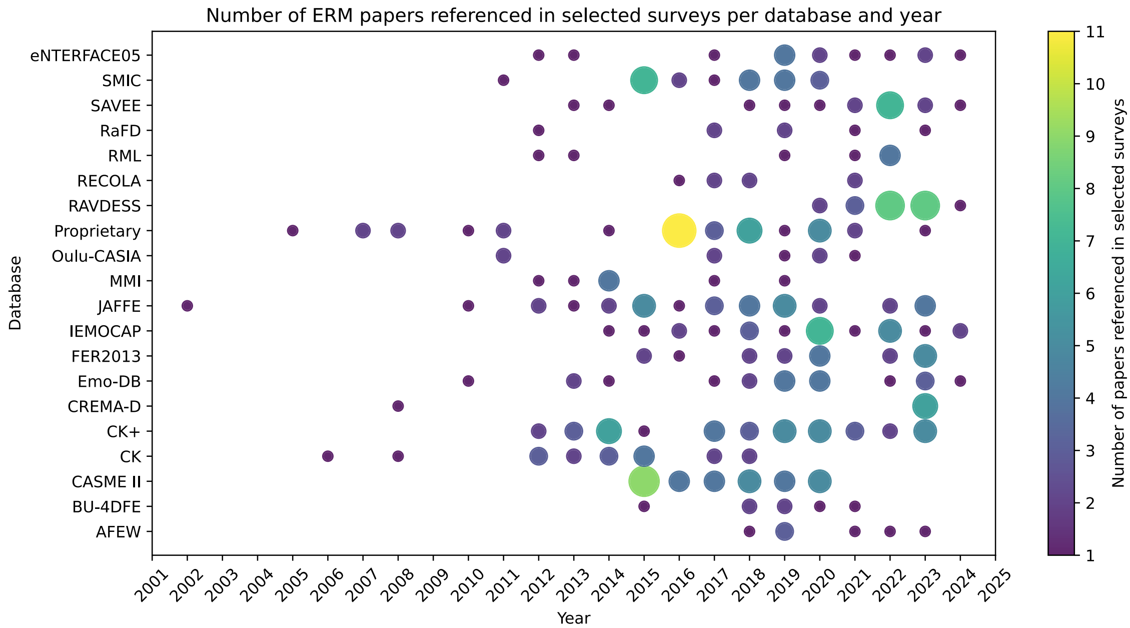


**FIGURE 3.** Number of ERM papers referenced in selected surveys per database and year.

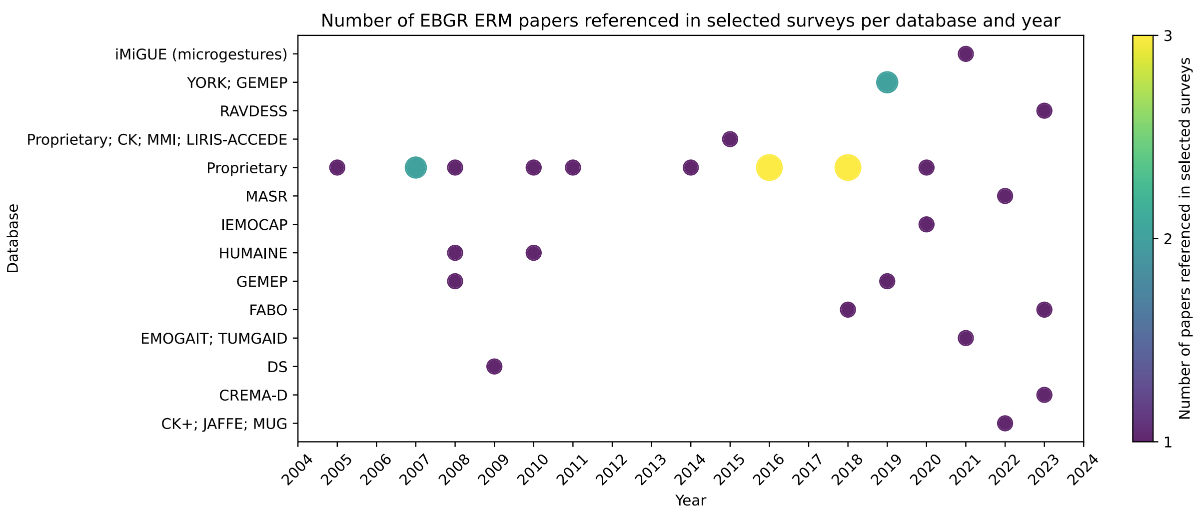


**FIGURE 4.** Number of body-based ERM papers referenced in selected surveys per database and year.

bridging theoretical capabilities with practical implementation needs, requiring not just technical innovation, but thoughtful consideration of real-world constraints, ethical implications, and end-user needs, as suggested by the evolving patterns observed across our analyzed data.

Based on our discussion, the following recommendations can be made to address the challenges and future directions of emotion recognition methodologies, particularly in the context of body-based emotion recognition: i) *Development of efficient DL models* – Transformer-based and other advanced architectures, while effective, require high computational resources. Researchers should focus on optimizing models to reduce computational costs while maintaining high accuracy, ensuring real-world applicability, especially in resource-constrained environments. ii) *Creation of large-scale and diverse databases* – To address the critical need for large-scale and diverse databases, in body-based emotion recognition, collaborative initiatives could leverage synthetic data generation to systematically expand database coverage, ensuring cultural, ethnical, and contextual diversity. By integrating models to map body language features (e.g., posture, gait, or gesture kinematics) to emotional states, researchers can automate labeling and maintain consistency across databases. Synthetic environments (e.g., Unity or Blender)

**TABLE 3.** Detailed analysis of body-based emotion recognition studies.

| Survey | Paper | Year | ML-based | DL-based | Architecture | Database |
|---|---|---|---|---|---|---|
| [1] | [54] | 2005 | ✓ | | DT + BayesNet | Proprietary |
| [1] | [56] | 2007 | ✓ | | 1NN | Proprietary |
| [1] | [7] | 2008 | ✓ | | BayesNet | HUMAINE |
| [1] | [61] | 2008 | ✓ | | Statistical ML | GEMEP |
| [1] | [64] | 2009 | ✓ | | Statistical Analysis (Chi-square) | DS |
| [1] | [67] | 2010 | ✓ | | BayesNet | HUMAINE |
| [1] | [71] | 2011 | ✓ | | Fuzzy Choquet Integral + Decision Level Fusion | Proprietary |
| [1] | [110] | 2015 | ✓ | | ANN | Proprietary; CK; MMI; LIRIS-ACCEDE |
| [46] | [59] | 2008 | | ✓ | Multilayer Perceptron, Self-organizing Maps | Proprietary |
| [46] | [58] | 2007 | ✓ | | Original non-linear source separation, Sparse Multivariate Regression | Proprietary |
| [46] | [99] | 2014 | ✓ | | Similarity index | Proprietary |
| [46] | [70] | 2010 | ✓ | | KNN, Naive Bayes, SVM | Proprietary |
| [46] | [50] | 2016 | ✓ | | LDA, Naive Bayes, SVM with PCA features | Proprietary |
| [46] | [135] | 2016 | ✓ | | Naive Bayes, Random Forest, SVM, SMO | Proprietary |
| [46] | [132] | 2016 | ✓ | | Decision Tree, SVM, Random Forest, Random Tree with PCA features | Proprietary |
| [46] | [188] | 2018 | ✓ | | Random Forest, Logistic Regression | Proprietary |
| [46] | [189] | 2018 | ✓ | | SVM, Decision Tree, Naive Bayes, Random Forest, Logistic Regression, Multilayer Perceptron | Proprietary |
| [46] | [190] | 2018 | ✓ | | KNN, SVM, LDA, Decision Tree, Genetic algorithm, Score-level Fusion, Rank-level Fusion | Proprietary |
| [4] | [314] | 2023 | | ✓ | Transformer | RAVDESS |
| [4] | [314] | 2023 | | ✓ | Transformer | CREMA-D |
| [3] | [51] | 2019 | | ✓ | Convolutional neural networks, feedforward deep convolution neural network (FDCNN) | YORK; GEMEP |
| [47] | [51] | 2019 | | ✓ | CNN | YORK; GEMEP |
| [47] | [250] | 2020 | | ✓ | CNN | Proprietary |
| [47] | [49] | 2021 | | ✓ | Graph convolutional neural network, spatial–temporal attention | EMOGAIT; TUMGAID |
| [47] | [299] | 2022 | | ✓ | Multi-task CNN | CK+; JAFFE; MUG |
| [47] | [48] | 2020 | | ✓ | graph convolutional network | IEMOCAP |
| [13] | [319] | 2023 | | ✓ | Coupling network | FABO |
| [10] | [174] | 2018 | | ✓ | MLP | FABO |
| [10] | [51] | 2019 | | ✓ | MLP | GEMEP |
| [10] | [277] | 2021 | | ✓ | BiLSTM (encoder)/ LSTM (decoder) | iMiGUE (microgestures) |
| [10] | [302] | 2022 | | ✓ | HPN/SAE | MASR |

combined with generative models (e.g., GANs or VAEs) can also, potentially, simulate diverse human avatars performing emotion-specific movements. Active learning pipelines could then prioritize ambiguous or rare cases for human annotation. This hybrid approach—merging synthetic data with curated real-world examples—would enable the creation of standardized, open-access databases that capture nuanced emotional expressions while mitigating privacy concerns and biases. iii) *Enhancement of multimodal fusion strategies* – While multimodal approaches integrating facial, speech, and body-based emotion recognition improve performance, better fusion mechanisms are needed to handle varying data availability and ensure robust emotion detection under different conditions. I.e., adaptive fusion architectures (e.g., hierarchical transformers [185], [339]) could dynamically weight modalities based on contextual reliability (e.g., prioritizing body gestures in low-light settings where facial cues are obscured) while integrating domain-specific knowledge to resolve ambiguities. iv) *Addressing cross-database generalization challenges* – Current models often rely on database-specific characteristics, limiting their applicability across diverse real-world scenarios. Future research should focus on improving domain adaptation and transfer learning techniques to enhance cross-database performance. v) *Consideration of ethical and interpretability issues* – Emotion recognition systems, especially those used in healthcare, cultural studies, ethnical research etc., must ensure fairness, transparency, and user privacy. Researchers should develop interpretable AI models and establish ethical guidelines for responsible deployment. vi) *Real-time implementation and latency reduction* – Many emotion recognition systems, particularly those used in interactive environments, require real-time processing. Researchers should prioritize the development of low-latency models that maintain accuracy while ensuring minimal processing delays. E.g., this can be achieved through model optimization techniques (such as quantization, pruning, or distillation) to reduce computational complexity, alongside edge computing frameworks that deploy lightweight models directly on IoT or wearable devices. Hybrid pipelines might offload feature extraction to edge devices and perform fusion on cloud servers, balancing latency and accuracy. vii) *Exploration of novel feature extraction techniques* – Emotion recognition performance can be improved by leveraging more effective feature extraction methods, such as attention mechanisms, spatio-temporal representations, and graph-based approaches for skeletal motion analysis. viii) *Cross-cultural, ethnical, and demographic considerations* – Emotional expressions vary across cultures, ethnicities, and demographics. Future research should ensure that models are trained on diverse databases to enhance robustness and fairness across different populations. ix) *Standardization of evaluation metrics and benchmarks* – A lack of consistency in evaluation metrics makes it difficult to compare results across studies. The field would benefit from the establishment of standardized benchmarks and performance evaluation protocols.

By addressing these recommendations, the EBGR field can progress toward more accurate, efficient, and ethically responsible systems.

