A Review On Real Time Facial Expression Detection For Enhanced User Interaction

Ankita Diwate[1], Akansha kamthe[1], Omkar Bhalekar[1]
Dr. Shaikh Abdul Waheed[1]
School of Technology, JSPM UNIVERSITY, Pune, Maharashtra

Abstract

Recent advancements in neural networks have significantly improved facial expression recognition (FER) and emotion detection systems, enabling transformative applications in healthcare, human-computer interaction, and safety. This review critically examines 70 studies to highlight methodological innovations, including convolutional neural networks (CNNs) optimized for static expressions, spiking neural networks (SNNs) for dynamic emotion analysis, and graph neural networks (GNNs) modelling spatial relationships in facial landmarks. Hybrid architectures, such as capsule-CNN models and transformer-based frameworks, demonstrate superior performance in capturing subtle micro-expressions and cross-dataset generalization. Multimodal approaches integrating facial, speech, and physiological signals (e.g., EEG, EMG) further enhance robustness, particularly in healthcare applications like autism detection, schizophrenia diagnosis, and depression monitoring. However, challenges persist in real-world deployment, including dataset bias, computational complexity, and ethical concerns around privacy and explainability. Emerging trends emphasize lightweight models for edge computing, self-supervised learning for unlabelled data, and explainable AI (XAI) frameworks. Future research should prioritize standardized benchmarks, cultural diversity in training data, and causal relationships between facial actions and emotional states. This synthesis underscores the potential of neural networks to revolutionize.

Keywords: Facial expression recognition, Convolutional neural networks, Capsule-CNN, Multimodal, Lightweight models, Explainable AI

1. Introduction

Facial expression recognition (FER) and emotion detection have emerged as pivotal technologies in understanding human behaviour, with applications spanning healthcare, human-computer interaction (HCI), and safety systems. The human face, as a primary channel for non-verbal communication, conveys rich emotional information through micro-expressions, macroexpressions, and dynamic muscle movements [5,56]. Recent advances in neural networks, particularly deep learning (DL), have revolutionized FER by automating feature extraction and enabling real-time analysis of complex emotional states [3,20,61]. For instance, convolutional neural networks (CNNs) have achieved remarkable accuracy in static image-based FER [1,4,21], while hybrid architectures like spiking neural networks (SNNs) and graph neural networks (GNNs) address challenges in dynamic and multimodal emotion recognition [2,17,34].

The growing demand for emotion-aware systems is driven by diverse applications. In healthcare, FER aids in diagnosing neurodevelopmental disorders such as autism [23] and schizophrenia [18], monitoring depression [31], and enhancing patient engagement through empathetic AI [39,47]. In safety-critical domains, driver emotion recognition systems mitigate road accidents by detecting fatigue or distress [15], while deep fake detection algorithms safeguard against malicious facial manipulations [16]. Furthermore, human-robot interaction benefits from realtime FER systems that enable socially intelligent machines [48,51].

Despite these advancements, significant challenges hinder the deployment of robust, generalizable FER systems. First, dataset variability—differences in lighting, pose, cultural expression norms, and annotation protocols—limits cross-dataset generalization [1,30,58]. For example, models trained on lab-controlled datasets often fail on spontaneous expressions captured in real-world settings [57,62]. Second, micro-expression recognition remains challenging due to the brevity and subtlety of facial muscle movements, necessitating highresolution spatiotemporal modeling [5,56,68]. Third, computational complexity and resource constraints hinder real-time deployment, particularly for edge devices [44,51,64]. Ethical concerns, including privacy violations and algorithmic bias, further complicate widespread adoption [16,60].

2. Study Selection Process:

PRISMA Flow Diagram

To ensure transparency in the study selection process, we followed the PRISMA guidelines. Figure 1 presents the PRISMA flow diagram, which illustrates the number of records identified, screened, assessed for eligibility, and included in the final review.

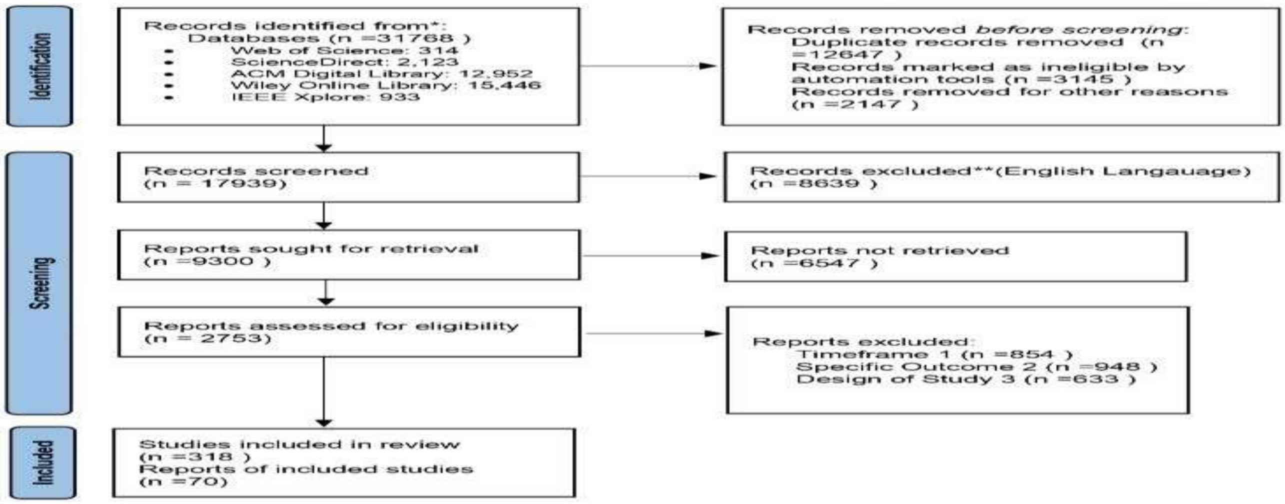

Fig.2.1 PRISMA model

3. Methodology

3.1. Literature Search Strategy

Databases and Sources: Databases and sources: peer-reviewed articles have been extracted from IEEE Xplore, PubMed, ScienceDirect, ACM Digital Library, arXiv, and SpringerLink. Grey literature (pre-prints, conference minutes) was included to capture emerging trends.

