Vol. 4 No. (1) (June 2026) 197-224

Journal of Data Insights

http://journalnew.unimus.ac.id/index.php/jodi

Systematic Review of Multimodal Emotion Recognition in the Wild: Integrating Facial Expressions, Speech, and Physiological Signals for Enhanced Context-Aware Applications

Muhammad Munsarif*[1], Norshuhani Zamin[2], Richmond Ampah-Mensah[3]
[1]Department of Informatics, Universitas Muhammadiyah Semarang, Indonesia
[2]College of Computing and Informatics, Institute of Informatics and Computing in Energy (IICE), Universiti Tenaga Nasional (UNITEN), Malaysia
[3]Department of Informatics, Universitas Muhammadiyah Semarang, Indonesia

DOI: https://doi.org/10.26714/jodi.v4i1.1197

Article Info

Article history:
Received June 04, 2026
Revised June 25, 2026
Accepted June 27, 2026

Keywords:
Facial Emotion Recognition; Multimodal Emotion Recognition; Deep Learning; Transfer Learning; Affective Computing; Human–Computer Interaction; Explainable AI.

Abstract

Facial Emotion Recognition (FER) has become an essential component of affective computing, human–computer interaction, healthcare, adaptive education, and assistive technologies. This systematic literature review synthesizes recent advances in multimodal emotion recognition in the wild by examining the integration of facial expressions, speech, physiological signals, deep learning architectures, and deployment technologies for robust context-aware FER systems. Following the PRISMA protocol, literature was retrieved from Scopus using FER-related deep learning keywords, yielding 202 initial records and 61 eligible studies for qualitative synthesis and methodological analysis. The findings indicate a clear transition from handcrafted features to deep learning approaches, particularly convolutional neural networks, transfer learning, hybrid architectures, attention mechanisms, and bio-inspired optimization techniques. Human–computer interaction remains the dominant application area, while healthcare, autism screening, education, assistive technologies, mining safety, and smart services are emerging research domains. Multimodal fusion combining facial images, speech, EEG, wearable sensors, and audio-visual signals has significantly improved contextual understanding and system robustness compared with unimodal approaches. Nevertheless, important challenges remain, including limited generalization in uncontrolled environments, dataset imbalance, cultural variability, micro-expression recognition, computational complexity, real-time implementation, and explainability. Future research should emphasize lightweight multimodal models, cross-cultural datasets, explainable AI, privacy-preserving learning, and comprehensive real-world validation.

✉ Correspondence Address:
E-mail: m.munsarif@unimus.ac.id

e-ISSN: 2988 - 2109

INTRODUCTION

Emotion recognition has become a rapidly growing research area in affective computing, artificial intelligence, computer vision, and human–computer interaction because it enables intelligent systems to identify and respond to human emotional states. Facial emotion recognition (FER) remains one of the most widely studied approaches because facial expressions provide rich visual cues for understanding emotions, intentions, and nonverbal communication. FER has been applied in healthcare, education, intelligent transportation, assistive technology, customer service, and human–robot interaction, where emotion-aware systems can improve personalization, safety, interaction quality, and decision-making (Akhand et al., 2021; Khan, 2022; Talaat, 2023).

The growth of this field is evident in the present systematic review, which identified 202 initial records from Scopus and retained 61 eligible studies for final synthesis. The publication trend shows a clear increase after 2020, with the highest number of reviewed studies appearing in 2023. Keyword analysis also indicates the dominance of deep learning-based approaches, with "Deep Learning," "Emotion Recognition," and "Facial Emotion Recognition" appearing as the most frequent terms. These patterns confirm that FER has evolved from conventional facial analysis into a mature research domain driven by CNNs, transfer learning, Vision Transformers, hybrid architectures, and context-aware intelligent systems.

Despite this progress, facial-only emotion recognition still faces major limitations in real-world environments. Facial expressions can be affected by illumination changes, pose variations, occlusions, low image quality, facial masks, spontaneous expressions, and cultural differences in emotional display (Khan, 2022; Mukhiddinov et al., 2023; Hongwei et al., 2025). These limitations reduce model generalization when systems are deployed outside controlled laboratory conditions. Therefore, high accuracy on benchmark datasets such as FER2013, CK+, JAFFE, KDEF, and RAF-DB does not always guarantee reliable performance in unconstrained settings.

Multimodal emotion recognition (MER) has emerged as a promising solution to these limitations by combining facial expressions with complementary signals such as speech, EEG, ECG, wearable sensors, gaze, posture, and audio-visual cues. Speech provides affective information through pitch, tone, rhythm, and intensity, while EEG captures neurophysiological responses that are less dependent on facial visibility. ECG and wearable sensor data reflect autonomic arousal and can support continuous emotion monitoring in healthcare, education, and safety-related applications. By integrating visual, vocal, and physiological cues, MER can compensate for the weaknesses of facial-only systems and produce more robust emotional representations (Hassouneh et al., 2020; Gupta et al., 2022; Xavier & John, 2025). Recent studies also show that emotion recognition research is moving toward three important directions. First, multimodal fusion strategies are increasingly used to combine visual, auditory, and physiological features through early fusion, feature-level fusion, decision-level fusion, attention-based fusion, and hybrid fusion. Second, emerging technologies such as IoT, wearable devices, smart glasses, fog computing, and edge computing are expanding FER and MER into real-time and context-aware applications (Talaat, 2023; Mukhiddinov et al., 2023; Yalcin & Alisawi, 2026). Third, recent models increasingly integrate CNNs, transformers, transfer learning, attention mechanisms, recurrent networks, and optimization techniques to improve robustness, generalization, and computational efficiency.

Several previous reviews have examined FER from the perspectives of CNN architectures, benchmark datasets, transfer learning, handcrafted versus deep features, and classification performance (Akhand et al., 2021; Alsharekh, 2022; Khan, 2022; Tsalera et al., 2022). However, these reviews remain limited in three aspects. First, they focus mainly on facial-image-based emotion recognition and provide limited synthesis of speech, EEG, ECG, wearable, and audio-visual modalities. Second, they do not sufficiently explain how different fusion strategies influence recognition performance across application contexts. Third, they give limited attention to real-world deployment challenges, including missing modalities, environmental noise, computational constraints, privacy, and ethical issues. A comparative table of previous reviews is presented in Section 2 to clarify these limitations and position the contribution of the present study.

To address these gaps, this study presents a systematic literature review of multimodal emotion recognition in the wild, focusing on the integration of facial expressions, speech signals, and physiological data for context-aware applications. The novelty of this review lies in its proposed modality–fusion–deployment taxonomy, which links five analytical dimensions: input modality, fusion mechanism, deep learning architecture, application domain, and real-world deployment constraint. Unlike previous reviews that mainly classify studies by model type or dataset, this review explains how modalities are combined, how fusion strategies are applied, where systems are deployed, and what challenges remain unresolved in real-world environments.

This review makes three main contributions. First, it profiles recent FER and MER research by analyzing publication trends, keywords, geographical distributions, datasets, methods, and application domains. Second, it develops a thematic taxonomy of multimodal emotion recognition covering modality configurations, fusion strategies, and deep learning architectures. Third, it identifies unresolved challenges and future research directions related to robustness, multimodal synchronization, missing modality handling, cross-dataset generalization, lightweight deployment, interpretability, privacy, and real-world validation.

METHODOLOGY

This study employed a Systematic Literature Review (SLR) approach following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines proposed by Moher et al. (2009). PRISMA was adopted because it provides a transparent and systematic framework for identifying, screening, assessing, and reporting literature in a review study, thereby improving methodological clarity, consistency, and reproducibility (Panic et al., 2013; Siddaway et al., 2019; ter Huurne et al., 2017). The review was conducted using a predefined internal protocol that specified the review objectives, research questions, search source, search terms, eligibility criteria, screening procedure, data extraction items, and synthesis strategy. The protocol was not prospectively registered in PROSPERO, OSF, or another public review registry because this study was conducted as an independent academic literature review. The literature search, screening, eligibility assessment, data extraction, and synthesis were conducted in June 2026. The complete PRISMA workflow adopted in this study is presented in Figure 1, which illustrates the sequential stages of identification, screening, eligibility assessment, and final inclusion of relevant studies.

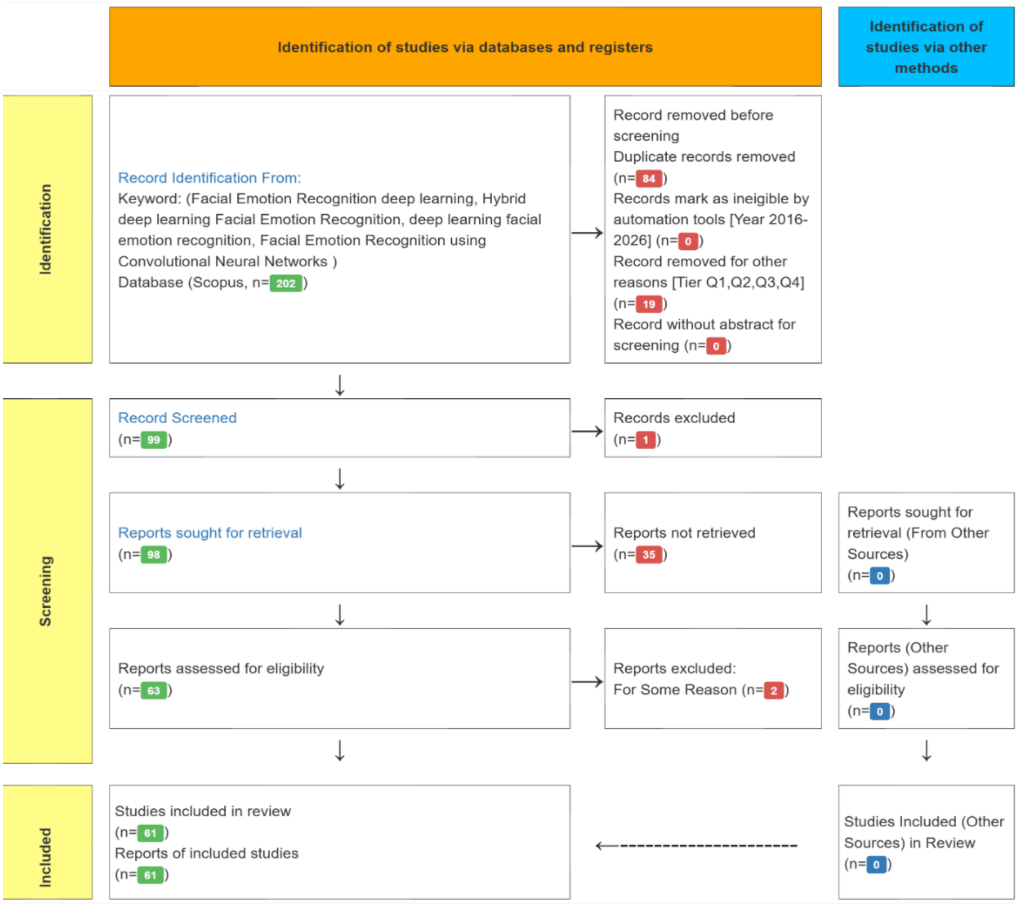

Figure 1. PRISMA Statement

The review process began with the identification stage, during which a comprehensive literature search was conducted in the Scopus database. Scopus was selected as the primary source because of its extensive multidisciplinary coverage, rigorous indexing standards, comprehensive metadata, and reduced likelihood of duplicate records compared with alternative databases (Lasda Bergman, 2012; Rocha et al., 2020). Furthermore, Scopus provides higher-quality scholarly records and minimizes the inclusion of non-peer-reviewed or predatory publications, which are often encountered in broader search engines such as Google Scholar (Hariningsih et al., 2024). To ensure comprehensive retrieval of relevant studies, several keyword combinations were employed, including "Facial Emotion Recognition deep learning," "Hybrid deep learning Facial Emotion Recognition," "deep learning facial emotion recognition," and "Facial Emotion Recognition using Convolutional Neural Networks." This initial search yielded 202 records.

Following identification, duplicate and irrelevant records were removed to improve the quality of the dataset. A total of 84 duplicate articles were excluded, while 19 records were removed because they did not satisfy the predefined journal quality criteria, including publication tier requirements (Q1–Q4 indexed journals). No articles were excluded based on publication year or missing abstracts, as all retrieved records satisfied the specified publication period (2016–2026) and contained sufficient bibliographic information. After this initial filtering process, 99 records remained for further evaluation.

