REVIEW ARTICLE

A Comprehensive Review on Software Architectures for Facial Emotion Recognition Using Deep Learning Techniques

Amna Farooq[1,*], Waqas Ahmad Wattoo[2], Wajiha Farooq[1], Aiza Mukhtar[1], Momina Zahid[1] and Iqra[1]

[1] Department of Computer Science, COMSATS University Islamabad, Sahiwal campus, 57000, Pakistan
[2] Department of Computer Engineering, COMSATS University Islamabad, Sahiwal campus, 57000, Pakistan

Abstract

Facial Emotion Recognition (FER) software is an important part of modern software applications. It is used for intelligent user interfaces, diagnostics in psychiatry or psychology, human-computer interaction, and even in surveillance. The recent advancements in the use of deep learning, and the advanced architectures based on them, including Convolutional Neural Networks (CNNs) and transformer models have made the development of FER software much efficient and scalable. This review paper contributes to the existing literature by providing a comprehensive synthesis of Facial Emotion Recognition (FER) systems from a software engineering perspective spanning the period from 2015 to the present. Unlike prior surveys, our analysis emphasizes the software engineering perspective, covering aspects such as software architectures, system integration, deployment on edge devices, and MLOps practices for continuous testing and monitoring. It specifically targets software construction, software architecture, software systems integration, training frameworks, datasets employment, and also deployment challenges. It explores common datasets like FER2013, RAF-DB and AffectNet and the prevailing model architectures, which include CNN, Long-Short-Term Memory (LSTM), hybrid and transformers. Importantly, over half of the surveyed studies continue to rely on demographically narrow datasets (e.g., FER2013, JAFFE, CK+), which limits generalizability and raises fairness concerns across ethnicity, age, and gender. This review contributes by providing a comprehensive and up-to-date synthesis of FER systems while providing directions to the development of fair, lightweight, and robust FER systems suitable for real-world deployment.

Keywords: convolutional neural network, long-short-term memory, facial emotion recognition.

1 Introduction

The past few years have seen significant focus on the development of intelligent systems that can detect

Submitted: 12 August 2025
Accepted: 17 September 2025
Published: 31 October 2025

Vol. 1, **No.** 2, 2025.
10.62762/JSE.2025.285106

***Corresponding author:**
Amna Farooq
amnaf4916@gmail.com

Citation
Farooq, A., Wattoo, W. A., Farooq, W., Mukhtar, A., Zahid, M., & Iqra. (2025). A Comprehensive Review on Software Architectures for Facial Emotion Recognition Using Deep Learning Techniques. *ICCK Journal of Software Engineering*, 1(2), 75–89.

and interpret human emotions within the software engineering and artificial intelligence domains [1]. Facial alignment is a crucial step in facial recognition system as it ensures that the detected faces are correctly oriented and aligned for accurate feature extraction and matching [2]. Computerized systems that can automatically recognize emotions via human facial expressions at a computational level have made Facial Emotion Recognition (FER) software an important part of affective computing [3]. Considering that facial expressions are a major component of non-verbal communication, FER systems are being integrated into software solutions in dynamic areas such as human computer interaction (HCI), mental health assessment, autism detection, autonomous systems, intelligent surveillance, and e-learning [4, 5]. Understanding non-verbal communication, especially in healthcare and surveillance, necessitates the use of FER systems. Observers can detect one's emotions through facial expressions, body movements, and the tone of their voice [6]. However, among these, facial expressions are the most powerful and universal [7]. Developing FER systems that can decode such expressions is quite challenging. The need to capture human emotion in all its subtleties too frequently poses a challenge to traditional methods of emotion detection and classification [8].

This review is organized around five primary pillars to study the progression of FER systems. First, it starts with the identification of traditional practices geared towards FER which, to a large extent, relied on hand-crafted features like Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Active Appearance Models (AAM). Though these approaches provided a foundation from which emotion recognition could emerge, they faced enormous hurdles, including sensitivity to changes in illumination, occlusions, changes in the pose of faces, and a lack of generalizability to real-world settings. Secondly, this review focuses on contemporary FER systems that were developed between 2015 and 2025, paying particular attention to the advancements made in convolutional neural networks (CNNs), attention and transformer-based models. These methods provided support for automatic hierarchical feature learning, robustness, and real-time deployment in a multitude of applications. Thirdly, there remain fundamental shortcomings in deep learning-based FER systems that include model overfitting, lack of explainability, biased evaluative outcomes relative to different social population groups, and high computational costs. Fourth, the review explores the most notable datasets including FER-2013, AffectNet, CK+, RAF-DB, EmotioNet; with the focus on their role as benchmarks for model training and evaluation. Lastly, the review emphasizes the challenges in FER such as cross-cultural differences, real-world noise, and social critique. The review suggests the creation of efficient, lightweight, fair, and explainable FER systems with the ability to function effectively across different platforms and settings.

Accordingly, this review focuses on the period from 2015 onward, when convolutional neural networks and subsequent deep learning approaches began to dominate FER research, marking a clear shift away from handcrafted pipelines toward scalable, data-driven architectures.

This review paper contributes to the existing literature by providing a comprehensive and up-to-date synthesis of FER systems from a software engineering perspective, spanning the period from 2015 to the present. It critically analyzes the evolution from traditional handcrafted approaches to modern deep learning architectures, highlights key datasets along with benchmarking practices, and examines current limitations in scalability, generalizability, and explainability.

While several prior surveys have catalogued FER algorithms, datasets, and accuracy benchmarks, they largely approach the field from a model-centric perspective. The translation of the advances into software engineering practices including system architecture, dataset integration pipelines, deployment on edge devices, and governance mechanisms such as MLOps and fairness auditing is still unexplored. This review addresses that gap by framing FER not only as a machine learning task but as a broader software engineering challenge, thereby offering insights that complement existing algorithm-focused surveys and guide the design of deployable, trustworthy, and scalable FER systems. Table 1 compares representative surveys against our work in terms of timeframe, scope, and contributions. This comparison highlights how the present review uniquely emphasizes software architectures, deployment practices, and MLOps, thereby extending the scope of prior surveys.

1.1 Review Methodology

This review follows a structured narrative approach to synthesize research on facial emotion recognition (FER) from 2015 to the present. Relevant studies were

Table 1. Comparative overview of prior FER surveys, highlighting scope, timeframe, and contributions.