## V. CONCLUSION

This study conducted an analysis of SRoS papers on emotion recognition methods published between 2021 and 2024, focusing particularly on emotional body gesture recognition. Through systematic examination, an expected clear progression from traditional ML to sophisticated DL architectures was revealed, with recent work emphasizing transformer-based models and attention mechanisms. This evolution reflects a growing understanding of emotion recognition as a complex temporal-spatial phenomenon requiring sophisticated modeling approaches.

The analysis has identified several critical constraints currently facing the field. Database availability remains a critical constraint, particularly for body-based approaches requiring specialized motion data. The computational requirements of modern architectures present implementation challenges in resource-constrained environments, while cross-database generalization remains problematic.

These challenges suggest important directions for future research, including the development of efficient architectures that balance sophistication with practical implementation, creation of large body-based emotional databases, and investigation of robust fusion strategies for multimodal approaches. The integration of real-time processing capabilities in multimodal systems, coupled with the establishment of robust ethical guidelines for emotion recognition deployment, will be essential for the field's continued development. Additionally, the investigation of privacy-preserving techniques and culturally adaptive recognition systems has become important as these systems move toward real-world deployment.

In summary, the AffC field shows particularly promising directions in emotional body gesture recognition and multimodal approaches, though significant challenges remain. Success in addressing these challenges will require not just technical innovation, but thoughtful consideration of real-world constraints, ethical implications, and end-user needs in diverse application contexts. Furthermore, the advancement of emotion recognition systems must prioritize both performance and practical applicability, ensuring broad accessibility across diverse real-world scenarios.

## REFERENCES

[1] F. Noroozi, C. A. Corneanu, D. Kaminska, T. Sapinski, S. Escalera, and G. Anbarjafari, "Survey on emotional body gesture recognition," *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 505–523, Apr. 2021.

[2] M. Spezialetti, G. Placidi, and S. Rossi, "Emotion recognition for human–robot interaction: Recent advances and future perspectives," *Frontiers Robot. AI*, vol. 7, Dec. 2020, Art. no. 532279.

[3] S. C. Leong, Y. M. Tang, C. H. Lai, and C. K. M. Lee, "Facial expression and body gesture emotion recognition: A systematic review on the use of visual data in affective computing," *Comput. Sci. Rev.*, vol. 48, May 2023, Art. no. 100545.

[4] A. Geetha, T. Mala, D. Priyanka, and E. Uma, "Multimodal emotion recognition with deep learning: Advancements, challenges, and future directions," *Inf. Fusion*, vol. 105, May 2024, Art. no. 102218.

[5] A. Bera, T. Randhavane, R. Prinja, K. Kapsaskis, A. Wang, K. Gray, and D. Manocha, "How are you feeling? Multimodal emotion learning for socially-assistive robot navigation," in *Proc. 15th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Nov. 2020, pp. 644–651.

[6] S. S. Hosseini, M. Yamaghani, and S. Arabani, "A review of methods for detecting multimodal emotions in sound, image and text," *J. Appl. Res. Ind. Eng.*, vol. 11, no. 4, pp. 669–705, 2024. [Online]. Available: https://www.journal-aprie.com/article_190350.html

[7] G. Castellano, L. Kessous, and G. Caridakis, "Emotion recognition through multiple modalities: Face, body gesture, speech," in *Affect and Emotion in Human–Computer Interaction: From Theory to Applications*, C. Peter and R. Beale, Eds., Berlin, Germany: Springer, 2008, pp. 92–103.

[8] D. Liu, Z. Wang, L. Wang, and L. Chen, "Multi-modal fusion emotion recognition method of speech expression based on deep learning," *Frontiers Neurorobotics*, vol. 15, Jul. 2021, Art. no. 697634.

[9] D. Mamieva, A. B. Abdusalomov, A. Kutlimuratov, B. Muminov, and T. K. Whangbo, "Multimodal emotion detection via attention-based fusion of extracted facial and speech features," *Sensors*, vol. 23, no. 12, p. 5475, Jun. 2023.

[10] G. Udahemuka, K. Djouani, and A. M. Kurien, "Multimodal emotion recognition using visual, vocal and physiological signals: A review," *Appl. Sci.*, vol. 14, no. 17, p. 8071, Sep. 2024.

[11] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423–443, Feb. 2019.

[12] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Inf. Fusion*, vol. 37, pp. 98–125, Sep. 2017.

[13] S. Kalateh, L. A. Estrada-Jimenez, S. Nikghadam-Hojjati, and J. Barata, "A systematic review on multimodal emotion recognition: Building blocks, current state, applications, and challenges," *IEEE Access*, vol. 12, pp. 103976–104019, 2024.

[14] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 2000.

[15] M. Munezero, C. S. Montero, E. Sutinen, and J. Pajunen, "Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text," *IEEE Trans. Affect. Comput.*, vol. 5, no. 2, pp. 101–111, Apr. 2014.

[16] K. R. Scherer, "What are emotions? And how can they be measured?" *Social Sci. Inf.*, vol. 44, no. 4, pp. 695–729, Dec. 2005.

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

[18] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *J. Personality Social Psychol.*, vol. 17, no. 2, pp. 124–129, 1971.

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

[20] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, nos. 1–2, pp. 1–135, 2008.

[21] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2004, pp. 168–177.

[22] J. J. McAuley and J. Leskovec, "From amateurs to connoisseurs: Modeling the evolution of user expertise through online reviews," in *Proc. 22nd Int. Conf. World Wide Web*, May 2013, pp. 897–908.
[23] S. M. Mohammad, "Sentiment analysis: Detecting valence, emotions, and other affectual states from text," in *Emotion Measurement*. Amsterdam, The Netherlands: Elsevier, 2016, pp. 201–237.
[24] Y. Wang, W. Song, W. Tao, A. Liotta, D. Yang, X. Li, S. Gao, Y. Sun, W. Ge, W. Zhang, and W. Zhang, "A systematic review on affective computing: Emotion models, databases, and recent advances," *Inf. Fusion*, vols. 83–84, pp. 19–52, Jul. 2022.
[25] N. J. Shoumy, L.-M. Ang, K. P. Seng, D. M. M. Rahaman, and T. Zia, "Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals," *J. Netw. Comput. Appl.*, vol. 149, Jan. 2020, Art. no. 102447.
[26] P. Ekman, "Universals and cultural differences in facial expressions of emotion," in *Proc. Nebraska Symp. Motivat.*, 1971, pp. 207–283.
[27] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *Amer. Sci.*, vol. 89, no. 4, pp. 344–350, 2001.
[28] L. F. Barrett, "Are emotions natural kinds?" *Perspect. Psychol. Sci.*, vol. 1, no. 1, pp. 28–58, Mar. 2006.
[29] B. Martinez, M. F. Valstar, B. Jiang, and M. Pantic, "Automatic analysis of facial actions: A survey," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 325–347, Jul. 2019.
[30] J. A. Russell, "A circumplex model of affect," *J. Personality Social Psychol.*, vol. 39, no. 6, p. 1161, 1980.
[31] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychol.*, vol. 14, no. 4, pp. 261–292, Dec. 1996.
[32] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behav. Therapy Experim. Psychiatry*, vol. 25, no. 1, pp. 49–59, Mar. 1994.
[33] I. Bakker, T. van der Voordt, P. Vink, and J. de Boon, "Pleasure, arousal, dominance: Mehrabian and Russell revisited," *Current Psychol.*, vol. 33, no. 3, pp. 405–421, Sep. 2014.
[34] J. A. Russell and L. F. Barrett, "Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant," *J. Personality Social Psychol.*, vol. 76, no. 5, pp. 805–819, 1999.
[35] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," 2002, *arXiv:cs/0205070*.
[36] B. Liu, *Sentiment Analysis and Opinion Mining*. Cham, Switzerland: Springer, 2022.
[37] S. Mohammad, F. Bravo-Marquez, M. Salameh, and S. Kiritchenko, "SemEval-2018 task 1: Affect in tweets," in *Proc. 12th Int. Workshop Semantic Eval.*, 2018, pp. 1–17.
[38] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2013, pp. 1631–1642.
[39] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment strength detection for the social web," *J. Amer. Soc. for Inf. Sci. Technol.*, vol. 63, no. 1, pp. 163–173, Jan. 2012.
[40] D. Preoţiuc-Pietro, H. A. Schwartz, G. Park, J. C. Eichstaedt, M. L. Kern, L. Ungar, and E. Shulman, "Modelling valence and arousal in Facebook posts," in *Proc. 7th Workshop Comput. Approaches Subjectivity*, Mar. 2016, pp. 9–15.
[41] J. Wang, L.-C. Yu, K. R. Lai, and X. Zhang, "Dimensional sentiment analysis using a regional CNN-LSTM model," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, 2016, pp. 225–230.
[42] S. Rosenthal, N. Farra, and P. Nakov, "SemEval-2017 task 4: Sentiment analysis in Twitter," 2019, *arXiv:1912.00741*.
[43] P. J. Vaz, J. M. F. Rodrigues, and P. J. S. Cardoso, "Affective computing databases: In-depth analysis of systematic reviews and surveys," *IEEE Trans. Affect. Comput.*, vol. 16, no. 2, pp. 537–554, Apr. 2025.
[44] P. A. Baffour, H. Nunoo-Mensah, E. Keelson, and B. Kommey, "A survey on deep learning algorithms in facial emotion detection and recognition," *Inf., Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 7, no. 1, pp. 24–32, Jan. 2022.
[45] F. Z. Canal, T. R. Müller, J. C. Matias, G. G. Scotton, A. R. de Sa Junior, E. Pozzebon, and A. C. Sobieranski, "A survey on facial emotion recognition techniques: A state-of-the-art literature review," *Inf. Sci.*, vol. 582, pp. 593–617, Jan. 2022.
[46] S. Xu, J. Fang, X. Hu, E. Ngai, W. Wang, Y. Guo, and V. C. M. Leung, "Emotion recognition from gait analyses: Current research and future directions," *IEEE Trans. Computat. Social Syst.*, vol. 11, no. 1, pp. 363–377, Feb. 2024.
[47] K. Ezzameli and H. Mahersia, "Emotion recognition from unimodal to multimodal analysis: A review," *Inf. Fusion*, vol. 99, Nov. 2023, Art. no. 101847.
[48] J. Shi, C. Liu, C. T. Ishi, and H. Ishiguro, "Skeleton-based emotion recognition based on two-stream self-attention enhanced spatial–temporal graph convolutional network," *Sensors*, vol. 21, no. 1, p. 205, Dec. 2020.
[49] W. Sheng and X. Li, "Multi-task learning for gait-based identity recognition and emotion recognition using attention enhanced temporal graph convolutional network," *Pattern Recognit.*, vol. 114, Jun. 2021, Art. no. 107868.
[50] B. Li, C. Zhu, S. Li, and T. Zhu, "Identifying emotions from non-contact gaits information based on Microsoft kinects," *IEEE Trans. Affect. Comput.*, vol. 9, no. 4, pp. 585–591, Oct. 2018.
[51] R. Santhoshkumar and M. K. Geetha, "Deep learning approach for emotion recognition from human body movements with feedforward deep convolution neural networks," *Proc. Comput. Sci.*, vol. 152, pp. 158–165, Sep. 2019.
[52] M. Gargesha, P. Kuchi, and I. Torkkola, "Facial expression recognition using artificial neural networks," *Artif. Neural Comput. Syst.*, vol. 8, no. 4, pp. 1–6, 2002.
[53] Y. Guo, D. Tao, J. Yu, H. Xiong, Y. Li, and D. Tao, "Deep neural networks with relativity learning for facial expression recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2005, pp. 568–573.
[54] H. Gunes and M. Piccardi, "Affect recognition from face and body: Early fusion vs. late fusion," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, vol. 4, Apr. 2005, pp. 3437–3443.
[55] M. Yeasin, B. Bullot, and R. Sharma, "Recognition of facial expressions and measurement of levels of interest from video," *IEEE Trans. Multimedia*, vol. 8, no. 3, pp. 500–508, Jun. 2006.
[56] G. Castellano, S. D. Villalba, and A. Camurri, "Recognising human emotions from body movement and gesture dynamics," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.* Cham, Switzerland: Springer, 2007, pp. 71–82.
[57] B. Hernández, G. Olague, R. Hammoud, L. Trujillo, and E. Romero, "Visual learning of texture descriptors for facial expression recognition in thermal imagery," *Comput. Vis. Image Understand.*, vol. 106, nos. 2–3, pp. 258–269, May 2007.
[58] L. Omlor and M. A. Giese, "Extraction of spatio-temporal primitives of emotional body expressions," *Neurocomputing*, vol. 70, nos. 10–12, pp. 1938–1942, Jun. 2007.
[59] D. Janssen, W. I. Schöllhorn, J. Lubienetzki, K. Fölling, H. Kokenge, and K. Davids, "Recognition of emotions in gait patterns by means of artificial neural nets," *J. Nonverbal Behav.*, vol. 32, no. 2, pp. 79–92, Jun. 2008.
[60] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "IEMOCAP: Interactive emotional dyadic motion capture database," *Lang. Resour. Eval.*, vol. 42, no. 4, pp. 335–359, Dec. 2008.
[61] D. Glowinski, A. Camurri, G. Volpe, N. Dael, and K. Scherer, "Technique for automatic emotion recognition by body gesture analysis," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2008, pp. 1–6.
[62] J.-M. Sun, X.-S. Pei, and S.-S. Zhou, "Facial emotion recognition in modern distant education system using SVM," in *Proc. Int. Conf. Mach. Learn. Cybern.*, Jul. 2008, pp. 3545–3548.
[63] J. N. Bailenson, E. D. Pontikakis, I. B. Mauss, J. J. Gross, M. E. Jabon, C. A. C. Hutcherson, C. Nass, and O. John, "Real-time classification of evoked emotions using facial feature tracking and physiological responses," *Int. J. Hum.-Comput. Stud.*, vol. 66, no. 5, pp. 303–317, May 2008.
[64] M. Kipp and J.-C. Martin, "Gesture and emotion: Can basic gestural form features discriminate emotions?" in *Proc. 3rd Int. Conf. Affect. Comput. Intell. Interact. Workshops*, Sep. 2009, pp. 1–8.
[65] D. Bitouk, R. Verma, and A. Nenkova, "Class-level spectral features for emotion recognition," *Speech Commun.*, vol. 52, nos. 7–8, pp. 613–625, Jul. 2010.