The screening stage involved a detailed review of titles, abstracts, and metadata to assess their relevance to the research objectives. During this phase, one article was excluded because it did not meet the inclusion criteria established for this review. Consequently, 98 studies proceeded to the document retrieval stage. However, 35 articles could not be accessed in full text despite extensive retrieval efforts. As a result, 63 full-text articles were successfully obtained and advanced to the next phase for comprehensive evaluation.

The eligibility stage involved a full-text review of 63 articles. Each was assessed for relevance to facial and multimodal emotion recognition, deep learning methods, datasets, application domains, modality configuration, methodological clarity, and contribution to the research questions. Two articles were excluded for lacking sufficient focus on multimodal or deep-learning-based emotion recognition. In total, 61 studies met all criteria and were included for qualitative synthesis.

To enhance reliability, a basic quality appraisal was conducted using a predefined checklist covering five aspects: (1) clarity of research objective; (2) relevance to facial or multimodal emotion recognition; (3) clarity of dataset or data source; (4) clarity of method or model; and (5) availability of findings, limitations, or future directions. Each criterion was scored as 1 if adequately reported and 0 if unclear. Studies with low relevance or unclear methods were excluded or given lower priority. This appraisal supported consistency rather than statistical meta-analysis.

Screening, eligibility assessment, and data extraction were conducted systematically using the predefined criteria and checklist. Records were repeatedly checked against inclusion and exclusion criteria, and exclusion reasons were documented following PRISMA guidelines. Cohen's Kappa was not calculated, as no separate inter-rater agreement test was performed. Reliability was maintained through repeated checks of titles, abstracts, full texts, extracted data, and the final study list.

Data extraction used a structured form capturing author(s), publication year, country, research context, application domain, focus, method, dataset, modality, model architecture, fusion strategy, novelty, limitation, future direction, citation count, and journal suitability. The data were organized thematically for descriptive profiling and qualitative synthesis, focusing on trends, geographical distribution, keywords, methods, datasets, modalities, fusion strategies, emerging technologies, robustness, and research gaps.

Due to heterogeneity in datasets, models, modalities, applications, and metrics, no statistical meta-analysis was conducted. Instead, descriptive and thematic synthesis was used to identify patterns, trends, relationships, challenges, and future directions, including links among facial expressions, speech, physiological data, deep learning models, multimodal fusion, and real-world applications.

RESULTS AND ANALYSIS

This section presents the findings obtained from the systematic review of 61 eligible studies selected through the PRISMA framework. The analysis is structured according to the research objectives and research questions outlined in Section 4.1, focusing on publication trends, geographical distribution, methodological developments, application domains, technological integration, robustness strategies, and future research opportunities in multimodal emotion recognition. Specifically, the discussion of publication trends and geographical distribution addresses the research question on research profiling, while the analysis of methodological developments, technological integration, and robustness strategies responds to questions concerning model evolution, multimodal configurations, and system reliability. The examination of application domains explains how emotion recognition has been applied across different practical contexts, whereas the discussion of future research opportunities identifies unresolved challenges and potential directions for further development. The synthesis is supported by Figure 2 and Tables 3–8, which collectively illustrate the evolution of the field from conventional facial expression recognition (FER) toward context-aware multimodal emotion recognition systems (Alharbi et al., 2024; Alshahrani et al., 2024; Alshahrani et al., 2023).

A prominent finding across the reviewed literature is the substantial shift in research priorities over the last decade. Earlier studies predominantly focused on facial feature extraction, handcrafted descriptors, and classification accuracy using conventional machine learning and deep learning techniques. More recent studies increasingly emphasize multimodal fusion, transfer learning, explainable artificial intelligence (XAI), wearable sensing technologies, edge computing, and healthcare-oriented applications. This shift corresponds to the research questions that examine how methodological and technological developments support the transition from facial-only recognition to multimodal emotion recognition. It is particularly evident in studies integrating facial expressions with speech, physiological signals, EEG recordings, and contextual information to improve recognition performance in real-world environments (Alharbi et al., 2024; Alshahrani et al., 2024; Alshahrani et al., 2023; Alsubaie et al., 2024). Such developments are consistent with the broader evolution of affective computing, where emotions are increasingly understood as multimodal phenomena requiring contextual interpretation rather than solely visual analysis.

Analysis of Publication Trends and Scientific Growth

Understanding publication trends is essential for identifying the growth trajectory and scientific momentum of emotion recognition research. The temporal distribution of publications provides insight into how scholarly interest has evolved and highlights periods of accelerated development. Figure 2 presents the annual publication trend of the reviewed studies.

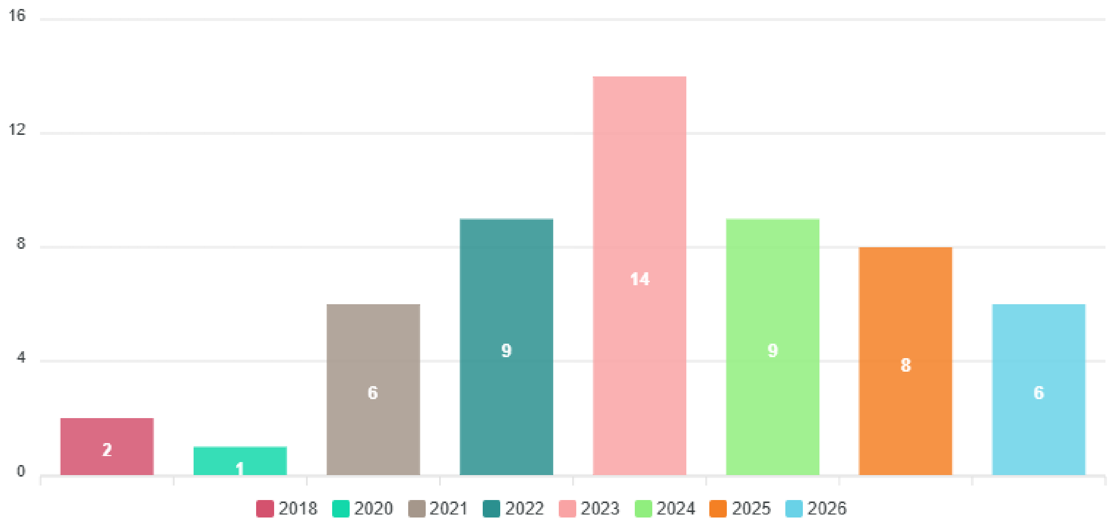

Figure 2. Annual publication trend of emotion recognition studies

As illustrated in Figure 2, publication activity increased significantly after 2020, with the highest concentration of studies appearing between 2022 and 2024. This trend reflects the growing importance of emotion recognition within the broader artificial intelligence ecosystem and aligns with the increasing demand for human-centered AI applications. The reviewed corpus shows a notable rise in studies addressing healthcare monitoring, autism support systems, multimodal learning, and intelligent interaction technologies during this period (Alharbi et al., 2024; Alsubaie et al., 2024; Alshahrani et al., 2023).

Several factors appear to have contributed to this growth. First, advances in deep learning architectures, including convolutional neural networks (CNNs), transformer-based models, and hybrid architectures, have substantially improved recognition performance under challenging conditions. Second, the expansion of remote learning, telemedicine, and intelligent monitoring systems has increased demand for automated emotion analysis technologies. Third, the availability of publicly accessible datasets and transfer learning frameworks has facilitated experimentation and benchmarking across diverse application domains (Alshahrani et al., 2024; Alharbi et al., 2024).

The publication trend also reveals a transition from laboratory-controlled FER systems toward real-world affective intelligence applications. Earlier studies primarily relied on benchmark datasets collected under controlled conditions. In contrast, recent studies increasingly investigate emotion recognition under unconstrained environments characterized by illumination variations, pose changes, facial occlusions, spontaneous expressions, and demographic diversity. This evolution is reflected in the growing adoption of transfer learning, domain adaptation, multimodal fusion, and context-aware learning strategies identified throughout the reviewed studies (Alshahrani et al., 2024; Alsubaie et al., 2024).

Furthermore, highly cited studies within the reviewed corpus frequently address multimodal integration, cross-domain generalization, healthcare applications, and context-aware emotion recognition. This observation suggests that the research community increasingly values solutions capable of addressing practical deployment challenges rather than merely improving benchmark accuracy. Consequently, future developments are expected to be driven by contextual intelligence, multimodal reasoning, and real-world applicability (Alharbi et al., 2024; Alshahrani et al., 2023).

Analysis of Geographic Distribution and Global Research Landscape

The geographical distribution of publications provides insight into the global development of emotion recognition research and highlights regional differences in research priorities. Understanding where research activities are concentrated helps reveal patterns of scientific investment, collaboration, and application focus. Figure 3 summarizes the country-level distribution of the reviewed studies.

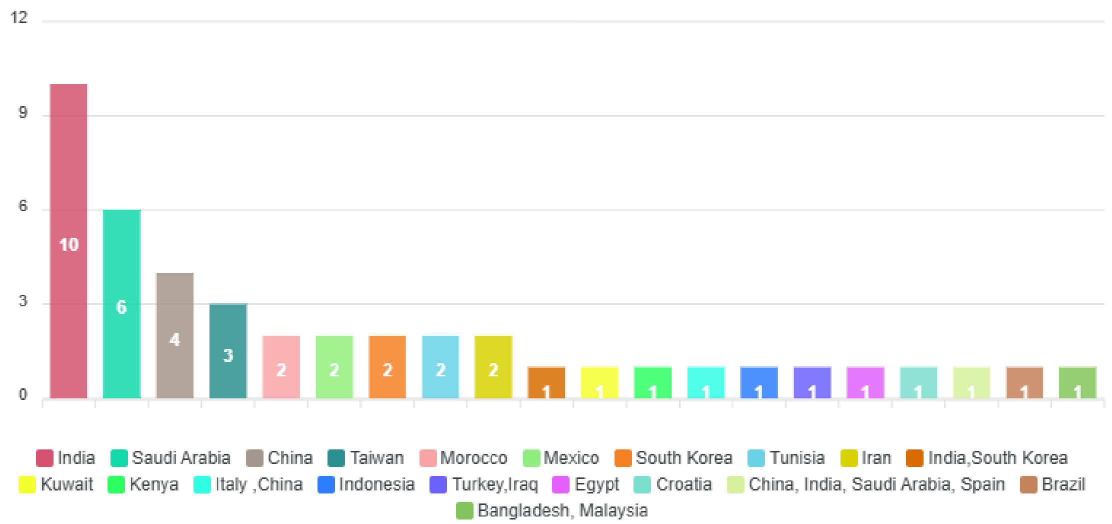

Figure 3. Geographic distribution of reviewed studies and regional research characteristics.

As shown in figure 3, emotion recognition research has become a highly international field, with substantial contributions originating from Asia, Europe, North America, and the Middle East. The reviewed studies demonstrate that research activity is no longer concentrated within a limited number of countries but has expanded globally due to the increasing accessibility of AI technologies, open-source frameworks, and cloud-based computational resources (Alharbi et al., 2024; Alshahrani et al., 2024).

India and China emerge as major contributors within the reviewed literature. Their strong publication output reflects increasing investments in artificial intelligence, expanding research infrastructure, and growing interest in human-centered computing applications. Studies from these countries frequently investigate deep learning optimization, healthcare monitoring, educational technologies, and multimodal emotion recognition systems. Similar patterns are evident in the reviewed studies focusing on intelligent healthcare systems and multimodal affective computing frameworks (Alsubaie et al., 2024; Alshahrani et al., 2023).

Although developing countries contribute a considerable number of publications, studies originating from developed countries often exhibit higher citation impact and broader international collaboration. This difference may be associated with stronger research funding mechanisms, access to advanced computational resources, and participation in global research networks. Nevertheless, contributions from emerging economies are becoming increasingly influential, particularly in healthcare-oriented and socially impactful applications such as autism support and mental health monitoring (Alharbi et al., 2024; Alsubaie et al., 2024).

The geographical analysis also reveals variations in research priorities. Studies conducted in developing regions frequently focus on healthcare applications, autism spectrum disorder (ASD) support, mental health assessment, educational monitoring, and social inclusion. Conversely, studies from developed countries tend to emphasize autonomous systems, human–robot interaction, intelligent transportation, and advanced affective computing frameworks. These findings indicate that emotion recognition technologies are increasingly shaped by local societal challenges and policy priorities (Alshahrani et al., 2023; Alharbi et al., 2024).

A recurring challenge identified across regions is the limited availability of culturally diverse datasets. Several reviewed studies highlight the difficulty of achieving robust performance across different ethnicities, age groups, and cultural backgrounds. This limitation directly affects model generalization and fairness, reinforcing the need for more inclusive datasets and cross-cultural evaluation protocols (Alshahrani et al., 2024; Alsubaie et al., 2024).

