



Facial Emotion Recognition – A Comprehensive Review of Deep Learning and Traditional Learning Approaches with Emerging Challenges

Sushilkumar Salve^{1*}, Shubham Kumar¹, Irfan Shaikh¹, Pranay Gawade¹, Sujit Ramesh Borey², and Puja Padiya³

¹ Department of Electronics and Telecommunication Engineering, Sinhgad Institute of Technology, Lonavala, Maharashtra, India

² Department of Electronics and Communication Engineering Visvesvaraya National Institute of Technology (VNIT), Nagpur, Maharashtra, India

³ Xavier Institute of Engineering, Mumbai, Maharashtra, India

* Correspondence: sushil.472@gmail.com

Citation:

Salve, S.; Kumar, S.; Shaikh, I.; Gawade, P.; Borey R.S.; Padiya, P. Facial emotion recognition – A comprehensive review of deep learning and traditional learning approaches with emerging challenges. *ASEAN J. Sci. Tech. Report.* **2026**, *29*(7), e262457. <https://doi.org/10.55164/ajstr.v29i7.262457>

Article history:

Received: October 30, 2025

Revised: April 11, 2026

Accepted: April 12, 2026

Available online: June 15, 2026

Publisher's Note:

This article is published and distributed under the terms of the Thaksin University.

Abstract: In the domains of computer vision and artificial intelligence, facial expression-based emotion recognition (FER) has emerged as a key research area and industrial challenge. Its growing importance is evident in applications such as intelligent surveillance, healthcare monitoring, interactive systems, and mental health assessment. This review systematically examines 95 research papers published between 2014 and 2024, focusing exclusively on studies employing facial imagery for FER. The reviewed works encompass traditional machine learning, deep learning, and hybrid approaches, and analyze their methodologies for preprocessing, data augmentation, and temporal feature extraction. Traditional machine learning methods such as SVM, KNN, and Random Forest achieved average accuracies ranging from 85% to 93%, with the best reaching 99% on the CK+ dataset. Deep learning architectures, including VGG16, ResNet-50, MobileNetV2, and DenseNet-161, reported accuracy improvements of up to 99.5% across benchmark datasets such as FER2013 (35,887 images), CK+, JAFFE, KDEF, and RAF-DB (~40,000 images). Hybrid models combining CNN feature extraction with classifiers such as SVM and GMM achieved 3–6% improvements in recognition accuracy over standalone CNN or SVM systems, reaching 99.69% on CK+ and 99.53% on JAFFE. By comparing benchmark datasets, algorithms, and performance metrics, this study highlights the significant evolution from handcrafted feature-based systems to fully automated deep learning pipelines. The findings provide a quantitative synthesis of the current state of FER research, offering both newcomers and advanced researchers a comprehensive understanding of the field's progress, challenges, and future opportunities.

Keywords: Facial expression recognition (FER); deep learning; hybrid models; temporal feature extraction; motion classification.

1. Introduction

A study shows that approximately 67% of human communication is nonverbal, leaving only 33% for spoken language [1, 2]. Facial expressions are the primary means of nonverbal communication, enabling people to interpret their emotions and intentions. Along with vocal tone, these expressions help identify states such as happiness, sadness, or anger. Because facial expressions

carry strong emotional meaning, their study has expanded rapidly in recent decades, with applications ranging from perceptual science and affective computing to computer animation and cognitive research. Furthermore, facial expression analysis has gained significance in mental health, psychology, and psychiatry [3]. The starting point of any Facial Emotion Recognition (FER) system is face detection, which isolates the face or its features from single images or video streams. Since images often contain complex backgrounds, preprocessing becomes essential [4]. While humans effortlessly identify faces and expressions even in such conditions, machines without proper training face challenges [5,6]. The primary goal of face detection is to distinguish useful facial regions from irrelevant elements. Its applications include surveillance video systems, gesture-based interactive recognition, automated photography, gender detection, tagging, and teleconferencing [7,8]. Many current methods operate primarily on grayscale inputs, though a limited number can process color images as well. Pixel-based approaches often fail to distinguish faces from similar regions such as hands, whereas window-based methods are restricted in analyzing multi-view characteristics [9,10]. For this reason, model matching remains the most prevalent technique for face detection in FER pipelines.

The main objective of FER is to build relationships between facial expressions and the internal feelings they represent [11]. Conventionally, FER systems involve two fundamental processes: emotion detection and feature extraction. Image preprocessing, including normalization, scaling, cropping, and face localization, is a critical step for ensuring accuracy [12]. Removing irrelevant background elements narrows the focus to facial regions. Feature extraction then follows, with common methods such as Discrete Wavelet Transform (DWT) and Linear Discriminant Analysis (LDA) being widely used [13]. Once features are obtained, Neural Networks (NN) and other machine learning models are typically employed for classification. This stage also benefits from preprocessing steps such as alignment and rotation correction [14]. Cascade classifiers, particularly those based on AdaBoost, are highly reliable for face detection [15]. FER often combines geometric and appearance-based approaches. Geometric techniques, for instance, map facial landmarks into vectors containing positional attributes such as coordinates, angles, and distances [16]. To better capture facial variations, spatial analysis of external features has also been employed [17]. Motion characteristics play a significant role in recognizing dynamic expressions, with classifiers designed to handle a diverse set of emotions [18].

In recent years, attention has shifted towards deep learning architectures for FER [19]. Convolutional Neural Networks (CNNs), in particular, have demonstrated better accuracy than classical methods such as Recurrent Neural Networks (RNNs) and Support Vector Machines (SVMs), though they are resource-intensive [20]. Large, well-annotated datasets are vital for improving training performance. FER research has also expanded to related tasks, such as predicting age and gender from facial cues [21,22]. Smart healthcare applications showcase the practical advantages of FER, in which intelligent cameras continuously capture patients' facial data and send it to cloud platforms for monitoring [23]. In addition, lightweight architectures designed for embedded devices, such as the Jetson Nano, have potential applications in law enforcement, particularly for suspect identification [24]. Among all the approaches, CNN-based systems are considered the most advanced, with specialized architectures created for emotion recognition in static images [25]. To further improve recognition rates, researchers suggest networks containing multiple convolutional and hidden layers [26]. Preprocessing remains vital for removing noise and extracting key features during pretraining [27]. Recent work has also examined how factors such as gender, activity, and occupation influence emotional expression in interactive settings, such as robotics labs and classrooms.

Figure 1 presents a pictorial view of the growing research interest in emotion recognition over 15 years, from 2010 to 2025. This is a plot of a distinct, strong, increasing trend, showing an exponential rise in publication numbers. This is an indication of the field's growing significance and the rapid development of computational models and applications that have attracted increasing interest from the scientific community. Figure 2 provides a conceptual delineation of the major uses of emotion recognition technology in the principal sectors of healthcare and teaching. As the data shows, healthcare usage, including patient monitoring and therapeutic assistants, is a principal field of application. In the teaching arena, technology is used through applications such as monitoring student participation and enabling personalized learning. This illustration highlights the real-world application and societal contribution of FER systems. In summary, the growing integration of emotion-aware systems into intelligent environments reflects a broader trend toward human-

centered artificial intelligence. As FER continues to advance through deep learning and hybrid models, its potential extends across smart healthcare, adaptive education, and intelligent surveillance systems. However, real-world deployment still faces challenges related to data privacy, cultural variability, and interpretability. Thus, this review aims to consolidate current findings, compare methodologies, and identify emerging opportunities and challenges to guide future research in developing reliable and ethically responsible FER systems.

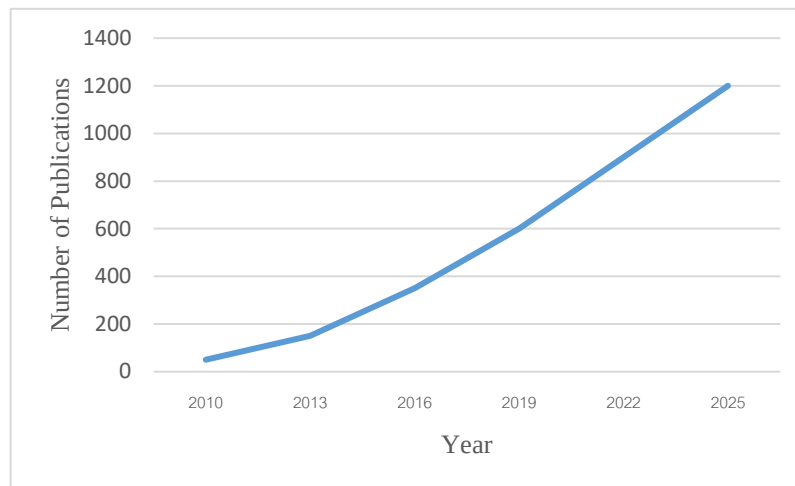


Figure 1. Growth of publications in emotion recognition [4,19,30].

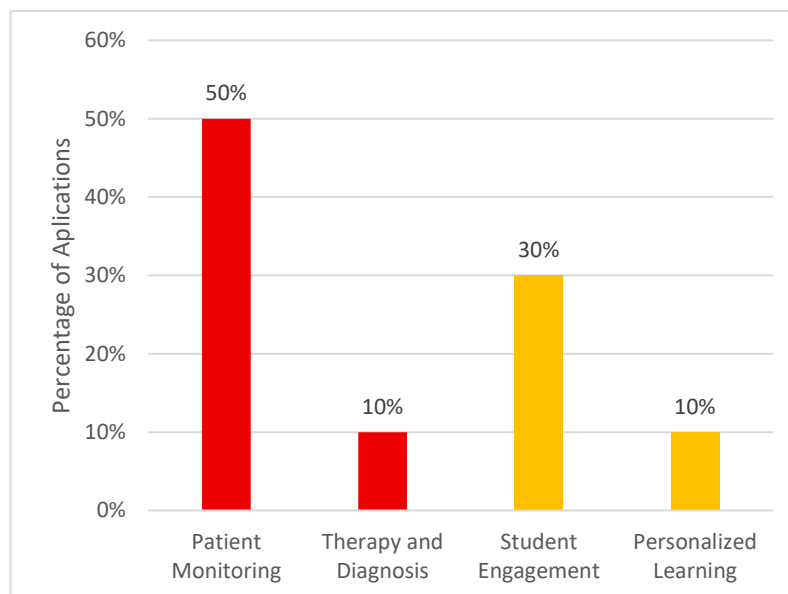


Figure 2. Global applications of emotion recognition in education/healthcare.[3, 23, 28]

2. Method of Review

The methodology adopted for this review follows a systematic literature review (SLR) framework designed to synthesize and critically evaluate recent advances in Facial Emotion Recognition (FER) across machine learning, deep learning, and hybrid paradigms. The study focuses on applications of FER in healthcare monitoring, intelligent surveillance, behavioral analytics, education, and adaptive user interfaces. The review process was conducted in accordance with established SLR guidelines and comprises three key stages: literature search, screening and selection, and analysis and synthesis.