[66] C.-H. Wu and W.-B. Liang, "Emotion recognition of affective speech based on multiple classifiers using acoustic-prosodic information and semantic labels," *IEEE Trans. Affect. Comput.*, vol. 2, no. 1, pp. 10–21, Jan. 2011.

[67] 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, nos. 1–2, pp. 33–48, Mar. 2010.

[68] H. Wang, H. Huang, Y. Hu, M. Anderson, P. Rollins, and F. Makedon, "Emotion detection via discriminative kernel method," in *Proc. 3rd Int. Conf. Pervasive Technol. Rel. Assistive Environ.*, Jun. 2010, pp. 1–7.

[69] K.-E. Ko and K.-B. Sim, "Development of a facial emotion recognition method based on combining AAM with DBN," in *Proc. Int. Conf. Cyberworlds*, Oct. 2010, pp. 87–91.

[70] M. Karg, K. Kuhnlenz, and M. Buss, "Recognition of affect based on gait patterns," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 40, no. 4, pp. 1050–1061, Aug. 2010.

[71] H. A. Vu, Y. Yamazaki, F. Dong, and K. Hirota, "Emotion recognition based on human gesture and speech information using RT middleware," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jun. 2011, pp. 787–791.

[72] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. Pietikäinen, "Facial expression recognition from near-infrared videos," *Image Vis. Comput.*, vol. 29, no. 9, pp. 607–619, Aug. 2011.

[73] S. Moore and R. Bowden, "Local binary patterns for multi-view facial expression recognition," *Comput. Vis. Image Understand.*, vol. 115, no. 4, pp. 541–558, Apr. 2011.

[74] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Recognising spontaneous facial micro-expressions," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1449–1456.

[75] W. Gu, C. Xiang, Y. V. Venkatesh, D. Huang, and H. Lin, "Facial expression recognition using radial encoding of local Gabor features and classifier synthesis," *Pattern Recognit.*, vol. 45, no. 1, pp. 80–91, Jan. 2012.

[76] J. Chen, D. Chen, Y. Gong, M. Yu, K. Zhang, and L. Wang, "Facial expression recognition using geometric and appearance features," in *Proc. 4th Int. Conf. Internet Multimedia Comput. Service*, Sep. 2012, pp. 29–33.

[77] M. Ilbeygi and H. Shah-Hosseini, "A novel fuzzy facial expression recognition system based on facial feature extraction from color face images," *Eng. Appl. Artif. Intell.*, vol. 25, no. 1, pp. 130–146, Feb. 2012.

[78] N. Perveen, S. Gupta, and K. Verma, "Facial expression recognition using facial characteristic points and Gini index," in *Proc. Students Conf. Eng. Syst.*, Mar. 2012, pp. 1–6.

[79] M. Tkalcic, A. Odic, A. Kosir, and J. Tasic, "Affective labeling in a content-based recommender system for images," *IEEE Trans. Multimedia*, vol. 15, no. 2, pp. 391–400, Feb. 2013.

[80] A. R. Rivera, J. R. Castillo, and O. O. Chae, "Local directional number pattern for face analysis: Face and expression recognition," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1740–1752, May 2013.

[81] Y. Wang, L. Guan, and A. N. Venetsanopoulos, "Kernel cross-modal factor analysis for information fusion with application to bimodal emotion recognition," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 597–607, Jun. 2012.

[82] M. Wöllmer, F. Weninger, T. Knaup, B. Schuller, C. Sun, K. Sagae, and L.-P. Morency, "YouTube movie reviews: Sentiment analysis in an audio-visual context," *IEEE Intell. Syst.*, vol. 28, no. 3, pp. 46–53, May 2013.

[83] L. Zhang, M. Jiang, D. Farid, and M. A. Hossain, "Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot," *Expert Syst. Appl.*, vol. 40, no. 13, pp. 5160–5168, Oct. 2013.

[84] S. E. Kahou et al., "Combining modality specific deep neural networks for emotion recognition in video," in *Proc. 15th ACM Int. Conf. Multimodal Interact.*, Dec. 2013, pp. 543–550.

[85] A. Milton, S. S. Roy, and S. T. Selvi, "SVM scheme for speech emotion recognition using MFCC feature," *Int. J. Comput. Appl.*, vol. 69, no. 9, pp. 34–39, May 2013.

[86] D. Ghimire and J. Lee, "Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines," *Sensors*, vol. 13, no. 6, pp. 7714–7734, Jun. 2013.

[87] Y. Li, S. Wang, Y. Zhao, and Q. Ji, "Simultaneous facial feature tracking and facial expression recognition," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2559–2573, Jul. 2013.

[88] D. Mushfieldt, M. Ghaziasgar, and J. Connan, "Robust facial expression recognition in the presence of rotation and partial occlusion," in *Proc. South Afr. Inst. Comput. Scientists Inf. Technologists Conf.*, Oct. 2013, pp. 186–193.

[89] F. S. Farahani, M. Sheikhan, and A. Farrokhi, "A fuzzy approach for facial emotion recognition," in *Proc. 13th Iranian Conf. Fuzzy Syst. (IFSC)*, Farah, Iran, Aug. 2013, pp. 1–4.

[90] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen, "A spontaneous micro-expression database: Inducement, collection and baseline," in *Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, Apr. 2013, pp. 1–6.

[91] Z. Xie and L. Guan, "Multimodal information fusion of audiovisual emotion recognition using novel information theoretic tools," in *Proc. IEEE Int. Conf. Multimedia Expo. (ICME)*, Jul. 2013, pp. 1–6.

[92] Z. Huang, M. Dong, Q. Mao, and Y. Zhan, "Speech emotion recognition using CNN," in *Proc. 22nd ACM Int. Conf. Multimedia*, Nov. 2014, pp. 801–804.

[93] Y. Jin, P. Song, W. Zheng, and L. Zhao, "A feature selection and feature fusion combination method for speaker-independent speech emotion recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 4808–4812.

[94] S. L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," *IEEE Trans. Affect. Comput.*, vol. 6, no. 1, pp. 1–12, Jan. 2015.

[95] R. Ghasemi and M. Ahmady, "Facial expression recognition using facial effective areas and fuzzy logic," in *Proc. Iranian Conf. Intell. Syst. (ICIS)*, Iran, Feb. 2014, pp. 1–4.

[96] S. Mohseni, N. Zarei, and S. Ramazani, "Facial expression recognition using anatomy based facial graph," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2014, pp. 3715–3719.

[97] A. C. Cruz, B. Bhanu, and N. S. Thakoor, "One shot emotion scores for facial emotion recognition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 1376–1380.

[98] M. T. Eskil and K. S. Benli, "Facial expression recognition based on anatomy," *Comput. Vis. Image Understand.*, vol. 119, pp. 1–14, Feb. 2014.

[99] G. Venture, H. Kadone, T. Zhang, J. Grèzes, A. Berthoz, and H. Hicheur, "Recognizing emotions conveyed by human gait," *Int. J. Social Robot.*, vol. 6, no. 4, pp. 621–632, Nov. 2014.

[100] A. Savran, H. Cao, A. Nenkova, and R. Verma, "Temporal Bayesian fusion for affect sensing: Combining video, audio, and lexical modalities," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1927–1941, Sep. 2015.

[101] K. Mistry, L. Zhang, S. C. Neoh, M. Jiang, A. Hossain, and B. Lafon, "Intelligent appearance and shape based facial emotion recognition for a humanoid robot," in *Proc. 8th Int. Conf. Softw., Knowl., Inf. Manage. Appl. (SKIMA)*, Dec. 2014, pp. 1–8.

[102] H. Ali, M. Hariharan, S. Yaacob, and A. H. Adom, "Facial emotion recognition using empirical mode decomposition," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1261–1277, Feb. 2015.

[103] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in *Proc. ACM Int. Conf. Multimodal Interact.*, Nov. 2015, pp. 435–442.

[104] H.-W. Ng, V. D. Nguyen, V. Vonikakis, and S. Winkler, "Deep learning for emotion recognition on small datasets using transfer learning," in *Proc. ACM Int. Conf. Multimodal Interact.*, Nov. 2015, pp. 443–449.

[105] G. Levi and T. Hassner, "Emotion recognition in the wild via convolutional neural networks and mapped binary patterns," in *Proc. ACM Int. Conf. Multimodal Interact.*, Nov. 2015, pp. 503–510.

[106] Y. Wang, J. See, R. C.-W. Phan, and Y.-H. Oh, "Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition," *PLoS ONE*, vol. 10, no. 5, May 2015, Art. no. e0124674.

[107] A. Azazi, S. L. Lutfi, I. Venkat, and F. Fernández-Martínez, "Towards a robust affect recognition: Automatic facial expression recognition in 3D faces," *Expert Syst. Appl.*, vol. 42, no. 6, pp. 3056–3066, Apr. 2015.

[108] Y.-J. Liu, J.-K. Zhang, W.-J. Yan, S.-J. Wang, G. Zhao, and X. Fu, "A main directional mean optical flow feature for spontaneous micro-expression recognition," *IEEE Trans. Affect. Comput.*, vol. 7, no. 4, pp. 299–310, Oct. 2016.

