



REVIEW

Comprehensive Review and Analysis on Facial Emotion Recognition: Performance Insights into Deep and Traditional Learning with Current Updates and Challenges

Amjad Rehman¹, Muhammad Mujahid¹, Alex Elyassih¹, Bayan ALGhofaily¹ and Saeed Ali Omer Bahaj^{2,*}

¹Artificial Intelligence & Data Analytics Lab, College of Computer & Information Sciences, Prince Sultan University, Riyadh, 11586, Saudi Arabia

²MIS Department College of Business Administration, Prince Sattam Bin Abdulaziz University, AlKharj, 11942, Saudi Arabia

*Corresponding Author: Saeed Ali Omer Bahaj. Email: saobahaj@gmail.com

Received: 03 September 2024 Accepted: 26 November 2024 Published: 03 January 2025

ABSTRACT

In computer vision and artificial intelligence, automatic facial expression-based emotion identification of humans has become a popular research and industry problem. Recent demonstrations and applications in several fields, including computer games, smart homes, expression analysis, gesture recognition, surveillance films, depression therapy, patient monitoring, anxiety, and others, have brought attention to its significant academic and commercial importance. This study emphasizes research that has only employed facial images for face expression recognition (FER), because facial expressions are a basic way that people communicate meaning to each other. The immense achievement of deep learning has resulted in a growing use of its much architecture to enhance efficiency. This review is on machine learning, deep learning, and hybrid methods' use of preprocessing, augmentation techniques, and feature extraction for temporal properties of successive frames of data. The following section gives a brief summary of assessment criteria that are accessible to the public and then compares them with benchmark results the most trustworthy way to assess FER-related research topics statistically. In this review, a brief synopsis of the subject matter may be beneficial for novices in the field of FER as well as seasoned scholars seeking fruitful avenues for further investigation. The information conveys fundamental knowledge and provides a comprehensive understanding of the most recent state-of-the-art research.

KEYWORDS

Face emotion recognition; deep learning; hybrid learning; CK+; facial images; machine learning; technological development

1 Introduction

Facial expressions play a crucial role in interpersonal communication as they enable us to comprehend the intended messages of others. Individuals often rely on the intonation of their speech and the facial expressions of others to deduce their emotional states, such as happy, sadness, or anger. Several studies have shown that nonverbal communication makes up around 67% of human communication



[1,2]. Verbal communication constitutes just 33% of human communication. Facial expressions play a crucial role in interpersonal communication since they are one of the many nonverbal cues that convey emotional significance. This is because they are one of the several signs that convey significance. As a result, it is unsurprising that the study of facial expressions has gained popularity in recent decades. Facial expressions have a wide range of possible uses in areas such as emotional computing, computer animation, cognitive science, and perceptual research. The identification of emotions via facial expressions is a recently growing and captivating subject of research in the fields of psychiatry, psychology, and mental health [3].

Automated emotion recognition from facial expressions is a critical element in various fields, such as smart home technology, healthcare systems, diagnosing emotion disorders in autism spectrum disorder and schizophrenia, human-computer interaction, human-relational intelligence, and social services programs. The research community is now focusing on this field because of the abundance of potential applications for facial emotion recognition [4]. The primary objective of facial emotion recognition is to create a link between different facial expressions and the corresponding emotional states they transmit. The conventional facial emotion recognition approach involves two core procedures: emotion detection and feature extraction. Preprocessing images is critical and involves several tasks such as cropping, scaling, normalization, and face recognition [5]. By removing the background and other non-facial elements, face detection decreases the accessible area inside the picture for the face. The extraction of features from the processed image is the essential element of a standard facial expression recognition system. The suggested strategy utilizes many techniques, such as linear discriminant analysis and discrete wavelet transform (DWT) [6].