2.1 Literature Search Strategy

Relevant publications were systematically retrieved from five major scientific databases: IEEE Xplore, ScienceDirect, SpringerLink, Wiley Online Library, and Google Scholar. To ensure comprehensive coverage, the following Boolean keyword combinations were used during the search:

("Facial Emotion Recognition" OR "Facial Expression Recognition") AND ("Machine Learning" OR "Deep Learning" OR "Hybrid Models" OR "Affective Computing") AND ("Datasets" OR "Feature Extraction" OR "Preprocessing").

The search covered 10 years (2014–2024) and captured both foundational and state-of-the-art studies. Additional manual searches were conducted in conference proceedings such as *IEEE ICMEAS* and *CVPR* to include the most recent advancements.

2.2 Inclusion and Exclusion Criteria

The retrieved articles were filtered using the following criteria:

Inclusion Criteria:

- Studies focused explicitly on facial emotion or facial expression recognition using machine learning, deep learning, or hybrid techniques.
- Peer-reviewed journal or conference papers published between 2014 and 2024.
- Articles that included quantifiable performance metrics (accuracy, precision, recall, or F1-score).
- Studies employing publicly available datasets such as FER2013, CK+, JAFFE, RAF-DB, MMI, and KDEF.

Exclusion Criteria:

- Non-English publications, theses, and unpublished manuscripts.
- Studies that focused solely on speech or gesture emotion recognition without facial analysis.
- Duplicate or overlapping studies were retrieved from multiple databases.

2.3 Screening and Filtration Process

An initial pool of approximately 230 publications was identified. After title and abstract screening, 120 studies were shortlisted for full-text evaluation. After applying the inclusion and exclusion criteria, 95 papers were selected for detailed analysis. Each paper was examined in terms of:

1. Dataset characteristics and preprocessing strategies.
2. Feature extraction and augmentation techniques.
3. Model architecture (traditional, deep, or hybrid).
4. Performance metrics and benchmarking protocols.

2.4 Analysis and Synthesis

The selected studies were categorized into three methodological groups—traditional machine learning, deep learning, and hybrid approaches—to enable structured comparison. Quantitative data, such as accuracy and dataset size, were tabulated for benchmarking, while qualitative aspects, such as robustness, computational cost, and ethical challenges, were synthesized thematically. Statistical trends and performance comparisons were then consolidated to provide a holistic overview of the evolution and current state of FER research.

2.5 Keywords and search strategy

Literature searching was supplemented with a combination of prespecified keywords and Boolean operators to achieve a relevant, focused retrieval of studies. The following key search terms were used:

2.2.1 "Facial Expression Recognition (FER)" AND "Internet of Things (IoT)".

2.2.2 "Deep Learning".

2.2.3 "Hybrid Models" AND "Affective Computing".

2.2.4 "Temporal Feature Extraction".

2.2.5 "Emotion Classification" AND "Benchmark Datasets."

The searches were performed across all fields (keyword, title, abstract, and full text, where available). Truncation and wildcard characters were used to capture variation in keywords (e.g., sensor, sensing, sensors).

3. Machine Learning Approaches

One common challenge is that while the main existing systems perform effectively on images from a regulated image environment, the model's performance may decline on difficult datasets due to increased image variation and partially visible facial regions. To address the concern, one approach uses the Histogram of Oriented Gradients (HOG) descriptor. The approach starts with the initial processing of the input image to identify the reference region for extracting the critical content, followed by the development of a face emotion classifier using the RF algorithm. This method's effectiveness and accuracy were demonstrated through analysis on the Japanese Female Facial Emotions (JAFPE) Database [32].

3.1 Introduction to different machine learning approaches

Within the domain of facial emotion recognition, several machine learning algorithms have been extensively explored by researchers. Among these, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF) are the most frequently used due to their simplicity, robustness, and effective classification performance on small and medium-sized datasets. For instance, one study applied facial expression recognition to classify consumer gratification with products and services. Their architecture involved three primary stages: data collection (including user satisfaction, age, gender, and facial expressions), classification using machine learning algorithms, and a final model assessment to confirm accuracy [33].

X, Liu et al. proposed a method for sentiment analysis of subjects by employing two new geometry-based attributes: the landmark curve and the vectorized landmark, extracted from multiple facial regions related to facial muscle activity [34]. N. Iromini et al. implemented an innovative geometric extraction technique that uses six different distance measurements to compute facial features that optimally reflect one emotion, followed by the application of a decision tree to the JAFPE dataset. This system could distinguish between seven discrete emotions and achieved an identification rate of 89.20% on the JAFPE database, whereas the COHEN database achieved an identification rate of 90.61% [36]. For face detection, investigators have applied Principal Component Analysis (PCA) to identify faces using a reduced set of attributes and have utilized the KNN algorithm to derive features from the JAFPE dataset, achieving 85.71% accuracy after image preprocessing [37].

Other studies have focused on rigorous validation and direct classifier comparisons. H. I. et al. applied one-way analysis of variance and fivefold cross-validation ($p < 0.01$) to confirm the precision of the derived features, ultimately training KNN and Decision Tree (DT) classifiers to identify multiple categories of facial emotional states. KNN recorded a mean accuracy of 98.03%, whereas DT averaged 97.21% rate [38]. A comprehensive examination of a separate experiment revealed that the approach performed best when an SVM classifier was used. The accuracy declined to 79.98% with KNN and rose to 82.96% when Multilayer Perceptron (MLP) was employed [39]. To enhance performance, Linear Discriminant Analysis (LDA) has been used to improve SVM classification, achieving 99% accuracy and a 98% F1-score. Similarly, B. Tamer Ghareeb et al. integrated SVM, Random Forest, KNN, and Naive Bayes algorithms, which exhibited varying degrees of accuracy [40].

Another enhancement technique proposed by B. Tamer et al. involves using Principal Component Analysis (PCA) approach on effective Haar Face features to reduce dimensionality, allowing the resulting Eigenvectors to be used with an SVM classifier to detect fundamental human emotions [41]. SVM methods and Active learning have also been adopted to recognize human facial emotions by categorizing facial action units into classifications [42]. A study used the Yale Face dataset, which includes 11 human facial states, such as normal, happy, sad, sleepy, and surprised. Using sensitivity, precision, F-measure, G-mean, and accuracy as performance evaluation metrics, the proposed BGA-RF method achieved a maximum precision of 96.03%. The experimental outcomes confirmed the method's efficacy, highlighting its enhanced accuracy compared to similar methods [43].

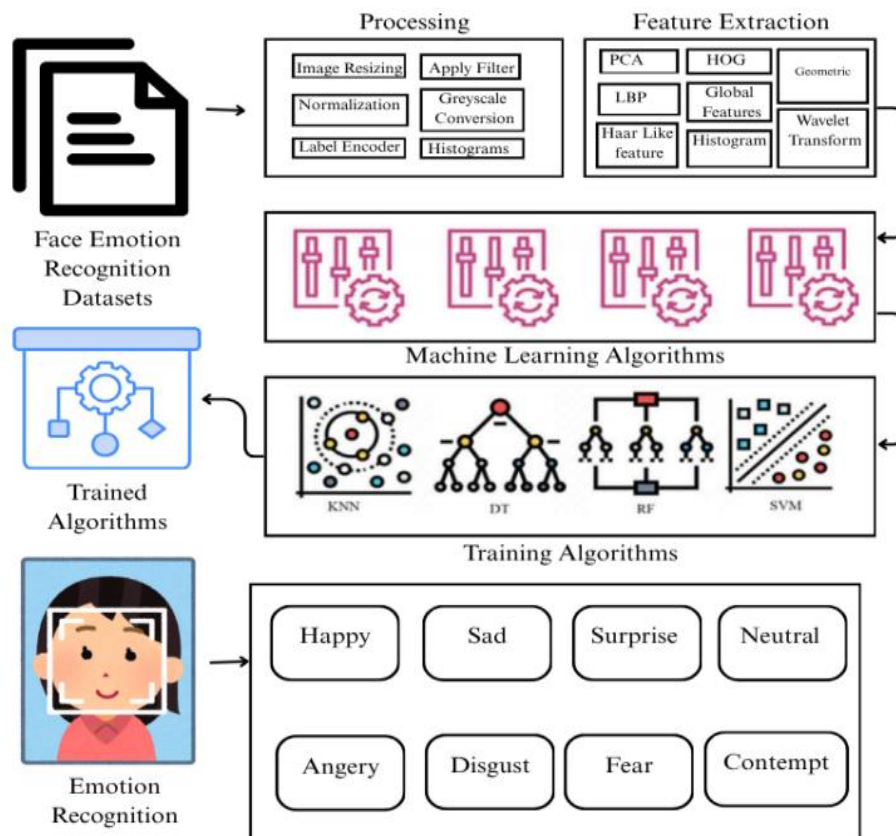


Figure 3. Computational Approaches to FER using Machine Learning.

Figure 3 shows a typical pipeline of an FER system with machine learning. This pipeline is activated by a labeled image dataset, which is prepared (e.g., resizing, normalization) to bring the images into a standard format. Then, discriminative face features are extracted through methods such as HOG, LBP, or PCA. These extracted face features are now passed to a Machine Learning Algorithm for training, e.g., SVM, KNN, or a Random Forest. This trained algorithm can now accept a new face image and return its class, indicating the emotion it belongs to, such as happy, sad, or angry.