Theoretical Foundation of Facial Emotion Recognition Studies

Table 1 shows that the theoretical foundation of Facial Emotion Recognition (FER) research is primarily built upon five major theories, namely Deep Learning Theory, Affective Computing Theory, Facial Emotion Recognition Theory, Computer Vision Theory, and Human–Computer Interaction (HCI) Theory. The reviewed studies consistently demonstrate that FER is a multidisciplinary research area that integrates computational intelligence, emotion understanding, visual analysis, and interactive system design. Accordingly, the overarching theoretical framework of this review can be conceptualized as Deep Learning-Based Affective Computing for Facial Emotion Recognition, which combines advances in deep learning with affective and human-centered computing approaches.

Table 1 : Theoretical Foundation of Facial Emotion Recognition Studies

Theoretical stream	Conceptual role in FER/MER	Relationship with other theories	Conceptual evolution	Representative studies
Emotion psychology and facial expression theory	Explains how emotions are expressed through facial movements and basic affective categories	Provides the conceptual basis for FER and supports affective computing	From basic emotion theory and FACS toward computational emotion interpretation	Li and Deng, 2018; Dias et al., 2022; Salagean et al., 2025
Affective computing theory	Frames emotion recognition as the ability of intelligent systems to sense, interpret,	Connects emotion psychology with HCI, AI, and multimodal sensing	From emotion detection toward context-aware affective intelligence	Hassouneh et al., 2020; Khan, 2022; Xavier and John, 2025

	and respond to human affect			
Computer vision theory	Supports face detection, landmark extraction, image preprocessing, feature representation, and visual pattern recognition	Provides the visual processing layer for FER systems	From handcrafted descriptors to automated visual feature learning	Tsalera et al., 2022; Anjani and Satyanarayana, 2021; Said and Barr, 2021
Deep learning theory	Enables automatic hierarchical feature learning from facial, speech, and physiological data	Operationalizes affective computing through CNNs, RNNs, LSTM, ViT, attention, and hybrid architectures	From CNN-based facial classification toward multimodal deep representation learning	Akhand et al., 2021; Alzahrani, 2024; Kim et al., 2024; Dada et al., 2023
Human–computer interaction theory	Positions FER/MER as part of adaptive, responsive, and human-centered intelligent systems	Links affective computing and deep learning to practical interaction contexts	From standalone emotion classification toward emotionally aware interactive systems	Ruiz-Garcia et al., 2018; Lasri et al., 2022; Gupta et al., 2022; Radocaj and Martinovic, 2025
Multimodal and deployment-oriented theory	Explains the integration of facial, speech, EEG, wearable, and contextual data for real-world use	Extends deep learning and affective computing into IoT, edge, assistive, and healthcare systems	From unimodal FER toward multimodal context-aware affective intelligence	Hassouneh et al., 2020; Talaat, 2023; Mukhiddinov et al., 2023; Xavier and John, 2025

Table 1 should be interpreted as a theoretical synthesis rather than a bibliographic classification. The reviewed studies show that FER and multimodal emotion recognition are built on the interaction of five main theoretical streams: emotion psychology, affective computing, computer vision, deep learning, and human–computer interaction. Emotion psychology, including basic emotion theory, Facial Action Coding System, and categorical emotion models, provides the conceptual foundation for understanding how emotional states are expressed through facial movements. Computer vision theory translates these visible facial patterns into computational representations through face detection, landmark extraction, image enhancement, and feature representation. Affective computing then expands this visual interpretation into a broader intelligent-system perspective, where machines are expected not only to classify facial expressions but also to recognize, interpret, and respond to human affective states.

The strongest conceptual evolution identified in the reviewed literature is the interaction between affective computing and deep learning. Affective computing defines the goal of emotion-aware systems, while deep learning provides the computational mechanism for learning complex emotional representations from facial images, speech signals, EEG, physiological data, and contextual cues. CNNs support spatial feature learning, recurrent networks and LSTM support temporal modeling, transfer learning improves generalization across limited datasets, and transformer-based attention mechanisms help models focus on salient emotional cues. This interaction shows that FER has evolved from static facial expression classification toward multimodal, context-aware affective intelligence. Therefore, the theoretical foundation of this review is best understood as Deep Learning-Based Affective Computing, where emotion psychology explains the phenomenon, computer vision processes visual evidence, deep learning performs representation learning, and HCI guides the deployment of emotion recognition in human-centered intelligent systems.

Research Scope Based on Keyword analysis

Figure 4 illustrates the distribution of keywords, indicating that research in the field of Facial Emotion Recognition (FER) is dominated by themes such as Deep Learning, Emotion Recognition, and Convolutional Neural Networks (CNN). Deep Learning emerges as the central theme with 97 occurrences, reflecting the dominance of neural network-based approaches in the development of FER systems. Related sub-themes, including Transfer Learning, Feature Extraction, and Machine Learning, further support this trend, indicating a strong research focus on improving model accuracy and efficiency through advanced computational techniques.

The relationships among these themes are clearly evident. Deep Learning and CNN serve as the primary methodological foundations, while Emotion Recognition represents the main application domain. Computer Vision and Artificial Intelligence further reinforce the interdisciplinary nature of this research area by connecting engineering, psychology, and computational science. Regional research trends may also be influenced by specific application needs, such as the use of FER in healthcare or education, particularly in regions that emphasize assistive technologies and AI-supported learning environments.

Overall, these trends reflect global priorities in developing AI systems that are more responsive to human needs, including emotion detection for social, educational, healthcare, security, and human–machine interaction applications. Future challenges include improving model generalization, reducing algorithmic bias, enhancing robustness across diverse populations, and integrating FER with multimodal technologies such as speech, physiological signals, and body movement. Therefore, these themes are expected to remain highly relevant as the demand for more human-centered, ethical, and context-aware AI systems continues to increase

Figure 4. Research Scope

Analysis of Dominant Research Contexts and Application Domains

Beyond geographical trends, it is important to understand where emotion recognition technologies are being applied and which domains receive the greatest research attention. Identifying dominant application contexts helps clarify how the field is evolving from a purely technical discipline into a multidisciplinary area with broad societal impact. Table 2 presents the classification of research contexts identified in the reviewed literature.

As presented in Table 2, Human–Computer Interaction (HCI) remains the dominant application domain among the reviewed studies. This finding reflects the fundamental objective of emotion recognition technologies: enabling intelligent systems to perceive and respond appropriately to human emotional states. Numerous reviewed studies position emotion recognition as a core component of adaptive and affective interaction systems (Alshahrani et al., 2023; Alharbi et al., 2024).

Within the HCI domain, the reviewed studies report applications in intelligent tutoring systems, virtual assistants, social robots, gaming environments, adaptive interfaces, and user experience analysis. These applications aim to improve interaction quality by incorporating emotional awareness into system behavior.

Table 2. Classification of research contexts and application domains.

Dataset	Count	Authors
FER2013	13	Saurav et al., 2021; Hung and Chang, 2021; Boughanem et al., 2022; Gupta et al., 2022; Khan, 2022; Kumari and Bhatia, 2023; Bakariya et al., 2023; Yal?in and Alisawi, 2024; Alsubai et al., 2024; Bohi et al., 2024; Salagean et al., 2025; Fekri-Ershad, 2026; Lu et al., 2026
AffectNet	3	Mukhiddinov et al., 2023; Almubarak and Alsulaiman, 2025; Yalcin and Alisawi, 2026
FER2013;CK+;KDEF	2	Alsharekh, 2022; Alonazi et al., 2023
KDEF	2	Jain et al., 2023; Radocaj and Martinovic, 2025
CK+;KDEF;FER2013	1	Ramirez-Quintana et al., 2025
FER2013;CK+	1	Helaly et al., 2023
FER2013Plus	1	Pruthviraja et al., 2024
JAFFE	1	Lasri et al., 2022
JAFFE;CK+;MMI	1	M et al., 2026
KDEF;JAFFE	1	Akhand et al., 2021

The prevalence of HCI-related studies demonstrates that emotion recognition continues to serve as a foundational technology for affective computing systems (Alshahrani et al., 2023; Alshahrani et al., 2024). Healthcare represents one of the fastest-growing application areas identified in the review. Numerous studies investigate emotion recognition for autism spectrum disorder screening, Alzheimer's disease monitoring, depression detection, stress assessment, and mental health evaluation. The increasing representation of healthcare applications highlights the potential of emotion-aware technologies to support preventive healthcare, personalized interventions, and continuous patient monitoring (Alharbi et al., 2024; Alsubaie et al., 2024).Educational applications also constitute an emerging research direction. Several reviewed studies employ emotion recognition systems to monitor student engagement, cognitive workload, attention levels, and emotional responses during online learning activities. The rapid adoption of digital learning environments has accelerated interest in emotion-aware educational technologies capable of supporting adaptive and personalized learning experiences (Alshahrani et al., 2024; Alharbi et al., 2024).The diversification of application domains demonstrates that emotion recognition is evolving beyond a traditional computer vision problem. Instead, it is becoming a key enabling technology for human-centered AI systems, where emotional information functions as an important contextual signal for adaptive decision-making across healthcare, education, assistive technologies, and intelligent interaction systems (Alshahrani et al., 2023; Alsubaie et al., 2024).

Dataset Classification for Face Emotion Recognition

Table 3 summarizes the distribution of datasets employed in the reviewed Facial Emotion Recognition (FER) studies. The results indicate that FER2013 is the most frequently utilized standalone dataset, appearing in thirteen studies (Saurav et al., 2021; Hung and Chang, 2021; Boughanem et al., 2022; Gupta et al., 2022; Khan, 2022; Kumari and Bhatia, 2023; Bakariya et al., 2023; Yalcin and Alisawi, 2024; Alsubai et al., 2024; Bohi et al., 2024; Salagean et al., 2025; Fekri-Ershad, 2026; Lu et al., 2026). This finding underscores the continued prominence of FER2013 as a benchmark dataset in FER research, particularly for the development and evaluation of deep learning-based emotion recognition models due to its accessibility, standardized structure, and widespread acceptance within the research community.

In addition to FER2013, AffectNet is employed in three studies (Mukhiddinov et al., 2023; Almubarak and Alsulaiman, 2025; Yalcin and Alisawi, 2026), whereas KDEF appears as a standalone dataset in two studies (Jain et al., 2023; Radocaj and Martinovic, 2025). Furthermore, Table 3 reveals that several studies adopt combined dataset configurations, including FER2013; CK+; KDEF, CK+; KDEF; FER2013, FER2013; CK+, JAFFE; CK+; MMI, and KDEF; JAFFE. This observation suggests that CK+ is more frequently incorporated as part of multi-dataset evaluation frameworks rather than being used independently. The integration of multiple datasets reflects an increasing emphasis on improving model robustness and assessing generalization capabilities across diverse facial expression scenarios and experimental settings.

Overall, these findings demonstrate that dataset selection constitutes a critical aspect of FER research, as it directly affects model training, evaluation reliability, and generalization performance. The predominance of FER2013 highlights its enduring role as a standard benchmark for comparative analysis, while the utilization of datasets such as AffectNet, KDEF, JAFFE, MMI, and various combined configurations indicates ongoing efforts to evaluate FER models under more diverse and challenging conditions. Nevertheless, the persistent reliance on a limited number of established benchmark datasets also emphasizes the necessity for future research to develop more comprehensive, representative, and real-world datasets capable of capturing broader variations in facial expressions, illumination conditions, head poses, demographic characteristics, cultural backgrounds, and spontaneous emotional behaviors.

Table 3. Dataset Classification

Dataset	Count	Authors
FER2013	13	Saurav et al., 2021; Hung and Chang, 2021; Boughanem et al., 2022; Gupta et al., 2022; Khan, 2022; Kumari and Bhatia, 2023; Bakariya et al., 2023; Yal?in and Alisawi, 2024; Alsubai et al., 2024; Bohi et al., 2024; Salagean et al., 2025; Fekri-Ershad, 2026; Lu et al., 2026
AffectNet	3	Mukhiddinov et al., 2023; Almubarak and Alsulaiman, 2025; Yalcin and Alisawi, 2026
FER2013;CK+;KDEF	2	Alsharekh, 2022; Alonazi et al., 2023
KDEF	2	Jain et al., 2023; Radocaj and Martinovic, 2025
CK+;KDEF;FER2013	1	Ramirez-Quintana et al., 2025
FER2013;CK+	1	Helaly et al., 2023
FER2013Plus	1	Pruthviraja et al., 2024
JAFFE	1	Lasri et al., 2022
JAFFE;CK+;MMI	1	M et al., 2026
KDEF;JAFFE	1	Akhand et al., 2021

Overall, Table 3 confirms that FER remains a relevant and promising research area, particularly in the development of intelligent systems capable of recognizing human emotions more effectively. With the continuous advancement of deep learning, FER research is expected to expand further into applications such as human–computer interaction, affective computing, sentiment analysis, education, healthcare, and intelligent assistive systems. Therefore, FER will continue to be an important research topic in the coming years (Khan, 2022; Alsharekh, 2022).