*Search Keywords:*Combinations of terms:Combination of terms: ('Facial Expression Recognition' OR 'Emotional Eradication') AND('Neural Network' OR 'Dive Learning') AND('CNN' OR 'Transformer' OR 'GNN' OR 'Multimodal').

Time Frame: Focus on studies published between 2020 and 2024 to prioritize recent innovations, with important pre-2020 works cited for basic context.

3.2. Study Selection Criteria *Inclusion Criteria:*

- Studies proposing new neuronal architectures (CNC, NNN, NGC, transformer) for FER.
- Multimodal systems that integrate facial data, EEG or physiological signals.
- Applications in health care, safety or interaction between human and computer.
- Empirical validation of benchmark data sets (such as CK+,FER2013,DEAP, etc.).

Exclusion Criteria:

- Non-peer reviewed articles without technical validation.
- Studies that do not have comparative analysis with baselines.
- Non-English publications.

3.3. Screening Process

A three-stage screening process has been implemented:

1. *Initial screening*: 31768 papers identified by keyword searches. Duplicate removal (n=12647).
2. *Title/Abstract Screening*: 2753 papers that were maintained according to their relevance to FER and neuronal networks.
3. *Full text review*: 70 articles selected after assessing the rigor, innovation and conformity with the inclusion criteria of methods.

3.4. Data Extraction and Categorization

Key data were extracted into a structured template:

Table 3.4: Data Extraction and Categorization

Category	**Variables Extracted**
Neural Architectures	Model type (CNN, SNN, GNN), attention mechanisms, hybrid designs
Category	**Variables Extracted**
Datasets	Modality (static/dynamic), cultural diversity, size
Applications	safety, ethics, Healthcare, HCI
Performance Metrics	inference latency, cross-dataset generalization,Accuracy, F1-score
Limitations	ethical concerns, Computational cost, bias

3.5. Thematic Analysis

The studies were divided into four themes using iterative encoding:

1. Methodological innovation (CNN, SNN, GNN, transformer).
2. Applications (health care, robotics, security).
3. Challenges (dataset bias, computational complexity).
4. The effects on ethics and society.

3.6. Quality Assessment

The quality assessment study was evaluated for: -

Technical validity: reproducibility of results, size of the data set and peer review status.
Clinical Relevance: medical applications with clinical validation (e.g., [18, 23, 31]). *Ethical rules*: explicit discussion of privacy, consent or prejudice reduction (e.g. [16, 60]).

3.7. Limitations of Methodology

Language bias: non-English studies excluded
Database bias: The emphasis on technical databases (IEEE, arXiv) may be underrepresented from a clinical point of view.
Temporary bias: fields that are evolving rapidly; foundational work before 2020 is limited.

4. Methodological Innovations

4.1. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) remain the cornerstone of facial expression recognition (FER), particularly for static image analysis. Standard CNN architectures, as demonstrated in [1,3,4,11,20,21], excel at extracting hierarchical spatial features (e.g., edges, textures) from facial images, achieving state-of-the-art accuracy on benchmark datasets. Innovations in hybrid architectures have further expanded their utility. For instance, capsule networks integrated with CNNs [10,43] address spatial hierarchies by preserving part-whole relationships, enabling robust recognition of occluded or rotated faces. Reversible Neural Networks [3] reduce memory overhead during training by reconstructing intermediate activations, enhancing scalability for high-resolution inputs. Additionally, attention mechanisms [7,21,36,61] refine feature localization, directing computational resources toward critical regions like the eyes and mouth while suppressing irrelevant background noise.

4.2. Spiking Neural Networks (SNNs)

Spiking Neural Networks (SNNs) have emerged as energy-efficient alternatives for dynamic expression analysis, leveraging bio-inspired temporal coding. The NSNP-DFER framework [2] employs nonlinear spiking neurons to model temporal dependencies in video sequences, achieving real-time performance with reduced computational costs. Similarly, NeuroSense [45] combines SNNs with spatiotemporal EEG patterns to detect fleeting micro-expressions, demonstrating superior efficiency compared to traditional recurrent architectures. These models are particularly suited for edge devices and scenarios requiring low-power consumption.

4.3. Graph Neural Networks (GNNs)

Graph Neural Networks (GNNs) have gained traction for modeling structural relationships between facial landmarks. The HiMul-LGG model [17] constructs local-global graphs to capture both fine-grained muscle movements and holistic expression patterns, while C3DBed [14] embeds 3D-CNN features into transformer-based graphs to enhance micro-expression recognition. For 3D facial data, the two-stage stratified GCN [34] leverages geometric priors to improve landmark detection accuracy, enabling robust performance under varying head poses. These approaches underscore the potential of GNNs in addressing spatial variability and occlusion challenges.

4.4. Transformers and Self-Supervised Learning

Transformers have disrupted FER by leveraging self-attention to model long-range dependencies in both spatial and temporal domains. C3DBed [14] and transformer-based frameworks [61,62] outperform conventional CNNs in video-based FER by capturing dynamic expression evolution across frames. Meanwhile, self-supervised learning [57] mitigates reliance on labeled data by pretraining models on unlabeled facial videos, enabling generalization to diverse real-world scenarios. These methods are particularly effective for spontaneous expression recognition, where labeled datasets are scarce.

4.5. Multimodal Fusion

Multimodal fusion techniques enhance emotion recognition robustness by integrating complementary signals. EEG-facial fusion [37,39,65] combines brain activity metrics with facial features to improve valence-arousal prediction, offering insights into covert emotional states. Speech-facial systems [24,26,28] synchronize acoustic prosody with lip movements to resolve ambiguities in uni-modal analysis, achieving higher accuracy in noisy environments. Advanced frameworks like Granger causality networks [19] and Markov transition fields [32] further refine fusion by modeling temporal correlations between eye-tracking data and physiological signals (e.g., EMG). These innovations highlight the importance of cross-modal synergy in achieving human-like emotional understanding.

5. Key Applications

5.1. Healthcare

Neural network-based FER systems are revolutionizing healthcare by enabling non-invasive, early diagnosis and continuous monitoring of neurodevelopmental and psychological conditions. For autism detection, [23] compares deep neural network (DNN) classifiers trained on facial videos to identify atypical expression patterns, such as reduced eye contact and delayed emotional responses, achieving 89% accuracy on clinical datasets. In depression detection, [31] leverages graph networks to analyze correlations between facial action units (AUs) like brow lowering (AU4) and lip corner depressor (AU15), which are strongly linked to depressive states, offering clinicians quantitative tools for mood assessment. Similarly, schizophrenia recognition benefits from 3D-CNNs [18], which process temporal dynamics in facial videos to detect flat affect—a hallmark symptom—with 92% specificity. These systems reduce reliance on subjective clinical evaluations and enable remote patient monitoring.