Table 1. Comprehensive evaluation of machine learning for FER system.

Reference	Databases	Techniques	Properties	Emotion
[31]	FEER	SVM	LDA, PCA	6
[32]	JAFPE	RF	Histograms of oriented gradient (HOG)	7
[33]	Private	SVM	No	7
[34]	CK+, MUG	SVM	Geometric	8
[35]	CK+, JAFPE	CART	Geometric and global features	7
[36]	JAFPE	KNN	PCA	7
[37]	Private	KNN	Haar-like features	6
[38]	CK+	SVM	LBP, Gabor filter, PCA	7
[39]	CK+	SVM	PCA	7
[40]	AffectNet	SVM	Grid search	8
[41]	FER 13	LR	No	5
[42]	CK+	SVM	PCA and LBP	7
[43]	Yale faces	RF	Histograms of oriented gradient (HOG)	11

Table 1 provides a comparative overview of machine learning techniques used for facial emotion recognition (FER) across multiple benchmark datasets. The studies used algorithms such as SVM, KNN, RF, CART, and LR, often combined with feature extraction methods such as PCA, LBP, Gabor filters, and geometric features. The number of recognized emotions ranged from 5 to 11, depending on the dataset and feature type. Notably, SVM emerged as the most widely used classifier due to its strong generalization across datasets such as CK+, JAFFE, and AffectNet. Overall, this comparison highlights how the choice of features and classifiers significantly influences FER system performance. Note: The “Emotion” column indicates the number of discrete emotional classes recognized in each study (e.g., six basic emotions — happiness, sadness, anger, fear, surprise, and disgust — or extended sets including neutral and contempt)

4. Deep Learning Approaches

Deep learning has witnessed remarkable progress in Facial Emotion Recognition (FER), with researchers achieving high accuracy using various pre-trained models and novel architectures. For instance, one methodology that used pre-configured models found that DenseNet-161 achieved the highest accuracy on both the JAFFE and Karolinska Directed Emotional Faces (KDEF) test sets, reaching 99.52% and 96.51%, respectively, using a 10-fold cross-validation method [44]. Z. Y. Huang et al. have developed a high-performing model, the Facial Emotion Recognition using Convolutional Neural Networks (FERC) model, which uses an Expressional Vector (EV) and a two-level CNN to identify five distinct facial gestures, attaining an accuracy rate of 96% on a database of 10,000 images [45]. In a comprehensive evaluation across 25 cutting-edge techniques on 5 standard datasets (including JAFFE and CK+), one proposed methodology was found to perform better than all other advanced methods on every dataset [46]. For specific real-time applications, deep learning models have been rigorously compared. A study proposed by S. Gupta et al. focused on measuring online learners' engagement by analyzing facial expressions compared to Inception-V3 (89.11% accuracy), VGG19 (90.14%), and ResNet-50. The results showed that ResNet-50 surpassed the other models, achieving 92.32% accuracy in real-time facial emotion categorization on widely used benchmark datasets, namely the Real-world Affective Faces Database (RAF-DB), CK+, and FER-2013 [47].

4.1 Different methodologies in deep learning

A major focus in FER research is creating efficient, lightweight networks that do not sacrifice accuracy. MobileNetV2, a CNN model known for its speed, produces quick and accurate real-time output when integrated with supervised transfer learning, an island loss, and a joint supervision strategy, surpassing other approaches on the CK+, JAFFE, and FER2013 datasets [48]. Building on this, a lightweight model called A-MobileNet integrates an attention module into MobileNetV1 to enhance local feature extraction. By combining softmax loss and center loss to tune the model's parameters, this strategy improves accuracy while maintaining the original MobileNet series' model size, demonstrating superior efficiency on the FERPlus and RAF-DB datasets [49].

Beyond lightweight models, researchers have improved existing architectures. C. Dong et al. proposed modifications to the VGG16 network by replacing its three dense layers with one dense layer and two convolutional layers, and switching to local-based adaptive pooling before training on the Static Facial Expressions in the Wild (SFEW) and RAF-DB datasets [51]. Innovations in residual networks include a residual masking network, which merges a deep residual network with a U-Net-like design to retain high accuracy on the Versatile Elastic Multi-modal (VEMO) and FER2013 datasets [52]. B. Li et al.'s method offers a way to extract attributes from ResNet-50 that exhibit superior performance in experimental simulations compared to the most popular existing models [53]. Furthermore, a deep learning approach leveraging attentional convolutional networks has been shown to surpass previous models by precisely identifying and focusing on important facial attributes across datasets such as CK+, FER-2013, JAFFE, and FERF [54].

A key challenge is that existing models are often inadequate for real-world situations involving inconsistent lighting conditions, partially hidden faces, and random noise in images. To solve this, L. Xiong et al. have proposed integrating a Context Transformer (CoT) module with CNN and Vision Transformer (ViT) architectures [55]. This improves the model's resilience to illumination and its capability to uncover intricate dependencies among local characteristics in intricate situations [56]. Advanced techniques proposed by

Z. Gao et al. and M. Verma et al. include a hierarchical attention architecture incorporating a progressive feature fusion methodology, a selective kernel network, and an innovative feature descriptor called the Cross-Centroid Ripple Pattern (CRIP) [57,58]. For specialized conditions, researchers have developed a method for detecting faces using drones to improve identification in poor lighting or aerial scenarios [59]. A method for identifying faces emotion from masked facial images using CNN feature analysis and low-light image augmentation [60].

Many projects have focused on building deep learning frameworks for specific tasks. E. Pranav et al. in their study have aimed to build a Deep Convolutional Neural Network (DCNN) that can recognize a set number of distinct emotions, often using manually compiled photo collections for training and verification [61]. A real-time face emotion classification and recognition using a deep learning model study demonstrated this by training a CNN framework on the JAFFE and FER-2013 datasets, achieving 98.64% and 70.01% accuracy, respectively [62]. Real-time recognition systems have applications in security, healthcare, and safety assessments, with some systems using a sequential approach that combines Haar cascade detection for facial identification with a Keras CNN model for emotion analysis. These deep learning approaches are attracting considerable attention from the research community and are becoming crucial for classification, enabling innovative biomedical applications and even systems that detect individuals with autism instantaneously [63]. Even with extensive research, facial expressions remain intricate and dynamic, hindering accurate identification. Based on both laboratory and real-world findings, it is evident that deep learning holds significant potential for the future of facial emotion recognition. The field is constantly advancing, with researchers now refining models that build on ConvNeXt and Vision Transformer, using improved hyperparameters on new datasets [64].

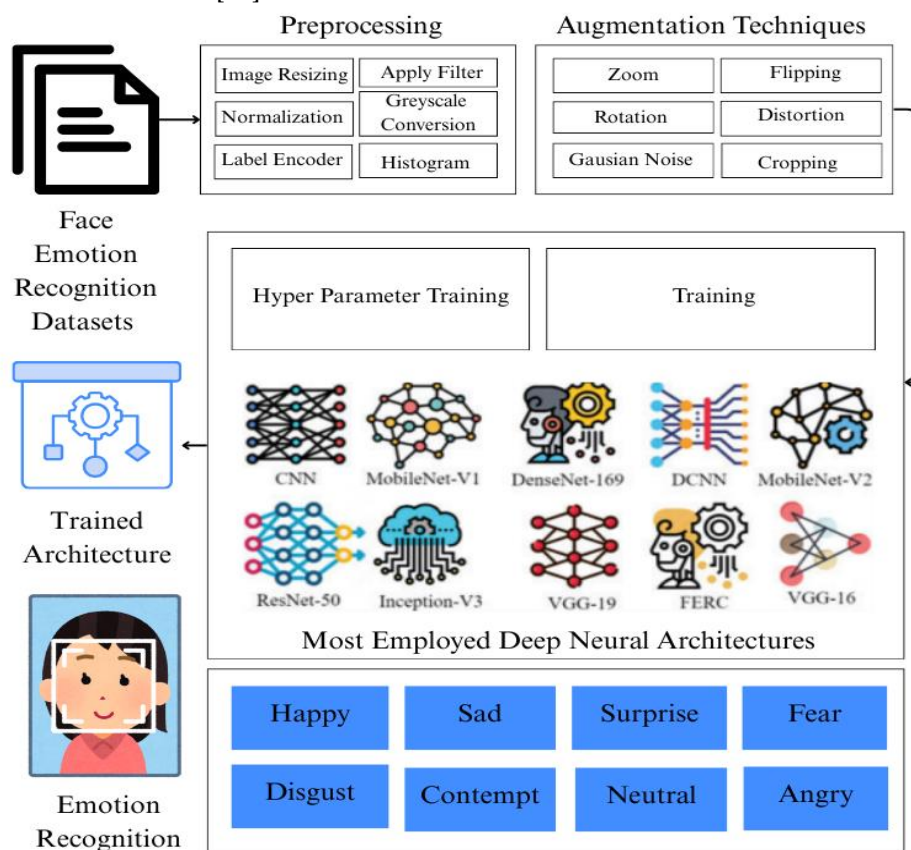


Figure 4. Computational Approaches to FER using Deep Learning

Figure 4 illustrates an end-to-end, complete pipeline of an FER system that is deep learning-based. A large image Dataset is first made ready by preprocessing (e.g., normalization, resizing) and augmentation (e.g., rotation, flipping) to make it robust. Such prepared images are then fed directly to a deep neural network, e.g., a CNN, ResNet, or even MobileNet, to train and perform hyperparameter tuning. Finally, the trained

architecture is ready to classify emotions from new facial images. Note: The “Emotion” column indicates the number of discrete emotional classes recognized in each study (e.g., six basic emotions — happiness, sadness, anger, fear, surprise, and disgust — or extended sets including neutral and contempt)

Table 2. Comprehensive evaluation of deep learning for FER system

Reference	Datasets	Methods	Emotions
[44]	NIST, CMU, and Caltech face	FERC	7
[46]	FERPlus and RAF-DB	MobileNet	7
[47]	FER2013 and CK+	VGG16	7
[49]	JAFPE,CK+ and FER2013	MobileNet V2	7
[51]	Random dataset, FER-2013	MobileNet-V1	7
[52]	CK	ResNet-50	7
[53]	KDEF	VGG16	6
[54]	KDEF and JAFPE	DenseNet-161	7

Table 2 summarizes the performance of various deep learning models for facial emotion recognition (FER) across different datasets. The studies primarily focused on CNN-based architectures such as VGG16, ResNet-50, MobileNet (V1 and V2), DenseNet-161, and FERC, across well-known datasets including FER2013, CK+, JAFPE, KDEF, RAF-DB, and NIST. Most approaches focused on recognizing six to seven basic emotions, demonstrating the robustness of convolutional neural networks in learning spatial and hierarchical facial features. Among these, lightweight models such as MobileNet variants offered efficient performance suitable for real-time FER applications, while deeper architectures such as VGG16 and DenseNet-161 achieved higher accuracy on more complex datasets.