Application Areas of Facial Emotion Recognition Studies

Table 4 presents the distribution of application areas in Facial Emotion Recognition (FER) studies. The findings indicate that Human–Computer Interaction (HCI) constitutes the most dominant application area, represented by 16 studies (Saurav et al., 2021; Anjani and Satyanarayana, 2021; Lu et al., 2021; Said and Barr, 2021; Akhand et al., 2021; Khan, 2022; Kumari and Bhatia, 2023; Wang et al., 2023; Bakariya et al., 2023; Mukhiddinov et al., 2023; Alonazi et al., 2023; Pruthviraja et al., 2024; Salagean et al., 2025; Ramirez-Quintana et al., 2025; Almubarak and Alsulaiman, 2025; Yalcin and Alisawi, 2026). This dominance suggests that FER research is increasingly positioned within the broader agenda of developing intelligent, adaptive, and emotion-aware interactive systems. In this context, FER is not merely treated as a classification task, but also as a core component of human-centered computing, particularly in real-time emotion recognition, companion robotics, intelligent learning environments, and affective user interfaces.

Table 4 . Application Area of FER Studies

Application Area	Count	Authors
Human–Computer Interaction	16	Saurav et al., 2021; Anjani and Satyanarayana, 2021; Lu et al., 2021; Said and Barr, 2021; Akhand et al., 2021; Khan, 2022; Kumari and Bhatia, 2023; Wang et al., 2023; Bakariya et al., 2023; Mukhiddinov et al., 2023; Alonazi et al., 2023; Pruthviraja et al., 2024; Salagean et al., 2025; Ramirez-Quintana et al., 2025; Almubarak and Alsulaiman, 2025; Yalcin and Alisawi, 2026
Affective Computing Systems	12	Ruiz-Garcia et al., 2018; Zhu et al., 2022; Tsalera et al., 2022; AlEisa et al., 2023; Dada et al., 2023; Rao et al., 2024; Kim et al., 2024; Alsubai et al., 2024; Bhati et al., 2025; Hongwei et al., 2025; Radocaj and Martinovic, 2025; Lu et al., 2026
Mental Health	7	Hassouneh et al., 2020; Boughanem et al., 2022; Kumari and Bhatia, 2022; Bohi et al., 2024; Yalcin and Alisawi, 2024; Xavier and John, 2025; Alabi et al., 2026
Education	5	Lasri et al., 2022; Gupta et al., 2022; Talaat, 2023; Sultana et al., 2023; El et al., 2026
Autonomous Systems	2	Jain et al., 2023; Raj and Demirkol, 2025
Security And Surveillance	2	Alsharekh, 2022; M et al., 2026

The second most prominent application area is Affective Computing Systems, comprising 12 studies (Ruiz-Garcia et al., 2018; Zhu et al., 2022; Tsalera et al., 2022; AlEisa et al., 2023; Dada et al., 2023; Rao et al., 2024; Kim et al., 2024; Alsubai et al., 2024; Bhati et al., 2025; Hongwei et al., 2025; Radocaj and Martinovic, 2025; Lu et al., 2026). This pattern demonstrates the strong conceptual alignment between FER and affective computing, in which facial expressions are employed as computational indicators for recognizing, interpreting, and responding to human emotional states.

In addition, Mental Health emerges as a significant application domain, with 7 studies (Hassouneh et al., 2020; Boughanem et al., 2022; Kumari and Bhatia, 2022; Bohi et al., 2024; Yalcin and Alisawi, 2024; Xavier and John, 2025; Alabi et al., 2026). This finding reflects the growing relevance of FER in mental health assessment, emotional state monitoring, and assistive diagnostic support. Meanwhile, Education is represented by 5 studies, indicating the use of FER for learner engagement detection, online learning analytics, and affective learning support. Other application areas, including Autonomous Systems and Security and Surveillance, are represented by 2 studies each, suggesting more specialized but still relevant applications of FER in intelligent mobility, driver monitoring, and smart surveillance contexts.

Overall, Table 4 indicates that FER research is strongly oriented toward human-centered applications, particularly in HCI, affective computing, mental health, and education. These findings imply that FER has substantial potential to support the development of intelligent systems that are more responsive to human emotions, behavioral signals, and social needs. Future studies should therefore focus on enhancing model accuracy, robustness, generalizability, and interpretability across diverse real-world environments. In addition, the integration of multimodal data, such as speech, EEG, physiological signals, and body movement, may further strengthen the practical reliability and contextual sensitivity of FER systems.

Model Architectures in FER Studies

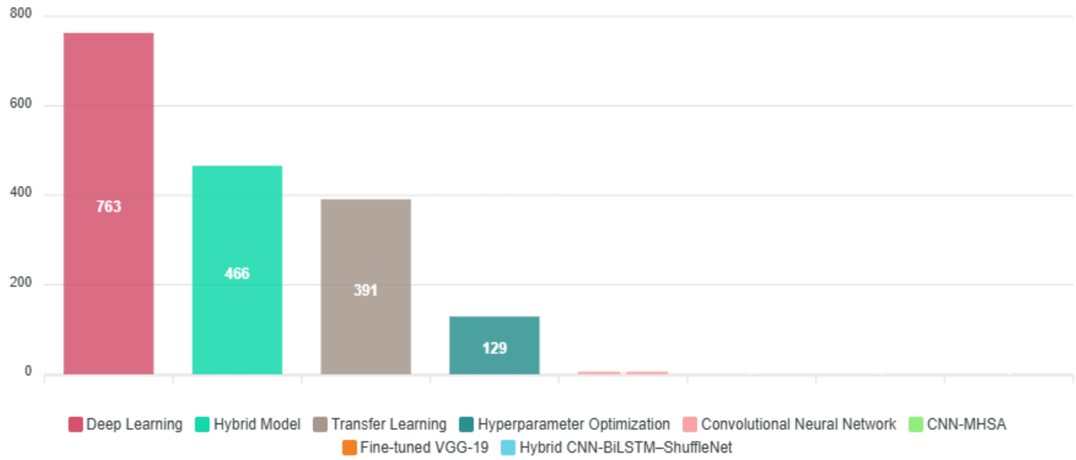

Figure .5 Methods and Model Architectures in FER Studies

Figure 5 illustrates the distribution of methods and model architectures used in Facial Emotion Recognition (FER) studies. The figure indicates that Deep Learning is the most dominant approach, with the highest citation value of 763, followed by Hybrid Model with 466, Transfer Learning with 391, and Hyperparameter Optimization with 129. This pattern shows that FER research has been strongly shaped by deep learning-based methods, mainly because these approaches can automatically learn meaningful facial patterns from image data without relying heavily on handcrafted feature extraction. The strong position of deep learning in this field is also reflected in highly cited studies such as Akhand et al. (2021) and Gupta et al. (2022), which demonstrate the important role of deep neural networks in improving FER performance.

The prominence of Hybrid Model and Transfer Learning also suggests that recent FER research is moving toward more adaptive and flexible model designs. Hybrid models are widely used because they allow researchers to combine the strengths of different architectures, such as CNN, LSTM, BiLSTM, Transformer, and attention mechanisms. This combination helps models capture richer facial representations and improves their ability to classify emotional expressions more accurately. Transfer learning, on the other hand, is often used to overcome the limitations of FER datasets, especially when the available data are limited, imbalanced, or collected under controlled conditions. Studies such as Hassouneh et al. (2020), Jain et al. (2023), and Yalcin and Alisawi (2024) show that these approaches are important for improving both model accuracy and robustness.

Figure 5 also shows that several more specific architectures, such as Convolutional Neural Network, CNN-MHSA, Fine-tuned VGG-19, and Hybrid CNN-BiLSTM–ShuffleNet, appear with lower citation values compared with broader categories such as Deep Learning, Hybrid Model, and Transfer Learning. This does not mean that these architectures are less important. Rather, it indicates that they represent more specialized methodological developments within the broader deep learning ecosystem. These models contribute to the refinement of FER techniques by addressing particular challenges, such as feature representation, attention to facial regions, temporal learning, and computational efficiency.

Overall, Figure 5 confirms that the methodological development of FER studies is largely driven by the interaction between Deep Learning, Hybrid Model, Transfer Learning, and Hyperparameter Optimization. These approaches provide an important foundation for developing FER systems that are more accurate, robust, and applicable in real-world environments. Future studies are expected to move further toward advanced hybrid architectures, attention-based learning, multimodal fusion, explainable AI, and lightweight models that can be deployed in practical applications such as human–computer interaction, affective computing, assistive technology, healthcare, education, and intelligent surveillance systems (Alsharekh, 2022; Ramirez-Quintana et al., 2025).

Strategies for Improving FER Model Robustness and Generalization

Figure 6 illustrates the distribution of strategies employed to enhance the robustness and generalization capability of Facial Emotion Recognition (FER) models. The figure shows that Transfer Learning is the most dominant strategy, being identified in 32 reviewed studies. This finding indicates that transfer learning has become a central methodological approach in FER research. Its dominance can be attributed to its ability to transfer visual representations learned from large-scale datasets to more specific FER tasks, particularly when the available datasets are limited, imbalanced, or collected under controlled conditions. By utilizing pre-trained models, researchers can improve recognition accuracy, reduce training complexity, and strengthen model performance in situations where task-specific training data are insufficient (Sultana et al., 2023; Lu et al., 2021; Jain et al., 2023).

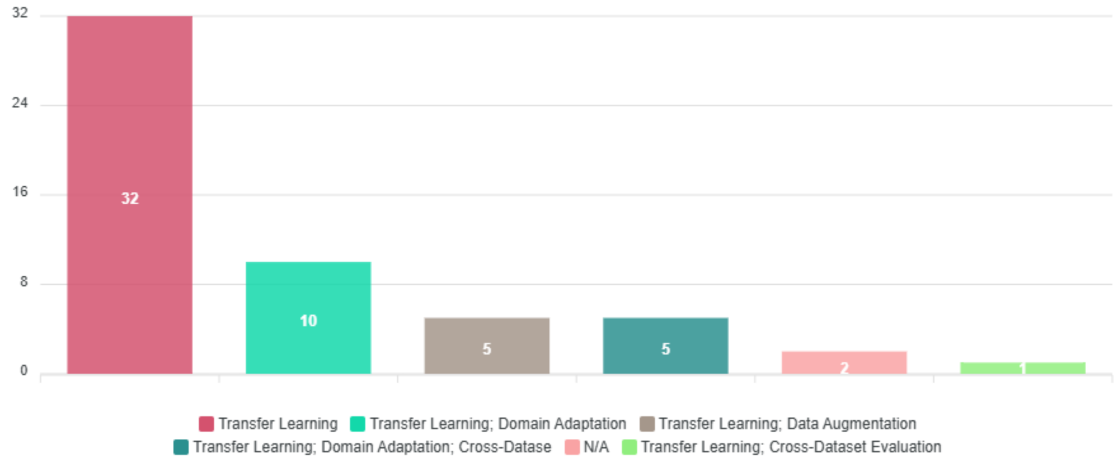

Figure 6. Strategies for Improving FER Model Robustness and Generalization

In addition to the standalone use of transfer learning, Figure 6 also reveals the emergence of combined strategies aimed at improving model adaptability. The combination of Transfer Learning and Domain Adaptation was identified in 10 reviewed studies, suggesting that researchers increasingly recognize the importance of reducing the performance gap between benchmark datasets and real-world deployment environments. This strategy is particularly relevant because FER models often experience performance degradation when exposed to variations in illumination, facial pose, demographic characteristics, cultural expression patterns, and spontaneous emotional responses.

Furthermore, the combination of Transfer Learning and Data Augmentation was reported in five reviewed studies. This pattern indicates that data augmentation is commonly used as a complementary strategy to increase the diversity of facial image data through transformations such as rotation, scaling, cropping, illumination adjustment, and facial orientation variation. Although this strategy appears less frequently than transfer learning and domain adaptation, its presence remains important because it directly addresses the problem of limited data variability in FER datasets. Similarly, the combination of Transfer Learning, Domain Adaptation, and Cross-Dataset Evaluation was also identified in five reviewed studies, reflecting a growing concern with evaluating whether FER models can maintain reliable performance across datasets with different characteristics.