The next phase will categorize the obtained data using machine learning techniques, such as neural networks (NN), to analyze emotion facial photos, pre-processing is required for both face recognition and rotation correction [7]. AdaBoost [8] have demonstrated the reliability of cascade classifiers, such as the one they use, in performing the previously mentioned task. It is customary to include artistic and geometric elements. The first phase involves extracting the locations of different facial features. Afterwards, authors interconnect these positions to construct a feature vector that encompasses the geometric details of the face, such as the angle, coordinates, and location [9]. To accurately replicate the many appearances of a face, the use of external features rely on extensive geographical study [10]. An essential part of expression recognition is leveraging the characteristics of the motion data. The last stage involves using the discovered attributes to build a robust classifier capable of recognizing a diverse range of facial expressions [11].

The first step in the Facial Expression Recognition technique is face detection, which entails recognizing a face or its characteristics in a video or single picture. The images showcase elaborate surroundings instead of just portraying people [12]. While humans possess the ability to reliably discern facial expressions and other traits from an image, robots without extensive training experience difficulties in doing the same [13,14]. Face detection primarily aims to differentiate photographs of faces from irrelevant background elements. The face detection domains include gesture recognition, video surveillance, automated cameras, gender identification, facial feature recognition, face recognition, tagging, and teleconferencing [15,16]. The primary prerequisite for these systems is the ability to recognize features as inputs. It captures color pictures in any area where a color sensor is available for image acquisition. Consequently, the bulk of face recognition algorithms now in use rely mostly on gray scale images, with just a limited number capable of handling color shots. To improve their performance, these systems use either window-based or pixel-based approaches, which are the two main types of face recognition methods. The pixel-based technique is slow in accurately distinguishing a person's face from their palms or other areas of skin [17,18]. In contrast, the window-based technique lacks

the ability to analyze characteristics from many viewpoints. In facial expression recognition model matching techniques are the most commonly used methods for detecting faces.

The research examines novel deep learning architectures and techniques to identify and detect facial expressions [19]. The paper also discusses the advantages of CNN systems in comparison to other prominent architectures such as RNNs and SVMs. It also addresses their shortcomings, contributions, and model efficacy. A substantial sample with accurate expression IDs may significantly enhance the rate of expression identification throughout the training phase [20]. Facial expressions are a crucial aspect of nonverbal communication since they convey an individual's internal emotions. The authors of the research [21] proposed a successful deep learning approach for determining age and gender based on facial movements and expressions in images of faces [22]. Intelligent cameras equipped with integrated facial recognition systems can consistently capture images of a patient's face in a smart home under standard lighting conditions and transmit them to a cloud server. In this manner, physicians and nurses may consistently transmit notifications [23]. The inherent capabilities of FER technology, compatible with resource-limited devices such as the Jetson Nano, might significantly assist law enforcement in identifying suspects via facial expression analysis [24]. Previous research on FER has shown that deep learning techniques, particularly those using CNNs, are the most sophisticated. Research outlines the use of a bespoke CNN architecture for the recognition of fundamental face emotions in static images [25]. To enhance the efficacy of automated facial emotion recognition (FER), a methodology including two hidden layers and four convolutional layers should be used [26]. The preprocessing technique eliminates noise from the supplied image. During the pretraining phase, which includes feature extraction [27], face recognition may become evident. The research aimed to investigate the relationships among gender, occupations, activities, and emotional expression inside a robotics workshop, akin to a dynamic classroom [28].

Applications:

Several sectors have seen a surge in demand for facial recognition technology (FER), which has greatly improved human-computer interaction. In addition, the user brings up marketing strategies, security protocols, and health care systems. Facial Emotion Recognition, or FER, improves HCI interactions by making them more accurate and transparent [29]. It lets computers correctly detect and understand human emotions and behaviors. For example, educational apps may now adjust the difficulty level based on students' facial expressions, making learning more engaging and participatory. FER provides better and more engaging experiences by creating systems that adapt in real-time to users' emotions [30].

Facial Expression Recognition, or FER, is a powerful resource for mental health problem detection, sickness treatment, and monitoring in the healthcare industry. Contributing significantly to the field of mental healthcare, the system is capable of independently detecting emotional signals that might indicate the existence of melancholy, worry, or stress. Subsequent treatment sessions benefit greatly from these non-verbal cues. For example, people with mental illness or other emotional disorders may benefit from using facial expression recognition (FER) to monitor their emotional state and spot signs of anxiety. Importantly, many individuals need care for an extended period of time.