[109] S. Poria, E. Cambria, and A. Gelbukh, "Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 2539–2544.

[110] M. Gavrilescu, "Recognizing emotions from videos by studying facial expressions, body postures and hand gestures," in *Proc. 23rd Telecommun. Forum Telfor (TELFOR)*, Nov. 2015, pp. 720–723.

[111] S. Biswas and J. Sil, "An efficient expression recognition method using contourlet transform," in *Proc. 2nd Int. Conf. Perception Mach. Intell.*, Feb. 2015, pp. 167–174.

[112] M. Szwoch and P. Pieniazek, "Facial emotion recognition using depth data," in *Proc. 8th Int. Conf. Hum. Syst. Interact. (HSI)*, Jun. 2015, pp. 271–277.

[113] A. Nicolai and A. Choi, "Facial emotion recognition using fuzzy systems," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2015, pp. 2216–2221.

[114] L. Zhang, K. Mistry, M. Jiang, S. C. Neoh, and M. A. Hossain, "Adaptive facial point detection and emotion recognition for a humanoid robot," *Comput. Vis. Image Understand.*, vol. 140, pp. 93–114, Nov. 2015.

[115] J. C. Kim and M. A. Clements, "Multimodal affect classification at various temporal lengths," *IEEE Trans. Affect. Comput.*, vol. 6, no. 4, pp. 371–384, Oct. 2015.

[116] S. Haq, T. Jan, A. Jehangir, M. Asif, A. Ali, and N. Ahmad, "Bimodal human emotion classification in the speaker-dependent scenario," *Pak. Acad. Sci.*, vol. 52, no. 1, pp. 27–38, 2015.

[117] A. C. L. Ngo, R. C. Phan, and J. See, "Spontaneous subtle expression recognition: Imbalanced databases and solutions," in *Proc. 12th Asian Conf. Comput. Vis. Comput. Vis. (ACCV)*, Mar. 2015, pp. 33–48.

[118] Y. Wang, J. See, R. C. Phan, and Y.-H. Oh, "LBP with six intersection points: Reducing redundant information in LBP-TOP for micro-expression recognition," in *Proc. 12th Asian Conf. Comput. Vis. Comput. Vis.*, 2015, pp. 525–537.

[119] X. Huang, S.-J. Wang, G. Zhao, and M. Piteikäinen, "Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 1–9.

[120] Y.-H. Oh, A. C. Le Ngo, J. See, S.-T. Liong, R. C.-W. Phan, and H.-C. Ling, "Monogenic Riesz wavelet representation for micro-expression recognition," in *Proc. IEEE Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2015, pp. 1237–1241.

[121] S. Y. Park, S. H. Lee, and Y. M. Ro, "Subtle facial expression recognition using adaptive magnification of discriminative facial motion," in *Proc. 23rd ACM Int. Conf. Multimedia*, Oct. 2015, pp. 911–914.

[122] S. Liong, J. See, R. C. Phan, A. C. L. Ngo, Y.-H. Oh, and K. Wong, "Subtle expression recognition using optical strain weighted features," in *Proc. Comput. Vision-ACCV Workshops*, 2015, pp. 644–657.

[123] S. Ghosh, E. Laksana, L.-P. Morency, and S. Scherer, "Representation learning for speech emotion recognition," in *Proc. Interspeech*, Sep. 2016, pp. 3603–3607.

[124] G. Trigeorgis, F. Ringeval, R. Brueckner, E. Marchi, M. A. Nicolaou, B. Schuller, and S. Zafeiriou, "Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 5200–5204.

[125] J. Jeon, J.-C. Park, Y. Jo, C. Nam, K.-H. Bae, Y. Hwang, and D.-S. Kim, "A real-time facial expression recognizer using deep neural network," in *Proc. 10th Int. Conf. Ubiquitous Inf. Manage. Commun.*, Jan. 2016, pp. 1–4.

[126] Y. Guo, D. Tao, J. Yu, H. Xiong, Y. Li, and D. Tao, "Deep neural networks with relativity learning for facial expression recognition," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2016, pp. 1–6.

[127] D. H. Kim, W. J. Baddar, and Y. M. Ro, "Micro-expression recognition with expression-state constrained spatio-temporal feature representations," in *Proc. 24th ACM Int. Conf. Multimedia*, Oct. 2016, pp. 382–386.

[128] L.-A. Perez-Gaspar, S.-O. Caballero-Morales, and F. Trujillo-Romero, "Multimodal emotion recognition with evolutionary computation for human–robot interaction," *Expert Syst. Appl.*, vol. 66, pp. 42–61, Dec. 2016.

[129] K. Bahreini, R. Nadolski, and W. Westera, "Towards multimodal emotion recognition in e-learning environments," *Interact. Learn. Environments*, vol. 24, no. 3, pp. 590–605, Apr. 2016.

[130] J. Chen, Z. Chen, Z. Chi, and H. Fu, "Facial expression recognition in video with multiple feature fusion," *IEEE Trans. Affect. Comput.*, vol. 9, no. 1, pp. 38–50, Jan. 2018.

[131] G. Ali, M. A. Iqbal, and T.-S. Choi, "Boosted NNE collections for multicultural facial expression recognition," *Pattern Recognit.*, vol. 55, pp. 14–27, Jul. 2016.

[132] Z. Zhang, Y. Song, L. Cui, X. Liu, and T. Zhu, "Emotion recognition based on customized smart bracelet with built-in accelerometer," *PeerJ*, vol. 4, p. e2258, Jul. 2016.

[133] Z. Zhang, L. Cui, X. Liu, and T. Zhu, "Emotion detection using Kinect 3D facial points," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Oct. 2016, pp. 407–410.

[134] N. Mehta and S. Jadhav, "Facial emotion recognition using log Gabor filter and PCA," in *Proc. Int. Conf. Comput. Commun. Control Autom. (ICCUBEA)*, Aug. 2016, pp. 1–5.

[135] S. Li, L. Cui, C. Zhu, B. Li, N. Zhao, and T. Zhu, "Emotion recognition using Kinect motion capture data of human gaits," *PeerJ*, vol. 4, p. e2364, Sep. 2016.

[136] Y.-D. Zhang, Z.-J. Yang, H.-M. Lu, X.-X. Zhou, P. Phillips, Q.-M. Liu, and S.-H. Wang, "Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation," *IEEE Access*, vol. 4, pp. 8375–8385, 2016.

[137] S.-T. Liong, J. See, R. C.-W. Phan, Y.-H. Oh, A. Cat Le Ngo, K. Wong, and S.-W. Tan, "Spontaneous subtle expression detection and recognition based on facial strain," *Signal Process., Image Commun.*, vol. 47, pp. 170–182, Sep. 2016.

[138] H. Zheng, X. Geng, and Z. Yang, "A relaxed K-SVD algorithm for spontaneous micro-expression recognition," in *Proc. 14th Pacific Rim Int. Conf. Artif. Intell., PRICAI Trends Artif. Intell.*, Phuket, Thailand, Aug. 2016, pp. 692–699.

[139] A. C. Le Ngo, J. See, and R. C.-W. Phan, "Sparsity in dynamics of spontaneous subtle emotions: Analysis and application," *IEEE Trans. Affect. Comput.*, vol. 8, no. 3, pp. 396–411, Jul. 2017.

[140] A. M. Badshah, J. Ahmad, N. Rahim, and S. W. Baik, "Speech emotion recognition from spectrograms with deep convolutional neural network," in *Proc. Int. Conf. Platform Technol. Service (PlatCon)*, Feb. 2017, pp. 1–5.

[141] S. Mirsamadi, E. Barsoum, and C. Zhang, "Automatic speech emotion recognition using recurrent neural networks with local attention," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 2227–2231.

[142] Z. Meng, P. Liu, J. Cai, S. Han, and Y. Tong, "Identity-aware convolutional neural network for facial expression recognition," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2017, pp. 558–565.

[143] K. Zhang, Y. Huang, Y. Du, and L. Wang, "Facial expression recognition based on deep evolutional spatial–temporal networks," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4193–4203, Sep. 2017.

[144] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, "End-to-end multimodal emotion recognition using deep neural networks," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 8, pp. 1301–1309, Dec. 2017.

[145] D. Nguyen, K. Nguyen, S. Sridharan, A. Ghasemi, D. Dean, and C. Fookes, "Deep spatio-temporal features for multimodal emotion recognition," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 1215–1223.

[146] G. Yolcu, I. Oztel, S. Kazan, C. Oz, K. Palaniappan, T. E. Lever, and F. Bunyak, "Deep learning-based facial expression recognition for monitoring neurological disorders," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2017, pp. 1652–1657.

[147] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order," *Pattern Recognit.*, vol. 61, pp. 610–628, Jan. 2017.

[148] V. V. Salunke and C. G. Patil, "A new approach for automatic face emotion recognition and classification based on deep networks," in *Proc. Int. Conf. Comput., Commun., Control Autom. (ICCUBEA)*, Aug. 2017, pp. 1–5.

[149] L. Tan, K. Zhang, K. Wang, X. Zeng, X. Peng, and Y. Qiao, "Group emotion recognition with individual facial emotion CNNs and global image based CNNs," in *Proc. 19th ACM Int. Conf. Multimodal Interact.*, Nov. 2017, pp. 549–552.

[150] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, "Emotion recognition using facial expressions," *Proc. Comput. Sci.*, vol. 108, pp. 1175–1184, Aug. 2017.

[151] M. Peng, C. Wang, T. Chen, G. Liu, and X. Fu, "Dual temporal scale convolutional neural network for micro-expression recognition," *Frontiers Psychol.*, vol. 8, p. 1745, Oct. 2017.