Overall, Figure 6 confirms that transfer learning remains the primary strategy for improving FER model robustness and generalization. However, the presence of combined strategies suggests that transfer learning alone may not be sufficient to address the complexity of real-world FER applications. Future studies should therefore place greater emphasis on integrating transfer learning with domain adaptation, data augmentation, cross-dataset evaluation, attention mechanisms, multimodal fusion, and explainable AI. Such integration is essential for developing FER systems that are more robust against illumination changes, facial occlusion, head-pose variation, dataset bias, cultural diversity, and spontaneous expressions. These advances will strengthen the applicability of FER in human–computer interaction, affective computing, education, healthcare, assistive technology, and other human-centered intelligent systems.

Research Objectives in Facial Emotion Recognition Studies

Table 5 summarizes the main research objectives identified in Facial Emotion Recognition (FER) studies. The most dominant objective is Improving FER Accuracy, which appears in 30 reviewed studies. This finding indicates that accuracy enhancement remains the primary concern in FER research, particularly because reliable emotion recognition depends on the model's ability to extract discriminative facial features and classify emotional expressions correctly.

This objective is reflected in studies by Hung and Chang (2021), Said and Barr (2021), Lu et al. (2021), Anjani and Satyanarayana (2021), Akhand et al. (2021), Dias et al. (2022), Kumari and Bhatia (2022), Boughanem et al. (2022), Tsalera et al. (2022), Sultana et al. (2023), Dada et al. (2023), AlEisa et al. (2023), Agung et al. (2024), Aghabeigi et al. (2024), Bohi et al. (2024), Pruthviraja et al. (2024), Alzahrani (2024), Rao et al. (2024), Kim et al. (2024), Hongwei et al. (2025), Salagean et al. (2025), Ramirez-Quintana et al. (2025), Almubarak and Alsulaiman (2025), M et al. (2026), and Fekri-Ershad (2026). The prominence of this objective is closely related to the rapid development of deep learning, transfer learning, and hybrid architectures, which have significantly improved feature representation and classification performance.

The second major objective is Applying FER in Real-World Applications, reported in 16 studies. This pattern suggests a gradual shift from purely benchmark-oriented evaluation toward practical implementation in real-world settings. Studies by Hassouneh et al. (2020), Lasri et al. (2022), Alsharekh (2022), Gupta et al. (2022), Zhu et al. (2022), Talaat (2023), Mukhiddinov et al. (2023), Bakariya et al. (2023), Jain et al. (2023), Wang et al. (2023), Raj and Demirkol (2025), Radocaj and Martinovic (2025), Yalcin and Alisawi (2026), El et al. (2026), Lu et al. (2026), and Alabi et al. (2026) demonstrate that FER is increasingly applied in education, healthcare, human–computer interaction, human–robot interaction, assistive technology, autonomous systems, and intelligent services. This development shows that FER is not only a technical classification problem, but also an enabling technology for human-centered intelligent systems.

Table 5. Research Objectives in Facial Emotion Recognition Studies

Research Objectives	Count	Authors
Improving FER Accuracy	30	Hung and Chang, 2021; Said and Barr, 2021; Lu et al., 2021; Anjani and Satyanarayana, 2021; Akhand et al., 2021; Dias et al., 2022; Kumari and Bhatia, 2022; Boughanem et al., 2022; Tsalera et al., 2022; Sherly et al., 2023; Kumari and Bhatia, 2023; Kumari and Bhatia, 2023; Sultana et al., 2023; Helaly et al., 2023; Alonazi et al., 2023; Dada et al., 2023; AlEisa et al., 2023; Agung et al., 2024; Aghabeigi et al., 2024; Bohi et al., 2024; Pruthviraja et al., 2024; Alzahrani, 2024; Rao et al., 2024; Kim et al., 2024; Hongwei et al., 2025; Salagean et al., 2025; Ramirez-Quintana et al., 2025; Almubarak and Alsulaiman, 2025; M et al., 2026; Fekri-Ershad, 2026
Applying FER In Real-World Applications	16	Hassouneh et al., 2020; Lasri et al., 2022; Alsharekh, 2022; Gupta et al., 2022; Zhu et al., 2022; Talaat, 2023; Mukhiddinov et al., 2023; Bakariya et al., 2023; Jain et al., 2023; Wang et al., 2023; Raj and Demirkol, 2025; Radocaj and Martinovic, 2025; Yalcin and Alisawi, 2026; El et al., 2026; Lu et al., 2026; Alabi et al., 2026
Addressing Data Variations	4	Li and Deng, 2018; Bisogni et al., 2023; Yal?in and Alisawi, 2024; Xavier and John, 2025
Developing Efficient Models	1	Saurav et al., 2021

Another important objective is Addressing Data Variations, which appears in 4 studies, including Li and Deng (2018), Bisogni et al. (2023), Yalcin and Alisawi (2024), and Xavier and John (2025). Although less frequent, this objective is highly relevant because FER models often encounter performance degradation when exposed to variations in illumination, facial pose, occlusion, spontaneous expressions, demographic diversity, and dataset characteristics. In this context, addressing data variation is essential for improving model robustness and ensuring that FER systems can operate reliably beyond controlled experimental environments.

Meanwhile, Developing Efficient Models is represented by only one study, namely Saurav et al. (2021). This limited representation indicates that computational efficiency has received relatively less attention than accuracy improvement and application development. However, efficiency remains an important research direction, especially for real-time FER systems and deployment in resource-constrained environments such as mobile devices, embedded systems, edge computing platforms, and IoT-based applications.

Overall, Table 5 shows that the current direction of FER research is primarily driven by the need to improve recognition accuracy, followed by the effort to expand FER into practical application domains. At the same time, the smaller number of studies addressing data variation and model efficiency reveals important opportunities for future investigation. Future FER research should therefore move beyond accuracy-oriented development by emphasizing robustness, efficiency, generalizability, and real-world deployability. Such a direction is essential for advancing FER as a reliable human-centered technology in education, healthcare, affective computing, assistive systems, and intelligent human–machine interaction.

DISCUSSION

The findings of this review suggest that the development of Facial Emotion Recognition (FER) should not be interpreted merely as a progression of model accuracy, but as a broader transition in the conceptual role of emotion recognition within intelligent systems. Earlier FER research was largely shaped by the assumption that facial expressions could serve as sufficient visual indicators of emotional states. However, the reviewed literature shows that this assumption is increasingly being challenged by real-world deployment requirements. Illumination changes, facial occlusion, pose variation, cultural differences, spontaneous expressions, and dataset bias reveal that facial information alone is often insufficient for reliable affective interpretation. Therefore, the main implication of the reviewed evidence is that FER is gradually moving from a facial-image classification problem toward a context-aware affective computing problem.

This transition has important theoretical implications. The strong presence of deep learning in the reviewed studies confirms that hierarchical representation learning has become the dominant mechanism for extracting emotional features. However, the significance of deep learning lies not only in its ability to improve classification performance, but also in its role as a bridge between affective computing and real-world intelligent systems. Affective computing defines the goal of recognizing and responding to human emotions, while deep learning provides the computational capacity to model complex visual, auditory, and physiological patterns. This interaction explains why recent FER studies increasingly incorporate transfer learning, attention mechanisms, transformer-based architectures, and multimodal fusion. These methods reflect an effort to make emotion recognition systems more adaptive, generalizable, and context-sensitive rather than simply more accurate on benchmark datasets.

The dominance of Human–Computer Interaction (HCI) as an application context also carries deeper conceptual meaning. It indicates that FER is increasingly positioned as an enabling technology for emotionally responsive interaction, not merely as a standalone recognition tool. In HCI, the value of FER depends on whether emotional predictions can support adaptive system behavior, improve user experience, enhance accessibility, or assist decision-making. This interpretation extends earlier reviews that primarily discussed FER in terms of algorithmic performance. The present synthesis suggests that the practical relevance of FER depends on its ability to operate meaningfully within interaction scenarios such as education, assistive technology, healthcare monitoring, social robotics, and intelligent services.

The growing use of FER in healthcare and mental health applications further supports this shift toward human-centered affective intelligence. Emotional states in contexts such as autism support, Alzheimer's monitoring, stress detection, and learner engagement are often subtle, dynamic, and influenced by physiological or behavioral conditions. In these settings, facial expressions may not fully represent the user's emotional state. This explains why multimodal approaches are increasingly important: speech, EEG, ECG, wearable sensors, gaze, and contextual data can provide complementary cues when facial signals are weak, ambiguous, or unreliable. Thus, multimodal emotion recognition should be understood not simply as a technical enhancement, but as a response to the theoretical limitation of assuming that emotion can be inferred from the face alone.

The continued reliance on benchmark datasets also reveals a central tension in FER research. Benchmark datasets are useful because they support comparison across models, but they may also encourage narrow optimization toward controlled evaluation settings. This creates a gap between benchmark performance and real-world reliability. The reviewed evidence implies that future FER progress should not be measured only through incremental improvements in accuracy, but also through robustness across cultures, demographic groups, lighting conditions, poses, occlusions, and spontaneous emotional expressions. This issue is especially important for socially sensitive applications, where biased or unreliable emotion predictions may produce harmful consequences.

Transfer learning and domain adaptation can be interpreted as methodological responses to this generalization problem. Their increasing use indicates that researchers recognize the limitations of training FER models from small or homogeneous datasets. However, transfer learning alone cannot fully solve the problem of ecological validity if the target deployment environment differs substantially from the source data. Therefore, future work should combine transfer learning with cross-dataset evaluation, multimodal validation, data diversity, and explainability mechanisms. Such integration is necessary to ensure that FER systems are not only technically accurate but also reliable, interpretable, and fair in real-world use.

Another important interpretation concerns the emergence of optimization-based and lightweight FER models. These approaches reflect the growing pressure to move FER from laboratory experiments to deployable systems. Real-world applications in smart glasses, mobile devices, IoT platforms, classroom monitoring, healthcare systems, and assistive technologies require models that are not only accurate but also computationally efficient, low-latency, and privacy-aware. This shows that FER research is entering a deployment-oriented phase in which model performance must be evaluated together with efficiency, interpretability, usability, and ethical acceptability.

Overall, this review indicates that the future of FER lies in the convergence of multimodal sensing, adaptive deep learning, context-aware system design, and responsible AI principles. The field is no longer defined solely by the question of which model achieves the highest recognition accuracy. Instead, the more important question is how FER systems can produce reliable and meaningful emotional interpretations across diverse users, environments, and application contexts. This perspective positions FER as a foundational component of human-centered artificial intelligence and highlights the need for future studies to prioritize robustness, fairness, explainability, and real-world validation.

CONCLUSION

This systematic literature review synthesized 61 eligible studies on Facial Emotion Recognition (FER) and multimodal emotion recognition selected through the PRISMA framework. The findings show that FER research has shifted from conventional facial-image classification toward deep learning-based and increasingly multimodal emotion recognition systems. This shift is supported by the dominance of CNN-based models, transfer learning, hybrid architectures, attention mechanisms, and optimization-based approaches identified across the reviewed studies. However, the synthesis also shows that methodological progress is not limited to improving classification accuracy. Recent studies increasingly address robustness, generalization, real-time deployment, multimodal fusion, and application-specific requirements.

The conclusion that FER is moving toward more human-centered applications is supported by the distribution of reviewed studies across Human–Computer Interaction, affective computing systems, healthcare, mental health, education, assistive technology, autonomous systems, and safety-related domains. These findings indicate that FER is no longer used only as a technical image-recognition task, but increasingly as a supporting component in systems designed to interpret human emotional states for practical decision-making and adaptive interaction. For example, studies on autism support, Alzheimer's monitoring, learner engagement detection, smart glasses for visually impaired users, driver emotion monitoring, and workplace safety demonstrate that FER is being applied in contexts where emotional understanding has direct social, educational, clinical, or safety-related value.

The review also shows that multimodal emotion recognition has become an important response to the limitations of facial-only systems. Facial expressions remain valuable, but they are vulnerable to illumination variation, pose changes, occlusion, cultural differences, spontaneous expressions, and dataset bias. The integration of speech signals, EEG, physiological measurements, wearable sensors, and audio-visual cues provides complementary information that can improve contextual understanding and system reliability. Therefore, the contribution of multimodal FER lies not only in improving model performance, but also in enabling emotion recognition systems to operate more effectively in complex real-world environments.