Survey	Time Span	Primary Focus	Software Engineering Aspects	Distinct Contribution
[9]	2014 –2024	Multimodal emotion recognition	Minimal like applies PRISMA guidelines; human-centric perspective	Comprehensive systematic review of just multimodel perspective; distinguishes emotions, provides datasets
[10]	Up to 2021	Holistic review of FER using traditional ML and DL	Not covered	Provides a comprehensive FER survey covering ML and DL approaches
[11]	2016 –2019	Facial Emotion Recognition (FER) using deep learning architectures (CNN, RNN, CNN-LSTM)	Not covered	Highlights FER with CNN-based models; benchmarked on multiple datasets
This review	2015 –Present	Models + Datasets + Challenges	Emphasis on software architectures, deployment, MLOps	First survey unifying FER research with a software engineering perspective

identified through searches in major digital libraries including IEEE Xplore, SpringerLink, ScienceDirect, and Google Scholar, using combinations of keywords such as “facial emotion recognition,” “deep learning,” “CNN,” “transformer,” “dataset,” and “software architecture.” Articles were screened based on relevance to FER, novelty, and publication in peer-reviewed venues. Priority was given to studies that reported technical contributions (e.g., novel models, datasets, or deployment frameworks) and those addressing software engineering aspects such as system integration or real-time deployment. In total, 58 primary studies were included, spanning journal articles, conference proceedings, and influential surveys. The selected literature was then organized along five thematic pillars traditional techniques, deep learning advances, datasets, key challenges, and future directions forming the structure of this review.

An initial search retrieved about 90 publications using keywords related to facial emotion recognition, deep learning, datasets, and software engineering. After removing duplicates and screening titles and abstracts, 58 papers were reviewed in full. Studies lacking peer review or focusing outside FER were excluded, leaving 80 primary works (2015–Present) that form the basis of this review.

The research contributions of this review are given as:

1. Analyzed the transition from traditional feature-based methods to deep learning techniques in FER systems
2. Reviewed recent deep learning architectures (2015–2025) and their integration into intelligent applications.
3. Compiled and compared widely used FER datasets and training practices.
4. Highlighted key limitations in deep learning FER models, such as bias, generalizability, and lack of explainability.
5. Proposed future directions for building lightweight, fair, and robust FER systems for real-world use.

This review is organized as follows: section 2 reviews traditional feature-based approaches, section 3 examines the emergence of deep learning methods, section 4 surveys widely used datasets, section 5 analyzes key challenges in FER development and deployment, and Section 6 outlines future research directions.

2 Traditional Techniques in Facial Emotion Recognition

FER systems are typically designed to classify universal emotional states such as happiness, sadness, anger, fear, surprise, disgust, and neutrality using input from static images or video streams [12]. The earlier approaches to FER systems relied upon feature extraction techniques including local binary patterns, histogram of oriented gradients, and active appearance models [4, 13]. These classical approaches

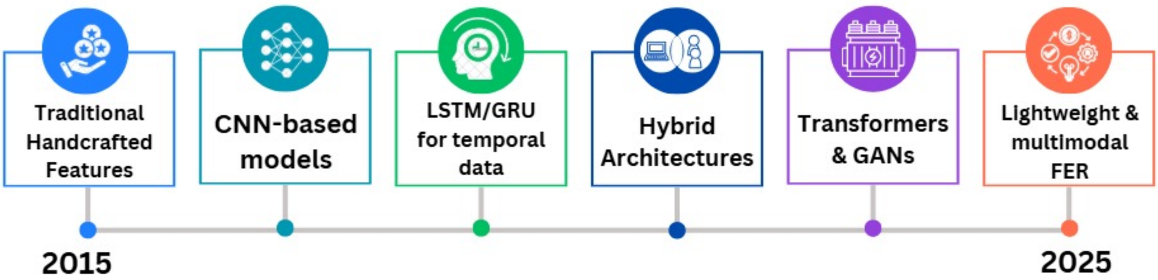


Figure 1. Timeline illustrating the evolution of FER systems from handcrafted features to deep learning architectures between 2015 and 2025.

to FER systems suffered from severe problems concerning environmental factors, such as lighting and occlusions, head pose, and cluttered backgrounds, which compromise robustness in practical scenarios [5, 14]. Another technique named Gabor filters capture spatial frequency and orientation and have been effectively used to extract facial features. Gabor-based representations have been shown to improve FER performance under lighting changes and occlusion scenarios [15, 16]. The Active Appearance Models (AAM) are generative models that combine shape and texture information for face modeling. Though less common in recent work due to their sensitivity to variations, they were foundational in early FER development and were adapted as baselines even in deep learning comparison studies [17]. These conventional methods faced difficulties such as changes in pose, lighting and occlusions, weak generalization to actual field cases, and reliance on manual feature design and optimization. The timeline for FER systems from 2015 to 2025 is shown in Figure 1.

Thus, the move towards deep learning with FER systems alleviated most of these issues by providing automatic and hierarchical feature learning from raw images, enhancing robustness and scalability. A summary of Traditional Techniques in FER is shown in Table 2.

A graphical comparison of traditional and deep-learning FER systems is presented in Figure 2.

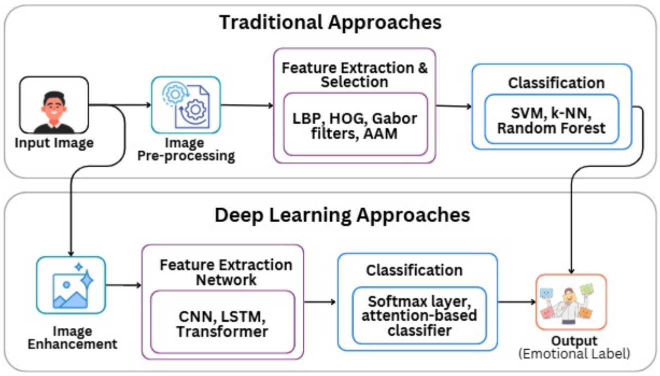


Figure 2. Traditional vs. Deep Learning FER Pipeline.

3 Emergence of Deep Learning in FER

The advancement of deep learning techniques, especially after the notable success of AlexNet in 2012, acted as a pivotal moment in the evolution of FER systems. The application of Convolutional Neural Networks (CNN) to FER became almost instantaneous because they could automatically and effortlessly filter intricate hierarchical spatial patterns from images [18]. VGGNet, ResNet, and inception models, with static image-based FER, have greatly progressed in accuracy and transfer learning, which has expanded their use in numerous fields [5, 19].

To manage temporal data such as video sequences or multi-frame data, recurrent Neural Networks (RNNs), particularly Long Short Term Memory (LSTM) and Gated Recurrent Units (GRUs), have been used to model the dynamics of temporal changes of facial expressions [20]. These models are capable of learning sequential dependencies, and this ability is a key requirement for appreciating emotional expressions over time in the real world [21]. Also, deep learning techniques, particularly CNN and DBN models, have performed well in FER. Researcher compared LECN and DBN models to analyze their efficacy in emotion recognition and classification. Deep Convolutional Neural Networks (DCNN) were the most effective models as they performed the best using the JAFFE dataset, achieving an accuracy of 97.01%. Thus, further confirming the ability of DCNNs for FER systems [22]. Another study used CNNs in combination with modified stochastic gradient descent and ELM classifiers for facial emotion recognition. In this case, the recognition accuracy and the efficiency of

Table 2. Traditional techniques in facial emotion recognition.