[152] Q. Zhen, D. Huang, H. Drira, B. B. Amor, Y. Wang, and M. Daoudi, "Magnifying subtle facial motions for effective 4D expression recognition," *IEEE Trans. Affect. Comput.*, vol. 10, no. 4, pp. 524–536, Oct. 2019.
[153] M. Jazouli, A. Majda, and A. Zarghili, "A $P recognizer for automatic facial emotion recognition using Kinect sensor," in *Proc. Intell. Syst. Comput. Vis. (ISCV)*, Apr. 2017, pp. 1–5.
[154] Y. Luo, L. Zhang, Y. Chen, and W. Jiang, "Facial expression recognition algorithm based on reverse co-salient regions (RCSR) features," in *Proc. 4th Int. Conf. Inf. Sci. Control Eng. (ICISCE)*, Jul. 2017, pp. 326–329.
[155] R. Jiang, A. T. S. Ho, I. Cheheb, N. Al-Maadeed, S. Al-Maadeed, and A. Bouridane, "Emotion recognition from scrambled facial images via many graph embedding," *Pattern Recognit.*, vol. 67, pp. 245–251, Jul. 2017.
[156] M. S. Hossain and G. Muhammad, "An emotion recognition system for mobile applications," *IEEE Access*, vol. 5, pp. 2281–2287, 2017.
[157] V. P. Gonçalves, E. P. Costa, A. Valejo, G. P. R. Filho, T. M. Johnson, G. Pessin, and J. Ueyama, "Enhancing intelligence in multimodal emotion assessments," *Int. J. Speech Technol.*, vol. 46, no. 2, pp. 470–486, Mar. 2017.
[158] Y. Wang, J. See, Y.-H. Oh, R. C.-W. Phan, Y. Rahulamathavan, H.-C. Ling, S.-W. Tan, and X. Li, "Effective recognition of facial micro-expressions with video motion magnification," *Multimedia Tools Appl.*, vol. 76, no. 20, pp. 21665–21690, Oct. 2017.
[159] F. Xu, J. Zhang, and J. Z. Wang, "Microexpression identification and categorization using a facial dynamics map," *IEEE Trans. Affect. Comput.*, vol. 8, no. 2, pp. 254–267, Apr. 2017.
[160] X. Li, X. Hong, A. Moilanen, X. Huang, T. Pfister, G. Zhao, and M. Pietikäinen, "Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods," *IEEE Trans. Affect. Comput.*, vol. 9, no. 4, pp. 563–577, Oct. 2018.
[161] M. Chen, X. He, J. Yang, and H. Zhang, "3-D convolutional recurrent neural networks with attention model for speech emotion recognition," *IEEE Signal Process. Lett.*, vol. 25, no. 10, pp. 1440–1444, Oct. 2018.
[162] Z. Zhao, Y. Zhao, Z. Bao, H. Wang, Z. Zhang, and C. Li, "Deep spectrum feature representations for speech emotion recognition," in *Proc. Joint Workshop 4th Workshop Affect. Social Multimedia Comput. 1st Multi-Modal Affect. Comput. Large-Scale Multimedia Data*, Oct. 2018, pp. 27–33.
[163] S. Sahu, R. Gupta, G. Sivaraman, W. AbdAlmageed, and C. Espy-Wilson, "Adversarial auto-encoders for speech based emotion recognition," 2018, *arXiv:1806.02146*.
[164] J. Han, Z. Zhang, Z. Ren, F. Ringeval, and B. Schuller, "Towards conditional adversarial training for predicting emotions from speech," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 6822–6826.
[165] Z. Chen, D. Huang, Y. Wang, and L. Chen, "Fast and light manifold CNN based 3D facial expression recognition across pose variations," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 229–238.
[166] A. Kartali, M. Roglić, M. Barjaktarović, M. Ðurić-Jovicić, and M. M. Janković, "Real-time algorithms for facial emotion recognition: A comparison of different approaches," in *Proc. 14th Symp. Neural Netw. Appl. (NEUREL)*, Nov. 2018, pp. 1–4.
[167] D. Yang, A. Alsadoon, P. W. C. Prasad, A. K. Singh, and A. Elchouemi, "An emotion recognition model based on facial recognition in virtual learning environment," *Proc. Comput. Sci.*, vol. 125, pp. 2–10, Aug. 2018.
[168] R. V. Puthanidam and T.-S. Moh, "A hybrid approach for facial expression recognition," in *Proc. 12th Int. Conf. Ubiquitous Inf. Manage. Commun.*, Jan. 2018, pp. 1–8.
[169] I. Tautkute, T. Trzcinski, and A. Bielski, "I know how you feel: Emotion recognition with facial landmarks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 1878–1880.
[170] D. Y. Choi, D. H. Kim, and B. C. Song, "Recognizing fine facial micro-expressions using two-dimensional landmark feature," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 1962–1966.
[171] Y. Gan, "Facial expression recognition using convolutional neural network," in *Proc. 2nd Int. Conf. Vis., Image Signal Process.*, Aug. 2018, pp. 1–5.
[172] V. Khryashchev, L. Ivanovsky, and A. Priorov, "Deep learning for real-time robust facial expression analysis," in *Proc. Int. Conf. Mach. Vis. Appl.*, Apr. 2018, pp. 66–70.
[173] R. V. Darekar and A. P. Dhande, "Emotion recognition from Marathi speech database using adaptive artificial neural network," *Biologically Inspired Cognit. Archit.*, vol. 23, pp. 35–42, Jan. 2018.
[174] S. T. Ly, G.-S. Lee, S.-H. Kim, and H.-J. Yang, "Emotion recognition via body gesture: Deep learning model coupled with keyframe selection," in *Proc. Int. Conf. Mach. Learn. Mach. Intell.*, Sep. 2018, pp. 27–31.
[175] S.-H. Wang, P. Phillips, Z.-C. Dong, and Y.-D. Zhang, "Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm," *Neurocomputing*, vol. 272, pp. 668–676, Jan. 2018.
[176] A. Ruiz-Garcia, M. Elshaw, A. Altahhan, and V. Palade, "A hybrid deep learning neural approach for emotion recognition from facial expressions for socially assistive robots," *Neural Comput. Appl.*, vol. 29, no. 7, pp. 359–373, Apr. 2018.
[177] Z. Xia, X. Feng, X. Hong, and G. Zhao, "Spontaneous facial micro-expression recognition via deep convolutional network," in *Proc. 8th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Nov. 2018, pp. 1–6.
[178] H.-Q. Khor, J. See, R. C. W. Phan, and W. Lin, "Enriched long-term recurrent convolutional network for facial micro-expression recognition," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 667–674.
[179] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur, "X-vectors: Robust DNN embeddings for speaker recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 5329–5333.
[180] Y. Gu, K. Yang, S. Fu, S. Chen, X. Li, and I. Marsic, "Multimodal affective analysis using hierarchical attention strategy with word-level alignment," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, p. 2225.
[181] X. Wang, J. Huang, J. Zhu, M. Yang, and F. Yang, "Facial expression recognition with deep learning," in *Proc. 10th Int. Conf. Internet Multimedia Comput. Service*, 2018, pp. 1–4.
[182] P. Buitelaar et al., "MixedEmotions: An open-source toolbox for multimodal emotion analysis," *IEEE Trans. Multimedia*, vol. 20, no. 9, pp. 2454–2465, Sep. 2018.
[183] S.-T. Liong, J. See, R. C.-W. Phan, K. Wong, and S.-W. Tan, "Hybrid facial regions extraction for micro-expression recognition system," *J. Signal Process. Syst.*, vol. 90, no. 4, pp. 601–617, Apr. 2018.
[184] Y. Yao, D. Huang, X. Yang, Y. Wang, and L. Chen, "Texture and geometry scattering representation-based facial expression recognition in 2D+3D videos," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 1s, pp. 1–23, Mar. 2018.
[185] Y. Zong, X. Huang, W. Zheng, Z. Cui, and G. Zhao, "Learning from hierarchical spatiotemporal descriptors for micro-expression recognition," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3160–3172, Nov. 2018.
[186] K. Slimani, M. Kas, Y. El Merabet, R. Messoussi, and Y. Ruichek, "Facial emotion recognition: A comparative analysis using 22 LBP variants," in *Proc. 2nd Medit. Conf. Pattern Recognit. Artif. Intell.*, Mar. 2018, pp. 88–94.
[187] K. C. Kirana, S. Wibawanto, and H. W. Herwanto, "Facial emotion recognition based on Viola–Jones algorithm in the learning environment," in *Proc. Int. Seminar Appl. Technol. Inf. Commun.*, Sep. 2018, pp. 406–410.
[188] J. C. Quiroz, E. Geangu, and M. H. Yong, "Emotion recognition using smart watch sensor data: Mixed-design study," *JMIR Mental Health*, vol. 5, no. 3, Aug. 2018, Art. no. e10153.
[189] M. Chiu, J. Shu, and P. Hui, "Emotion recognition through gait on mobile devices," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, vol. 32, May 2018, pp. 800–805.
[190] F. Ahmed, B. Sieu, and M. L. Gavrilova, "Score and rank-level fusion for emotion recognition using genetic algorithm," in *Proc. IEEE 17th Int. Conf. Cognit. Informat. Cognit. Comput. (ICCI CC)*, Jul. 2018, pp. 46–53.
[191] S.-T. Liong, J. See, K. Wong, and R. C.-W. Phan, "Less is more: Micro-expression recognition from video using apex frame," *Signal Process., Image Commun.*, vol. 62, pp. 82–92, Mar. 2018.
[192] H. K. Palo and M. N. Mohanty, "Wavelet based feature combination for recognition of emotions," *Ain Shams Eng. J.*, vol. 9, no. 4, pp. 1799–1806, Dec. 2018.

[193] Z. Xia, X. Hong, X. Gao, X. Feng, and G. Zhao, "Spatiotemporal recurrent convolutional networks for recognizing spontaneous micro-expressions," *IEEE Trans. Multimedia*, vol. 22, no. 3, pp. 626–640, Mar. 2020.

[194] K. Ali and C. E. Hughes, "Facial expression recognition using disentangled adversarial learning," 2019, *arXiv:1909.13135*.

[195] M. S. Hossain and G. Muhammad, "Emotion recognition using deep learning approach from audio–visual emotional big data," *Inf. Fusion*, vol. 49, pp. 69–78, May 2018.

[196] M. S. Akhtar, D. S. Chauhan, D. Ghosal, S. Poria, A. Ekbal, and P. Bhattacharyya, "Multi-task learning for multi-modal emotion recognition and sentiment analysis," 2019, *arXiv:1905.05812*.

[197] R. Ranjan and B. C. Sahana, "An efficient facial feature extraction method based supervised classification model for human facial emotion identification," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2019, pp. 1–6.

[198] I. Lasri, A. R. Solh, and M. E. Belkacemi, "Facial emotion recognition of students using convolutional neural network," in *Proc. 3rd Int. Conf. Intell. Comput. Data Sci. (ICDS)*, Oct. 2019, pp. 1–6.

[199] D. K. Jain, P. Shamsolmoali, and P. Sehdev, "Extended deep neural network for facial emotion recognition," *Pattern Recognit. Lett.*, vol. 120, pp. 69–74, Apr. 2019.

[200] A. N. Navaz, S. M. Adel, and S. S. Mathew, "Facial image pre-processing and emotion classification: A deep learning approach," in *Proc. IEEE/ACS 16th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Nov. 2019, pp. 1–8.

[201] M. M. T. Zadeh, M. Imani, and B. Majidi, "Fast facial emotion recognition using convolutional neural networks and Gabor filters," in *Proc. 5th Conf. Knowl. Based Eng. Innov. (KBEI)*, Feb. 2019, pp. 577–581.

[202] A. Ghofrani, R. M. Toroghi, and S. Ghanbari, "Realtime face-detection and emotion recognition using MTCNN and miniShuffleNet V2," in *Proc. 5th Conf. Knowl. Based Eng. Innov. (KBEI)*, Feb. 2019, pp. 817–821.

[203] I. Oztel, G. Yolcu, and C. Oz, "Performance comparison of transfer learning and training from scratch approaches for deep facial expression recognition," in *Proc. 4th Int. Conf. Comput. Sci. Eng. (UBMK)*, Sep. 2019, pp. 1–6.

[204] M. Hu, H. Wang, X. Wang, J. Yang, and R. Wang, "Video facial emotion recognition based on local enhanced motion history image and CNN-CTSLSTM networks," *J. Vis. Commun. Image Represent.*, vol. 59, pp. 176–185, Feb. 2019.

[205] E. Avots, T. Sapiński, M. Bachmann, and D. Kamińska, "Audiovisual emotion recognition in wild," *Mach. Vis. Appl.*, vol. 30, no. 5, pp. 975–985, Jul. 2019.

[206] M. Chen and Y. Hao, "Label-less learning for emotion cognition," in *Proc. IEEE Trans. Neural Netw. Learn. Syst.*, Jul. 2020, vol. 31, no. 7, pp. 2430–2440.

[207] I. Kansizoglou, L. Bampis, and A. Gasteratos, "An active learning paradigm for online audio-visual emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 13, no. 2, pp. 756–768, Apr. 2022.

[208] J. Li, Y. Wang, J. See, and W. Liu, "Micro-expression recognition based on 3D flow convolutional neural network," *Pattern Anal. Appl.*, vol. 22, no. 4, pp. 1331–1339, Nov. 2019.

[209] G. Yolcu, I. Oztel, S. Kazan, C. Oz, K. Palaniappan, T. E. Lever, and F. Bunyak, "Facial expression recognition for monitoring neurological disorders based on convolutional neural network," *Multimedia Tools Appl.*, vol. 78, no. 22, pp. 31581–31603, Nov. 2019.