In surveillance systems, the field of security uses facial expression recognition (FER) to detect when a person displays certain micro expressions or facial expressions that might be a sign of potential danger. This technology's goal is to make you healthier by giving you the power to protect yourself before things like airplane detection operations or accidents happen [31]. For example, FER may be used to assess the overarching emotional states communicated during an event, which can then be discussed. It enhances the comprehension of spoken language. To instantly boost their effectiveness,

marketing firms use facial emotion recognition (FER) technology to assess customer emotional reactions to products and services. Improving customer service, customizing marketing techniques, and handling questions arising from emotional statements are all areas where this shines.

Review Motivations:

Facial expression recognition has emerged as a crucial component of human-computer interaction, exerting a substantial influence in domains like mental health surveillance, security, statistical analysis of behavior, and tailored user experience. Accurate interpretation of human emotions as conveyed via facial expressions is crucial for the development of intelligent systems capable of adapting and responding to users' emotional states.

Conventional machine learning techniques are extensively used for the identification of face emotions, drawing on artifacts and fuzzy learning algorithms. A comprehensive understanding of these methods is frequently necessary to effectively manage the intricate and varied range of human emotions. However, deep learning techniques, particularly convolutional neural networks (CNNs), have shown strong performance at acquiring supplementary processing characteristics straight from unprocessed data. The capability to acquire high-resolution photographs of face profiles has resulted in substantial improvements in precision and dependability.

Nevertheless, deep learning models have several constraints, like the need for extensive data labelling and substantial processing resources. Hybrid methodologies that integrate machine learning and deep learning capabilities have arisen as viable avenues to address these obstacles. Through the integration of artefacts with deep learning layers or learning models, hybrid models have the potential to enhance generalization and scalability, particularly in tasks that involve limited data or intricate emotion identification. Due to the complex nature of facial expression identification, it is necessary to conduct a thorough evaluation of various techniques in order to determine the most efficient ones, analyze their constraints, and direct future research. This paper presents a meticulous evaluation of machine learning, free-learning, and hybrid approaches, emphasizing their advantages and disadvantages, and delineates their implementation and efficacy in the domain of face emotion identification.

Review Contributions:

In recent years, there has been a rapid and substantial growth in the field of face emotion recognition research, specifically in the domains of datasets, detection, and recognition. Most scholarly review studies have concentrated on specific methodologies that used deep learning or machine learning techniques. Previous studies did not adequately explore the recognition of facial emotions [30,31]. Although there are many new insights, technological advancements such as recurrent models, vision transformers, and self-supervised learning provide new information that strengthens the focus of our work on FER networks. So, there is a need to conduct a more thorough review of new approaches.

The aim of this work is to analyse and assess the existing research on face emotion recognition analysis, with a particular emphasis on machine learning and hybrid methods. Datasets, preprocessing, augmentation, feature extraction, and approaches used in emotion categorization are the primary areas of concentration. This review paper has some important contributions:

- This paper provides a comprehensive analysis of FER frameworks, including preprocessing, feature extraction, and augmentation approaches. Furthermore, this paper compares the sophisticated techniques used for FER and their comparative analyses. Also, this paper contains an elaborate overview of current FER datasets.

- We have compared and validated state-of-the-art deep networks in relation to FER classification methods. We have examined deep learning, hybrid, and conventional machine learning models independently; we have provided comprehensive evaluations of all three. Their inclusion serves to provide researchers and medical experts with aspirations with little experience a comprehensive view of the subject's corpus of study.

More importantly, [Sections 2–4](#) provide the required context on the use of machine, deep, and hybrid models in the identification of facial expressions. [Section 5](#) presents an in-depth examination of the FER datasets. The results of the continuing inquiry are explored in [Section 6](#), while a thorough evaluation is provided in [Section 7](#), which closes the summary.