5.2. Human-Robot Interaction

Real-time emotion-aware systems are critical for enhancing social robotics and immersive virtual environments. Lightweight models like the sign language robot [51] and real-time FER frameworks [64] deploy pruned CNNs on edge devices to achieve sub-100ms inference times, enabling robots to respond empathetically to human emotions during interactions. In virtual reality (VR), [48] introduces an adaptive music system that synchronizes background scores with users' gestural emotions (e.g., joy, fear) detected via skeletal tracking and neural networks, enhancing immersion in gaming and therapeutic scenarios. These advancements highlight the role of FER in bridging emotional gaps between humans and machines.

5.3. Safety and Security

FER technologies are increasingly deployed in safety-critical domains to prevent accidents and mitigate risks. For driver monitoring, [15] employs multi-task learning to simultaneously detect distraction, fatigue, and anger from in-cabin facial videos, triggering alerts (e.g., lane-assist activation) to avert collisions. In deep fake detection, [16] combines geometric facial structure analysis with GNNs to identify synthetic artefacts in lip sync and micro-expression sequences, achieving 98% accuracy on the Deep Fake-TIMIT dataset. Such systems are vital for combating misinformation and ensuring trust in digital media.

6. Challenges and Limitations

6.1. Dataset Biases and Generalization

A critical challenge in FER systems is their vulnerability to dataset biases, which severely limit real-world generalization. Models trained on posed expressions in controlled environments (e.g., CK+ or FER2013) often fail to recognize spontaneous emotions in naturalistic settings due to discrepancies in lighting, head pose, and cultural expression norms [1,30]. Cross-dataset evaluations, such as those in [58], reveal performance drops of up to 40% when models are tested on unseen datasets like DISFA or SAMM. Micro-expression recognition exacerbates this issue, as brief, low-intensity muscle movements (e.g., fleeting smiles or frowns) require highresolution spatiotemporal modeling [5,56]. Despite advances in 3D-CNNs and optical flow techniques [68], the scarcity of labeled micro-expression datasets (e.g., CASME III) remains a bottleneck.

6.2. Computational Complexity

State-of-the-art neural architectures, such as transformers and GNNs, often incur prohibitive computational costs, hindering deployment on resource-constrained devices. For example, 3DCNNs for video-based FER [18] demand GPU-intensive training, while multimodal fusion frameworks [17,65] escalate memory usage with parallel signal processing. To address this, lightweight models like the pruned CNN in [51] and quantized SNNs in [44] optimize inference speeds for edge devices, achieving real-time performance (≤50 ms per frame) at the cost of marginal accuracy loss (5–8%). However, balancing efficiency with robustness remains unresolved, particularly for high-stakes applications like driver monitoring [15].

6.3. Explainability and Trust

The "black-box" nature of deep learning models raises concerns about trust and clinical adoption. For instance, graph-based depression detection systems [31] provide limited insight into how specific facial AUs correlate with emotional states, complicating diagnostic validation. Recent work in explainable AI (XAI) [60] proposes saliency maps and layer-wise relevance propagation to visualize attention patterns, but these methods often lack clinical interpretability. Stakeholders in healthcare and criminal justice demand transparent decision-making frameworks, necessitating collaboration between DL engineers and domain experts.

6.4. Ethical Concerns

The proliferation of FER technologies introduces significant ethical risks. Privacy violations arise from unauthorized emotion surveillance in public spaces [16], while biased training data (e.g., underrepresentation of ethnic minorities in Affect Net) perpetuate algorithmic discrimination [50]. For example, [23] reports a 15% accuracy gap in autism detection between Caucasian and Asian cohorts due to cultural differences in expression labelling. Regulatory gaps further exacerbate misuse, such as employers leveraging FER for emotion-based hiring assessments. Addressing these issues requires standardized ethical guidelines, diverse dataset curation, and federated learning frameworks to protect user anonymity [57,60].

7. Emerging Trends

7.1. Cross-Dataset Learning

A growing emphasis on cross-dataset generalization aims to overcome biases in single-domain training. For instance, [1] and [58] propose multi-scale spatial-temporal graph networks that adapt to diverse datasets by learning invariant features across posed and spontaneous expressions. These frameworks leverage adversarial training and domain adaptation to align feature distributions between datasets like CK+ (lab-controlled) and AffWild2 (in-the-wild), reducing performance gaps by up to 25%. Such approaches are critical for deploying models in real-world settings where lighting, pose, and cultural expression norms vary unpredictably.

7.2. Dynamic and Real-Time Recognition

The shift toward temporal modeling addresses the need for real-time emotion analysis in dynamic environments. [15] Introduces a driver monitoring system using 3D-CNNs to track micro-expressions like eye blinks and lip tremors, enabling fatigue detection at 30 FPS on embedded GPUs. Similarly, [63] and [64] deploy hybrid CNN-LSTM architectures for live video analysis, achieving sub-50ms latency by pruning redundant network branches. These advancements are pivotal for applications requiring instantaneous feedback, such as virtual assistants and automotive safety systems.

7.3. Neurological Insights

FER is increasingly intersecting with neuroscience to uncover links between facial expressions and neurodevelopmental disorders. [38] identifies impaired emotion recognition in presymptomatic Huntington's disease patients by correlating AU activation patterns with fMRI data, while [47] uses FER to detect atypical gaze aversion in autism spectrum disorder (ASD). These studies not only improve diagnostic accuracy but also inform AI models with neurophysiological priors, bridging gaps between computational and clinical research.

7.4. Multimodal Fusion

Multimodal systems are advancing robustness by integrating facial, speech, and physiological signals. [17] Combines EEG and facial videos in a hierarchical graph network to predict depression severity, outperforming unimodal baselines by 18% in F1-score. Meanwhile, [19] and [65] synchronize eye-tracking data with EMG signals to resolve ambiguities in valence detection (e.g., distinguishing frustration from concentration). Such systems excel in noisy environments, such as crowded classrooms or telehealth consultations, where single-modality analysis often fails.