From a theoretical perspective, this review contributes an integrated understanding of FER by linking affective computing, computer vision, deep learning, multimodal learning, and Human–Computer Interaction. The synthesis indicates that affective computing provides the conceptual goal of recognizing and responding to emotional states, while deep learning provides the computational mechanism for extracting complex emotional representations from facial, auditory, and physiological data. This relationship explains the current movement from static facial expression recognition toward context-aware multimodal affective systems.

From a practical perspective, the findings suggest that future FER systems should not be evaluated only by benchmark accuracy. Greater attention should be given to dataset diversity, annotation quality, cultural representation, spontaneous expression recognition, real-time performance, model efficiency, explainability, privacy, and ethical deployment. This is particularly important for sensitive domains such as healthcare, education, assistive technology, and public safety, where inaccurate or biased emotion recognition may produce negative consequences.

Despite its contributions, this review has several limitations. The literature search was limited to Scopus, which may have excluded relevant studies indexed in Web of Science, IEEE Xplore, ACM Digital Library, PubMed, or other databases. The review also focused on studies published between 2016 and 2026 and included only accessible peer-reviewed journal publications. In addition, the synthesis was qualitative and descriptive rather than statistical, so direct meta-analytic comparison of model performance was not conducted.

Overall, this review demonstrates that the future development of FER depends on the integration of robust deep learning models, multimodal sensing, diverse datasets, explainable methods, and real-world validation. Rather than pursuing accuracy improvement alone, future research should prioritize reliable, fair, interpretable, and deployable emotion recognition systems that can function across diverse users, environments, and application contexts.

Future Research Directions

Based on the synthesis of the reviewed studies, future research on Facial Emotion Recognition (FER) should be organized into short-term, medium-term, and long-term priorities. This prioritization is important because the field faces different levels of challenges, ranging from immediate methodological limitations to broader issues of real-world deployment, ethical governance, and context-aware multimodal intelligence.

In the short term, future studies should prioritize improving dataset quality, benchmarking consistency, and model robustness. The review shows that many FER studies still depend heavily on established benchmark datasets, such as FER2013, CK+, JAFFE, KDEF, MMI, and AffectNet. Although these datasets support comparative evaluation, they also present limitations related to noisy annotations, class imbalance, posed expressions, limited cultural diversity, and weak representation of spontaneous real-world emotions. Therefore, immediate research should focus on cleaner annotation protocols, cross-dataset validation, balanced class distribution, and evaluation under challenging conditions such as illumination variation, occlusion, head-pose changes, and low-resolution images. In parallel, transfer learning, data augmentation, domain adaptation, and explainable AI should be used to improve robustness and interpretability without requiring major changes in existing FER pipelines.

In the medium term, research should move toward integrated multimodal emotion recognition frameworks. The findings indicate that facial information alone is often insufficient for reliable emotional interpretation in real-world environments. Future studies should therefore combine facial expressions with speech signals, EEG, ECG, wearable sensors, gaze behavior, body posture, and contextual information. However, this integration should not be limited to simply adding more modalities. Greater attention should be given to adaptive fusion strategies, modality weighting, missing-modality handling, temporal modeling, and synchronization across heterogeneous data sources. Medium-term research should also prioritize lightweight and efficient architectures that can operate on mobile devices, smart glasses, embedded systems, and edge-computing platforms. This direction is particularly important for healthcare monitoring, education, assistive technology, human–robot interaction, and safety-related applications.

In the long term, FER research should develop toward trustworthy, context-aware, and human-centered affective intelligence. This agenda requires moving beyond benchmark accuracy toward systems that are fair, explainable, privacy-preserving, culturally sensitive, and reliable across diverse populations and environments. Long-term studies should investigate cross-cultural FER, federated and privacy-preserving learning, ethical governance, bias mitigation, real-world clinical validation, and continuous adaptation in dynamic environments. In addition, FER should be integrated into broader intelligent ecosystems such as digital health platforms, adaptive learning systems, smart environments, autonomous vehicles, and socially assistive robots. Such development will allow FER to evolve from facial-image classification into a responsible multimodal technology capable of supporting meaningful, safe, and socially beneficial human–AI interaction.

Overall, the future development of FER should progress through three stages: first, strengthening datasets and robustness; second, building multimodal and deployable systems; and third, establishing trustworthy and context-aware affective intelligence. This staged research agenda provides a more structured direction for advancing FER from controlled experimental settings toward reliable real-world applications.

REFERENCES

[1] F. Aghabeigi, S. Nazari, and N. Osati Eraghi, "An efficient facial emotion recognition using convolutional neural network with local sorting binary pattern and whale optimization algorithm," *Int. J. Data Sci. Anal.*, vol. 20, pp. 2275–2290, 2025, doi: 10.1007/s41060-024-00601-1.

[2] E. S. Agung, A. P. Rifai, and T. Wijayanto, "Image-based facial emotion recognition using convolutional neural network on Emognition dataset," *Sci. Rep.*, vol. 14, Art. no. 14429, 2024, doi: 10.1038/s41598-024-65276-x.

[3] M. A. H. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, "Facial emotion recognition using transfer learning in the deep CNN," *Electronics*, vol. 10, Art. no. 1036, 2021, doi: 10.3390/electronics10091036.

[4] A. O. Alabi, B.-G. Song, and N. Kang, "A comprehensive investigation of auditory-induced emotions combining psychoacoustic metrics, facial emotion recognition, and explainable deep learning," *Expert Systems with Applications*, vol. 322, Art. no. 132385, 2026, doi: 10.1016/j.eswa.2026.132385.

[5] H. N. AlEisa *et al.*, "Henry gas solubility optimization with deep learning based facial emotion recognition for human computer interface," *IEEE Access*, vol. 11, pp. 62233–62241, 2023, doi: 10.1109/ACCESS.2023.3284457.

[6] M. Almubarak and F. A. Alsulaiman, "An ensemble learning approach for facial emotion recognition based on deep learning techniques," *Electronics*, vol. 14, Art. no. 3415, 2025, doi: 10.3390/electronics14173415.

[7] M. Alonazi *et al.*, "Automated facial emotion recognition using the pelican optimization algorithm with a deep convolutional neural network," *Electronics*, vol. 12, Art. no. 4608, 2023, doi: 10.3390/electronics12224608.

[8] M. F. Alsharekh, "Facial emotion recognition in verbal communication based on deep learning," *Sensors*, vol. 22, Art. no. 6105, 2022, doi: 10.3390/s22166105.

[9] S. Alsubai, A. Alqahtani, A. Alanazi, M. Sha, and A. Gumaei, "Facial emotion recognition using deep quantum and advanced transfer learning mechanism," *Front. Comput. Neurosci.*, vol. 18, Art. no. 1435956, 2024, doi: 10.3389/fncom.2024.1435956.

[10] A. A. Alzahrani, "Bioinspired image processing enabled facial emotion recognition using equilibrium optimizer with a hybrid deep learning model," *IEEE Access*, vol. 12, pp. 22219–22229, 2024, doi: 10.1109/ACCESS.2024.3359436.

[11] D. D. Anjani Suputri Devi and Ch. Satyanarayana, "An efficient facial emotion recognition system using novel deep learning neural network-regression activation classifier," *Multimed. Tools Appl.*, vol. 80, pp. 17543–17568, 2021, doi: 10.1007/s11042-021-10547-2.

[12] B. Bakariya, A. Singh, H. Singh, P. Raju, R. Rajpoot, and K. K. Mohbey, "Facial emotion recognition and music recommendation system using CNN-based deep learning techniques," *Evolving Systems*, vol. 15, pp. 641–658, 2024, doi: 10.1007/s12530-023-09506-z.

[13] V. S. Bhati, N. Tiwari, and M. Chawla, "A generalized zero-shot deep learning classifier for emotion recognition using facial expression images," *IEEE Access*, vol. 13, pp. 18687–18700, 2025, doi: 10.1109/ACCESS.2025.3533580.

[14] C. Bisogni, L. Cimmino, M. De Marsico, F. Hao, and F. Narducci, "Emotion recognition at a distance: The robustness of machine learning based on hand-crafted facial features vs deep learning models," *Image and Vision Computing*, vol. 136, Art. no. 104724, 2023, doi: 10.1016/j.imavis.2023.104724.

[15] A. Bohi, Y. E. Boudouri, and I. Sfeir, "A novel deep learning approach for facial emotion recognition: Application to detecting emotional responses in elderly individuals with Alzheimer's disease," *Neural Comput. & Applic.*, vol. 37, pp. 5235–5253, 2025, doi: 10.1007/s00521-024-10938-0.

[16] 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, pp. 5693–5718, 2023, doi: 10.1007/s00371-022-02690-0.

[17] E. G. Dada, D. O. Oyewola, S. B. Joseph, O. Emebo, and O. O. Oluwagbemi, "Facial emotion recognition and classification using the convolutional neural network-10 (CNN-10)," *Applied Computational Intelligence and Soft Computing*, vol. 2023, pp. 1–19, 2023, doi: 10.1155/2023/2457898.

[18] W. Dias *et al.*, "Cross-dataset emotion recognition from facial expressions through convolutional neural networks," *Journal of Visual Communication and Image Representation*, vol. 82, Art. no. 103395, 2022, doi: 10.1016/j.jvcir.2021.103395.

[19] U. Dudekula and P. Nalluri, "Analysis of facial emotion recognition rate for real-time application using NVIDIA Jetson Nano in deep learning models," *IJEECS*, vol. 30, p. 598, 2023, doi: 10.11591/ijeecs.v30.i1.pp598-605.

[20] H. El Kabtane, F. Z. Ennaji, K. Elkabtane, M. Knouzi, and A. Ourdou, "Toward emotionally intelligent e-learning: Real-time engagement detection through fine-tuned deep facial emotion recognition," *Journal of the Franklin Institute*, vol. 363, Art. no. 108477, 2026, doi: 10.1016/j.jfranklin.2026.108477.

[21] S. Fekri-Ershad, "Encoding using three-channel deep convolutional neural network and decoding in multi-layer perceptron for facial emotion recognition," *Neural Comput. & Applic.*, vol. 38, Art. no. 57, 2026, doi: 10.1007/s00521-025-11797-z.

[22] S. Gupta, P. Kumar, and R. K. Tekchandani, "Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models," *Multimed. Tools Appl.*, vol. 82, pp. 11365–11394, 2023, doi: 10.1007/s11042-022-13558-9.

[23] A. Hassouneh, A. M. Mutawa, and M. Murugappan, "Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods," *Informatics in Medicine Unlocked*, vol. 20, Art. no. 100372, 2020, doi: 10.1016/j.imu.2020.100372.

[24] R. Helaly, S. Messaoud, S. Bouaafia, M. A. Hajjaji, and A. Mtibaa, "DTL-I-ResNet18: Facial emotion recognition based on deep transfer learning and improved ResNet18," *SIViP*, vol. 17, pp. 2731–2744, 2023, doi: 10.1007/s11760-023-02490-6.

[25] M. Hongwei, W. Xue, and H. Kai, "Facial emotion recognition method based on convolutional neural network," *IET Computers & Digital Techniques*, vol. 2025, Art. no. 1845378, 2025, doi: 10.1049/cdt2/1845378.

[26] J. C. Hung and J.-W. Chang, "Multi-level transfer learning for improving the performance of deep neural networks: Theory and practice from the tasks of facial emotion recognition and named entity recognition," *Applied Soft Computing*, vol. 109, Art. no. 107491, 2021, doi: 10.1016/j.asoc.2021.107491.

[27] D. K. Jain *et al.*, "An automated hyperparameter tuned deep learning model enabled facial emotion recognition for autonomous vehicle drivers," *Image and Vision Computing*, vol. 133, Art. no. 104659, 2023, doi: 10.1016/j.imavis.2023.104659.

[28] A. R. Khan, "Facial emotion recognition using conventional machine learning and deep learning methods: Current achievements, analysis and remaining challenges," *Information*, vol. 13, Art. no. 268, 2022, doi: 10.3390/info13060268.

[29] J. H. Kim, A. Poulose, and D. S. Han, "CVGG-19: Customized visual geometry group deep learning architecture for facial emotion recognition," *IEEE Access*, vol. 12, pp. 41557–41578, 2024, doi: 10.1109/ACCESS.2024.3377235.

[30] N. Kumari and R. Bhatia, "Deep learning based efficient emotion recognition technique for facial images," *Int. J. Syst. Assur. Eng. Manag.*, vol. 14, pp. 1421–1436, 2023, doi: 10.1007/s13198-023-01945-w.