2 Machine Learning Based Approaches

Several research scholars utilized mostly Support Vector Machine (SVM), K Nearest Neighbor (KNN) and Random Forest (RF) approaches for the face emotion recognitions. The study used facial expression recognition and machine learning to categorize customer satisfaction with products and services. The recommended framework has three key steps. The initial step is collecting user satisfaction, gender, age, and facial expression data. Machine learning algorithms classify data. The last step is model assessment, which verifies model accuracy. The classification model may categorize user satisfaction by gender, age, and facial expression [32]. The two primary aspects of this endeavor are as follows: The initial stage involves the implementation of a revolutionary geometrically based extraction method. This approach calculates six distances to determine the facial characteristics that most effectively convey an emotion. The second stage involves the application of a decision tree, an automated supervised learning technique, to the dataset JAFFE [33]. The goal is to create a face expression categorization system that can take in six distances as input, which can be either the standard Euclidean, Manhattan distance. The system is capable of distinguishing between seven discrete emotions, including neutral. The JAFFE database achieved a recognition rate of 89.20%, while the COHEN database had a recognition rate of 90.61% [34]. For face analysis, researchers used PCA to recognize faces with few attributes and KNN to extract features. Researchers employed the Japanese Female Facial Expression (JAFFE) dataset, which included 10 female expressers and seven facial emotions: pleasure, sorrow, surprise, wrath, disgust, fear, and neutral. 210 images: 140 warm-ups, 70 final tests. Image preprocessing includes scaling, normalization, and gray-scale conversion. PCA identified image features. The KNN algorithm can recognize neutral, happy, sad, shocked, angry, disgusted, afraid, or disturbed facial emotions in pre-processed photos. They tested the approach using MATLAB script codes and three metrics: inaccurate, right, and erroneous categorization. This research included 60 correct, 10 incorrect identifications, and 85.71% accuracy [35].

The study used fivefold cross-validation and a one-way analysis of variance with a significance level of $p < 0.01$ to ensure that the features obtained are accurate. Lastly, system employs these cross-validated features to train KNN and DT classifiers to map six different facial emotions. To measure how well the classifier does at identifying different emotions, KNN uses four distance measures. On average, KNN has a 98.03% accuracy rate in emotion classification, whereas DT achieves a 97.21% rate. Due to an optical flow algorithm's ability to correctly detect facial emotions, a wide variety of real-time systems are now within reach [36]. According to a detailed analysis of the experiment data, the provided technique performs best when trained using SVM as the classifier; the recognition rate decreases to 79.99% when trained with KNN and increases to 82.97% when trained with MLP [37]. The Eigenvectors that result from extracting the effective Haar face features using the principal component analysis (PCA) approach are then used to reduce dimensionality. One can detect fundamental human

emotions like joy, sadness, and fear by using a kernel function and SVM classifier [38]. The authors combined SVM, Naive Bayes, Random Forest, and KNN algorithms. SVM (36%), Logistic Regression (64.2%), and Naive Bayes (38.1%) exhibit varied degrees of accuracy. Random Forest was 51.7% accurate [39]. Data preparation, grid search application for optimization, and classification were the three main steps of the FER. Linear discriminant analysis (LDA) classifies the data into eight categories: neutral, joyful, sad, surprised, fear, disgust, angry, and contempt. Applying LDA had clearly enhanced SVM's classification performance. The proposed improved SVM approach achieves 99% accuracy and 98% F1-score. The authors of [40] devised a method for determining the sentiments of individuals by utilising two novel geometry features: vectorised landmark and landmark curve. These characteristics are determined by examining a variety of facial regions that are associated with distinct components of facial muscle activity. The optimal features and parameters are selected across a variety of characteristics with SVM. Active learning and SVM methods were employed to identify human facial emotions by the authors of [41], who also categorised facial action units into classifications. Fig. 1 illustrates the SVM structure.