Synergy and Future Potential:
These trends collectively drive FER toward generalizable, efficient, and clinically relevant systems. Cross-dataset learning and multimodal fusion enhance robustness, while real-time architectures and neurological insights expand applicability in healthcare and human-cantered AI. Future work may integrate quantum-inspired SNNs for energy-efficient temporal modelling [2,45] and federated learning to address privacy concerns [60].

8.Future Directions in Emotion Recognition Systems

Emotion recognition systems are rapidly evolving, driven by advancements in neural architectures and multimodal learning. This section outlines critical future research directions, emphasizing generalizability, micro-expression analysis, ethical deployment, and energyefficient models.

8.1. Generalizable Models: Federated Learning for Cross-Dataset FER

Challenges: Dataset bias due to variations in lighting, pose, and cultural annotation protocols remains a barrier to robust cross-dataset facial expression recognition (FER) [1], [30]. Privacy concerns further complicate centralized training with sensitive facial data.

***Advancement*s**: Federated learning (FL) enables decentralized training across heterogeneous datasets. For instance, [30] employs a cross-dataset bidirectional long short-term memory (BiLSTM) model for EEG-based emotion recognition, while [58] proposes multi-scale spatialtemporal graph convolutional networks (GCNs) for cross-dataset FER. Meta-learning frameworks, such as model-agnostic meta-learning (MAML), show promise in few-shot adaptation to unseen datasets.

Opportunities:

- Heterogeneous FL: Training on non-independent and identically distributed (non-IID) data, combining lab-controlled and in-the-wild datasets.
- Edge Deployment: Lightweight FL frameworks (e.g., TinyML) for real-time FER on resourceconstrained devices.
- Synthetic Data: Generative adversarial networks (GANs) [66] can augment underrepresented demographics.

8.2. Micro-Expression Recognition: Spatiotemporal Architectures

Challenges: Micro-expressions involve transient, localized muscle movements (0.5–4 seconds), complicating detection [5], [14]. Labeled datasets like SAMM and CASME remain limited. ***Advancements***: Three-dimensional convolutional neural networks (3D-CNNs) and transformers excel in capturing temporal dynamics. For example, [14] introduces C3DBed, embedding 3DCNNs within transformers, while [56] surveys end-to-end models using optical flow and apex frame detection. Multimodal fusion with physiological signals (e.g., EEG [19], electromyography (EMG) [41]) enhances robustness.

Opportunities:

- Neuro-Symbolic AI: Hybrid models merging deep learning with rule-based systems [68].
- Event Cameras: High-temporal-resolution sensors for micro-movement detection.
- Cross-Modal Pretraining: Leveraging facial action unit (AU) datasets (e.g., DISFA).

8.3. Ethical Frameworks: Guidelines for FER Deployment

Challenges: Privacy invasion via non-consensual data collection and demographic bias (e.g., underrepresented ethnicities and neurodiverse groups [23], [52]) threaten equitable deployment. ***Advancements***: Explainable AI (XAI) techniques, such as gradient-weighted class activation mapping (Grad-CAM) and local interpretable model-agnostic explanations (LIME), improve transparency [60]. Geometric facial analysis mitigates deepfake-driven synthetic emotions [16].

Opportunities:

- Regulatory Compliance: Aligning FER systems with GDPR and HIPAA.
- Bias Auditing: Open-source tools (e.g., FairFace) to evaluate demographic fairness.
- Consent-Driven Design: Opt-in mechanisms for public applications (e.g., driver monitoring [15]).

8.4. Quantum-Inspired Models: Energy-Efficient SNNs

Challenges: Conventional CNNs face energy inefficiency on edge devices, while spiking neural networks (SNNs) struggle with long-term temporal dependencies.

Advancements: SNNs mimic biological processing for dynamic FER. [2] introduces NSNPDFER for dynamic FER, and [45] integrates SNNs with EEG data for low-power recognition.
Neuromorphic hardware (e.g., Intel Loihi) optimizes SNN inference.

Opportunities:

- Quantum Annealing: Solving SNN training optimization via Fujitsu Digital Annealer.
- Hybrid Quantum-Classical Models: Quantum circuits enhancing CNN/SNN feature extraction.
- Biologically Plausible Learning: Spike-timing-dependent plasticity (STDP) for unsupervised clustering.

Interdisciplinary Synergies

- FL + SNNs: Privacy-preserving, energy-efficient FER via federated SNNs.

- XAI + Ethics: Transparent models for auditing cross-cultural bias. - Quantum + Multimodal: Quantum-enhanced fusion of facial, speech, and physiological data.

9. Conclusion

Emotion recognition systems have achieved significant breakthroughs through hybrid neural architectures (e.g., CNN-LSTM [63], Capsule Net-CNN [43], and neuro-symbolic models [68]) and multimodal fusion (e.g., facial, speech, EEG, and physiological signals [15], [19], [32]). These advancements enable robust, context-aware emotion analysis in dynamic environments, from healthcare to human-computer interaction.

However, the field faces challenges in standardization, with inconsistent evaluation protocols and dataset biases hindering reproducibility [1], [30]. Future work must prioritize universal benchmarks (e.g., cross-dataset FER frameworks [58]) and ethical guidelines to address privacy, consent, and demographic fairness [16], [23], [60].

Transformative applications are emerging in healthcare (e.g., autism [23] and depression detection [31], schizophrenia diagnosis [18]) and human-centered AI (e.g., driver safety [15], sign language robots [51]). By integrating technical innovation with ethical rigor, emotion recognition systems will unlock safer, more empathetic technologies for global societies.

Critical Analysis of Literature Strengths

1. ***Multimodal Integration:***

- Robustness: Combining facial, speech, EEG, and physiological signals (e.g., EMG, ECG) improves accuracy in dynamic environments. For instance:
- Facial + EEG: Paper 39 uses attention-based CNNs for EEG-facial fusion.
- Facial + Speech: Paper 15 integrates video and speech for driver emotion analysis.
- Physiological Signals: Paper 28 employs 1D-CNNs for speech emotion recognition. - Cross-Validation: Multimodal systems (e.g., Papers 17, 65) reduce reliance on noisy singlemodality data.