[31] N. Kumari and R. Bhatia, "Saliency map and deep learning based efficient facial emotion recognition technique for facial images," *Multimed. Tools Appl.*, vol. 83, pp. 36841–36864, 2023, doi: 10.1007/s11042-023-16220-0.

[32] I. Lasri, A. Riadsolh, and M. Elbelkacemi, "Facial emotion recognition of deaf and hard-of-hearing students for engagement detection using deep learning," *Educ. Inf. Technol.*, vol. 28, pp. 4069–4092, 2023, doi: 10.1007/s10639-022-11370-4.
[33] S. H. Lee, "Facial data visualization for improved deep learning based emotion recognition," *Journal of Information Science Theory and Practice*, vol. 7, pp. 32–39, 2019, doi: 10.1633/JISTAP.2019.7.2.3.
[34] S. Li and W. Deng, "Blended emotion in-the-wild: Multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning," *Int. J. Comput. Vis.*, vol. 127, pp. 884–906, 2019, doi: 10.1007/s11263-018-1131-1.
[35] S.-Y. Lin, C.-M. Wu, S.-L. Chen, T.-L. Lin, and Y.-W. Tseng, "Continuous facial emotion recognition method based on deep learning of academic emotions," *Sensors and Materials*, vol. 32, p. 3243, 2020, doi: 10.18494/SAM.2020.2863.
[36] T.-T. Lu, S.-C. Yeh, C.-H. Wang, and M.-R. Wei, "Cost-effective real-time recognition for human emotion-age-gender using deep learning with normalized facial cropping preprocess," *Multimed. Tools Appl.*, vol. 80, pp. 19845–19866, 2021, doi: 10.1007/s11042-021-10673-x.
[37] Y. Lu, Z. Zhao, L. Yan, and X. Shi, "Deep learning-based bimodal speech and facial expression recognition of miners' unsafe emotions," *PLoS One*, vol. 21, Art. no. e0348906, 2026, doi: 10.1371/journal.pone.0348906.
[38] S. M, U. M, T. K, D. V, and M. C, "SiaCon-DetNet with HySHO: A cutting-edge transformer-based deep learning framework for emotion-aware facial recognition," *Sci. Rep.*, vol. 16, Art. no. 14131, 2026, doi: 10.1038/s41598-026-41890-9.
[39] J. L. Mazher Iqbal *et al.*, "Facial emotion recognition using geometrical features based deep learning techniques," *Int. J. Comput. Commun. Control*, vol. 18, 2023, doi: 10.15837/ijccc.2023.4.4644.
[40] L. Mednini and Z. Noubigh, "Deep learning-based facial emotion recognition for detecting brand hate," *JTDE*, vol. 13, pp. 244–267, 2025, doi: 10.18080/jtde.v13n1.1067.
[41] N. Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)," *SN Appl. Sci.*, vol. 2, Art. no. 446, 2020, doi: 10.1007/s42452-020-2234-1.
[42] M. Mukhiddinov, O. Djuraev, F. Akhmedov, A. Mukhamadiyev, and J. Cho, "Masked face emotion recognition based on facial landmarks and deep learning approaches for visually impaired people," *Sensors*, vol. 23, Art. no. 1080, 2023, doi: 10.3390/s23031080.
[43] P. Naga Bhushanam and S. Selva Kumar, "Retraction note: Modelling an efficient hybridized approach for facial emotion recognition using unconstraint videos and deep learning approaches," *Soft Comput.*, 2026, doi: 10.1007/s00500-026-11251-9.
[44] D. Pruthviraja, U. Mohan Kumar, S. Parameswaran, V. Guna Chowdary, and V. Bharadwaj, "Deep convolutional neural network architecture for facial emotion recognition," *PeerJ Computer Science*, vol. 10, Art. no. e2339, 2024, doi: 10.7717/peerj-cs.2339.
[45] P. Radočaj and G. Martinović, "Emotion recognition in autistic children through facial expressions using advanced deep learning architectures," *Applied Sciences*, vol. 15, Art. no. 9555, 2025, doi: 10.3390/app15179555.
[46] R. Raj and I. Demirkol, "An improved facial emotion recognition system using convolutional neural network for the optimization of human robot interaction," *Sci. Rep.*, vol. 15, Art. no. 38940, 2025, doi: 10.1038/s41598-025-22835-0.
[47] J. A. Ramirez-Quintana, J. J. Muñoz-Pacheco, G. Ramirez-Alonso, J. A. Medrano-Hermosillo, and A. D. Corral-Saenz, "Lightweight convolutional neural network with efficient channel attention mechanism for real-time facial emotion recognition in embedded systems," *Sensors*, vol. 25, Art. no. 7264, 2025, doi: 10.3390/s25237264.
[48] T. P. Rao *et al.*, "Oppositional brain storm optimization with deep learning based facial emotion recognition for autonomous intelligent systems," *IEEE Access*, vol. 12, pp. 44278–44285, 2024, doi: 10.1109/ACCESS.2024.3374893.
[49] 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. & Applic.*, vol. 29, pp. 359–373, 2018, doi: 10.1007/s00521-018-3358-8.
[50] Y. Said and M. Barr, "Human emotion recognition based on facial expressions via deep learning on high-resolution images," *Multimed. Tools Appl.*, vol. 80, pp. 25241–25253, 2021, doi: 10.1007/s11042-021-10918-9.
[51] G. L. Sălăgean, M. Leba, and A. C. Ionica, "Leveraging symmetry and addressing asymmetry challenges for improved convolutional neural network-based facial emotion recognition," *Symmetry*, vol. 17, Art. no. 397, 2025, doi: 10.3390/sym17030397.
[52] S. Saurav, R. Saini, and S. Singh, "EmNet: A deep integrated convolutional neural network for facial emotion recognition in the wild," *Appl. Intell.*, vol. 51, pp. 5543–5570, 2021, doi: 10.1007/s10489-020-02125-0.
[53] A. Sherly Alphonse, S. Abinaya, and S. Abirami, "Alibaba and forty thieves algorithm and novel prioritized Prewitt pattern (PPP)-based convolutional neural network (CNN) using hyperspherically compressed

weights for facial emotion recognition," *Journal of Visual Communication and Image Representation*, vol. 97, Art. no. 103948, 2023, doi: 10.1016/j.jvcir.2023.103948.
[54] A. Sultana, S. K. Dey, and Md. A. Rahman, "Facial emotion recognition based on deep transfer learning approach," *Multimed. Tools Appl.*, vol. 82, pp. 44175–44189, 2023, doi: 10.1007/s11042-023-15570-z.
[55] F. M. Talaat, "Real-time facial emotion recognition system among children with autism based on deep learning and IoT," *Neural Comput. & Applic.*, vol. 35, pp. 12717–12728, 2023, doi: 10.1007/s00521-023-08372-9.
[56] E. Tsalera, A. Papadakis, M. Samarakou, and I. Voyiatzis, "Feature extraction with handcrafted methods and convolutional neural networks for facial emotion recognition," *Applied Sciences*, vol. 12, Art. no. 8455, 2022, doi: 10.3390/app12178455.
[57] K.-J. Wang, C.-N. Hsu, and L. Sanjaya, "Empowering facial emotion recognition in service industry – A two-stage convolutional neural network model," *Multimed. Tools Appl.*, vol. 83, pp. 33161–33184, 2023, doi: 10.1007/s11042-023-16717-8.
[58] S. P. Xavier and S. P. John, "Advanced emotion recognition from facial images and speech signals using complex-value spatio-temporal graph convolutional neural networks," *SIViP*, vol. 19, Art. no. 939, 2025, doi: 10.1007/s11760-025-04449-1.
[59] N. Yalcin and M. Alisawi, "Facial emotion recognition through a smart glasses prototype: Improving social interaction for visually impaired users through enhanced deep learning CBAM architectures," *Applied Sciences*, vol. 16, Art. no. 4415, 2026, doi: 10.3390/app16094415.
[60] N. Yalçin and M. Alisawi, "Introducing a novel dataset for facial emotion recognition and demonstrating significant enhancements in deep learning performance through pre-processing techniques," *Heliyon*, vol. 10, Art. no. e38913, 2024, doi: 10.1016/j.heliyon.2024.e38913.
[61] D. Zhu, Y. Fu, X. Zhao, X. Wang, and H. Yi, "Facial emotion recognition using a novel fusion of convolutional neural network and local binary pattern in crime investigation," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–14, 2022, doi: 10.1155/2022/2249417.
[1] F. Aghabeigi, S. Nazari, and N. Osati Eraghi, "An efficient facial emotion recognition using convolutional neural network with local sorting binary pattern and whale optimization algorithm," *Int. J. Data Sci. Anal.*, vol. 20, pp. 2275–2290, 2025, doi: 10.1007/s41060-024-00601-1.
[2] E. S. Agung, A. P. Rifai, and T. Wijayanto, "Image-based facial emotion recognition using convolutional neural network on Emognition dataset," *Sci. Rep.*, vol. 14, Art. no. 14429, 2024, doi: 10.1038/s41598-024-65276-x.
[3] M. A. H. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, "Facial emotion recognition using transfer learning in the deep CNN," *Electronics*, vol. 10, Art. no. 1036, 2021, doi: 10.3390/electronics10091036.
[4] A. O. Alabi, B.-G. Song, and N. Kang, "A comprehensive investigation of auditory-induced emotions combining psychoacoustic metrics, facial emotion recognition, and explainable deep learning," *Expert Systems with Applications*, vol. 322, Art. no. 132385, 2026, doi: 10.1016/j.eswa.2026.132385.
[5] H. N. AlEisa *et al.*, "Henry gas solubility optimization with deep learning based facial emotion recognition for human computer interface," *IEEE Access*, vol. 11, pp. 62233–62241, 2023, doi: 10.1109/ACCESS.2023.3284457.
[6] M. Almubarak and F. A. Alsulaiman, "An ensemble learning approach for facial emotion recognition based on deep learning techniques," *Electronics*, vol. 14, Art. no. 3415, 2025, doi: 10.3390/electronics14173415.
[7] M. Alonazi *et al.*, "Automated facial emotion recognition using the pelican optimization algorithm with a deep convolutional neural network," *Electronics*, vol. 12, Art. no. 4608, 2023, doi: 10.3390/electronics12224608.
[8] M. F. Alsharekh, "Facial emotion recognition in verbal communication based on deep learning," *Sensors*, vol. 22, Art. no. 6105, 2022, doi: 10.3390/s22166105.
[9] S. Alsubai, A. Alqahtani, A. Alanazi, M. Sha, and A. Gumaei, "Facial emotion recognition using deep quantum and advanced transfer learning mechanism," *Front. Comput. Neurosci.*, vol. 18, Art. no. 1435956, 2024, doi: 10.3389/fncom.2024.1435956.
[10] A. A. Alzahrani, "Bioinspired image processing enabled facial emotion recognition using equilibrium optimizer with a hybrid deep learning model," *IEEE Access*, vol. 12, pp. 22219–22229, 2024, doi: 10.1109/ACCESS.2024.3359436.
[11] D. D. Anjani Suputri Devi and Ch. Satyanarayana, "An efficient facial emotion recognition system using novel deep learning neural network-regression activation classifier," *Multimed. Tools Appl.*, vol. 80, pp. 17543–17568, 2021, doi: 10.1007/s11042-021-10547-2.
[12] B. Bakariya, A. Singh, H. Singh, P. Raju, R. Rajpoot, and K. K. Mohbey, "Facial emotion recognition and music recommendation system using CNN-based deep learning techniques," *Evolving Systems*, vol. 15, pp. 641–658, 2024, doi: 10.1007/s12530-023-09506-z.
[13] V. S. Bhati, N. Tiwari, and M. Chawla, "A generalized zero-shot deep learning classifier for emotion recognition using facial expression images," *IEEE Access*, vol. 13, pp. 18687–18700, 2025, doi: 10.1109/ACCESS.2025.3533580.