[210] Y. Li, T. Zhao, and T. Kawahara, "Improved end-to-end speech emotion recognition using self attention mechanism and multitask learning," in *Proc. Interspeech*, Sep. 2019, pp. 2803–2807.

[211] Y. Xie, R. Liang, Z. Liang, C. Huang, C. Zou, and B. Schuller, "Speech emotion classification using attention-based LSTM," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 11, pp. 1675–1685, Nov. 2019.

[212] Y. Liu, H. Du, L. Zheng, and T. Gedeon, "A neural micro-expression recognizer," in *Proc. 14th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2019, pp. 1–4.

[213] L. Zhou, Q. Mao, and L. Xue, "Dual-inception network for cross-database micro-expression recognition," in *Proc. 14th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2019, pp. 1–5.

[214] Y. S. Gan, S.-T. Liong, W.-C. Yau, Y.-C. Huang, and L.-K. Tan, "OFF-ApexNet on micro-expression recognition system," *Signal Process., Image Commun.*, vol. 74, pp. 129–139, May 2019.

[215] S.-T. Liong, Y. S. Gan, J. See, H.-Q. Khor, and Y.-C. Huang, "Shallow triple stream three-dimensional CNN (STSTNet) for micro-expression recognition," in *Proc. 14th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2019, pp. 1–5.

[216] M. Peng, C. Wang, T. Bi, Y. Shi, X. Zhou, and T. Chen, "A novel apex-time network for cross-dataset micro-expression recognition," in *Proc. 8th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Sep. 2019, pp. 1–6.

[217] N. V. Quang, J. Chun, and T. Tokuyama, "CapsuleNet for micro-expression recognition," in *Proc. 14th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2019, pp. 1–7.

[218] Y. Huang, K. Tian, A. Wu, and G. Zhang, "Feature fusion methods research based on deep belief networks for speech emotion recognition under noise condition," *J. Ambient Intell. Humanized Comput.*, vol. 10, no. 5, pp. 1787–1798, May 2019.

[219] L. Kerkeni, Y. Serrestou, K. Raoof, M. Mbarki, M. A. Mahjoub, and C. Cleder, "Automatic speech emotion recognition using an optimal combination of features based on EMD-TKEO," *Speech Commun.*, vol. 114, pp. 22–35, Nov. 2019.

[220] Y. Zhu, M. Tong, Z. Jiang, S. Zhong, and Q. Tian, "Hybrid feature-based analysis of video's affective content using protagonist detection," *Expert Syst. Appl.*, vol. 128, pp. 316–326, Aug. 2019.

[221] F. Makhmudkhujaev, M. Abdullah-Al-Wadud, M. T. B. Iqbal, B. Ryu, and O. Chae, "Facial expression recognition with local prominent directional pattern," *Signal Process., Image Commun.*, vol. 74, pp. 1–12, May 2019.

[222] M. S. Bouzakraoui, A. Sadiq, and A. Y. Alaoui, "Appreciation of customer satisfaction through analysis facial expressions and emotions recognition," in *Proc. 4th World Conf. Complex Syst. (WCCS)*, Apr. 2019, pp. 1–5.

[223] F. Wang, J. Lv, G. Ying, S. Chen, and C. Zhang, "Facial expression recognition from image based on hybrid features understanding," *J. Vis. Commun. Image Represent.*, vol. 59, pp. 84–88, Feb. 2019.

[224] J. Fu, Q. Mao, J. Tu, and Y. Zhan, "Multimodal shared features learning for emotion recognition by enhanced sparse local discriminative canonical correlation analysis," *Multimedia Syst.*, vol. 25, no. 5, pp. 451–461, Oct. 2019.

[225] K. P. Rao, M. V. P. C. S. Rao, and N. H. Chowdary, "An integrated approach to emotion recognition and gender classification," *J. Vis. Commun. Image Represent.*, vol. 60, pp. 339–345, Apr. 2019.

[226] S. T. Ly, N.-T. Do, G.-S. Lee, S.-H. Kim, and H.-J. Yang, "Multimodal 2D and 3D for in-the-wild facial expression recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 2927–2934.

[227] A. Alelaiwi, "Multimodal patient satisfaction recognition for smart healthcare," *IEEE Access*, vol. 7, pp. 174219–174226, 2019.

[228] X. Huang, S.-J. Wang, X. Liu, G. Zhao, X. Feng, and M. Pietikäinen, "Discriminative spatiotemporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition," *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, pp. 32–47, Jan. 2019.

[229] L. Lo, H.-X. Xie, H.-H. Shuai, and W.-H. Cheng, "MER-GCN: Micro-expression recognition based on relation modeling with graph convolutional networks," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, Aug. 2020, pp. 79–84.

[230] D. Liu, X. Ouyang, S. Xu, P. Zhou, K. He, and S. Wen, "SAANet: Siamese action-units attention network for improving dynamic facial expression recognition," *Neurocomputing*, vol. 413, pp. 145–157, Nov. 2020.

[231] Y. Fu, X. Wu, X. Li, Z. Pan, and D. Luo, "Semantic neighborhood-aware deep facial expression recognition," *IEEE Trans. Image Process.*, vol. 29, pp. 6535–6548, 2020.

[232] S. Zhao, Y. Ma, Y. Gu, J. Yang, T. Xing, P. Xu, R. Hu, H. Chai, and K. Keutzer, "An end-to-end visual-audio attention network for emotion recognition in user-generated videos," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 1, pp. 303–311.

[233] M. Hao, W.-H. Cao, Z.-T. Liu, M. Wu, and P. Xiao, "Visual-audio emotion recognition based on multi-task and ensemble learning with multiple features," *Neurocomputing*, vol. 391, pp. 42–51, May 2020.

[234] D. Priyasad, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Attention driven fusion for multi-modal emotion recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 3227–3231.

[235] L. Pepino, P. Riera, L. Ferrer, and A. Gravano, "Fusion approaches for emotion recognition from speech using acoustic and text-based features," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 6484–6488.

[236] T. Mittal, U. Bhattacharya, R. Chandra, A. Bera, and D. Manocha, "M3ER: Multiplicative multimodal emotion recognition using facial, textual, and speech cues," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 2, pp. 1359–1367.

[237] E. Pranav, S. Kamal, C. S. Chandran, and M. H. Supriya, "Facial emotion recognition using deep convolutional neural network," in *Proc. 6th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Mar. 2020, pp. 317–320.

[238] A. Jaiswal, A. K. Raju, and S. Deb, "Facial emotion detection using deep learning," in *Proc. Int. Conf. Emerg. Technol. (INCET)*, Jun. 2020, pp. 1–5.

[239] N. Naik and M. A. Mehta, "An improved method to recognize hand-over-face gesture based facial emotion using convolutional neural network," in *Proc. IEEE Int. Conf. Electron., Comput. Commun. Technol. (CONECCT)*, Jul. 2020, pp. 1–6.

[240] S. Begaj, A. O. Topal, and M. Ali, "Emotion recognition based on facial expressions using convolutional neural network (CNN)," in *Proc. Int. Conf. Comput., Netw., Telecommun. Eng. Sci. Appl. (CoNTESA)*, Dec. 2020, pp. 58–63.

[241] S. Liu, D. Li, Q. Gao, and Y. Song, "Facial emotion recognition based on CNN," in *Proc. Chin. Autom. Congr. (CAC)*, Nov. 2020, pp. 398–403.

[242] S. Engoor, S. SendhilKumar, C. H. Sharon, and G. S. Mahalakshmi, "Occlusion-aware dynamic human emotion recognition using landmark detection," in *Proc. 6th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Mar. 2020, pp. 795–799.

[243] A. John, M. Abhishek, A. S. Ajayan, S. Sanoop, and V. R. Kumar, "Real-time facial emotion recognition system with improved preprocessing and feature extraction," in *Proc. 3rd Int. Conf. Smart Syst. Inventive Technol. (ICSSIT)*, Aug. 2020, pp. 1328–1333.

[244] L. Cai, J. Dong, and W. Min, "Multi-modal emotion recognition from speech and facial expression based on deep learning," in *Proc. Chin. Autom. Congr. (CAC)*, Aug. 2020, pp. 5726–5729.

[245] G. Wei, L. Jian, and S. Mo, "Multimodal(audio, facial and gesture) based emotion recognition challenge," in *Proc. 15th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Nov. 2020, pp. 908–911.

[246] Z. Lian, Y. Li, J.-H. Tao, J. Huang, and M.-Y. Niu, "Expression analysis based on face regions in real-world conditions," *Int. J. Autom. Comput.*, vol. 17, no. 1, pp. 96–107, Feb. 2020.

[247] E. Pei, D. Jiang, and H. Sahli, "An efficient model-level fusion approach for continuous affect recognition from audiovisual signals," *Neurocomputing*, vol. 376, pp. 42–53, Feb. 2020.

[248] S. Rasipuram, J. H. Bhat, and A. Maitra, "Multi-modal expression recognition in the wild using sequence modeling," in *Proc. 15th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Nov. 2020, pp. 629–631.

[249] J. Comas, D. Aspandi, and X. Binefa, "End-to-end facial and physiological model for affective computing and applications," in *Proc. 15th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Nov. 2020, pp. 93–100.

[250] T. S. Ashwin and R. M. R. Guddeti, "Impact of inquiry interventions on students in e-learning and classroom environments using affective computing framework," *User Model. User-Adapted Interact.*, vol. 30, no. 5, pp. 759–801, Nov. 2020.

[251] A. Agrawal and N. Mittal, "Using CNN for facial expression recognition: A study of the effects of kernel size and number of filters on accuracy," *Vis. Comput.*, vol. 36, no. 2, pp. 405–412, Feb. 2020.

[252] D. K. Jain, Z. Zhang, and K. Huang, "Multi angle optimal pattern-based deep learning for automatic facial expression recognition," *Pattern Recognit. Lett.*, vol. 139, pp. 157–165, Nov. 2020.

[253] D. Issa, M. F. Demirci, and A. Yazici, "Speech emotion recognition with deep convolutional neural networks," *Biomed. Signal Process. Control*, vol. 59, May 2020, Art. no. 101894.

[254] A. Caroppo, A. Leone, and P. Siciliano, "Comparison between deep learning models and traditional machine learning approaches for facial expression recognition in ageing adults," *J. Comput. Sci. Technol.*, vol. 35, no. 5, pp. 1127–1146, Oct. 2020.

[255] C. Wang, M. Peng, T. Bi, and T. Chen, "Micro-attention for micro-expression recognition," *Neurocomputing*, vol. 410, pp. 354–362, Oct. 2020.

[256] H.-X. Xie, L. Lo, H.-H. Shuai, and W.-H. Cheng, "AU-assisted graph attention convolutional network for micro-expression recognition," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 2871–2880.

[257] S. Zhou and H. Beigi, "A transfer learning method for speech emotion recognition from automatic speech recognition," 2020, *arXiv:2008.02863*.

[258] L. Chen, W. Su, Y. Feng, M. Wu, J. She, and K. Hirota, "Two-layer fuzzy multiple random forest for speech emotion recognition in human–robot interaction," *Inf. Sci.*, vol. 509, pp. 150–163, Jan. 2020.

[259] Y. Yan, Z. Zhang, S. Chen, and H. Wang, "Low-resolution facial expression recognition: A filter learning perspective," *Signal Process.*, vol. 169, Apr. 2020, Art. no. 107370.

[260] W. Wei, Q. Jia, Y. Feng, G. Chen, and M. Chu, "Multi-modal facial expression feature based on deep-neural networks," *J. Multimodal User Interfaces*, vol. 14, no. 1, pp. 17–23, Mar. 2020.