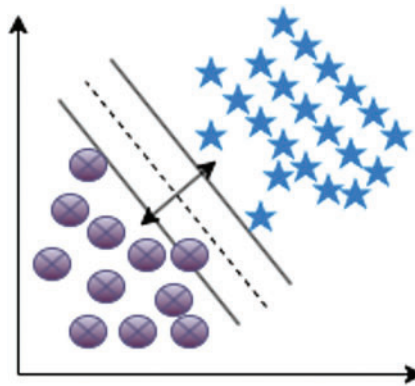


Figure 1: SVM

Normal, happy, sad, sleepy, wink, left light, surprised, no glasses, right light, glasses, and centre light are eleven human facial emotions included in the photographs of the faces taken from the Yale Face dataset. Precision, sensitivity (recall), accuracy, F-measure (F1-score), and G-mean are the performance evaluation criteria for the suggested system. The proposed BGA-RF method achieves an accuracy of up to 96.03%. In addition, the proposed BGA-RF has shown superior accuracy when compared to its peers. The experiment results demonstrate that the suggested BGA-RF method is effective in recognizing human facial expressions from pictures [42]. Most of these systems work well with controlled-environment picture datasets. These datasets become less successful with harder datasets because of increased picture variance and incomplete faces. This research method is based on the Histogram of Orientated Gradient (HOG) descriptor. The technique begins by preprocessing the input picture to establish the datum region to extract the most important information. The next step was training a face emotion classifier using RF. Researchers used JAFFE, the Japanese Female Facial Emotions Database, to analyze the methods. The experiment demonstrated the effectiveness and accuracy of the suggested method for face emotion recognition [43]. There has already been a performance demonstration and review. Both were based on the uncertainty matrix and the classifications correctness. SVM, PCA, and LDA features work well together to accurately identify and classify facial expressions [44].

Overview abstract of machine learning approaches for the FER system is shown in Fig. 2. This involves data collection, preprocessing, feature extraction using several methodologies, application of machine learning classifiers, followed by model fine-tuning and training. Once the models are completely trained, they are evaluated for emotion recognition. In addition, Table 1 presents the critical analysis of machine learning for FER system.