[14] C. Bisogni, L. Cimmino, M. De Marsico, F. Hao, and F. Narducci, "Emotion recognition at a distance: The robustness of machine learning based on hand-crafted facial features vs deep learning models," *Image and Vision Computing*, vol. 136, Art. no. 104724, 2023, doi: 10.1016/j.imavis.2023.104724.
[15] A. Bohi, Y. E. Boudouri, and I. Sfeir, "A novel deep learning approach for facial emotion recognition: Application to detecting emotional responses in elderly individuals with Alzheimer's disease," *Neural Comput. & Applic.*, vol. 37, pp. 5235–5253, 2025, doi: 10.1007/s00521-024-10938-0.
[16] 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, pp. 5693–5718, 2023, doi: 10.1007/s00371-022-02690-0.
[17] E. G. Dada, D. O. Oyewola, S. B. Joseph, O. Emebo, and O. O. Oluwagbemi, "Facial emotion recognition and classification using the convolutional neural network-10 (CNN-10)," *Applied Computational Intelligence and Soft Computing*, vol. 2023, pp. 1–19, 2023, doi: 10.1155/2023/2457898.
[18] W. Dias *et al.*, "Cross-dataset emotion recognition from facial expressions through convolutional neural networks," *Journal of Visual Communication and Image Representation*, vol. 82, Art. no. 103395, 2022, doi: 10.1016/j.jvcir.2021.103395.
[19] U. Dudekula and P. Nalluri, "Analysis of facial emotion recognition rate for real-time application using NVIDIA Jetson Nano in deep learning models," *IJEECS*, vol. 30, p. 598, 2023, doi: 10.11591/ijeecs.v30.i1.pp598-605.
[20] H. El Kabtane, F. Z. Ennaji, K. Elkabtane, M. Knouzi, and A. Ourdou, "Toward emotionally intelligent e-learning: Real-time engagement detection through fine-tuned deep facial emotion recognition," *Journal of the Franklin Institute*, vol. 363, Art. no. 108477, 2026, doi: 10.1016/j.jfranklin.2026.108477.
[21] S. Fekri-Ershad, "Encoding using three-channel deep convolutional neural network and decoding in multi-layer perceptron for facial emotion recognition," *Neural Comput. & Applic.*, vol. 38, Art. no. 57, 2026, doi: 10.1007/s00521-025-11797-z.
[22] S. Gupta, P. Kumar, and R. K. Tekchandani, "Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models," *Multimed. Tools Appl.*, vol. 82, pp. 11365–11394, 2023, doi: 10.1007/s11042-022-13558-9.
[23] A. Hassouneh, A. M. Mutawa, and M. Murugappan, "Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods," *Informatics in Medicine Unlocked*, vol. 20, Art. no. 100372, 2020, doi: 10.1016/j.imu.2020.100372.
[24] R. Helaly, S. Messaoud, S. Bouaafia, M. A. Hajjaji, and A. Mtibaa, "DTL-I-ResNet18: Facial emotion recognition based on deep transfer learning and improved ResNet18," *SIViP*, vol. 17, pp. 2731–2744, 2023, doi: 10.1007/s11760-023-02490-6.
[25] M. Hongwei, W. Xue, and H. Kai, "Facial emotion recognition method based on convolutional neural network," *IET Computers & Digital Techniques*, vol. 2025, Art. no. 1845378, 2025, doi: 10.1049/cdt2/1845378.
[26] J. C. Hung and J.-W. Chang, "Multi-level transfer learning for improving the performance of deep neural networks: Theory and practice from the tasks of facial emotion recognition and named entity recognition," *Applied Soft Computing*, vol. 109, Art. no. 107491, 2021, doi: 10.1016/j.asoc.2021.107491.
[27] D. K. Jain *et al.*, "An automated hyperparameter tuned deep learning model enabled facial emotion recognition for autonomous vehicle drivers," *Image and Vision Computing*, vol. 133, Art. no. 104659, 2023, doi: 10.1016/j.imavis.2023.104659.
[28] A. R. Khan, "Facial emotion recognition using conventional machine learning and deep learning methods: Current achievements, analysis and remaining challenges," *Information*, vol. 13, Art. no. 268, 2022, doi: 10.3390/info13060268.
[29] J. H. Kim, A. Poulose, and D. S. Han, "CVGG-19: Customized visual geometry group deep learning architecture for facial emotion recognition," *IEEE Access*, vol. 12, pp. 41557–41578, 2024, doi: 10.1109/ACCESS.2024.3377235.
[30] N. Kumari and R. Bhatia, "Deep learning based efficient emotion recognition technique for facial images," *Int. J. Syst. Assur. Eng. Manag.*, vol. 14, pp. 1421–1436, 2023, doi: 10.1007/s13198-023-01945-w.
[31] N. Kumari and R. Bhatia, "Saliency map and deep learning based efficient facial emotion recognition technique for facial images," *Multimed. Tools Appl.*, vol. 83, pp. 36841–36864, 2023, doi: 10.1007/s11042-023-16220-0.
[32] I. Lasri, A. Riadsolh, and M. Elbelkacemi, "Facial emotion recognition of deaf and hard-of-hearing students for engagement detection using deep learning," *Educ. Inf. Technol.*, vol. 28, pp. 4069–4092, 2023, doi: 10.1007/s10639-022-11370-4.
[33] S. H. Lee, "Facial data visualization for improved deep learning based emotion recognition," *Journal of Information Science Theory and Practice*, vol. 7, pp. 32–39, 2019, doi: 10.1633/JISTAP.2019.7.2.3.
[34] S. Li and W. Deng, "Blended emotion in-the-wild: Multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning," *Int. J. Comput. Vis.*, vol. 127, pp. 884–906, 2019, doi: 10.1007/s11263-018-1131-1.

[35] S.-Y. Lin, C.-M. Wu, S.-L. Chen, T.-L. Lin, and Y.-W. Tseng, "Continuous facial emotion recognition method based on deep learning of academic emotions," *Sensors and Materials*, vol. 32, p. 3243, 2020, doi: 10.18494/SAM.2020.2863.

[36] T.-T. Lu, S.-C. Yeh, C.-H. Wang, and M.-R. Wei, "Cost-effective real-time recognition for human emotion-age-gender using deep learning with normalized facial cropping preprocess," *Multimed. Tools Appl.*, vol. 80, pp. 19845–19866, 2021, doi: 10.1007/s11042-021-10673-x.

[37] Y. Lu, Z. Zhao, L. Yan, and X. Shi, "Deep learning-based bimodal speech and facial expression recognition of miners' unsafe emotions," *PLoS One*, vol. 21, Art. no. e0348906, 2026, doi: 10.1371/journal.pone.0348906.

[38] S. M, U. M, T. K, D. V, and M. C, "SiaCon-DetNet with HySHO: A cutting-edge transformer-based deep learning framework for emotion-aware facial recognition," *Sci. Rep.*, vol. 16, Art. no. 14131, 2026, doi: 10.1038/s41598-026-41890-9.

[39] J. L. Mazher Iqbal *et al.*, "Facial emotion recognition using geometrical features based deep learning techniques," *Int. J. Comput. Commun. Control*, vol. 18, 2023, doi: 10.15837/ijccc.2023.4.4644.

[40] L. Mednini and Z. Noubigh, "Deep learning-based facial emotion recognition for detecting brand hate," *JTDE*, vol. 13, pp. 244–267, 2025, doi: 10.18080/jtde.v13n1.1067.

[41] N. Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)," *SN Appl. Sci.*, vol. 2, Art. no. 446, 2020, doi: 10.1007/s42452-020-2234-1.

[42] M. Mukhiddinov, O. Djuraev, F. Akhmedov, A. Mukhamadiyev, and J. Cho, "Masked face emotion recognition based on facial landmarks and deep learning approaches for visually impaired people," *Sensors*, vol. 23, Art. no. 1080, 2023, doi: 10.3390/s23031080.

[43] P. Naga Bhushanam and S. Selva Kumar, "Retraction note: Modelling an efficient hybridized approach for facial emotion recognition using unconstraint videos and deep learning approaches," *Soft Comput.*, 2026, doi: 10.1007/s00500-026-11251-9.

[44] D. Pruthviraja, U. Mohan Kumar, S. Parameswaran, V. Guna Chowdary, and V. Bharadwaj, "Deep convolutional neural network architecture for facial emotion recognition," *PeerJ Computer Science*, vol. 10, Art. no. e2339, 2024, doi: 10.7717/peerj-cs.2339.

[45] P. Radočaj and G. Martinović, "Emotion recognition in autistic children through facial expressions using advanced deep learning architectures," *Applied Sciences*, vol. 15, Art. no. 9555, 2025, doi: 10.3390/app15179555.

[46] R. Raj and I. Demirkol, "An improved facial emotion recognition system using convolutional neural network for the optimization of human robot interaction," *Sci. Rep.*, vol. 15, Art. no. 38940, 2025, doi: 10.1038/s41598-025-22835-0.

[47] J. A. Ramirez-Quintana, J. J. Muñoz-Pacheco, G. Ramirez-Alonso, J. A. Medrano-Hermosillo, and A. D. Corral-Saenz, "Lightweight convolutional neural network with efficient channel attention mechanism for real-time facial emotion recognition in embedded systems," *Sensors*, vol. 25, Art. no. 7264, 2025, doi: 10.3390/s25237264.

[48] T. P. Rao *et al.*, "Oppositional brain storm optimization with deep learning based facial emotion recognition for autonomous intelligent systems," *IEEE Access*, vol. 12, pp. 44278–44285, 2024, doi: 10.1109/ACCESS.2024.3374893.

[49] 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. & Applic.*, vol. 29, pp. 359–373, 2018, doi: 10.1007/s00521-018-3358-8.

[50] Y. Said and M. Barr, "Human emotion recognition based on facial expressions via deep learning on high-resolution images," *Multimed. Tools Appl.*, vol. 80, pp. 25241–25253, 2021, doi: 10.1007/s11042-021-10918-9.

[51] G. L. Sălăgean, M. Leba, and A. C. Ionica, "Leveraging symmetry and addressing asymmetry challenges for improved convolutional neural network-based facial emotion recognition," *Symmetry*, vol. 17, Art. no. 397, 2025, doi: 10.3390/sym17030397.

[52] S. Saurav, R. Saini, and S. Singh, "EmNet: A deep integrated convolutional neural network for facial emotion recognition in the wild," *Appl. Intell.*, vol. 51, pp. 5543–5570, 2021, doi: 10.1007/s10489-020-02125-0.

[53] A. Sherly Alphonse, S. Abinaya, and S. Abirami, "Alibaba and forty thieves algorithm and novel prioritized Prewitt pattern (PPP)-based convolutional neural network (CNN) using hyperspherically compressed weights for facial emotion recognition," *Journal of Visual Communication and Image Representation*, vol. 97, Art. no. 103948, 2023, doi: 10.1016/j.jvcir.2023.103948.

[54] A. Sultana, S. K. Dey, and Md. A. Rahman, "Facial emotion recognition based on deep transfer learning approach," *Multimed. Tools Appl.*, vol. 82, pp. 44175–44189, 2023, doi: 10.1007/s11042-023-15570-z.

[55] F. M. Talaat, "Real-time facial emotion recognition system among children with autism based on deep learning and IoT," *Neural Comput. & Applic.*, vol. 35, pp. 12717–12728, 2023, doi: 10.1007/s00521-023-08372-9.

[56] E. Tsalera, A. Papadakis, M. Samarakou, and I. Voyiatzis, “Feature extraction with handcrafted methods and convolutional neural networks for facial emotion recognition,” *Applied Sciences*, vol. 12, Art. no. 8455, 2022, doi: 10.3390/app12178455.

[57] K.-J. Wang, C.-N. Hsu, and L. Sanjaya, “Empowering facial emotion recognition in service industry – A two-stage convolutional neural network model,” *Multimed. Tools Appl.*, vol. 83, pp. 33161–33184, 2023, doi: 10.1007/s11042-023-16717-8.

[58] S. P. Xavier and S. P. John, “Advanced emotion recognition from facial images and speech signals using complex-value spatio-temporal graph convolutional neural networks,” *SIViP*, vol. 19, Art. no. 939, 2025, doi: 10.1007/s11760-025-04449-1.

[59] N. Yalcin and M. Alisawi, “Facial emotion recognition through a smart glasses prototype: Improving social interaction for visually impaired users through enhanced deep learning CBAM architectures,” *Applied Sciences*, vol. 16, Art. no. 4415, 2026, doi: 10.3390/app16094415.

[60] N. Yalçin and M. Alisawi, “Introducing a novel dataset for facial emotion recognition and demonstrating significant enhancements in deep learning performance through pre-processing techniques,” *Heliyon*, vol. 10, Art. no. e38913, 2024, doi: 10.1016/j.heliyon.2024.e38913.

[61] D. Zhu, Y. Fu, X. Zhao, X. Wang, and H. Yi, “Facial emotion recognition using a novel fusion of convolutional neural network and local binary pattern in crime investigation,” *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–14, 2022, doi: 10.1155/2022/2249417.