[261] P. Zarbakhsh and H. Demirel, "4D facial expression recognition using multimodal time series analysis of geometric landmark-based deformations," *Vis. Comput.*, vol. 36, no. 5, pp. 951–965, May 2020.

[262] E. Agrawal and J. Christopher, "Emotion recognition from periocular features," in *Proc. Mach. Learn.*, Silchar, India, Jul. 2020, pp. 194–208.

[263] S. Nagarajan, S. S. S. Nettimi, L. S. Kumar, M. K. Nath, and A. Kanhe, "Speech emotion recognition using cepstral features extracted with novel triangular filter banks based on bark and ERB frequency scales," *Digit. Signal Process.*, vol. 104, Sep. 2020, Art. no. 102763.

[264] D. Gera and S. Balasubramanian, "Landmark guidance independent spatio-channel attention and complementary context information based facial expression recognition," *Pattern Recognit. Lett.*, vol. 145, pp. 58–66, May 2021.

[265] M. Behzad, N. Vo, X. Li, and G. Zhao, "Towards reading beyond faces for sparsity-aware 3D/4D affect recognition," *Neurocomputing*, vol. 458, pp. 297–307, Oct. 2021.

[266] J. Yu, C. Zhang, Y. Song, and W. Cai, "ICE-GAN: Identity-aware and capsule-enhanced GAN with graph-based reasoning for micro-expression recognition and synthesis," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2021, pp. 1–8.

[267] D. Krishnani, P. Shivakumara, T. Lu, U. Pal, D. Lopresti, and G. H. Kumar, "A new context-based feature for classification of emotions in photographs," *Multimedia Tools Appl.*, vol. 80, no. 10, pp. 15589–15618, Apr. 2021.

[268] L.-N. Do, H.-J. Yang, H.-D. Nguyen, S.-H. Kim, G.-S. Lee, and I.-S. Na, "Deep neural network-based fusion model for emotion recognition using visual data," *J. Supercomput.*, vol. 77, no. 10, pp. 1–18, Oct. 2021.

[269] D. Nguyen, D. T. Nguyen, R. Zeng, T. T. Nguyen, S. N. Tran, T. Nguyen, S. Sridharan, and C. Fookes, "Deep auto-encoders with sequential learning for multimodal dimensional emotion recognition," *IEEE Trans. Multimedia*, vol. 24, pp. 1313–1324, 2022.

[270] E. M. Benssassi and J. Ye, "Investigating multisensory integration in emotion recognition through bio-inspired computational models," *IEEE Trans. Affect. Comput.*, vol. 14, no. 2, pp. 906–918, Apr. 2023.

[271] J. He, X. Yu, B. Sun, and L. Yu, "Facial expression and action unit recognition augmented by their dependencies on graph convolutional networks," *J. Multimodal User Interfaces*, vol. 15, no. 4, pp. 429–440, Dec. 2021.

[272] S. Aslan, U. Güdükbay, and H. Dibeklioğlu, "Multimodal assessment of apparent personality using feature attention and error consistency constraint," *Image Vis. Comput.*, vol. 110, Jun. 2021, Art. no. 104163.

[273] M. G. Huddar, S. S. Sannakki, and V. S. Rajpurohit, "Attention-based multi-modal sentiment analysis and emotion detection in conversation using RNN," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 6, no. 6, pp. 112–121, Jun. 2021.

[274] X. Huang, M. Ren, Q. Han, X. Shi, J. Nie, W. Nie, and A.-A. Liu, "Emotion detection for conversations based on reinforcement learning framework," *IEEE MultimediaMag.*, vol. 28, no. 2, pp. 76–85, Apr. 2021.

[275] M. Liu and J. Tang, "Audio and video bimodal emotion recognition in social networks based on improved alexnet network and attention mechanism," *J. Inf. Process. Syst.*, vol. 17, no. 4, pp. 754–771, 2021.

[276] R.-H. Huan, J. Shu, S.-L. Bao, R.-H. Liang, P. Chen, and K.-K. Chi, "Video multimodal emotion recognition based on bi-GRU and attention fusion," *Multimedia Tools Appl.*, vol. 80, no. 6, pp. 8213–8240, Mar. 2021.

[277] X. Liu, H. Shi, H. Chen, Z. Yu, X. Li, and G. Zhao, "IMiGUE: An identity-free video dataset for micro-gesture understanding and emotion analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10626–10637.
[278] C. Wu and F. Guo, "TSNN: Three-stream combining 2D and 3D convolutional neural network for micro-expression recognition," *IEEJ Trans. Electr. Electron. Eng.*, vol. 16, no. 1, pp. 98–107, Jan. 2021.
[279] P. Kumawat and A. Routray, "Applying TDNN architectures for analyzing duration dependencies on speech emotion recognition," in *Proc. Interspeech*, Aug. 2021, pp. 3410–3414.
[280] C. Luna-Jiménez, R. Kleinlein, D. Griol, Z. Callejas, J. M. Montero, and F. Fernández-Martínez, "A proposal for multimodal emotion recognition using aural transformers and action units on RAVDESS dataset," *Appl. Sci.*, vol. 12, no. 1, p. 327, Dec. 2021.
[281] G. Agarwal and H. Om, "Performance of deer hunting optimization based deep learning algorithm for speech emotion recognition," *Multimedia Tools Appl.*, vol. 80, no. 7, pp. 9961–9992, Jun. 2021.
[282] N. Hajarolasvadi, E. Bashirov, and H. Demirel, "Video-based person-dependent and person-independent facial emotion recognition," *Signal, Image Video Process.*, vol. 15, no. 5, pp. 1049–1056, Jul. 2021.
[283] J. Wei, X. Yang, and Y. Dong, "User-generated video emotion recognition based on key frames," *Multimedia Tools Appl.*, vol. 80, no. 9, pp. 14343–14361, Apr. 2021.
[284] Z. Farhoudi and S. Setayeshi, "Fusion of deep learning features with mixture of brain emotional learning for audio-visual emotion recognition," *Speech Commun.*, vol. 127, pp. 92–103, Sep. 2020.
[285] M. Murugappan and A. Mutawa, "Facial geometric feature extraction based emotional expression classification using machine learning algorithms," *PLoS ONE*, vol. 16, no. 2, Feb. 2021, Art. no. e0247131.
[286] J. M. López-Gil and N. Garay-Vitoria, "Photogram classification-based emotion recognition," *IEEE Access*, vol. 9, pp. 136974–136984, 2021.
[287] A. I. Middya, B. Nag, and S. Roy, "Deep learning based multimodal emotion recognition using model-level fusion of audio–visual modalities," *Knowl.-Based Syst.*, vol. 244, May 2022, Art. no. 108580.
[288] H. Filali, J. Riffi, I. Aboussaleh, A. M. Mahraz, and H. Tairi, "Meaningful learning for deep facial emotional features," *Neural Process. Lett.*, vol. 54, no. 1, pp. 387–404, Feb. 2022.
[289] M. Sharafi, M. Yazdchi, R. Rasti, and F. Nasimi, "A novel spatio-temporal convolutional neural framework for multimodal emotion recognition," *Biomed. Signal Process. Control*, vol. 78, Sep. 2022, Art. no. 103970.
[290] C. Zhu, T. Ding, and X. Min, "Emotion recognition of college students based on audio and video image," *Traitement du Signal*, vol. 39, no. 5, pp. 1475–1481, Nov. 2022.
[291] S. Chamishka, I. Madhavi, R. Nawaratne, D. Alahakoon, D. De Silva, N. Chilamkurti, and V. Nanayakkara, "A voice-based real-time emotion detection technique using recurrent neural network empowered feature modelling," *Multimedia Tools Appl.*, vol. 81, no. 24, pp. 35173–35194, Oct. 2022.
[292] B. Yalamanchili, S. K. Samayamantula, and K. R. Anne, "Neural network-based blended ensemble learning for speech emotion recognition," *Multidimensional Syst. Signal Process.*, vol. 33, no. 4, pp. 1323–1348, Dec. 2022.
[293] L.-Y. Liu, W.-Z. Liu, J. Zhou, H.-Y. Deng, and L. Feng, "ATDA: Attentional temporal dynamic activation for speech emotion recognition," *Knowl.-Based Syst.*, vol. 243, May 2022, Art. no. 108472.
[294] N. Kim, S. Cho, and B. Bae, "SMaTE: A segment-level feature mixing and temporal encoding framework for facial expression recognition," *Sensors*, vol. 22, no. 15, p. 5753, Aug. 2022.
[295] C. Wang, Y. Ren, N. Zhang, F. Cui, and S. Luo, "Speech emotion recognition based on multi-feature and multi-lingual fusion," *Multimedia Tools Appl.*, vol. 81, no. 4, pp. 4897–4907, 2021.
[296] Q. Sun, L. Liang, X. Dang, and Y. Chen, "Deep learning-based dimensional emotion recognition combining the attention mechanism and global second-order feature representations," *Comput. Electr. Eng.*, vol. 104, Dec. 2022, Art. no. 108469.
[297] A. A. Albraikan, J. S. Alzahrani, R. Alshahrani, A. Yafoz, R. Alsini, A. M. Hilal, A. Alkhayyat, and D. Gupta, "Intelligent facial expression recognition and classification using optimal deep transfer learning model," *Image Vis. Comput.*, vol. 128, Dec. 2022, Art. no. 104583.
[298] R. Hammouche, A. Attia, S. Akhrouf, and Z. Akhtar, "Gabor filter bank with deep autoencoder based face recognition system," *Expert Syst. Appl.*, vol. 197, Jul. 2022, Art. no. 116743.
[299] S. Zaghbani and M. S. Bouhlel, "Multi-task CNN for multi-cue affects recognition using upper-body gestures and facial expressions," *Int. J. Inf. Technol.*, vol. 14, no. 1, pp. 531–538, Feb. 2022.
[300] A. A. Abdelhamid, E.-S.-M. El-Kenawy, B. Alotaibi, G. M. Amer, M. Y. Abdelkader, A. Ibrahim, and M. M. Eid, "Robust speech emotion recognition using CNN+LSTM based on stochastic fractal search optimization algorithm," *IEEE Access*, vol. 10, pp. 49265–49284, 2022.
[301] 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.*, vol. 12, no. 1, p. 527, Jan. 2022.
[302] J. Wu, Y. Zhang, S. Sun, Q. Li, and X. Zhao, "Generalized zero-shot emotion recognition from body gestures," *Appl. Intell.*, vol. 52, no. 8, pp. 1–19, Jun. 2022.
[303] E. Morais, R. Hoory, W. Zhu, I. Gat, M. Damasceno, and H. Aronowitz, "Speech emotion recognition using self-supervised features," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 6922–6926.
[304] H. M. Shahzad, S. M. Bhatti, A. Jaffar, and M. Rashid, "A multi-modal deep learning approach for emotion recognition," *Intell. Autom. Soft Comput.*, vol. 36, no. 2, pp. 1561–1570, 2023.
[305] G. K. Sahoo, S. K. Das, and P. Singh, "Performance comparison of facial emotion recognition: A transfer learning-based driver assistance framework for in-vehicle applications," *Circuits, Syst., Signal Process.*, vol. 42, no. 7, pp. 4292–4319, Jul. 2023.
[306] S. Kumar, V. Sagar, and D. Punetha, "A comparative study on facial expression recognition using local binary patterns, convolutional neural network and frequency neural network," *Multimedia Tools Appl.*, vol. 82, no. 16, pp. 24369–24385, Jul. 2023.
[307] W. Guo, X. Zhao, S. Zhang, and X. Pan, "Learning inter-class optical flow difference using generative adversarial networks for facial expression recognition," *Multimedia Tools Appl.*, vol. 82, no. 7, pp. 10099–10116, Mar. 2023.
[308] K. L. Lakshmi, P. Muthulakshmi, A. A. Nithya, R. B. Jeyavathana, R. Usharani, N. S. Das, and G. N. R. Devi, "Recognition of emotions in speech using deep CNN and RESNET," *Soft Comput.*, pp. 1–17, Mar. 2023. [Online]. Available: https://doi.org/10.1007/s00500-023-07969-5
[309] S. C. Pravin, V. B. Sivaraman, and J. Saranya, "Deep ensemble models for speech emotion classification," *Microprocessors Microsystems*, vol. 98, Apr. 2023, Art. no. 104790.
[310] D. Nguyen, D. T. Nguyen, S. Sridharan, S. Denman, T. T. Nguyen, D. Dean, and C. Fookes, "Meta-transfer learning for emotion recognition," *Neural Comput. Appl.*, vol. 35, no. 14, pp. 10535–10549, May 2023.
[311] G. Tang, Y. Xie, K. Li, R. Liang, and L. Zhao, "Multimodal emotion recognition from facial expression and speech based on feature fusion," *Multimedia Tools Appl.*, vol. 82, no. 11, pp. 16359–16373, May 2023.
[312] G. Castellano, B. De Carolis, and N. Macchiarulo, "Automatic facial emotion recognition at the COVID-19 pandemic time," *Multimedia Tools Appl.*, vol. 82, no. 9, pp. 12751–12769, Apr. 2023.
[313] K. Chauhan, K. K. Sharma, and T. Varma, "Improved speech emotion recognition using channel-wise global head pooling (CwGHP)," *Circuits, Syst., Signal Process.*, vol. 42, no. 9, pp. 5500–5522, Sep. 2023.
[314] E. Ghaleb, J. Niehues, and S. Asteriadis, "Joint modelling of audio-visual cues using attention mechanisms for emotion recognition," *Multimedia Tools Appl.*, vol. 82, no. 8, pp. 11239–11264, Mar. 2023.
[315] H. Gao, M. Wu, Z. Chen, Y. Li, X. Wang, S. An, J. Li, and C. Liu, "SSA-ICL: Multi-domain adaptive attention with intra-dataset continual learning for facial expression recognition," *Neural Netw.*, vol. 158, pp. 228–238, Jan. 2023.
[316] S. Hossain, S. Umer, R. K. Rout, and M. Tanveer, "Fine-grained image analysis for facial expression recognition using deep convolutional neural networks with bilinear pooling," *Appl. Soft Comput.*, vol. 134, Feb. 2023, Art. no. 109997.
[317] B. Mocanu, R. Tapu, and T. Zaharia, "Multimodal emotion recognition using cross modal audio-video fusion with attention and deep metric learning," *Image Vis. Comput.*, vol. 133, May 2023, Art. no. 104676.
[318] F. M. Alamgir and M. S. Alam, "Hybrid multi-modal emotion recognition framework based on InceptionV3DenseNet," *Multimedia Tools Appl.*, vol. 82, no. 26, pp. 40375–40402, Nov. 2023.