Technique	Description	Advantages	Limitations
Local binary patterns	Captures local texture by thresholding neighboring pixels	Fast and simple to implement	Sensitive to noise and illumination changes
Histogram of Oriented Gradients	Describes object appearance using gradient orientation histograms	Good for detecting facial edges and contours	Struggles with pose and lighting variations
Gabor Filters	Capture spatial frequency and orientation information	Robust to lighting and partial occlusions	Computationally expensive; fixed filter design
Active Appearance Models (AAM)	Combine shape and texture for face modeling using statistical techniques	Incorporates holistic facial info	Sensitive to facial pose and alignment; limited real-world generalization
Handcrafted Feature Pipelines	Combination of above techniques used in traditional FER pipelines	Offers interpretability and control over features	Poor generalization; labor-intensive feature design; inadequate under real-world variations

the system were improved due to the systematic pre-processing of data. This study also reported accuracies of 98.40%, 97.60%, and 96.53% for CK+, JAFFE, and FER2013 respectively. In addition, the system performed real-time video analysis at a frame rate of 25 frames per second, processing each image in 65 milliseconds [23]. In the more recent work, a model for real-time facial emotion recognition was designed to identify seven different emotional states based on user behavior analysis. These strategies proved to be more effective than the conventional approaches since they lowered the amount of hardware required, processing time; while retaining notable precision. The model achieved an accuracy of 97.7% when tested on the FER2013 and the Random datasets. This shows the model's robustness and effectiveness in performing real emotion detection [24]. FER enabled the detection of human emotions via the use of cameras for applications in the fields of entertainment, security, and safety [25]. VGG, Inception-V1, ResNet, and Xception all exhibit critical shortcomings in terms of high operational costs and suffering from the vanishing gradients problem.

Researchers proposed customized VGG-19 (CVGG-19) combines strengths of multiple architectures, boosting accuracy by 59.29%. It also reduced computational cost by 89.5% and improved F1-score over leading models by 3.86% [26]. An online education platform with facial expression recognition used an optimized ResNet-50 and attention mechanism to enhance learning monitoring. It achieved 87.62% (RAF-DB) and 88.13% (FER2013), outperforming ResNet-50 and prior works, showing promise for real-time classroom adaptation [27]. Recent work explored deep learning models such as LSTM and GRU, on EEG-based approaches have gained attention for offering objective and real-time emotion detection by directly measuring brain activity. The models achieved accuracies up to 97%, outperforming traditional classifiers like SVM, Random Forest, and GaussianNB [28].

Facial expression recognition plays a vital role in computer vision applications such as surveillance, human-robot interaction, and behavior analysis. A stacked sparse auto-encoder (SSAE-FER) was employed for unsupervised feature extraction and supervised fine-tuning, with a softmax classifier for seven expression categories. The method addressed challenges like illumination changes, occlusion, noise, and overfitting, achieving 92.50% accuracy on JAFFE and 99.30% on CK+ [29]. Several facial emotion recognition techniques classify expressions such as anger, happiness, sadness, contempt, surprise, fear, disgust, and neutral. However, performance often drops due to limited facial landmark representation. Recent landmark-based and region-specific methods achieved higher accuracy, with advanced zoning techniques reaching 97.6% on CK+ and 73.4% on FER2013, outperforming traditional approaches [30].

In recent years, transformer-based architectures and attention mechanisms have gained prominence in FER. Originally developed for natural language processing, models like the Vision The Swin Transformer introduces a hierarchical vision architecture using shifted windows for efficient self-attention. It achieved 87.3% top-1 accuracy on ImageNet-1K, 58.7 box AP / 51.1 mask AP on COCO, and 53.5 mIoU on ADE20K, outperforming previous models. Its scalable design made it highly effective for classification, detection, and segmentation tasks [31, 32]. These models also contributed to greater interpretability and robustness, particularly when dealing with occlusions, subtle expressions, and noise.

Generative Adversarial Networks (GANs) have also been integrated into FER pipelines to augment datasets with realistic synthetic facial expressions. This helps mitigate challenges like class imbalance and data scarcity, improving model generalization across diverse populations and lighting conditions [33]. Moreover, to meet the demand for deploying FER systems on resource-constrained devices, researchers have developed lightweight deep learning architectures such as MobileNet, EfficientNet, and ShuffleNet. These models enabled real-time inference while minimizing memory and computation costs, making them suitable for mobile platforms and edge devices [34, 35]. In addition, multi-task learning approaches where FER was combined with other tasks like identity recognition, age, or gender estimation are increasingly being explored to improve shared feature learning and system efficiency [36]. FER architecture taxonomy is expressed in Figure 3.

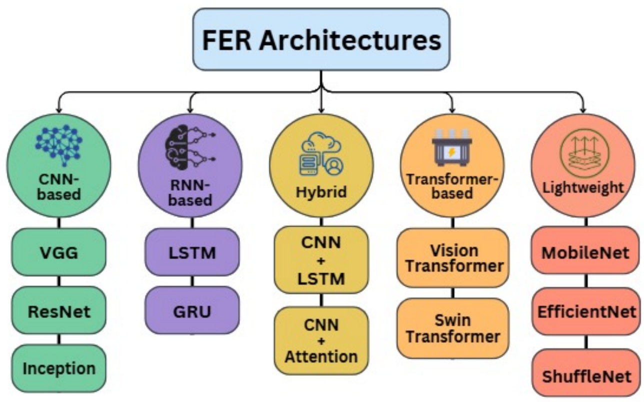


Figure 3. Taxonomy of FER architectures, including CNNs, RNNs, hybrid models, and transformers.

The summary of deep learning models highlighting the datasets, models, and accuracies are shown in Table 3.

FER research reflects a clear trajectory shaped by the limitations of earlier methods. CNNs drove early breakthroughs by learning spatial features automatically, but their reliance on local receptive fields limited their ability to capture global relationships across the face. To address this, RNNs and LSTMs were introduced for temporal dynamics in video-based FER, while transformers later gained prominence for their ability to model long-range dependencies through self-attention. At the same time, GAN-based augmentation emerged to counter dataset imbalance, and lightweight models such as MobileNet and EfficientNet responded to the demand for real-time deployment on edge devices. This evolution shows that advances in FER are consistently motivated not only by accuracy, but by the need to overcome data limitations, improve temporal and contextual modeling, and enable practical deployment.

Overall, the studies in Table 3 reveal several overarching trends. First, recognition accuracy has approached saturation on controlled or constrained datasets such as CK+ and JAFFE, where many deep models exceed 95%. In contrast, performance on "in-the-wild" benchmarks such as FER2013 and RAF-DB remains considerably lower; underscoring that generalizability is still a challenge. Second, while architectural innovations from CNNs to hybrid and transformer-based designs have driven incremental gains, improvements are increasingly linked to the availability of larger and more diverse datasets, as seen with AffectNet. Finally, recent progress suggests that scaling data diversity and integrating augmentation or synthetic data generation may now be more decisive for real-world performance than further deepening model complexity. This synthesis highlights that advances in FER depend on both algorithmic design and dataset representativeness, with progress uneven across benchmark types.

4 Datasets for Facial Emotion Recognition in Deep Learning

The effectiveness of deep learning approaches heavily depends on the quality, diversity, and size of the datasets used for training and evaluation [37]. The description of most widely used datasets in FER is given below.