5. Hybrid Approaches

Hybrid learning in FER often involves integrating the powerful feature extraction of deep networks with the effectiveness of traditional classifiers. For instance, the Hybrid facial emotion recognition using CNN-based features study used a Deep Belief Network (DBN) to create a comprehensive video feature representation by averaging segment-level features, and then leveraged a Support Vector Machine (SVM) for classification. The approach was particularly useful for complex scenarios, such as interpreting emotions from concealed faces [65]. It's known that concealing the mouth complicates emotion recognition, and to address this, H. M. Shahzad et al. used prominent deep learning models such as VGG-16 and AlexNet for feature extraction, followed by multiple classifiers to improve predictive accuracy [66]. Another method employed facial recognition software to derive extract Network Difference Features (NDF) by analyzing connections between adjacent locations, before using a random forest classifier to sort expressions into seven groups [67].

5.1 Different methodologies in deep learning

More intricate integrations involve combining various kinds of neural networks. To capture the geometric interactions among facial landmarks, one approach integrates Attention Long Short-Term Memory (ALSTM) networks with Spatial Attention Convolutional Neural Networks (SACNNs) to develop robust integrated features for expression identification [68]. The concept of Deep Hybrid Learning (DHL) involves combining deep learning and transfer learning methods, and one study compared models such as CNN, LSTM, Inception, and ResNet to categorize facial emotions as happy or angry [69]. Another successful strategy is the fusion of shallow and deep features. Cross-database studies have demonstrated the immense potential of this approach, yielding promising results compared with other cutting-edge models [70]. A specific example of this is a combined method that uses the energy-entropy-based Dual-Tree M-Band Wavelet Transform (DTMBWT) for high-precision feature extraction, which is then classified using a Gaussian mixture model (GMM) [71].

Several advanced hybrid systems have been developed for unique applications. One state-of-the-art system sorts facial motion flows by combining CNN-based classification of facial geometry motion and SVM-based classification of facial motion, achieving much higher performance than individual classifiers [72].

Another model introduces a deep learning approach that uses one CNN to assess a primary emotion (such as neutral or angry) and a subsequent CNN to determine the subordinate emotion [73]. To achieve results comparable to human assessors, a Gaussian Mixture Model (GMM) and a Deep Neural Network (DNN) were tested in .in under both low- and high-noise conditions [74]. Furthermore, in scenarios where facial expressions may be fundamentally misleading, a multimodal emotion recognition paradigm is highly effective at accurately determining emotions, with performance that surpasses or compares favorably with that reported in the existing literature [75]. A comprehensive investigation of the efficacy of Hybrid learning for the FER system is presented in Table 3.

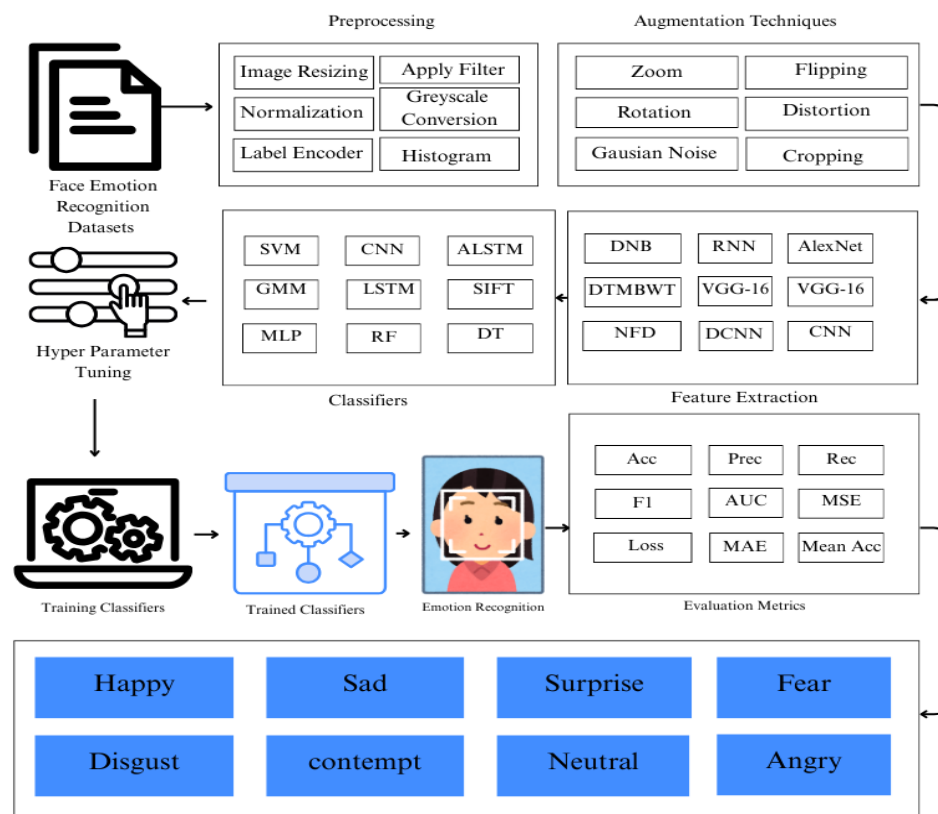


Figure 5. Computational Approaches to FER using Hybrid Learning

Figure 5 illustrates a hybrid FER pipeline that divides feature extraction from classification. Following initial Preprocessing and Augmentation of a prepared Dataset, a deep network (e.g., a CNN or VGG-16) serves as a Feature Extractor. These higher-level features are passed through a distinct, sometimes traditional, Classifier (e.g., SVM, Random Forest) to produce an emotion prediction. The model is optimized through training and confirmed using evaluation metrics.

Table 3. Comprehensive evaluation of hybrid learning for FER system.

Reference	Datasets	Feature extraction	Segmentation	Emotions
[65]	BAUM-1s, RML, and MMI	Deep belief network (DBN)	SVM	6
[66]	M-LFW-FER	AlexNet and VGG-16, CNN	SVM	7
[67]	SFEW, RAF	Neighborhood difference features NFD	RF	7
[68]	FER2013, CK+, and JAFFE	Spatial attention convolutional neural network (SACNN)	Long short-term memory networks with attention mechanism (ALSTMs)	7
[69]	Multisource	CNN	LSTM	6
[70]	CK+	CNN	-	7
[71]	JAFFE	Dual-tree M-band Wavelet Transform (DTMBWT)	Gaussian Mixture Model (GMM)	7
[72]	CK+, BU4D	CNN	SVM	7
[73]	JAFFE, FER2013	CNN	CNN	7
[74]	ESD	Gaussian Mixture Model and Deep Neural Network (GMM-DNN)	SVMs and MLP	6
[75]	LUMED-2, DEAP	CNN	DT	3

Table 4. Benchmarking Datasets for Facial Emotion Recognition Systems.

Dataset	Description
MUG [76]	The database contains details of 86 Caucasian individuals aged 20-35 years. Among them, 35 are women, and 51 are men, with some participants being clean-shaven. Apart from seven entries found in the latter portion of the dataset, no subjects were related to sports. The few irregularities observed were stray facial hairs. In some images in the second part, a hand was partially covering the face.
RAF-DB [77]	The Real-world Affective Faces Database (RAF-DB) is designed for facial emotion analysis. It contains 39,672 images annotated by 40 independent evaluators, covering both basic and compound emotional expressions. The dataset includes individuals of varying ages, genders, and ethnicities, as well as a wide range of lighting conditions. It also accounts for different types of obstructions, such as glasses, self-obstruction, or facial hair, and incorporates subsequent processing steps, such as filters and special effects.
MMI [78]	This dataset contains more than 2,900 videos and high-definition static images of 75 individuals. Event coding tracks the presence of Action Units (AUs) in video sequences, while incomplete-frame encoding indicates whether an AU is at the onset, peak, offset, or neutral stage within a given frame. In addition, limited annotations are provided for multimodal entertainment purposes. The researchers can freely access the database.
FER2013 [79]	The FER+ dataset classifies images into neutral, surprised, angry, disgusted, sad, contemptuous, and fearful categories. It serves an enhanced version of the initial FER dataset. In addition to the standard FER labels, FER+ includes supplementary annotations. These annotations are generated by averaging responses from 10 independent crowdsourced evaluators, resulting in more reliable ground-truth labels for emotions in static images than in the original FER dataset. Using multiple annotators per image also allows researchers to derive a probabilistic model of each face's emotions.
KDEF [80]	The dataset consists of colour facial images without visible hairlines, taken from the Karolinska Directed Emotional Faces (KDEF) database. It includes 490 photographs representing 70 individuals, evenly divided into 35 males and 35 females. As reported by Lundqvist et al. (1998), the facial configurations captured cover a range of intensities, from neutral to a full emotional display, including disgust, happiness, surprise, fear, and sadness. All images are presented in a frontal view.

A comparative examination of learning techniques uncovers the most important trade-off between hand and algorithmic methods. Classical machine learning involves handcrafted feature engineering, so it is efficient for smaller datasets, where domain knowledge can guide simpler algorithms with lower computational expense. Conversely, deep learning learns features automatically using complex neural networks that perform better on enormous datasets but require substantial computational power and large amounts of data to learn well. Hybrid learning addresses the gap between these two paradigms by combining techniques from both in synergistic models, solving complex problems more accurately. Among the above approaches, data augmentation is an essential technique that artificially enlarges training data to enhance model generalization, avoid overfitting, and improve overall robustness. Therefore, the comparative study of various learning approaches is shown in Table 5.