[319] L. Chen, M. Li, M. Wu, W. Pedrycz, and K. Hirota, "Coupled multimodal emotional feature analysis based on broad-deep fusion networks in human–robot interaction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 7, pp. 9663–9673, Jul. 2024.
[320] S. N. Khanbebin and V. Mehrdad, "Improved convolutional neural network-based approach using hand-crafted features for facial expression recognition," *Multimedia Tools Appl.*, vol. 82, no. 8, pp. 11489–11505, Mar. 2023.
[321] H. Boughanem, H. Ghazouani, and W. Barhoumi, "Multichannel convolutional neural network for human emotion recognition from in-the-wild facial expressions," *Vis. Comput.*, vol. 39, no. 11, pp. 5693–5718, Nov. 2023.
[322] M. R. Ahmed, S. Islam, A. K. M. M. Islam, and S. Shatabda, "An ensemble 1D-CNN-LSTM-GRU model with data augmentation for speech emotion recognition," *Expert Syst. Appl.*, vol. 218, May 2023, Art. no. 119633.
[323] S. Sekkate, M. Khalil, and A. Adib, "A statistical feature extraction for deep speech emotion recognition in a bilingual scenario," *Multimedia Tools Appl.*, vol. 82, no. 8, pp. 11443–11460, Mar. 2023.
[324] H. M. Shahzad, S. M. Bhatti, A. Jaffar, M. Rashid, and S. Akram, "Multi-modal CNN features fusion for emotion recognition: A modified xception model," *IEEE Access*, vol. 11, pp. 94281–94289, 2023.
[325] A. Aguilera, D. Mellado, and F. Rojas, "An assessment of in-the-wild datasets for multimodal emotion recognition," *Sensors*, vol. 23, no. 11, p. 5184, May 2023.
[326] Y. Lei and H. Cao, "Audio-visual emotion recognition with preference learning based on intended and multi-modal perceived labels," *IEEE Trans. Affect. Comput.*, vol. 14, no. 4, pp. 2954–2969, Oct. 2023.
[327] N. Banskota, A. Alsadoon, P. W. C. Prasad, A. Dawoud, T. A. Rashid, and O. H. Alsadoon, "A novel enhanced convolution neural network with extreme learning machine: Facial emotional recognition in psychology practices," *Multimedia Tools Appl.*, vol. 82, no. 5, pp. 6479–6503, Feb. 2023.
[328] B. Pan, K. Hirota, Z. Jia, L. Zhao, X. Jin, and Y. Dai, "Multimodal emotion recognition based on feature selection and extreme learning machine in video clips," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 3, pp. 1903–1917, Mar. 2023.
[329] Y. Yaddaden, "An efficient facial expression recognition system with appearance-based fused descriptors," *Intell. Syst. Appl.*, vol. 17, Feb. 2023, Art. no. 200166.
[330] M. Dirik, "Optimized anfis model with hybrid metaheuristic algorithms for facial emotion recognition," *Int. J. Fuzzy Syst.*, vol. 25, no. 2, pp. 485–496, Mar. 2023.
[331] H. Arabian, T. A. Alshirbaji, J. G. Chase, and K. Moeller, "Emotion recognition beyond pixels: Leveraging facial point landmark meshes," *Appl. Sci.*, vol. 14, no. 8, p. 3358, Apr. 2024.
[332] H.-J. Nam and H.-J. Park, "Speech emotion recognition under noisy environments with SNR down to −6 dB using multi-decoder wave-U-Net," *Appl. Sci.*, vol. 14, no. 12, p. 5227, Jun. 2024.
[333] E. A. Alkhamali, A. Allinjawi, and R. B. Ashari, "Combining transformer, convolutional neural network, and long short-term memory architectures: A novel ensemble learning technique that leverages multi-acoustic features for speech emotion recognition in distance education classrooms," *Appl. Sci.*, vol. 14, no. 12, p. 5050, Jun. 2024.
[334] N. Simić, S. Suzić, N. Milošević, V. Stanojev, T. Nosek, B. Popović, and D. Bajović, "Enhancing emotion recognition through federated learning: A multimodal approach with convolutional neural networks," *Appl. Sci.*, vol. 14, no. 4, p. 1325, Feb. 2024.
[335] D. Kang, D. Kim, D. Kang, T. Kim, B. Lee, D. Kim, and B. C. Song, "Beyond superficial emotion recognition: Modality-adaptive emotion recognition system," *Expert Syst. Appl.*, vol. 235, Jan. 2024, Art. no. 121097.
[336] X. Han, F. Chen, and J. Ban, "FMFN: A fuzzy multimodal fusion network for emotion recognition in ensemble conducting," *IEEE Trans. Fuzzy Syst.*, vol. 33, no. 1, pp. 168–179, Jan. 2025.
[337] S. Hazmoune and F. Bougamouza, "Using transformers for multimodal emotion recognition: Taxonomies and state of the art review," *Eng. Appl. Artif. Intell.*, vol. 133, Jul. 2024, Art. no. 108339.
[338] P. V. V. Paiva, J. J. G. Ramos, M. Gavrilova, and M. A. G. D. Carvalho, "SkelETT—Skeleton-to-emotion transfer transformer," *IEEE Access*, vol. 13, pp. 23344–23358, 2025.
[339] B. Zhao, M. Gong, and X. Li, "Hierarchical multimodal transformer to summarize videos," *Neurocomputing*, vol. 468, pp. 360–369, Jan. 2022.

**PEDRO J. VAZ** received the B.Sc. degree in electrical and electronics engineering and the M.Sc. degree in electrical and computer engineering from Universidade do Algarve, Faro, Portugal, in 2009 and 2022, respectively, where he is currently pursuing the Ph.D. degree in informatics engineering. From 2009 to 2010, he was a Software Engineer with the Maine Avenue Technologies, Madrid, Spain. From 2011 to 2015, he was an Electrical Engineer and a Senior Electronics Design Engineer with Stanley Black and Decker GmbH, Idstein, Germany, from 2015 to 2016. From 2018 to 2021, he was a High School Teacher, lecturing classes in computer science, electricity, and electronics. From 2021 to 2023, he was a Researcher under the Project Green Spaces SMART Irrigation Control (GSSIC), from 2023 to 2024. He has been a Ph.D. Scholarship Holder, under the Sustainable Horizons European Universities (SHEs) Program. He has been an FCT Ph.D. Grant Holder, since 2024. Since 2021, he has been an Invited Assistant with Universidade do Algarve, lecturing on computer programming classes. Starting in January 2024, he joined the Research Centre NOVA LINCS, Algarve. His research interests include machine learning, affective computing, data science, electronics, and audio signal processing. Details at https://tinyurl.com/554krh5n

**JOÃO M. F. RODRIGUES** received the bachelor's degree in electrical engineering, the master's degree in systems and computer engineering, the Ph.D. degree in electronic and computer engineering with a specialty in computer science, and the Habilitation degree in electrical and computer engineering, in 1993, 1998, 2008, and 2021, respectively. He is currently the Vice-Rector for Transfer, Innovation, and Digital University, Universidade do Algarve (UAlg), Portugal. He is a Coordinator Professor with Tenure, Institute of Engineering (ISE), UAlg. Since 1994, he has taught curricular units in computer science and computer vision. In addition to his academic and administrative roles, he is a member of the Research Centre NOVA LINCS, Algarve. He has extensive research experience, having participated in more than 30 nationally or internationally funded scientific projects, several of which he served as a coordinator. He has co-authored more than 200 scientific articles and is a member of the editorial board of several international journals. His main research interests include computer vision, human-computer interaction, and human-centered artificial intelligence and affective computing, focusing on adaptive interfaces, emotion recognition, human-sense modeling, and human–machine cooperation. Details at https://tinyurl.com/56fvsuwb

**PEDRO J. S. CARDOSO** received the B.Sc. degree in mathematics/computer science from the University of Coimbra, Portugal, the M.Sc. degree in computational mathematics from the University of Minho, Portugal, and the Ph.D. degree in mathematics/operations research from the University of Seville, Spain. He has been a Coordinator Professor and a Researcher with Universidade do Algarve for more than two decades. He is also an Integrated Member of the NOVA LINCS Research Center. His extensive expertise in computer science has led him to specialize in machine learning, data science, and multiple objective meta-heuristics. He has a long-standing career as a Lecturer in these fields. Throughout his career, he has been involved in more than a dozen scientific and development projects and has published more than 90 articles. Additionally, he has edited more than a dozen books. His primary research interests include machine learning, particularly its real-world applications, affective computing, anomaly detection, and energy efficiency. Details at https://tinyurl.com/26hdrntd