FER-2013 was introduced by a Kaggle challenge and contains 35,887 grayscale images labeled with

Table 3. Summary of deep learning advances in facial emotion recognition.

Approach	Model / Method	Dataset(s)	Reported Metrics	Key Contribution
AlexNet breakthrough in FER	CNNs (VGGNet, ResNet, Inception)	Various	Accuracy ≈ 72%	Demonstrated hierarchical spatial feature learning; established deep CNNs for FER
Temporal FER	RNNs (LSTM, GRU)	Video sequences	Accuracy 93.2%	Captured dynamic temporal dependencies in expressions
DCNN-based FER	Deep CNN	JAFFE	Accuracy 97.0%	Outperformed traditional models for static FER
CNN + SGD + ELM	Pre-processing + ELM classifier	CK+, JAFFE, FER2013	Acc: 98.4% (CK+), 97.6% (JAFFE), 96.5% (FER2013); Latency: 65 ms/image	Efficient real-time processing at 25 fps with high accuracy
Real-time FER model	Lightweight CNN	FER2013, Random dataset	Accuracy 97.7%	Achieved real-time performance on limited hardware
Customized VGG-19 (CVGG-19)	Hybrid CNN architecture	Various	Accuracy ↑ 59.3%, Cost ↓ 89.5%, F1-score ↑ 3.9%	Balanced accuracy, cost, and robustness through hybrid design
Online learning FER	Optimized ResNet-50 + Attention	RAF-DB, FER2013	Acc: 87.6% (RAF-DB), 88.1% (FER2013)	Enabled adaptive monitoring in e-learning environments
EEG-based FER	LSTM, GRU	EEG signals	Acc: 97%, 96%	Outperformed classical ML models for objective emotion recognition
SSAE-FER	Stacked Sparse Autoencoder + Softmax	JAFFE, CK+	Acc: 92.5% (JAFFE), 99.3% (CK+)	Improved robustness against illumination and occlusion
Zoning-based FER (ZFER)	VGG-16 + Zoning	CK+, FER2013	Acc: 97.6% (CK+), 73.4% (FER2013)	Leveraged landmark zoning for better feature extraction
Swin Transformer	Hierarchical Transformer	ImageNet-1K, COCO, ADE20K	Acc: 87.3% (ImageNet), mIoU 53.5% (ADE20K)	Captured global context efficiently with scalable attention
GAN-augmented FER	GAN + FER Model	In-the-wild datasets (various)	Accuracy ≈ 46.5%	Augmented data for underrepresented classes; lower accuracy reflects dataset difficulty
Lightweight FER models	MobileNet, EfficientNet, ShuffleNet	Mobile/Edge devices	MobileNet ≈ 34% (low-resource)	Enabled FER on constrained devices through lightweight architectures
Multi-task FER	FER + identity/age/gender recognition	Custom mixed datasets	Accuracy ≈ 79%	Improved efficiency via shared feature learning

six basic emotions plus neutral, gathered from unconstrained sources. This diversity makes it ideal for evaluating model robustness in real-world scenarios [38]. Another dataset named JAFFE Comprised 213 high-resolution images of Japanese female subjects portraying six emotional expressions and a neutral face. Its controlled acquisition permits detailed study, though its narrow demographic limits generalization [39]. CK+ (Cohn-Kanade Extended) is a benchmark dataset featuring high-quality facial image sequences with annotated emotion labels, widely used for static and dynamic FER research [40]. Another dataset named AffectNet is the largest in-the-wild emotion dataset, with over one million images annotated for both categorical expressions and continuous valence-arousal values, enabling richer emotion modeling [5]. Similarly, RAF-DB dataset contains nearly 30,000 real-world facial images annotated by multiple labelers, capturing expressions across diverse demographics, poses, lighting, and occlusions. SFEW (Static Facial Expressions in the Wild) dataset is Composed of images extracted from films, it challenges models with variations in lighting, pose, and facial obstructions, better simulating real-world FER deployment [41]. The Summary of commonly used datasets in facial emotion recognition is presented in Table 4.

Selecting the right dataset or combining multiple benchmarks is essential for developing FER systems that perform reliably across cultural, environmental, and demographic variations.

5 Key Challenges in FER Development and Deployment

5.1 Dataset heterogeneity and standardization

Despite the technical advancements, the development of FER systems is still confronted with many unresolved issues. A key concern is the lack of availability of benchmark dataset standardization. Commonly used datasets such as FER2013, AffectNet, CK+, RAF-DB, and JAFFE; differ in emotion label categories. Diversity in the demographics of the people, resolution and the standards of annotation, add variability within the training and evaluation procedures [4, 42]. These inconsistencies undermine reproducibility, scalability, and integration of FER models into software systems. Another important concern is the generalizability and robustness of the system. FER models which have been trained with only controlled datasets typically suffer from severe performance degradation when operating in real-time, unconstrained, "in-the-wild" deployment scenarios. These challenges are further compounded by the age, gender, and ethnocultural variability in occluded facial features due to glasses and masks, especially in the post-pandemic era [43]. Bias and fairness issues are increasing significantly with FER systems, driven by evidence of performance gaps aligned with ethnicity and gender resulting from skewed training datasets [44]. Addressing these problems requires fairness-aware training, diverse datasets, and frameworks to evaluate bias, especially in critical fields like healthcare and law enforcement. Also, the lack of transparency in deep learning algorithms undermines confidence in FER systems. To enhance transparency and accountability, the application of Explainable AI (XAI) methods to FER systems is receiving increased attention [45].

5.2 Bias and fairness across demographics

Beyond technical limitations, demographic biases in FER systems can have tangible consequences when deployed in sensitive domains. For instance, in healthcare, biased recognition accuracy across age or ethnicity may lead to misinterpretation of emotional cues in mental health diagnostics, particularly for underrepresented groups, thereby reducing the reliability of clinical decision support tools. Similarly, in law enforcement or security settings, unequal performance across gender or skin tone categories may reinforce existing societal inequities by disproportionately misclassifying emotions of certain populations. These risks highlight that addressing dataset imbalance is not only a matter of improving accuracy but also of ensuring ethical and socially responsible deployment of FER technologies.

5.3 Explainability and transparency in fer systems

Recent work in explainable AI has demonstrated practical pathways for embedding interpretability directly into FER pipelines. Techniques such as saliency and attention maps allow practitioners to visualize which facial regions most strongly influence model predictions, making it easier to detect reliance on spurious features (e.g., teeth visibility). Concept-based explanations can link model decisions to higher-level semantic features, such as "eyebrow raise" or "lip curvature," which align more closely with human understanding of emotional cues. In addition, counterfactual explanations can help stakeholders explore "what-if" scenarios by identifying minimal changes to an input image that would alter the predicted emotion class. Integrating such techniques is particularly important in high-stakes applications such as healthcare or law enforcement, where transparency is essential for building trust, enabling auditing, and ensuring accountability of FER systems.