Table 5. Comparative study of various learning approaches.

Reference	Datasets	Classification	Performance (%)	Comments
[33]	Private	SVM-SMOTE	86.01	The results were confirmed in subsequent presentations, and the initial experiment validated the method for collecting the machine learning dataset. The study demonstrated the technique's ability to determine user satisfaction accurately.
[34]	JAFFE and CK+	CART	89.21	Facial distance was calculated using Minkowski, Manhattan, or Euclidean metrics. A decision tree was then used to analyze these three distance measurements. On the JAFFE dataset, the Minkowski distance yielded the best results.
[36]	JAFFE	KNN	85.71	The algorithm's effectiveness was assessed using MATLAB scripts, with accuracy and the number of correct versus incorrect classifications used to evaluate performance. The final results showed that 60 images were correctly classified and 10 were not, yielding an overall accuracy of 85.71%.
[37]	Private	KNN	98.02	This study used the Lucas-Kanade optical flow method to monitor marker placements continuously. The significance of distance-based features extracted from six different emotions was then examined using a single-factor analysis of variance.
[38]	CK+	SVM	93.53	This work presents an algorithm for identifying eight archetypal facial emotions. Principal Component Analysis (PCA) was applied to reduce the feature space. These authors trained

Table 5. Comparative study of various learning approaches. (continue)

Reference	Datasets	Classification	Performance (%)	Comments
[38]	CK+	SVM	93.53	and evaluated three different classifiers for categorization employing a 10-fold cross-validation approach.
[39]	CK+	SVM	91.5	The outcomes clearly confirmed the effectiveness of the recommended technique for reliably extracting facial expressions from video.
[46]	FERPlus and RAF-DB	MobileNet	88.11	According to experimental findings, the enhanced model outperforms other advanced techniques, including the lightweight MobileNet series. The model achieved 88.11% accuracy on FERPlus and 84.5% on the RAF-DB dataset.
[47]	FER2013 dataset and CK+	VGG16	57.34	The authors modified the pre-trained VGG16 model by substituting two entirely connected layers with convolutional layers. This was done to minimize network parameters and alter the pooling strategy.
[51]	FER-2013, Random dataset	MobileNet-V1	97.9	A key contribution of this research is the accurate and effective retrieval of seven distinct behaviors. The method successfully created a seven-class emotion recognition system using photographic inputs.
[53]	KDEF	VGG16	88	The CNN model, validated using performance metrics, achieved 88% accuracy. The results demonstrate that this network design significantly surpasses previous approaches.
[63]	FER-2013, CK+ and RAF-DB	ResNet-50	92.32	The proposed method was evaluated in an online learning scenario involving twenty students. The system successfully used automated facial expression recognition to differentiate between “engaged” and “disengaged” states.
[65]	BAUM-1s, RML, and MMI	SVM	55.85, 73.73, 71.43	The authors first trained a VGG16 model on ImageNet, then separately modified the temporal and spatial CNN networks using video data. A deep DBN was used to study multiple spatio-temporal CNN

Table 5. Comparative study of various learning approaches. (continue)

Reference	Datasets	Classification	Performance (%)	Comments
[65]	BAUM-1s, RML, and MMI	SVM	55.85, 73.73, 71.43	characteristics for effective integration jointly.
[66]	M-LFW-FER	SVM	65.5	The presence of a mask reduces realism by obscuring key facial features. A mask can also alter facial characteristics, making detection more difficult for an algorithm. Furthermore, a machine's accuracy may be limited by its inability to understand the complexity of human emotions fully.
[67]	RAF, SFEW	RF	57.7, 59	Previous approaches often required significant computational power due to their reliance on deep models and were limited to specific emotions. In contrast, the suggested technology is adaptable and can recognize complex expressions, with performance depending on the training.
[69]	Multisource	LSTM	81.42	Experimental results showed that VGG19 had the lowest and most consistent accuracy during training. Among the transfer learning methods tested, the InceptionResNet Model achieved the peak precision.
[70]	CK+	Scale-invariant feature transform (SIFT)	94.82	This work demonstrated a method for extracting features from individual image frames to recognize various facial expressions. By integrating SIFT with features from a deep learning CNN model, composite characteristics are created. An SVM is then used for expression analysis, considering all attributes simultaneously.
[71]	JAFFE	Gaussian mixture model (GMM)	99.53	This study introduces a method to enhance the accuracy of high-performance FER. To achieve the desired performance level, a hybrid approach is recommended, using the DTMBWT algorithm, which relies on the Gray-level Co-occurrence Matrix (GLCM) and energy.

Table 5. Comparative study of various learning approaches. (continue)

Reference	Datasets	Classification	Performance (%)	Comments
[72]	CK+, BU4D	SVM	99.69, 94.69	While the outlined techniques are useful for large datasets, they tend to generate an excessive number of interest points. A current limitation is the system's reliance on Western databases; the final goal is to incorporate an Eastern dataset.
[73]	JAFFE, FER2013	CNN	94.12, 97.07	This research employs a CNN for identifying the main emotion in an image (joyful or angry) and for estimating a subsequent emotion. When optimized on the JAFFE and FER2013 datasets, the model demonstrated greater effectiveness than traditional methods.
[74]	ESD	MLP and SVMs	83.97	A key limitation was the use of a synthetic Emirati speech sample that lacked naturalness, reflecting the general difficulty of capturing genuine emotions on video. Most available databases of this type operate similarly.
[75]	LUMED-2, DEAP	DT	74.2, 53.8	The EEG modality has low subject-independent accuracy because of its non-stationary nature, but this can be beneficial when there are few output classes. The system only classified two states—high/low valence and arousal—to determine emotional state.

6. Discussion

While many existing reviews have emphasized facial emotion recognition (FER) in various contexts, they often miss certain innovative approaches and fail to cover the full range of classification models available. The primary objective of the study is to analyze deep learning models capable of accurately detecting diverse facial emotions. Model training, testing, and validation were carried out using a manually designed image collection protocol. A referenced study highlighted the evolution of a CNN-based deep learning data-driven structure that successfully classified emotions from images with high precision. This work offers a thorough overview of deep learning, machine learning, and hybrid fuzzy systems applied to FER. It also explores a wide array of applications and the challenges associated with FER systems. The research aims to conduct a detailed literature review of recent publications and identify the most widely used architectures for emotional state detection in facial expression datasets, along with their preprocessing methods. The study initially examines classical machine learning techniques, including Random Forests (RF), Support Vector Machines (SVM), Decision Trees (DT), and K-Nearest Neighbors (KNN). The results varied depending on the datasets and the hyperparameter tuning. For example, RF achieved 99.59% accuracy that utilized the CK+ dataset for both training and validation, while SVM reached 57.10% accuracy on the RAF dataset. This study also summarizes several convolutional neural network (CNN) architectures, with a particular emphasis on deep learning models such as CNN, MobileNet, ResNet, and VGG-16. Both lightweight and intricate network designs were

evaluated in the FER context. A key challenge in comparing model performance was the use of different datasets—some containing image sequences and others containing single frames—limiting the consistency of training and testing.

Hybrid models outperform conventional CNN architectures primarily because they **combine** deep feature extraction with robust statistical classification. While CNNs automatically learn hierarchical spatial features, their softmax classifiers often struggle with small or imbalanced datasets. Integrating traditional classifiers such as SVMs, RFs, or GMMs enhances decision boundary precision and reduces overfitting, particularly when training data is limited. Moreover, hybrid systems leverage handcrafted features to capture subtle geometric and textural variations that CNNs miss. However, hybrid methods are not without limitations. They often involve higher computational complexity due to multi-stage training, and performance depends heavily on the feature fusion strategy. Additionally, their deployment in real-time applications remains challenging because of increased processing time and memory requirements. Future work should focus on optimizing lightweight hybrid architectures that maintain accuracy without compromising computational efficiency.