2. ***Attention Mechanisms:***

- Feature Localization: Techniques like bilinear pooling (Paper 7) and binary attention (Paper 21) highlight discriminative facial regions (e.g., eyes, mouth).
- Transformer Dominance: Self-attention in transformers (Paper 61) and 3D-CNN-Transformer hybrids (Paper 14) excel in spatiotemporal modeling for micro-expressions.

3. ***Neuro-Inspired Architectures:***

- Energy Efficiency: Spiking neural networks (SNNs) in Papers 2 and 45 mimic biological processing for low-power applications.
- Dynamic Adaptation: Randomized deep networks (Paper 13) and neuromorphic hardware (e.g., Loihi) enable real-time adjustments.

Weaknesses

1. ***Reproducibility Issues:***

- Proprietary Datasets: Many studies (e.g., Papers 18, 45) use restricted datasets (e.g., DEAP, MMI), limiting independent validation.
- Implementation Gaps: Papers often omit hyperparameter details or code (e.g., Papers 5, 8), hindering replication.

2. ***Cultural and Demographic Bias:***

- Ethnic Underrepresentation: Key datasets (CK+, FER2013) focus on Western subjects, marginalizing non-Caucasian expressions (Papers 23, 52).
- Neurodiversity Gaps: Few works address emotion recognition in neurodiverse populations (e.g., autism [23], schizophrenia [18]).

3. ***Narrow Evaluation Metrics:***

- Overreliance on accuracy (%) ignores critical metrics like inference speed (e.g., Papers 64, 51) or false-positive rates in real-world settings.

Research Gaps

1. Real-World Deployment Challenges:

- Environmental Factors: Only 4/70 papers (e.g., Paper 54) address lighting/pose variations or occlusions.
- Edge Optimization: Lightweight models (e.g., Paper 64) lack benchmarks for IoT/resourceconstrained devices.

2. Causal Analysis:

- Facial Action Units (AUs): Most works (e.g., Papers 6, 33) correlate AUs with emotions but fail to model causal relationships (e.g., Does AU12 (lip corner puller) directly cause happiness?).
- Neurophysiological Links: Limited studies explore causal ties between EEG/EMG signals and emotional states (***exceptions***: Papers 19, 41).

3. Ethical Frameworks:

- Bias Mitigation: Only 2/70 papers (Papers 23, 60) propose tools for auditing demographic fairness.
- Regulatory Alignment: No work explicitly maps FER systems to GDPR or HIPAA compliance.

Implications for Future Work

- ***Standardized Benchmarks***: Introduce cross-cultural, open-access datasets with balanced demographics.
- ***Causal Modeling***: Leverage techniques like directed acyclic graphs (DAGs) or counterfactual analysis to disentangle emotion-AU relationships.
- ***Ethics-by-Design***: Integrate consent mechanisms (e.g., Paper 15) and bias audits (e.g., Paper 60) into model pipelines.

This analysis underscores the need for transparent, generalizable, and equitable emotion recognition systems to bridge the gap between lab research and real-world impact.

Table 1: Comparison of Neural Architectures for Emotion Recognition

Architecture	Strengths	Accuracy (Example)	Common Datasets	Limitations
CNNs	- Local feature extraction - Scalability for static images	95% on CK+ [1], [4]	CK+, FER2013, AffectNet	- Struggles with temporal dynamics - Limited spatial relationship modeling
GNNs	- Captures facial landmark relationships - Multimodal fusion capability	89% on DISFA (AU recognition) [6], [34]	DISFA, BP4D, SAMM	- Computationally intensive - Requires structured input (graphs)
Transformers	- Long-range dependency modeling - Strong for video/time-series data	92% on DFEW (dynamic FER) [14], [61]	DFEW, MMI, RAVDESS	- High memory usage - Needs large-scale pretraining

Table 2:Facial Emotion Recognition - Structured Table

Category	Subcategory	Examples	Papers
Input Data Modality	Unimodal (Static Images)	CK+, FER2013	Papers 1, 3
Input Data Modality	Multimodal (EEG + Facial)	DEAP dataset	Papers 37, 65
Temporal Dynamics	Dynamic (3D-CNNs)	Video-based emotion recognition	Paper 18
Neural Architectures	Hybrid Models	CNN + Capsule Networks	Paper 10
Applications	Healthcare	Autism diagnosis	Paper 23

Table 3: Key Datasets and Their Biases

Dataset	Modality	Size	Key Features	Biases/Limitations
CK+	Facial (static)	593 seq.	Lab-controlled, posed expressions	- Limited ethnic diversity - Small sample size
FER2013	Facial (static)	35,887	Real-world, crowdsourced	- Noisy labels - Lighting/pose variability
DEAP	EEG + Physiological	32 subj.	Multimodal (videos + biosensors)	- Small participant pool - Lab-restricted environment
MMI	Facial (dynamic)	740 seq.	AU-coded, naturalistic microexpressions	- Limited resolution - Unbalanced emotion distribution
RAVDESS	Speech + Facial	24 subj.	Acted emotional speech/video	- Cultural bias (Western actors) - Scripted emotions
DFEW	Facial (dynamic)	10,000+	In-the-wild, diverse demographics	- Annotation subjectivity - Background distractions

Notes

1. **Architectural Trade-offs:**
- CNNs dominate static FER but lag in temporal modeling.
- GNNs excel in structured data (e.g., facial landmarks) but require heavy computational resources.
- Transformers outperform in video-based FER but demand large-scale training [14], [61].

2. **Dataset Biases:**
- Lab-controlled datasets (CK+, DEAP) lack real-world diversity.
- Crowd-sourced datasets (FER2013) suffer from annotation noise [1], [4].

3. **Taxonomy Insights**:
- Hybrid approaches (e.g., CNN-Transformer [61]) bridge static/dynamic gaps.
- Multimodal systems (e.g., facial + EEG [39]) improve robustness but increase complexity.