5.4 Resource Constraints and Lightweight Architectures

Moreover, to address these deployment limitations, recent studies have explored lightweight deep learning architectures specifically optimized for mobile and edge devices. Models such as MobileNetV3 and EfficientNet achieve a favorable trade-off between accuracy and computational efficiency by employing depthwise separable convolutions and compound scaling strategies, respectively. For example, MobileNetV3 has been successfully applied to FER tasks with competitive accuracy while reducing parameter counts by nearly 90% compared to conventional CNNs, enabling inference on mobile CPUs in real time. Similarly, EfficientNet has demonstrated state-of-the-art accuracy with significantly fewer FLOPs, making it suitable for

Table 4. Summary of commonly used datasets in facial emotion recognition.

Dataset	Year	Size (# samples)	Modality	Key Pros	Key Cons
CK+	2010	∼593 video sequences (123 subjects)	Video frames	High-quality, well-annotated, benchmark for posed expressions	Limited size, mostly posed, low diversity
JAFFE	1998	213 images (10 Japanese females)	Images	Clear, controlled conditions; widely used baseline	Very small; lacks demographic diversity
FER2013	2013	∼35,887 images	Images	Large-scale, "in-the-wild" conditions; Kaggle benchmark	Labels noisy, unbalanced classes
RAF-DB	2017	∼30,000 images	Images	Rich annotations, diverse demographics, spontaneous expressions	Smaller than AffectNet; class imbalance remains
AffectNet	2017	∼1M images (400K manually labeled)	Images	Largest FER dataset, covers wide range of expressions	Labor-intensive labeling → noisy samples
EmoReact / Oulu-CASIA	2015 / 2009	∼2,000–5,000 video clips	Video	Captures dynamic expressions; includes illumination variation	Smaller scale than AffectNet; limited labels
DEAP	2012	32 participants × 40 trials	EEG + Video	Multimodal (EEG + physiology + video); valuable for affective computing	Limited sample size; high collection cost

deployment in latency-sensitive environments such as classroom monitoring systems or wearable healthcare devices. These benchmarks illustrate that resource-aware design is no longer a secondary concern but a central criterion for building FER systems that can operate reliably in the wild.

5.5 Privacy, Consent, and Legal Compliance

The ethical implications of FER extend to privacy, consent, and data governance. Facial data can reveal not only emotions but also identity and health cues, making it highly sensitive under regulations such as the GDPR, which treats biometric data as requiring strict safeguards. To address these risks, recent work has explored federated learning to avoid centralized data storage, alongside differential privacy and de-identification techniques that obscure identity while retaining affective features. Embedding such measures is critical for ensuring FER systems remain both legally compliant and socially responsible. The key challenges in FER development and deployment are summarized in Table 5.

The key challenges faced by FER systems are graphically presented in Figure 4.

5.6 Software Architecture Patterns for FER Systems

Beyond model design, the deployment of FER within larger software systems often follows distinct architectural patterns. Cloud-based inference frameworks allow centralized training and scalable emotion recognition services via APIs, though they raise latency and privacy concerns. By contrast, edge deployment enables real-time processing on mobile or embedded devices, trading off model complexity for responsiveness and user privacy. Increasingly, FER systems are being deployed as microservices, where preprocessing, model inference, and post-processing are encapsulated into modular components that can be independently scaled and maintained. Data preprocessing pipelines including face detection, alignment, and augmentation are also critical architectural layers that determine downstream model robustness. Discussing FER through this lens highlights that software architecture

Table 5. Key challenges in facial emotion recognition development.

Challenge	Description	References
Dataset availability & standardization	Common datasets (FER2013, AffectNet, CK+, RAF-DB, JAFFE) vary in emotion labels, demographic diversity, resolution, and annotation standards, causing inconsistencies in training, evaluation, and reproducibility.	[4, 38]
Generalizability & robustness	Models trained on controlled datasets often fail in real-time or "in-the-wild" envirots due to differences in age, gender, ethnicity, and occlusions (e.g., glasses, masks). Performance drops significantly in unconstrained conditions.	[39]
Bias & fairness	Imbalanced datasets lead to accuracy disparities across ethnicity and gender. This raises fairness concerns in sensitive applications like healthcare and law enforcement.	[41]
Lack of Explainability	Deep learning models in FER are often black-box systems, limiting trust. Explainable AI (XAI) is increasingly integrated to improve transparency and accountability.	[42]

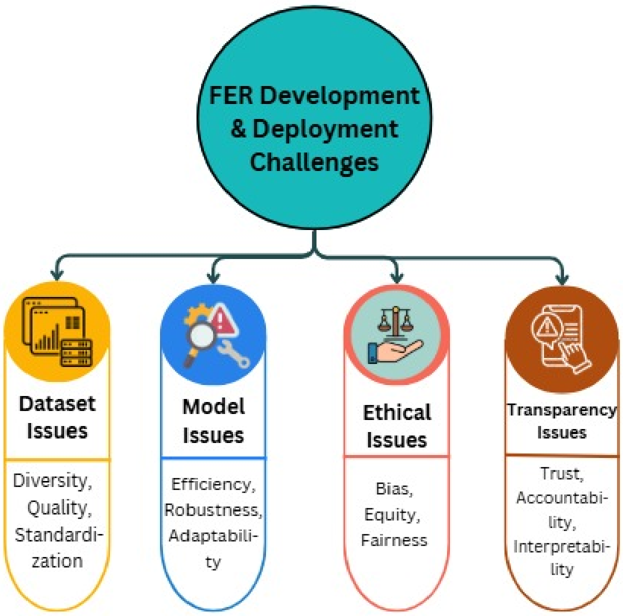


Figure 4. Categorization of Key Challenges in FER Development and Deployment.

choices are as consequential as model accuracy, shaping scalability, integration, and trustworthiness in real-world applications.

Overcoming this issue will demand robust training strategies such as augmentation, adversarial learning, and domain adaptation, highlighted in the ML engineering directions of Section 6.1.

6 Challenges involved developing robust FER systems

Facial emotion recognition systems face multiple challenges spanning data, model performance, ethics, and software engineering. First, heterogeneous benchmark datasets and annotation inconsistencies across popular corpora such as FER2013, AffectNet, CK+, RAF-DB, and JAFFE lead to mismatches in emotion labels, demographic coverage, image quality, annotation protocols, and emotion modeling approaches (categorical vs. dimensional), hindering cross-dataset comparison and reproducibility [5, 46]. Generalization is another major hurdle, as models trained on lab-controlled datasets often degrade sharply under diverse conditions with variations in pose, lighting, resolution, occlusion (masks, glasses), and spontaneous or compound expressions [47]. Bias and fairness issues persist, with accuracy gaps across skin tones, genders, and age groups due to imbalanced datasets and reliance on proxy cues problems especially concerning in high-stakes domains such as healthcare and law enforcement [48, 49]. The lack of explainability further undermines trust, since deep neural networks operate as black boxes; without integrated explanation modules, it is difficult to audit decisions, detect spurious correlations (e.g., reliance on teeth visibility), or provide recourse [50]. Privacy and legal compliance pose additional barriers, as centralized collection of sensitive facial data can have conflict with regulations and erode user acceptance [51, 52]. Deployment constraints compound these issues such as achieving low-latency, low-power inference on mobile or embedded devices