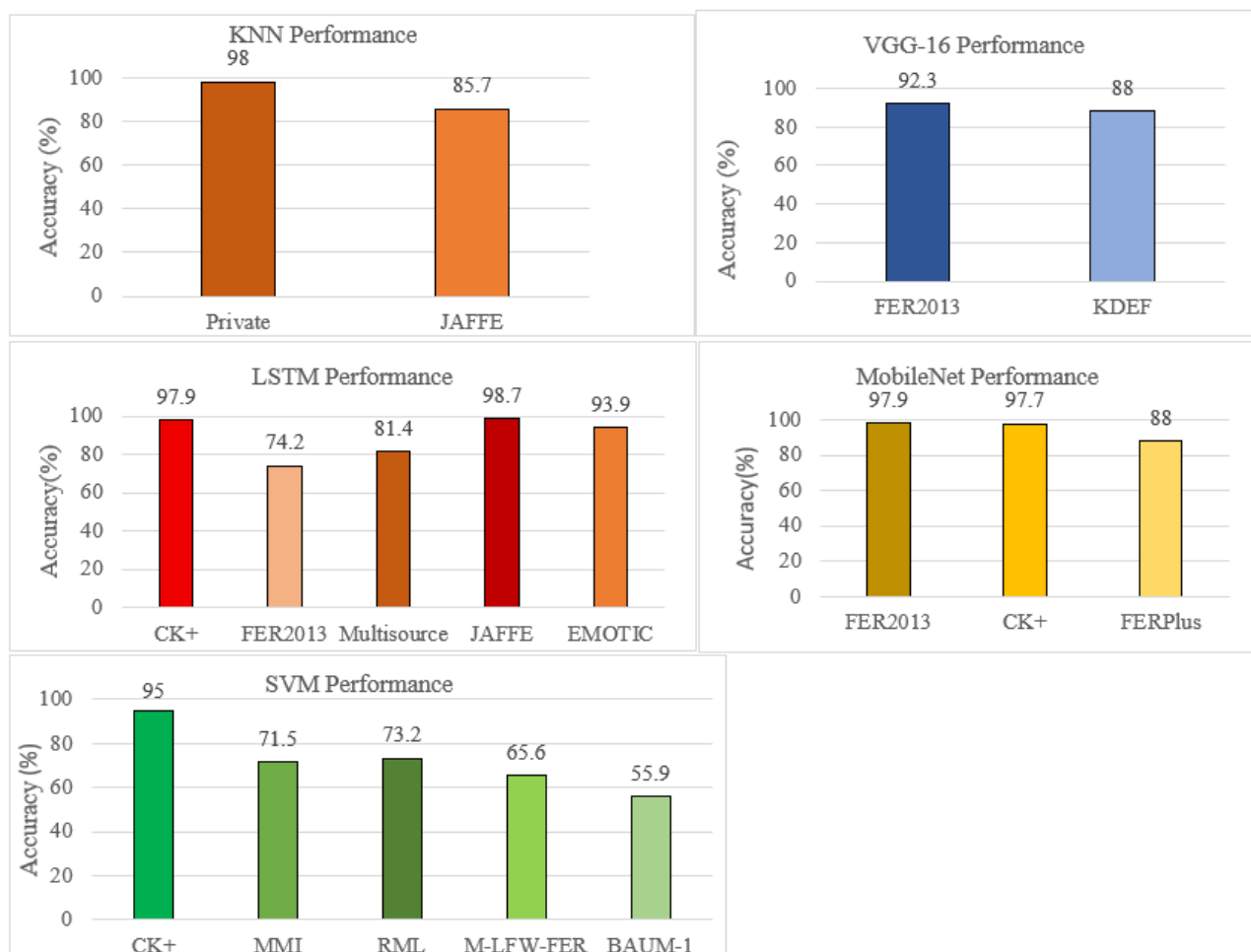


Figure 6. Employed approaches and their effectiveness using various FER datasets

Figure 6: Practical assessment has shown that deep learning- and hybrid-learning-based FER methods achieve very high accuracy. However, several challenges remain and require further research. As the complexity of these frameworks grows, data preparation requires large-scale datasets and significant computational resources. Running experiments and tests demands considerable memory, extended processing time, and a large collection of human-curated, labeled samples. The computational capability

required by the model for these approaches—especially those that rely on deep learning—is quite extensive. In addition, many of these approaches are designed to recognize only a fixed set of emotions, which limits their ability to detect other emotional states. Hence, designing a revised framework that is generally relevant and capable of encompassing the full spectrum of emotional states would be extremely valuable and could conceivably be extended to recognize more intricate and nuanced facial expressions.

6.1 Quantitative Performance and Computational Efficiency Analysis

A quantitative analysis of FER techniques indicates trade-offs between accuracy and computational cost. Conventional machine learning algorithms, such as Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN), have relatively low computational costs because they rely on handcrafted features. However, the performance of these algorithms deteriorates on large-scale, complex data sets.

On the other hand, the performance of deep learning algorithms such as VGG16, ResNet-50, DenseNet-161, and MobileNet in the FER system is relatively higher because these algorithms learn hierarchical features directly from raw facial images. The performance of these algorithms in the FER system has been observed to be above 95% for the CK+ and JAFFE data sets. However, these algorithms have relatively higher computational costs because they depend on the training process and the availability of GPUs and large-scale training datasets. Hybrid models are also used to achieve a balance between the trade-offs by using a combination of deep features and traditional classifiers such as SVM and GMM. Hybrid models have been known to improve classification accuracy by 3 to 6 percent compared to traditional models. However, it is important to note that the improvement comes at a price: the system becomes more complex as it undergoes a series of processes. Future facial emotion recognition (FER) systems should be designed to be lean and simple, achieving a balance between high accuracy and real-time performance.

7. Challenges and Constraints

There are still several limitations that affect the accuracy and real-world performance. Even though deep learning models like CNNs and ViTs can effectively recognize emotions from facial features, they are quite sensitive to environmental and contextual factors. Some of them are listed below:

1. **Partial Face Occlusion:** The accuracy of FER systems decreases significantly when hands, accessories, or other objects block parts of the face. In real-world scenarios, it is not always possible to capture the entire face clearly, which limits recognition performance.
2. **Head Pose Variations:** Variations in head posture—for instance, tilting, turning sideways, or looking away create strain on facial muscles and distort visible expressions, making recognition more difficult. Most FER systems perform optimally when the subject is facing forward in a stable position.
3. **Lighting Conditions:** Fluctuations in illumination can reduce the accuracy of facial detection and tracking. Both low light and excessive brightness can hide critical facial features or introduce noise, thereby affecting model performance.
4. **Cultural Variability:** Emotional expressions often differ across cultures, leading to diverse ways of perceiving and expressing emotions. This cultural variability makes it challenging for FER systems trained on a specific population to generalize well across global users.

Deployment Challenges: Although both manual and automated FER approaches can achieve good results, implementing them is not always straightforward. In dynamic or unstructured environments, they may produce inconsistent results depending on the dataset, features, or context.

8. Future Directions

Subsequent research in FER should have a primary focus on key points such as existing complications, reducing algorithmic bias, and enhancing system usability in real-world environments. A key priority is creating larger, more representative datasets that capture a broad spectrum of facial expressions across diverse regions, cultural contexts, and demographic groups. Additionally, incorporating multimodal techniques—combining facial features with speech tone, gestures, and body movements—can significantly improve recognition accuracy. For practical deployment, developing dynamic FER systems that are optimized for speed, storage efficiency, and operation in disorganized or dynamic settings will be essential. Tackling

population bias requires systematic strategies to ensure fair performance across diverse groups, resulting in more equitable, unbiased outcomes. The interpretability and trustworthiness of FER models can be enhanced by explainable AI (XAI) methods, which are especially valuable in low-resource and high-stakes applications. Further, advancing temporal modeling with architectures such as Transformers or enhanced LSTM networks will enable a better understanding of emotion sequences over time. Finally, emerging research highlights the potential of unsupervised and self-supervised learning techniques to decrease the necessity for extensive amounts of labeled training data, making FER systems more scalable and adaptable to new domains.

9. Conclusions

This review provided a comprehensive synthesis of Facial Emotion Recognition (FER) techniques based on machine learning, deep learning, and hybrid models, analyzing nearly a decade of research (2014–2024). Traditional approaches such as SVM, KNN, and Random Forests achieved competitive accuracy on small, controlled datasets such as JAFFE and CK+, typically ranging from 85–93%. Deep learning methods, particularly CNN-based architectures including VGG16, ResNet-50, and DenseNet-161, significantly enhanced performance, attaining up to 99.5% accuracy on large-scale datasets such as FER2013 and RAF-DB. Hybrid methods that integrate deep feature extraction with classical classifiers, such as SVMs or GMMs, demonstrated additional improvements of 3–6% over standalone models. While deep learning has redefined FER accuracy and robustness, challenges remain concerning data diversity, real-world generalization, and computational cost. Future work should emphasize the development of cross-cultural datasets, bias mitigation, and lightweight FER models suitable for real-time and edge deployment. Overall, this study consolidates key advancements, comparative performance, and open challenges to guide researchers and practitioners toward more inclusive, efficient, and interpretable FER systems.

Acknowledgement

I extend my profound thanks to my **team members** for their cooperation, commitment, and teamwork. Your collective effort and support have been invaluable in achieving our common goal.

Author Contributions: The authors confirm their contribution to the paper as follows:

Study conception and design: Sushilkumar Salve, Shubham Kumar, Irfan Shaikh.

Literature review and data collection: Pranay Gawade, Irfan Shaikh, Sushilkumar Salve

Analysis and interpretation of findings: Shubham Kumar, Pranay Gawade, Sushilkumar Salve.

Draft manuscript preparation and editing: Irfan Shaikh, Shubham Kumar, Sushilkumar Salve

Critical review of the manuscript and validation of content: Sushilkumar Salve, Sujit Ramesh Borey, Puja
All authors discussed the results, contributed to the writing, and approved the final version of the manuscript.

Funding Statement: “The author(s) received no specific funding for this study”.

Conflicts of Interest: “The authors declare that they have no conflicts of interest to report regarding the present study”.

References

- [1] Chu, W.-S.; De la Torre, F.; Cohn, J. F. Selective Transfer Machine for personalized facial expression analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 529–545. <https://doi.org/10.1109/TPAMI.2016.2547397>
- [2] AlBdairi, A. J. A.; Qasim, M. A.; Nsaif, A. K.; Al Bdairi, S. J. A.; Al-Obaidy, M. T.; Hassan, M. A.; Ali, A. R. Face recognition based on deep learning and FPGA for ethnicity identification. *Appl. Sci.* **2022**, *12*, 2605. <https://doi.org/10.3390/app12052605>
- [3] Simcock, G.; Reavley, N. J.; Vitis, L.; O’Sullivan, J. S.; Murray, A. L. Associations between facial emotion recognition and mental health in early adolescence. *Int. J. Environ. Res. Public Health* **2020**, *17*, 330. <https://doi.org/10.3390/ijerph17010330>