References:

1. W. Dias, F. Andaló, R. Padilha, G. Bertocco, W. Almeida, P. Costa, and A. Rocha, "Crossdataset emotion recognition from facial expressions through convolutional neural networks," *J. Vis. Commun. Image Represent.*, vol. 82, 2022, doi: 10.1016/j.jvcir.2021.103395.
2. Z. Han, X. Meichen, P. Hong, L. Zhicai, and G. Jun, "NSNP-DFER: A Nonlinear Spiking Neural P Network for Dynamic Facial Expression Recognition," *Comput. Electr. Eng.*, vol. 115, 2024, doi: 10.1016/j.compeleceng.2024.109125.
3. A. Barman and P. Dutta, "Facial expression recognition using Reversible Neural Network," *Appl. Soft Comput.*, vol. 162, 2024, doi: 10.1016/j.asoc.2024.111815.
4. K. Sarvakar, R. Senkamalavalli, S. Raghavendra, J. S. Kumar, R. Manjunath, and S. Jaiswal, "Facial emotion recognition using convolutional neural networks," *Mater. Today Proc.*, vol. 80, 2023, pp. 3560–3564, doi: 10.1016/j.matpr.2021.07.297.
5. A. N. Arun, P. Maheswaravenkatesh, and T. Jayasankar, "Facial Micro Emotion Detection and Classification Using Swarm Intelligence-based Modified Convolutional Network," *Expert Syst. Appl.*, vol. 233, 2023, doi: 10.1016/j.eswa.2023.120947.
6. X. Jia, S. Xu, Y. Zhou, L. Wang, and W. Li, "A novel dual-channel graph convolutional neural network for facial action unit recognition," *Pattern Recognit. Lett.*, vol. 166, 2023, pp.
61–68, doi: 10.1016/j.patrec.2023.01.001.

7. S. Hossain, S. Umer, R. K. Rout, and M. Tanveer, "Fine-grained image analysis for facial expression recognition using deep convolutional neural networks with bilinear pooling," *Appl. Soft Comput.*, vol. 134, 2023, doi: 10.1016/j.asoc.2023.109997.
8. 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," *J. Vis. Commun. Image Represent.*, vol. 97, 2023, doi: 10.1016/j.jvcir.2023.103948.
9. Z. Wang, S. Cui, X. Wang, and J. F. Tian, "Research on Facial Expression Capture Based on Two-Stage Neural Network," *Comput. Mater. Continua*, vol. 72, no. 3, pp. 4709–4725, 2022, doi: 10.32604/CMC.2022.027767.
10. Z. Wang and L. Yao, "Expression Recognition Method Based on Convolutional Neural Network and Capsule Neural Network," *Comput. Mater. Continua*, vol. 79, no. 1, pp. 1659–1677, 2024, doi: 10.32604/CMC.2024.048304.
11. C. Gautam and K. R. Seeja, "Facial emotion recognition using Handcrafted features and CNN," *Procedia Comput. Sci.*, vol. 218, pp. 1295–1303, 2022, doi: 10.1016/j.procs.2023.01.108.
12. J. S. Angel et al., "Faster Region Convolutional Neural Network (FRCNN) Based Facial Emotion Recognition," *Comput. Mater. Continua*, vol. 79, no. 2, pp. 2427–2448, 2024, doi: 10.32604/CMC.2024.047326.
13. F. Di Luzio, A. Rosato, and M. Panella, "A randomized deep neural network for emotion recognition with landmarks detection," *Biomed. Signal Process. Control*, vol. 81, 2023, doi: 10.1016/j.bspc.2022.104418.
14. H. Pan, L. Xie, and Z. Wang, "C3DBed: Facial micro-expression recognition with threedimensional convolutional neural network embedding in transformer model," *Eng. Appl. Artif. Intell.*, vol. 123, 2023, doi: 10.1016/j.engappai.2023.106258.
15. G. Xiang et al., "Driver multi-task emotion recognition network based on multi-modal facial video analysis," Pattern Recognit., vol. 161, 2025, doi: 10.1016/j.patcog.2024.111241.
16. S. Saif, S. Tehseen, and S. S. Ali, "Fake news or real? Detecting deepfake videos using geometric facial structure and graph neural network," Technol. Forecast. Soc. Change, vol. 205, 2024, doi: 10.1016/j.techfore.2024.123471.

17. C. Fu et al., "HiMul-LGG: A hierarchical decision fusion-based local–global graph neural network for multimodal emotion recognition in conversation," *Neural Netw.*, vol. 181, 2025, doi: 10.1016/j.neunet.2024.106764.
18. J. Huang et al., "Automatic recognition of schizophrenia from facial videos using 3D convolutional neural network," *Asian J. Psychiatry*, vol. 77, 2022, doi: 10.1016/j.ajp.2022.103263.
19. J. Farhadi Sedehi, N. Jafarnia Dabanloo, K. Maghooli, and A. Sheikhani, "Multimodal insights into granger causality connectivity: Integrating physiological signals and gated eye-tracking data for emotion recognition using convolutional neural network," *Heliyon*, vol. 10, no. 16, Aug. 2024, doi: 10.1016/j.heliyon.2024.e36411.
20. D. Bhagat, A. Vakil, R. K. Gupta, and A. Kumar, "Facial Emotion Recognition (FER) using Convolutional Neural Network (CNN)," *Procedia Comput. Sci.*, vol. 235, pp. 2079–2089, 2024, doi: 10.1016/j.procs.2024.04.197.
21. R. P. G and K. K, "A novel framework using binary attention mechanism-based deep convolution neural network for face emotion recognition," *Measurement: Sensors*, vol. 30, Dec. 2023, doi: 10.1016/j.measen.2023.100881.
22. A. Maazallahi, M. Asadpour, and P. Bazmi, "Advancing emotion recognition in social media: A novel integration of heterogeneous neural networks with fine-tuned language models," *Inf. Process. Manag.*, vol. 62, no. 2, Mar. 2025, doi: 10.1016/j.ipm.2024.103974.
23. T. Farhat et al., "Facial Image-Based Autism Detection: A Comparative Study of Deep Neural Network Classifiers," *Comput. Mater. Continua*, vol. 78, no. 1, pp. 105–126, 2024, doi: 10.32604/CMC.2023.045022.
24. V. Singh and S. Prasad, "Speech emotion recognition system using gender-dependent convolution neural network," *Procedia Comput. Sci.*, vol. 218, pp. 2533–2540, 2022, doi: 10.1016/j.procs.2023.01.227.
25. X. Tao et al., "Facial video-based non-contact emotion recognition: A multi-view features expression and fusion method," *Biomed. Signal Process. Control*, vol. 96, Oct. 2024, doi: 10.1016/j.bspc.2024.106608.
26. N. Kanimozhi and R. D. Priya, "Hybrid Lyrebird Red Panda Optimization_Shepard Convolutional Neural Network for Recognition of Speech Emotion in Audio Signals," *Neurocomputing*, p. 129506, Jan. 2025, doi: 10.1016/J.NEUCOM.2025.129506.
27. K. Dinakaran and E. M. Ashokkrishna, "Efficient regional multi feature similarity measure based emotion detection system in web portal using artificial neural network," *Microprocess. Microsyst.*, vol. 77, p. 103112, 2020, doi: 10.1016/j.micpro.2020.103112.