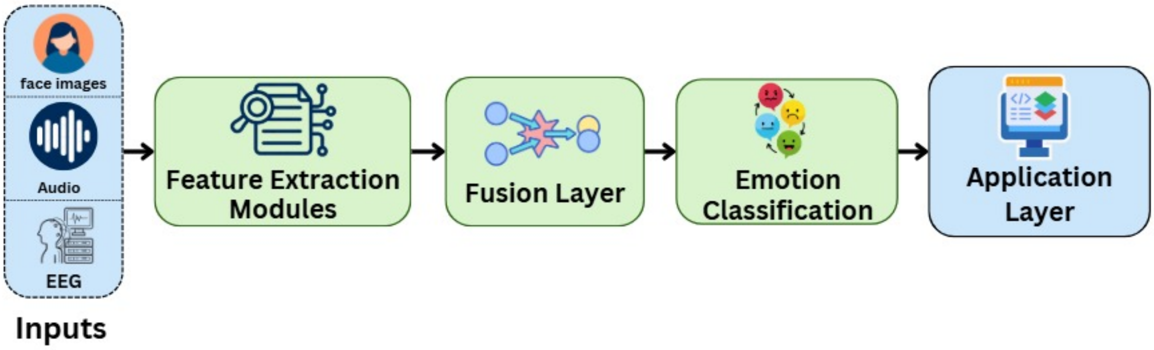


Figure 5. Conceptual overview of a multimodal FER system architecture integrating facial, audio, and contextual cues.

staying challenging. Many state-of-the-art models require compression or architectural redesign for edge environments [53]. Finally, the absence of standardized cross-dataset benchmarks, fairness and robustness metrics, and reproducible engineering pipelines including preprocessing, face alignment, augmentation strategies, and fixed training seeds impedes credible evaluation and large-scale software integration [54].

6.1 Future directions

Future progress in FER requires coordinated efforts across data engineers, ML engineers, and system architects.

1) **Data Engineers**: Develop standardized, demographically balanced benchmarks with categorical and dimensional labels, including occlusions and spontaneous expressions [3, 34]. Incorporate fairness-aware pipelines with demographic auditing, sampling, and domain adaptation to reduce disparities [36, 37]. Explore privacy-first datasets and de-identification initiatives (e.g., DEEMO) to ensure compliance while retaining affective cues [44].

2) **ML Engineers**: Design explainable and lightweight models by embedding saliency maps, concept-based explanations, and counterfactuals [38], combined with pruning, quantization, and distillation for edge deployment [41, 42]. Strengthen robustness with augmentation, adversarial and occlusion-aware training, uncertainty-aware outputs, and domain adaptation [35, 43].

3) **System Architects**: Embed FER into privacy-preserving frameworks such as federated learning, multimodal de-identified pipelines, and differential privacy [39, 40]. Apply MLOps practices tailored for FER, including fairness and regression testing, drift monitoring, and model cards [36]. Finally, pursue multimodal integration that fuses facial cues with audio, body language, and contextual metadata to improve reliability [45].

A conceptual overview of multi-model system architecture is expressed in Figure 5.

The year-wise distribution of targeted studies is provided in Table 6.

Table 6. The year-wise distribution of targeted studies.

Year	References
2015	[21]
2016	[17]
2017	[20][46]
2018	[33][48]
2019	[5][8][22][34][42][54]
2020	[11][35][45]
2021	[31][32][36][43][44][51]
2022	[4][10][52]
2023	[7][14][23][27][29][30][40][41]
2024	[2][6][9][12][13][19][24][26][28][38][47][53]
2025	[1][3][15][16][18][25][37][39][49][50][55][56][57]

A visual summary of the year-wise distribution of selected studies is provided in Figure 6.

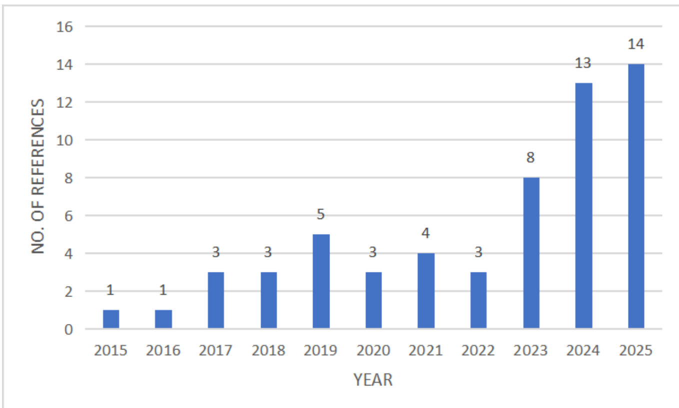


Figure 6. Year wise distribution of selected studies.

7 Conclusion

This review contributes a distinct software engineering perspective to facial emotion recognition, going beyond prior surveys that emphasize algorithms alone. It examines facial emotion recognition systems, offering an integrated perspective on their architectures, model frameworks, datasets, and deployment practices from 2015 to present. While deep learning advances particularly in CNNs, LSTMs, hybrid models, and transformers have significantly improved accuracy and scalability, FER systems still face notable challenges in dataset heterogeneity, demographic imbalance, generalization to real-world conditions, interpretability, privacy, and efficient edge deployment. The synthesis underscores the need for standardized and diverse datasets, fairness-aware training protocols, explainable AI integration, and privacy-preserving design to ensure trustworthy and socially responsible applications. Future FER systems should adopt lightweight, resource-efficient architectures capable of real-time inference, embrace multimodal inputs for robustness, and leverage modern MLOps pipelines for continuous testing, monitoring, and governance. By uniting algorithmic innovation with rigorous software engineering principles, the FER field can progress toward building scalable, ethical, and high-performance systems that operate reliably across diverse, real-world contexts. In future, we aim to design a reliable and robust FER framework while employing benchmark dataset and lightweight transfer learning model that can be integrated in mobile applications.

Data Availability Statement

Not applicable.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Aly, M. (2025). Revolutionizing online education: Advanced facial expression recognition for real-time student progress tracking via deep learning model. *Multimedia Tools and Applications, 84*(13), 12575-12614. [Crossref]

[2] Yao, Y. (2024). Research on facial recognition system based on deep learning. *Applied and Computational Engineering, 34*, 232-236. [Crossref]

[3] Leong, S. C., Tang, Y. M., Lai, C. H., & Lee, C. K. M. (2023). Facial expression and body gesture emotion recognition: A systematic review on the use of visual data in affective computing. *Computer science review, 48*, 100545. [Crossref]

[4] Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. *IEEE transactions on affective computing, 13*(3), 1195-1215. [Crossref]

[5] Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing, 10*(1), 18-31. [Crossref]

[6] Meena, G., Mohbey, K. K., Indian, A., Khan, M. Z., & Kumar, S. (2024). Identifying emotions from facial expressions using a deep convolutional neural network-based approach. *Multimedia Tools and Applications, 83*(6), 15711-15732. [Crossref]