- [4] Yang, H.; Ciftci, U.; Yin, L. Facial expression recognition by DE-expression residue learning. In *2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*; IEEE: Salt Lake City, UT, USA. 2018. <https://doi.org/10.1109/CVPR.2018.00231>
- [5] Majumder, A.; Behera, L.; Subramanian, V. K. Automatic facial expression recognition system using deep network-based data fusion. *IEEE Trans. Cybern.* **2018**, *48*, 103–114. <https://doi.org/10.1109/TCYB.2016.2625419>
- [6] Deshmukh, S.; Patwardhan, M.; Mahajan, A. Survey on real-time facial expression recognition techniques. *IET Biom.* **2016**, *5*, 155–163. <https://doi.org/10.1049/iet-bmt.2014.0104>
- [7] Ekweariri, A. N.; Yurtkan, K. Facial expression recognition using enhanced local binary patterns. In *2017 9th Int. Conf. Comput. Intell. Commun. Netw. (CICN)*; IEEE: Girne, Northern Cyprus. 2017, pp. 43–47. <https://doi.org/10.1109/CICN.2017.8319353>
- [8] Yasmin, S.; Pathan, R. K.; Biswas, M.; Khandaker, M. U.; Faruque, M. R. I. Development of a robust multi-scale featured local binary pattern for improved facial expression recognition. *Sensors* **2020**, *20*, 5391. <https://doi.org/10.3390/s20185391>
- [9] Yu, Z.; Dong, Y.; Cheng, J.; Sun, M.; Su, F. Research on face recognition classification based on improved GoogLeNet. *Secur. Commun. Netw.* **2022**, 1–6. <https://doi.org/10.1155/2022/7192306>
- [10] Cheng, E. J.; Fan, S. C.; Chen, Y. C.; Chen, G. Y.; Su, C. H.; Wang, C. Y.; Chu, Y. H. Deep sparse representation classifier for facial recognition and detection system. *Pattern Recognit. Lett.* **2019**, *125*, 71–77. <https://doi.org/10.1016/j.patrec.2019.03.006>
- [11] Li, L.; Mu, X.; Li, S.; Peng, H. A review of face recognition technology. *IEEE Access* **2020**, *8*, 139110–139120. <https://doi.org/10.1109/ACCESS.2020.3011028>
- [12] Mliki, H.; Fourati, N.; Smaoui, S.; Hammami, M. Automatic facial expression recognition system. In *2013 ACS Int. Conf. Comput. Syst. Appl. (AICCSA)*; IEEE: Ifrane, Morocco. 2013. <https://doi.org/10.1109/AICCSA.2013.6616505>
- [13] Bendjillali, R. I.; Beladgham, M.; Merit, K.; Taleb-Ahmed, A. Improved facial expression recognition based on DWT feature for deep CNN. *Electronics* **2019**, *8*, 324. <https://doi.org/10.3390/electronics8030324>
- [14] Shan, K.; Guo, J.; You, W.; Lu, D.; Bie, R. Automatic facial expression recognition based on a deep convolutional-neural-network structure. In *2017 IEEE 15th Int. Conf. Softw. Eng. Res., Manag. Appl. (SERA)*; IEEE: London, UK. 2017. <https://doi.org/10.1109/SERA.2017.7965717>
- [15] Owusu, E.; Zhan, Y.; Mao, Q. R. A neural-AdaBoost based facial expression recognition system. *Expert Syst. Appl.* **2014**, *41*, 3383–3390. <https://doi.org/10.1016/j.eswa.2013.11.041>
- [16] Durmusoglu, A.; Kahraman, Y. Facial expression recognition using geometric features. In *2016 Int. Conf. Syst., Signals Image Process. (IWSSIP)*; IEEE: Bratislava, Slovakia. 2016. <https://doi.org/10.1109/IWSSIP.2016.7502700>
- [17] Yu, Z.; Dong, Y.; Cheng, J.; Sun, M.; Su, F. Research on face recognition classification based on improved GoogLeNet. *Secur. Commun. Netw.* **2022**, 1–6. <https://doi.org/10.1155/2022/7192306>
- [18] Kumar, S.; Sagar, V.; Punetha, D. A comparative study on facial expression recognition using local binary patterns, convolutional neural network and frequency neural network. *Multimed. Tools Appl.* **2023**, *82*, 24369–24385. <https://doi.org/10.1007/s11042-023-14753-y>
- [19] Baffour, P. A.; Nunoo-Mensah, H.; Keelson, E.; Kommey, B. A survey on deep learning algorithms in facial emotion detection and recognition. *Inform. Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi* **2022**, *7*, 24–32. <https://doi.org/10.25139/inform.v7i1.4282>
- [20] Ge, H.; Zhu, Z.; Dai, Y.; Wang, B.; Wu, X. Facial expression recognition based on deep learning. *Comput. Methods Programs Biomed.* **2022**, *215*, Art. no. 106621. <https://doi.org/10.1016/j.cmpb.2022.106621>
- [21] Singh, S. K.; Thakur, R. K.; Kumar, S.; Anand, R. Deep learning and machine learning based facial emotion detection using CNN. In *2022 9th Int. Conf. Comput. Sustain. Glob. Dev. (INDIACom)*; IEEE: New Delhi, India. 2022. <https://doi.org/10.23919/INDIACom54597.2022.9763165>
- [22] Khattak, A.; Asghar, M. Z.; Ali, M.; Batool, U. An efficient deep learning technique for facial emotion recognition. *Multimed. Tools Appl.* **2022**, *81*, 1649–1683. <https://doi.org/10.1007/s11042-021-11298-w>

- [23] Bisogni, C.; Castiglione, A.; Hossain, S.; Narducci, F.; Umer, S. Impact of deep learning approaches on facial expression recognition in healthcare industries. *IEEE Trans. Ind. Inform.* **2022**, *18*, 5619–5627. <https://doi.org/10.1109/TII.2022.3141400>
- [24] Alsharekh, M. F. Facial emotion recognition in verbal communication based on deep learning. *Sensors* **2022**, *22*, 6105. <https://doi.org/10.3390/s22166105>
- [25] Borgalli, M. R. A.; Surve, S. Deep learning for facial emotion recognition using custom CNN architecture. *J. Phys. Conf. Ser.* **2022**, *2236*, 012004.
- [26] Saad, A. A.; Shah, M.; Ehsan, M. R. K.; Amirzada, A.; Mahmood, T. Automated facial expression recognition framework using deep learning. *J. Healthc. Eng.* **2022**, *2022*, 1–11. <https://doi.org/10.1155/2022/5707930>
- [27] Arora, T. K.; Garg, M.; Kaushik, B.; Kaushik, R.; Gupta, V.; Gulati, M. Optimal facial feature based emotional recognition using deep learning algorithm. *Comput. Intell. Neurosci.* **2022**, Art. no. 8379202.
- [28] Dukic, D.; Sovic Krzic, A. Real-time facial expression recognition using deep learning with application in the active classroom environment. *Electronics* **2022**, *11*, 1240. <https://doi.org/10.3390/electronics11081240>
- [29] Zhao, X.; Zhu, J.; Luo, B.; Gao, Y. Survey on facial expression recognition: History, applications, and challenges. *IEEE Multimed.* **2021**, *28*, 38–44. <https://doi.org/10.1109/MMUL.2021.3107862>
- [30] Al-Modwahi, A. A. M.; Sebetela, O.; Batleng, L. N.; Parhizkar, B.; Lashkari, A. H. Facial expression recognition intelligent security system for real time surveillance. In *World Congr. Comput. Sci., Comput. Eng., Appl. Comput. (WORLDCOMP'12)*; CSREA Press: Las Vegas, NV, USA. 2012.
- [31] Varma, S.; Shinde, M.; Chavan, S. S. Analysis of PCA and LDA features for facial expression recognition using SVM and HMM classifiers. In *Techno-Societal 2018*; Springer: Cham, Switzerland. 2020, pp. 109–119. https://doi.org/10.1007/978-3-030-16848-3_11
- [32] Saeed, V. A. A framework for recognition of facial expression using HOG features. *Int. J. Math.* **2024**, *2*, 1–8. <https://doi.org/10.59543/ijmscs.v2i.7815>
- [33] Koonsanit, K.; Nishiuchi, N. Classification of user satisfaction using facial expression recognition and machine learning. In *2020 IEEE Region 10 Conf. (TENCON)*; IEEE: Osaka, Japan. 2020, pp. 561–566. <https://doi.org/10.1109/TENCON50793.2020.9293912>
- [34] Liu, X.; Cheng, X.; Lee, K. GA-SVM-based facial emotion recognition using facial geometric features. *IEEE Sens. J.* **2021**, *21*, 11532–11542. <https://doi.org/10.1109/JSEN.2020.3028075>
- [35] Salmam, F. Z.; Madani, A.; Kissi, M. Facial expression recognition using decision trees. In *2016 13th Int. Conf. Comput. Graph., Imag. Visualizat. (CGiV)*; IEEE: Beni Mellal, Morocco. 2016, pp. 125–130. <https://doi.org/10.1109/CGiV.2016.33>
- [36] Iromini, N.; Asi, B. T.; Okuwa, C. O.; Gbadamosi, K. Facial expression recognition system using KNN. *Int. J. Sci. Eng. Environ. Technol.* **2021**, *6*, 31–38.
- [37] Murugappan, M.; Alshehri, A. A. M.; Ahamed, S. R.; Ananth, C.; P. S.; Kumar, M. M. S. Facial expression classification using KNN and decision tree classifiers. In *4th Int. Conf. Comput., Commun. Signal Process. (ICCCSP)*; IEEE: Chennai, India. 2020, pp. 1–6. <https://doi.org/10.1109/ICCCSP49186.2020.9315234>
- [38] Dino, H. I.; Abdulrazzaq, M. B. Facial expression classification based on SVM, KNN and MLP classifiers. In *2019 Int. Conf. Adv. Sci. Eng. (ICOASE)*; IEEE: Zakho-Duhok, Iraq. 2019, pp. 70–75. <https://doi.org/10.1109/ICOASE.2019.8723728>
- [39] Priya, M.; Ranjani, B.; Rajammal, K.; Murali, N.; Tamilselvi, M.; Kavitha, M. Automatic emotion detection using SVM-based optimal kernel function. In *2024 11th Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*; IEEE: New Delhi, India. 2024, pp. 1607–1611. <https://doi.org/10.23919/INDIACom61295.2024.10498982>
- [40] Alhussan, A. A.; Ghoneim, S. S. M.; Kamal, E. K. M.; Abdelkhalek, O. O. O. Facial expression recognition model depending on optimized SVM. *Comput., Mater. Contin.* **2023**, *76*, 499–515. <https://doi.org/10.32604/cmc.2023.039368>
- [41] Tamer Ghareeb, B.; Tarek, F.; Said, H. FER_ML: Facial emotion recognition using machine learning. *J. Comput. Commun.* **2023**, *2*, 40–49. <https://doi.org/10.21608/jocc.2023.282094>