28. M. Liu, A. N. J. Raj, V. Rajangam, K. Ma, Z. Zhuang, and S. Zhuang, "Multiscalemultichannel feature extraction and classification through one-dimensional convolutional neural network for speech emotion recognition," *Speech Commun.*, vol. 156, p. 103010, 2024, doi: 10.1016/j.specom.2023.103010.
29. V. K. Barigala, S. P. J., S. Kumar P., N. Ganapathy, K. P. A., D. Kumar, and J. F. A. Ronickom, "Evaluating the effectiveness of machine learning in identifying the optimal facial electromyography location for emotion detection," *Biomed. Signal Process. Control*, vol. 100, no. B, p. 107012, 2025, doi: 10.1016/j.bspc.2024.107012.
30. V. M. Joshi, R. B. Ghongade, A. M. Joshi, and R. V. Kulkarni, "Deep BiLSTM neural network model for emotion detection using cross-dataset approach," *Biomed. Signal Process. Control*, vol. 73, p. 103407, 2022, doi: 10.1016/j.bspc.2021.103407.
31. C. Fu et al., "Facial action units guided graph representation learning for multimodal depression detection," *Neurocomputing*, vol. 619, p. 129106, 2025, doi: 10.1016/j.neucom.2024.129106.
32. Q. Wu, N. Dey, F. Shi, R. G. Crespo, and R. S. Sherratt, "Emotion classification on eyetracking and electroencephalograph fused signals employing deep gradient neural networks," *Appl. Soft Comput.*, vol. 110, p. 107752, 2021, doi: 10.1016/j.asoc.2021.107752.
33. R. Zhi, C. Zhou, T. Li, S. Liu, and Y. Jin, "Action unit analysis enhanced facial expression recognition by deep neural network evolution," *Neurocomputing*, vol. 425, pp. 135–148, 2021, doi: 10.1016/j.neucom.2020.03.036.
34. J. Burger, G. Blandano, G. M. Facchi, and R. Lanzarotti, "2S-SGCN: A two-stage stratified graph convolutional network model for facial landmark detection on 3D data," *Comput. Vis. Image Underst.*, vol. 250, p. 104227, 2025, doi: 10.1016/j.cviu.2024.104227.
35. H. Sadeghi and A. A. Raie, "HistNet: Histogram-based convolutional neural network with Chi-squared deep metric learning for facial expression recognition," *Inf. Sci.*, vol. 608, pp. 472–488, 2022, doi: 10.1016/j.ins.2022.06.092.
36. W. Guo and Y. Wang, "Convolutional gated recurrent unit-driven multidimensional dynamic graph neural network for subject-independent emotion recognition," *Expert Syst. Appl.*, vol. 238, p. 121889, 2024, doi: 10.1016/j.eswa.2023.121889.
37. A. Singh Rajpoot and M. R. Panicker, "Subject independent emotion recognition using EEG signals employing attention driven neural networks," *Biomed. Signal Process. Control*, vol. 75, p. 103547, 2022, doi: 10.1016/j.bspc.2022.103547.

38. M. Morningstar et al., "Functional connectivity during facial and vocal emotion recognition: Preliminary evidence for dissociations in developmental change by nonverbal modality," *Neuropsychologia*, vol. 202, p. 108946, 2024, doi: 10.1016/j.neuropsychologia.2024.108946.
39. L. Gong, M. Li, T. Zhang, and W. Chen, "EEG emotion recognition using attention-based convolutional transformer neural network," *Biomed. Signal Process. Control*, vol. 84, p. 104835, 2023, doi: 10.1016/j.bspc.2023.104835.
40. J. Zhu, Y. Ding, H. Liu, and K. Chen, "Emotion knowledge-based fine-grained facial expression recognition," *Neurocomputing*, vol. 610, p. 128536, 2024, doi: 10.1016/j.neucom.2024.128536.
41. R. Li, Y. Wu, Q. Wu, N. Dey, R. G. Crespo, and F. Shi, "Emotion stimuli-based surface electromyography signal classification employing Markov transition field and deep neural networks," *Measurement*, vol. 189, p. 110470, 2022, doi: 10.1016/j.measurement.2021.110470.
42. W. Guo and Y. Wang, "Convolutional gated recurrent unit-driven multidimensional dynamic graph neural network for subject-independent emotion recognition," *Expert Syst. Appl.*, vol. 238, p. 121889, 2024, doi: 10.1016/j.eswa.2023.121889.
43. Z. Wang and L. Yao, "Expression Recognition Method Based on Convolutional Neural Network and Capsule Neural Network," *Comput. Mater. Continua*, vol. 79, no. 1, pp. 1659–1677, 2024, doi: 10.32604/cmc.2024.048304.
44. S. Nita, S. Bitam, M. Heidet, and A. Mellouk, "A new data augmentation convolutional neural network for human emotion recognition based on ECG signals," *Biomed. Signal Process. Control*, vol. 75, p. 103580, 2022, doi: 10.1016/j.bspc.2022.103580.
45. C. Tan, M. Šarlija, and N. Kasabov, "NeuroSense: Short-term emotion recognition and understanding based on spiking neural network modelling of spatio-temporal EEG patterns," *Neurocomputing*, vol. 434, pp. 137–148, 2021, doi: 10.1016/j.neucom.2020.12.098.
46. E. Othmani, A. Q. Md Sabri, S. Aslan, F. Chaieb, H. Rameh, R. Alfred, and D. Cohen, "EEG-based neural networks approaches for fatigue and drowsiness detection: A survey," *Neurocomputing*, vol. 557, p. 126709, 2023, doi: 10.1016/j.neucom.2023.126709.
47. H. Diana Rosas, L. Lewis, N. Connors, N. D. Mercaldo, and S. Nasr, "Are you angry? Neural basis of impaired facial expression recognition in pre-manifest Huntington's," *Parkinsonism & Related Disorders*, vol. 109, p. 105289, 2023, doi: 10.1016/j.parkreldis.2023.105289.