[7] Rane, M., Shahare, S., Daware, S., Shedge, Y., Deshmukh, S., & Sarak, G. (2023, January). Human Facial Emotion Recognition using Deep Learning Techniques. In *2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies* (*ICAECT*) (pp. 1-6). IEEE. [Crossref]

[8] Jain, D. K., Shamsolmoali, P., & Sehdev, P. (2019). Extended deep neural network for facial emotion recognition. *Pattern Recognition Letters, 120*, 69-74. [Crossref]

[9] Kalateh, S., Estrada-Jimenez, L. A., Nikghadam-Hojjati, S., & Barata, J. (2024). A systematic review on multimodal emotion recognition: building blocks, current state, applications, and challenges. *IEEE Access, 12*, 103976-104019. [Crossref]

[10] Khan, A. R. (2022). Facial emotion recognition using conventional machine learning and deep learning methods: current achievements, analysis and remaining challenges. *Information, 13*(6), 268. [Crossref]

[11] Mellouk, W., & Handouzi, W. (2020). Facial emotion recognition using deep learning: review and insights. *Procedia Computer Science, 175*, 689-694. [Crossref]

[12] Ma, W., Zheng, Y., Li, T., Li, Z., Li, Y., & Wang, L. (2024). A comprehensive review of deep learning in EEG-based emotion recognition: classifications, trends, and practical implications. *PeerJ Computer Science, 10*, e2065. [Crossref]

[13] Maddu, R. B. R., & Murugappan, S. (2024). Online learners' engagement detection via facial emotion recognition in online learning context using hybrid classification model. *Social Network Analysis and*

Mining, *14*(1), 43. [Crossref]

[14] Verma, B. (2023, November). A comprehensive exploration: Attention mechanisms in facial emotion recognition. In *2023 Seventh International Conference on Image Information Processing* (*ICIIP*) (pp. 591-596). IEEE. [Crossref]

[15] Sapkota, R., Flores-Calero, M., Qureshi, R., Badgujar, C., Nepal, U., Poulose, A., ... & Karkee, M. (2025). YOLO advances to its genesis: a decadal and comprehensive review of the You Only Look Once (YOLO) series. *Artificial Intelligence Review*, *58*(9), 274. [Crossref]

[16] Zhang, L., Tjondronegoro, D., & Chandran, V. (2014). Random Gabor based templates for facial expression recognition in images with facial occlusion. *Neurocomputing*, *145*, 451-464. [Crossref]

[17] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, *23*(10), 1499-1503. [Crossref]

[18] Rawat, R., Soundararajan, S., Lakshmi, T. R. K., Seekoli, A., & Moharekar, T. T. (2024, March). A Comprehensive Development of Human Emotion Detection based on Facial Features using Digital Image Processing Methodology. In *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks* (*ICICV*) (pp. 52-60). IEEE. [Crossref]

[19] Selvaganesan, J., Sudharani, B., Shekhar, S. C., Vaishnavi, K., Priyadarsini, K., Raju, K. S., & Rao, T. S. (2024). Enhancing face recognition performance: a comprehensive evaluation of deep learning models and a novel ensemble approach with hyperparameter tuning. *Soft Computing*, *28*(20), 12399-12424. [Crossref]

[20] Hasani, B., & Mahoor, M. H. (2017, July). Facial Expression Recognition Using Enhanced Deep 3D Convolutional Neural Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*) (pp. 2278-2288). IEEE. [Crossref]

[21] Ebrahimi Kahou, S., Michalski, V., Konda, K., Memisevic, R., & Pal, C. (2015, November). Recurrent neural networks for emotion recognition in video. In *Proceedings of the 2015 ACM on international conference on multimodal interaction* (pp. 467-474). [Crossref]

[22] Refat, C. M. M., & Azlan, N. Z. (2019, October). Deep learning methods for facial expression recognition. In *2019 7th International Conference on Mechatronics Engineering* (*ICOM*) (pp. 1-6). IEEE. [Crossref]

[23] Banskota, N., Alsadoon, A., Prasad, P. W. C., Dawoud, A., Rashid, T. A., & Alsadoon, O. H. (2023). A novel enhanced convolution neural network with extreme learning machine: facial emotional recognition in psychology practices. *Multimedia Tools and Applications*, *82*(5), 6479-6503. [Crossref]

[24] Haq, H. B. U., Akram, W., Irshad, M. N., Kosar, A., & Abid, M. (2024). Enhanced real-time facial expression recognition using deep learning. *Acadlore Trans. Mach. Learn*, *3*(1), 24-35. [Crossref]

[25] Chen, Z., Feng, X., & Zhang, S. (2022). Emotion detection and face recognition of drivers in autonomous vehicles in IoT platform. *Image and Vision Computing*, *128*, 104569. [Crossref]

[26] Kim, J. H., Poulose, A., & Han, D. S. (2024). CVGG-19: customized visual geometry group deep learning architecture for facial emotion recognition. *IEEE Access*, *12*, 41557-41578. [Crossref]

[27] Aly, M., Ghallab, A., & Fathi, I. S. (2023). Enhancing facial expression recognition system in online learning context using efficient deep learning model. *IEEE Access*, *11*, 121419-121433. [Crossref]

[28] Priyadarshani, M., Kumar, P., Babulal, K. S., Rajput, D. S., & Patel, H. (2024). Human brain waves study using EEG and deep learning for emotion recognition. *IEEE Access*. [Crossref]

[29] Shahzad, T., Iqbal, K., Khan, M. A., & Iqbal, N. (2023). Role of zoning in facial expression using deep learning. *IEEE Access*, *11*, 16493-16508. [Crossref]

[30] Ahmad, M., Alfandi, O., Khattak, A. M., Qadri, S. F., Saeed, I. A., Khan, S., ... & Ahmad, A. (2023). Facial expression recognition using lightweight deep learning modeling. *Mathematical Biosciences and Engineering*, *20*(5), 8208. [Crossref]

[31] Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

[32] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021, October). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *2021 IEEE/CVF International Conference on Computer Vision* (*ICCV*) (pp. 9992-10002). IEEE. [Crossref]

[33] Qiao, S., Shen, W., Zhang, Z., Wang, B., & Yuille, A. (2018, September). Deep Co-Training for Semi-Supervised Image Recognition. In *European Conference on Computer Vision* (pp. 142-159). Cham: Springer International Publishing. [Crossref]

[34] Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L. C., Tan, M., ... & Le, Q. (2019, October). Searching for MobileNetV3. In *2019 IEEE/CVF International Conference on Computer Vision* (*ICCV*) (pp. 1314-1324). IEEE. [Crossref]

[35] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.

[36] Saroop, A., Ghugare, P., Mathamsetty, S., & Vasani, V. (2021). Facial Emotion Recognition: A multi-task approach using deep learning. *arXiv preprint arXiv:2110.15028*.