- [42] Yao, L.; Wan, Y.; Ni, H.; Xu, B. Action unit classification for facial expression recognition using active learning and SVM. *Multimed. Tools Appl.* **2021**, *80*, 24287–24301. <https://doi.org/10.1007/s11042-021-10836-w>
- [43] Alzawali, M. I. H.; Abed, S. A.; Ghayyib, K. K.; Risan, M. R. Facial emotion images recognition based on binarized genetic algorithm-random forest. *Baghdad Sci. J.* **2024**, *21*, Art. no. 0780. <https://doi.org/10.21123/bsj.2024.9698>
- [44] Akhand, M. A. H.; Roy, S.; Siddique, N.; Kamal, M. A. S.; Shimamura, T. Facial emotion recognition using Transfer Learning in the Deep CNN. *Electronics* **2021**, *10*, 1036. <https://doi.org/10.3390/electronics10091036>
- [45] Huang, Z.-Y.; Liang, J.-K.; Luo, C.-F.; Liu, M.-Y.; Lin, S.-H.; Chang, C.-Y.; Wang, D.-Y.; Hsu, S.-M.; Chen, R.-C. A study on computer vision for facial emotion recognition. *Sci. Rep.* **2023**, *13*, 8425. <https://doi.org/10.1038/s41598-023-35446-4>
- [46] Prajapat, G. K.; Sharma, B. P.; Agrawal, A.; Soni, R.; Saini, S. K. Universal model for facial expression detection using convolutional neural network. *AIP Conf. Proc.* **2024**, *3111*, 030022. <https://doi.org/10.1063/5.0221449>
- [47] Gupta, S.; Kumar, P.; Tekchandani, R. K. Facial emotion recognition based on real-time learner engagement detection system in online learning context using deep learning models. *Multimed. Tools Appl.* **2023**, *82*, 11365–11394. <https://doi.org/10.1007/s11042-022-13558-9>
- [48] Hu, L.; Ge, Q. Automatic facial expression recognition based on MobileNetV2 in Real-time. *J. Phys. Conf. Ser.* **2020**, *1549*, 022136. <https://doi.org/10.1088/1742-6596/1549/2/022136>
- [49] Nan, Y.; Ju, J.; Hua, Q.; Zhang, H.; Wang, B. A-MobileNet: An approach of facial expression recognition. *Alex. Eng. J.* **2022**, *61*, 4435–4444. <https://doi.org/10.1016/j.aej.2021.09.066>
- [50] Ngwe, J. L.; Lim, K. M.; Lee, C. P.; Ong, T. S.; Alqahtani, A. PAtt-Lite: Lightweight patch and attention MobileNet for challenging facial expression recognition. *IEEE Access* **2024**, *12*, 79327–79341. <https://doi.org/10.1109/ACCESS.2024.3407108>
- [51] Dong, C.; Wang, R.; Hang, Y. Facial expression recognition based on improved VGG convolutional neural network. *J. Phys. Conf. Ser.* **2021**, *2083*, 032030. <https://doi.org/10.1088/1742-6596/2083/3/032030>
- [52] Pham, L.; Huynh Vu, T.; Tran, T. A. Facial expression recognition using residual masking network. In *2020 25th Int. Conf. Pattern Recognit. (ICPR)*; IEEE: Milan, Italy. 2021. <https://doi.org/10.1109/ICPR48806.2021.9411919>
- [53] Li, B.; Lima, D. Facial expression recognition via ResNet-50. *Int. J. Cognit. Comput. Eng.* **2021**, *2*, 57–64. <https://doi.org/10.1016/j.ijcce.2021.02.002>
- [54] Minaee, S.; Minaei, M.; Abdolrashidi, A. Deep-emotion: Facial expression recognition using attentional convolutional network. *Sensors* **2021**, *21*, 3046. <https://doi.org/10.3390/s21093046>
- [55] Xiong, L.; Zhang, J.; Zheng, X.; Wang, Y. Context transformer and adaptive method with visual transformer for robust facial expression recognition. *Appl. Sci.* **2024**, *14*, 1535. <https://doi.org/10.3390/app14041535>
- [56] Tao, H.; Duan, Q. Hierarchical attention network with progressive feature fusion for facial expression recognition. *Neural Netw.* **2024**, *170*, 337–348. <https://doi.org/10.1016/j.neunet.2023.11.033>
- [57] Gao, Z.; Gao, H.; Xiang, Y. Facial emotion recognition based on selective kernel network. *J. Flow Vis. Image Process* **2024**, *31*, 33–52. <https://doi.org/10.1615/JFlowVisImageProc.2023048881>
- [58] Verma, M.; Vipparthi, S. K. Cross-centroid ripple pattern for facial expression recognition. *Multimed. Tools Appl.* **2024**, *48*, 384. <https://doi.org/10.1007/s11042-024-19364-9>
- [59] Rostami, M.; Farajollahi, A.; Parvin, H. Deep learning-based face detection and recognition on drones. *J. Ambient Intell. Humaniz. Comput.* **2024**, *15*, 373–387. <https://doi.org/10.1007/s12652-022-03897-8>
- [60] Mukhiddinov, M.; Djuraev, O.; Akhmedov, F.; Mukhamadiyev, A.; Cho, J. Masked face emotion recognition based on facial landmarks and deep learning approaches for visually impaired people. *Sensors* **2023**, *23*, 1080. <https://doi.org/10.3390/s23031080>
- [61] Pranav, E.; Kamal, S.; Chandran, C. S.; Supriya, M. H. Facial emotion recognition using deep convolutional neural network. In *2020 6th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*; IEEE: Coimbatore, India. 2020. <https://doi.org/10.1109/ICACCS48705.2020.9074302>

- [62] Jaiswal, A.; Krishnama Raju, A.; Deb, S. Facial emotion detection using deep learning. In *2020 Int. Conf. Emerg. Technol. (INCET)*; IEEE: Belgaum, India. 2020.
- [63] Hussain, S. A.; Al Balushi, A. S. A. A real time face emotion classification and recognition using deep learning model. *J. Phys. Conf. Ser.* **2020**, *1432*, 012087. <https://doi.org/10.1088/1742-6596/1432/1/012087>
- [64] Akeh, L. J.; Kusuma, G. P. Mixed emotion recognition through facial expression using transformer based model. *Stat., Optim. Inf. Comput.* **2024**. <https://doi.org/10.19139/soic-2310-5070-2103>
- [65] Zhang, S.; Pan, X.; Cui, Y.; Zhao, X.; Liu, L. Learning affective video features for facial expression recognition via hybrid deep learning. *IEEE Access* **2019**, *7*, 32297–32304. <https://doi.org/10.1109/ACCESS.2019.2901521>
- [66] Shahzad, H. M.; Bhatti, S. M.; Jaffar, A.; Akram, S.; Alhajlah, M.; Mahmood, A. Hybrid facial emotion recognition using CNN-based features. *Appl. Sci.* **2023**, *13*, 5572. <https://doi.org/10.3390/app13095572>
- [67] Alreshidi, A.; Ullah, M. Facial emotion recognition using hybrid features. *Informatics* **2020**, *7*, 6. <https://doi.org/10.3390/informatics7010006>
- [68] Liu, C.; Hirota, K.; Ma, J.; Jia, Z.; Dai, Y. Facial expression recognition using hybrid features of pixel and geometry. *IEEE Access* **2021**, *9*, 18876–18889. <https://doi.org/10.1109/ACCESS.2021.3054332>
- [69] Mishra, R. K.; Urolagin, S.; ArulJothi, J. A.; Gaur, P. Deep hybrid learning for facial expression binary classifications and predictions. *Image Vis. Comput.* **2022**, *128*, 104573. <https://doi.org/10.1016/j.imavis.2022.104573>
- [70] Sun, X.; Lv, M. Facial expression recognition based on a hybrid model combining deep and shallow features. *Cognit. Comput.* **2019**, *11*, 587–597. <https://doi.org/10.1007/s12559-019-09654-y>
- [71] Kommineni, J.; Mandala, S.; Sunar, M. S.; Chakravarthy, P. M. Accurate computing official expression recognition using a hybrid feature extraction technique. *J. Supercomput.* **2021**, *77*, 5019–5044. <https://doi.org/10.1007/s11227-020-03468-8>
- [72] Kim, J. C.; Kim, M.-H.; Suh, H.-E.; Naseem, M. T.; Lee, C.-S. Hybrid approach for facial expression recognition using convolutional neural networks and SVM. *Appl. Sci.* **2022**, *12*, 5493. <https://doi.org/10.3390/app12115493>
- [73] Verma, G.; Verma, H. Hybrid-deep learning model for emotion recognition using facial expressions. *Rev. Socionetwork Strat.* **2020**, *14*, 171–180. <https://doi.org/10.1007/s12626-020-00061-6>
- [74] Shahin, I.; Nassif, A. B.; Hamsa, S. Emotion recognition using hybrid Gaussian mixture model and deep neural network. *IEEE Access* **2019**, *7*, 26777–26787. <https://doi.org/10.1109/ACCESS.2019.2901352>
- [75] Cimtay, Y.; Ekmekcioglu, E.; Caglar-Ozhan, S. Cross-subject multimodal emotion recognition based on hybrid fusion. *IEEE Access* **2020**, *8*, 168865–168878. <https://doi.org/10.1109/ACCESS.2020.3023871>
- [76] Aifanti, N.; Papachristou, C.; Delopoulos, A. The MUG facial expression database. In *11th Int. Workshop Image Anal. Multimed. Interact. Serv. (WIAMIS 10)*; IEEE: Desenzano del Garda, Italy. 2010, pp. 1–4.
- [77] Li, S.; Deng, W.; Du, J. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*; IEEE: Honolulu, HI, USA. 2017, pp. 2852–2861. <https://doi.org/10.1109/CVPR.2017.277>
- [78] Pantic, M.; Valstar, M.; Rademaker, R.; Maat, L. Web-based database for facial expression analysis. In *2005 IEEE Int. Conf. Multimed. Expo.*; IEEE: Amsterdam, The Netherlands. 2005. <https://doi.org/10.1109/ICME.2005.1521424>
- [79] Sambare, M. FER-2013: Facial expression recognition dataset; Kaggle, 2013. <https://www.kaggle.com/datasets/msambare/fer2013>
- [80] Lundqvist, D.; Flykt, A.; Öhman, A. Karolinska directed emotional faces; APA PsycTests, 2015. <https://doi.org/10.1037/t27732-000>