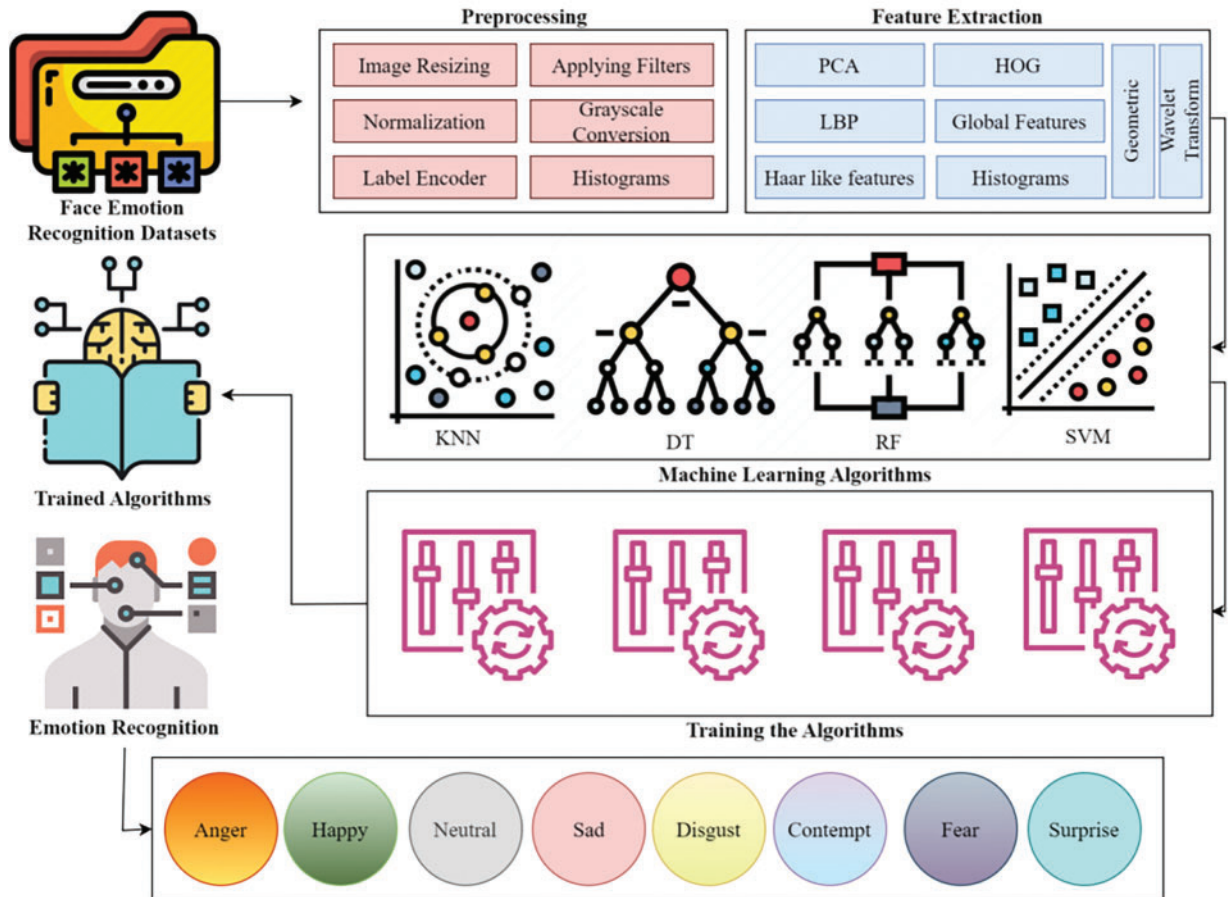


Figure 2: Overview abstract of machine learning approaches for the FER system

Table 1: Critical analysis of machine learning for FER system

Ref.	Datasets	Methods	Features	Emotions
[32]	Private	SVM	No	7
[33]	CK+, JAFFE	CART	Geometric and global features	7
[34]	JAFFE	KNN	PCA	7
[35]	Private	KNN	Haar-like features	6
[36]	CK+	SVM	LBP, Gabor filter, PCA	7
[37]	CK+	SVM	PCA	7

(Continued)

Table 1 (continued)

Ref.	Datasets	Methods	Features	Emotions
[38]	FER 13	LR	No	5
[29]	AffectNet	SVM	Grid search	8
[40]	CK+, MUG	SVM	Geometric	8
[41]	CK+	SVM	PCA and LBP	7
[42]	Yale faces	RF	Histograms of oriented gradients	11
[43]	JAFFE	RF	Histograms of oriented gradients	7
[44]	FECR	SVM	LDA, PCA	6
[45]	CK+	KNN	Histograms of oriented gradients	7
[46]	MMI, BBC, SPOS	SVM	LBP, Histograms of oriented gradients	2
[47]	CK+	SVM	Haar wavelet transform and PCA	7

3 Deep Learning Based Approaches

Computer vision and artificial intelligence researchers have made significant progress in the area of facial expression recognition (FER) over the last two decades. Human-computer interaction cannot occur without facial expression recognition (FER). Regarding expression recognition, it is an established fact that the backdrop or non-facial elements of an image are reliable. Efficient and instantaneous networks are becoming increasingly important in the mobile network era. However, many expression recognition networks fail to satisfy real-time demands because of their excessive parameter needs and computational complexity. Their proposed strategy addresses this challenge by integrating supervised transfer learning with an island loss and a joint supervision technique. This approach is crucial for face-related tasks. MobileNetv2, a recently created CNN model known for its speed and accuracy, generates rapid and precise real-time output when combined with a real-time framework. This approach outperforms other advanced techniques when evaluated on the CK+, JAFFE, and FER2013 facial expression datasets [48]. The authors propose a model called A-MobileNet, which is lightweight. Initially, authors will integrate the attention module into the MobileNetV1 model in order to enhance the process of extracting local features from face expressions. Authors combine the center loss and softmax loss to optimize the model parameters. This increases the distance between different classes while reducing the distance within each individual class. Their strategy significantly enhances recognition accuracy while maintaining the same number of model parameters as the original MobileNet series models. During the evaluation on the RAF-DB and FERPlus datasets, the A-MobileNet model demonstrated superior performance [49]. This strategy works well in certain countries, but worldwide adoption is difficult. Teamwork is the recommended remedy. This study aims to demonstrate that a comprehensive model for very accurate facial emotion identification is attainable. Mixing datasets from several nations into one coherent collection with uniform distribution is one technique. This study uses JAFFE and CK databases to achieve its purpose. They produce a convolutional neural network model after optimizing model complexity, training time, and processing resources [50]. This comparison is based on a subjective assessment approach that prioritizes the end user's happiness and approval. Both the laboratory and real-world findings demonstrate the method's excellent knowledge of emotions. Evidently, deep learning has enormous promise for facial emotion identification [51]. Mostly employed MobileNet Model for FER System, Used by Various Researchers is shown in Fig. 3.