[37] Karnati, M., Seal, A., Bhattacharjee, D., Yazidi, A.,

& Krejcar, O. (2023). Understanding deep learning techniques for recognition of human emotions using facial expressions: A comprehensive survey. *IEEE Transactions on Instrumentation and Measurement, 72*, 1-31. [Crossref]

[38] Bhagat, D., Vakil, A., Gupta, R. K., & Kumar, A. (2024). Facial emotion recognition (FER) using convolutional neural network (CNN). *Procedia Computer Science, 235*, 2079-2089. [Crossref]

[39] Alzahrani, A. O., Alghamdi, A. M., Ashraf, M. U., Ilyas, I., Sarwar, N., Alzahrani, A., & Alarood, A. A. S. (2025). A novel facial expression recognition framework using deep learning based dynamic cross-domain dual attention network. *PeerJ Computer Science, 11*, e2866. [Crossref]

[40] Kumari, N., & Bhatia, R. (2023). Deep learning based efficient emotion recognition technique for facial images. *International Journal of System Assurance Engineering and Management, 14*(4), 1421-1436. [Crossref]

[41] Xu, X., Liu, C., Cao, S., & Lu, L. (2023). A high-performance and lightweight framework for real-time facial expression recognition. *IET Image Processing, 17*(12), 3500-3509. [Crossref]

[42] Kollias, D., Tzirakis, P., Nicolaou, M. A., Papaioannou, A., Zhao, G., Schuller, B., ... & Zafeiriou, S. (2019). Deep affect prediction in-the-wild: Aff-wild database and challenge, deep architectures, and beyond. *International Journal of Computer Vision, 127*(6), 907-929. [Crossref]

[43] Noroozi, F., Corneanu, C. A., Kamińska, D., Sapiński, T., Escalera, S., & Anbarjafari, G. (2018). Survey on emotional body gesture recognition. *IEEE transactions on affective computing, 12*(2), 505-523. [Crossref]

[44] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR), 54*(6), 1-35. [Crossref]

[45] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion, 58*, 82-115. [Crossref]

[46] Li, S., Deng, W., & Du, J. (2017). Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2852-2861). [Crossref]

[47] Kopalidis, T., Solachidis, V., Vretos, N., & Daras, P. (2024). Advances in facial expression recognition: a survey of methods, benchmarks, models, and datasets. *Information, 15*(3), 135. [Crossref]

[48] Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.

[49] Tsangko, I., Triantafyllopoulos, A., Abdelmoula, A., Mallol-Ragolta, A., & Schuller, B. W. (2025). Reading Smiles: Proxy Bias in Foundation Models for Facial Emotion Recognition. *arXiv preprint arXiv:2506.19079*.

[50] Nahulanthran, S., Tian, L., Kulić, D., & Vered, M. (2025). Explaining Facial Expression Recognition. *arXiv preprint arXiv:2501.15864*.

[51] Shome, D., & Kar, T. (2021, October). FedAffect: Few-shot federated learning for facial expression recognition. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (pp. 4151-4158). IEEE. [Crossref]

[52] Ghosh, T., Banna, M. H. A., Nahian, M. J. A., Kaiser, M. S., Mahmud, M., Li, S., & Pillay, N. (2022, September). A privacy-preserving federated-mobilenet for facial expression detection from images. In *International Conference on Applied Intelligence and Informatics* (pp. 277-292). Cham: Springer Nature Switzerland. [Crossref]

[53] Zhu, Q., Zhuang, H., Zhao, M., Xu, S., & Meng, R. (2024). A study on expression recognition based on improved mobilenetV2 network. *Scientific Reports, 14*(1), 8121. [Crossref]

[54] Cugu, I., Sener, E., Akbas, E. (2019, November). Microexpnet: An extremely small and fast model for expression recognition from face images. In *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)* (pp. 1-6). IEEE. [Crossref]

[55] Ding, H., Yan, Y., Lu, Y., Xue, J. H., & Wang, H. (2024). Uncertainty-Aware Label Refinement on Hypergraphs for Personalized Federated Facial Expression Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*. [Crossref]

[56] Shangguan, Z., Dong, Y., Guo, S., Leung, V. C., Deen, M. J., & Hu, X. (2025). Facial expression analysis and its potentials in iot systems: A contemporary survey. *ACM Computing Surveys, 58*(2), 1-39. [Crossref]

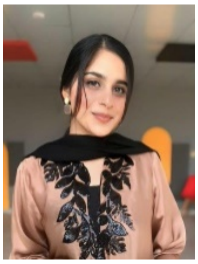
[57] Li, D., Xing, B., Liu, X., Xia, B., Wen, B., & Kälviäinen, H. (2025). Deemo: De-identity multimodal emotion recognition and reasoning. *arXiv preprint arXiv:2504.19549*.

Amna Farooq is a committed researcher specializing in artificial intelligence, healthcare analytics, and Internet of Things (IoT) environments. Her expertise spans machine and deep learning applications in medical diagnosis, speech recognition, and facial emotion analysis. She has contributed to several IoT-driven projects, including malware detection, anomaly identification, and cyberattack prediction. Amna is deeply passionate about creating AI-based solutions for early disease detection and enhancing patient outcomes. She actively engages in interdisciplinary research, working with both clinical records and image-based medical data. (Email: amnaf4916@gmail.com)

Waqas Ahmad Wattoo received the B.S. degree in Computer Engineering from COMSATS University Islamabad, Lahore, Pakistan, in 2012, the M.Sc. degree in Electrical and Electronics Engineering from Coventry University, United Kingdom, in 2013, and the Ph.D. degree in Electrical Engineering from Shanghai Jiao Tong University, China, in 2019. He is currently serving as an Assistant Professor/HoD in the Department of Computer Engineering at COMSATS University Islamabad, Sahiwal Campus, Pakistan. His research interests include electricity markets, renewable energy systems, and the application of artificial intelligence in energy forecasting, decentralized trading, and explainable decision support systems. (Email: waqas@cuisahiwal.edu.pk)

Wajiha Farooq is a lecturer at COMSATS University Islamabad, Sahiwal Campus, in the Department of Computer Science. Her research interests include underwater wireless sensor networks (UWSNs), machine learning, deep learning, and edge computing. She is currently exploring AI-based solutions for real-world sensing and classification problems. (Email: wajihafarooq980@gmail.com)

Aiza Mukhtar is a dedicated researcher specializing in AI, healthcare analytics, and IoT-based systems. Her work focuses on leveraging machine learning and deep learning for medical diagnostics and cybersecurity. Passionate about innovation, she aims to develop intelligent tools for early disease detection and improved healthcare. (Email: misaizarao1122@gmail.com)



Momina Zahid is a researcher focused on AI, healthcare analytics, and IoT, aiming to develop smart tools for early disease detection and better healthcare through machine learning and deep learning. Her work also addresses cybersecurity challenges in medical systems. She is committed to advancing innovative, data-driven solutions for real-world clinical applications. She also works as a UI/UX designer, combining technical expertise with creative design to develop user-centered solutions. (Email: mominazahid008@gmail.com)

Iqra is an accomplished researcher with expertise in artificial intelligence, healthcare analytics, and IoT-driven solutions. She focuses on applying machine learning and deep learning techniques to enhance medical diagnostics and strengthen cybersecurity. Driven by innovation, her goal is to design intelligent systems that enable early disease detection and advance healthcare outcomes. (Email: iqrabehzad234@gmail.com)