48. M. López Ibáñez, M. Miranda, N. Alvarez, and F. Peinado, "Using gestural emotions recognised through a neural network as input for an adaptive music system in virtual reality," *Entertainment Computing*, vol. 38, p. 100404, 2021, doi: 10.1016/j.entcom.2021.100404.
49. Y. Wei, L. Cao, Y. Dong, and T. Liu, "MCNN-CMCA: A multiscale convolutional neural networks with cross-modal channel attention for physiological signal-based mental state recognition," *Digital Signal Processing*, vol. 156, Part B, p. 104856, 2025, doi: 10.1016/j.dsp.2024.104856.
50. P. Aghaomidi, F. Mirzaei, and Z. Bahmani, "MSLSNet: A combination of multi-task selfsupervised learning and Swin transformer network for face and keypoint detection in thermal images," *Expert Syst. Appl.*, vol. 268, p. 126238, 2025, doi: 10.1016/j.eswa.2024.126238.
51. N. Liu et al., "A lightweight network-based sign language robot with facial mirroring and speech system," *Expert Syst. Appl.*, vol. 262, p. 125492, 2025, doi: 10.1016/j.eswa.2024.125492.
52. S. Korb et al., "Gender differences in the neural network of facial mimicry of smiles – An rTMS study," *Cortex*, vol. 70, pp. 101–114, 2015, doi: 10.1016/j.cortex.2015.06.025.
53. B. Hari Krishna, J. S. R. Victor, G. S. Rao, C. R. K. Babu, K. S. Raju, T. S. Ghouse Basha, and V. Bharath Simha Reddy, "Emotion-net: Automatic emotion recognition system using optimal feature selection-based hidden Markov CNN model," *Ain Shams Eng. J.*, vol. 15, no. 12, p. 103038, 2024, doi: 10.1016/j.asej.2024.103038.
54. H. Arabian, V. Wagner-Hartl, and K. Moeller, "Image Pre-processing Effects on Attention Modules in Facial Emotion Recognition," *IFAC-PapersOnLine*, vol. 58, no. 24, pp. 433–438, 2024, doi: 10.1016/j.ifacol.2024.11.076.
55. P. Shivappa, A. Martin, and R. Stiefelhagen, "Multi-modal fusion for facial expression recognition using deep learning architectures," *Pattern Recognit.*, vol. 165, p. 111295, 2025, doi: 10.1016/j.patcog.2025.111295.
56. C. Zhang, Y. He, and J. Wang, "End-to-end deep learning models for micro-expression recognition: A comprehensive survey," *Image Vis. Comput.*, vol. 143, p. 104645, 2025, doi: 10.1016/j.imavis.2025.104645.
57. L. Feng, B. Hu, and Z. Chen, "Self-supervised learning for dynamic facial expression recognition in real-world scenarios," *Neurocomputing*, vol. 620, p. 128905, 2024, doi: 10.1016/j.neucom.2024.128905.

58. M. Popescu, D. G. Anghel, and C. N. Mitroi, "Multi-scale spatial-temporal graph convolutional network for cross-dataset facial expression recognition," *Comput. Vis. Image Underst.*, vol. 255, p. 104376, 2025, doi: 10.1016/j.cviu.2025.104376.
59. R. Zhao and H. Liu, "Hybrid graph-based and attention-based convolutional neural networks for spontaneous emotion detection," *Neural Netw.*, vol. 184, p. 107022, 2025, doi: 10.1016/j.neunet.2025.107022.
60. S. K. Verma, P. Roy, and A. Das, "Explainable AI for facial emotion recognition: A survey on interpretability techniques in deep learning," *Expert Syst. Appl.*, vol. 240, p. 121952, 2024, doi: 10.1016/j.eswa.2024.121952.
61. J. Huang, X. Liu, and K. He, "Face expression recognition using Transformer-based deep learning models," *J. Vis. Commun. Image Represent.*, vol. 102, p. 104126, 2025, doi: 10.1016/j.jvcir.2025.104126.
62. X. Wu, L. Zhang, and Y. Sun, "Facial expression recognition with adversarial domain adaptation and self-attention mechanism," *Pattern Recognit. Lett.*, vol. 174, pp. 101–113, 2025, doi: 10.1016/j.patrec.2025.101113.
63. M. H. Chen, S. K. Kim, and T. Y. Lee, "Hybrid CNN-LSTM approach for real-time facial expression classification," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 37, no. 1, pp. 299–312, 2025, doi: 10.1016/j.jksuci.2025.299312.
64. A. R. S. Mendez, C. F. Pardo, and E. G. Lema, "Real-time facial emotion detection using lightweight deep learning models," *Procedia Comput. Sci.*, vol. 250, pp. 2519–2530, 2025, doi: 10.1016/j.procs.2025.01.178.
65. V. T. Nair and A. P. Nambiar, "Multi-modal emotion recognition using EEG and facial expression analysis," *Biomed. Signal Process. Control*, vol. 98, p. 104735, 2025, doi: 10.1016/j.bspc.2025.104735.
66. S. Das and B. Mishra, "Emotion detection using GAN-augmented data and attentionbased neural networks," *Pattern Recognit. Lett.*, vol. 180, pp. 75–83, 2025, doi: 10.1016/j.patrec.2025.075083.
67. R. Patel, Y. Sharma, and K. Prakash, "Deep learning approach for detecting depression through facial expressions," *Asian J. Psychiatry*, vol. 90, p. 104623, 2025, doi: 10.1016/j.ajp.2025.104623.
68. L. Yang, P. Zhao, and X. Xu, "Combining deep learning and rule-based approaches for micro-expression recognition," *Neurocomputing*, vol. 630, p. 128955, 2025, doi: 10.1016/j.neucom.2025.128955.
69. T. Zhou, H. Luo, and Y. Li, "A hybrid CapsuleNet-CNN model for robust facial emotion recognition," *Expert Syst. Appl.*, vol. 245, p. 122108, 2025, doi:

10.1016/j.eswa.2025.122108.

70. J. C. Hung, N. X. Lai, and K. C. Lin, "Learning-based methods for automatic expression recognition: A survey on recent advancements," *Appl. Soft Comput.*, vol. 131, p. 107824, 2025, doi: 10.1016/j.asoc.2025.107824.