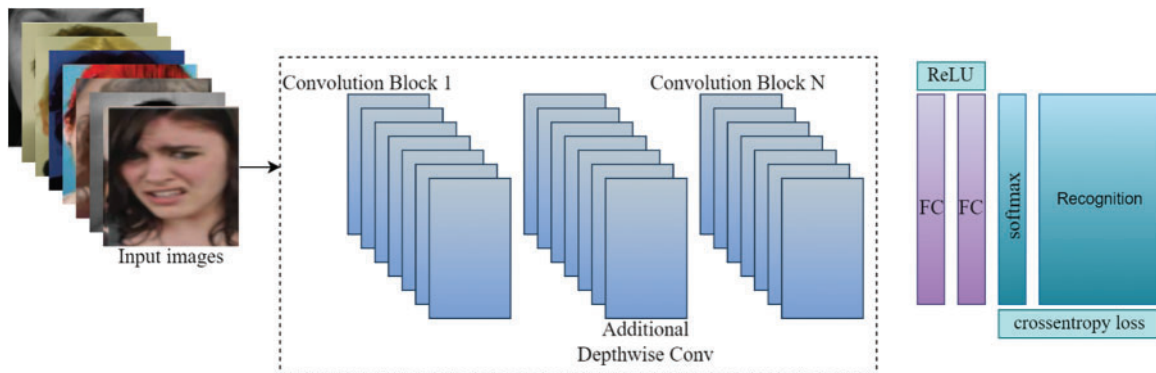


Figure 3: Mostly employed MobileNet model for FER system, used by various researchers

The authors examine the VGG16 network and propose improvements. To make things easier to grasp, replace the three fully connected layers in the network with two convolutional layers and one fully connected layer. Prior to training the network using the RAF-DB and SFEW facial expression datasets, switch the network's maximum pooling to local-based adaptive pooling. To recognize facial expressions, the network may then choose the most suitable characteristic information [52]. Accuracy and recognition rates increased by 7%. This study proposes a deep learning-based method that analyzes facial expressions to measure online learners' engagement in real time. Using facial expression detection, researchers can predict a user's "engaged" or "disengaged" status and compute their engagement index. Choose the best predictive classification model for real-time engagement detection by comparing deep learning models like Inception-V3, VGG19, and ResNet-50. To evaluate the proposed technique, they use benchmark datasets such as CK+, RAF-DB, and FER-2013. The suggested system performs 92.32% of the time for ResNet-50, 90.14% of the time for VGG19, and 89.11% of the time for Inception-V3. ResNet-50 surpasses all other models in real-time facial emotion classification with 92.3% accuracy [53]. Face recognition software verifies and distinguishes facial characteristics. However, in real time, haar cascade detection evaluates face traits. The sequential technique involves the detection of a human face from the camera, followed by an analysis of the collected input using Keras' convolutional neural network model, which analyzes data based on features and a database. The first step involves evaluating a human face to define emotions including happiness, neutrality, anger, sorrow, contempt, and surprise. The suggested study seeks identification, emotion categorization, and facial detection. Python, OpenCV, and a dataset enable this study's computer vision methods. To demonstrate the procedure's real-time applicability, a group of students evaluated their subjective feelings and observed physiological reactions to diverse facial expressions [54]. Authors offer a method for extracting features from deep residual networks (ResNet-50), which employ convolutional neural networks for facial expression detection. Compared to the most popular models for identifying facial emotions, this model performs better in the experimental simulations using the given data set [55].

The authors create a residual masking network by combining the widely used deep residual network with a design similar to Unet. The suggested technique preserves state-of-the-art accuracy on the well-known FER2013 and private VEMO datasets [56]. The proposed methodology was evaluated using a set of twenty-five cutting-edge techniques and five standard datasets: Japanese Female Facial Expressions, Extended Cohn-Kanade, and Karolinska. The publication, titled "Directed Emotional Faces, Real-world Affective Faces, and Facial Expression Recognition 2013," is about the study of

facial expressions and their recognition in real-world scenarios. The databases include seven primary emotions: disgust, shock, anger, sadness, fear, and neutrality. Researchers used four assessment criteria, namely accuracy, precision, recall, and F1-score, to compare the proposed strategy with alternative techniques. The findings indicate that the suggested approach performs better than all other advanced methods on all datasets [57]. The FER model identifies five distinct facial expressions using an expressional vector (EV). A database containing 10,000 photos, representing 154 persons, provided the data for the supervisory set. Researchers attained a 96% level of accuracy in discerning the mood, with a 24 value EV. At each iteration, the last layer of the two-level CNN modifies the exponent and weight values [58]. The common CNN architecture utilized for the classification and recognition of facial emotions is depicted in Fig. 4.

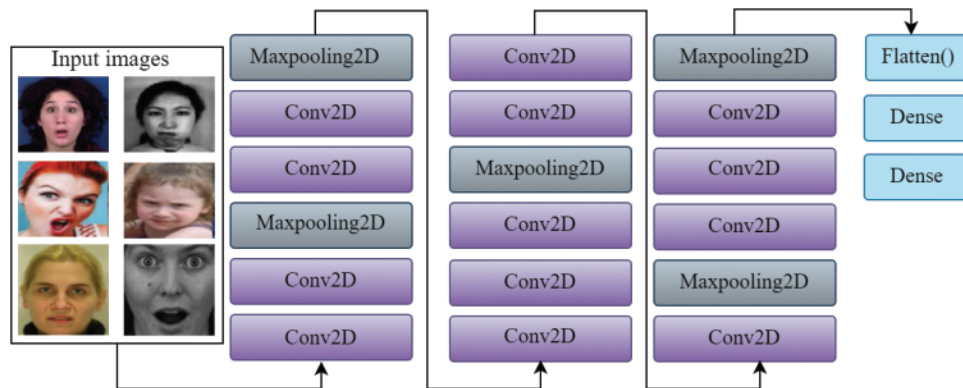


Figure 4: The common CNN architecture utilized for the classification and recognition of facial emotions

The methodology produced very precise outcomes on both datasets by using pre-trained models. DenseNet-161 earned the highest FER accuracy on the KDEF and JAFFE test sets, with values of 96.51% and 99.52%, respectively, using a 10-fold cross-validation method. The results demonstrate that the proposed FER technique outperforms the existing methods in identifying emotions. The profile views success on the KDEF dataset as promising due to its clear demonstration of the necessary competencies for real-world applications [59]. The authors provide a deep learning strategy based on attentional convolutional networks that significantly outperform previous models. This technique can accurately identify and focus on significant face characteristics within many datasets, including FER-2013, CK+, FERG, and JAFFE [60]. The primary objective of this project is to construct a Deep Convolutional Neural Network (DCNN) model capable of accurately identifying five distinct emotional emotions shown on the face. In order to train, test, and verify the model, they used a manual compilation of photographs [61]. Another study demonstrated the use of convolutional neural networks (CNNs) to train a deep learning framework for the purpose of identifying emotions in photographs. They rigorously tested the recommended approach by applying it to the Facial Emotion Recognition Challenge (FERC-2013) and the Japanese Female Facial Emotion datasets (JAFFE). The proposed model achieved an accuracy rate of 70% on the FERC-2013 dataset and 98.65% on the JAFFE dataset [62].

Deep learning methodologies are garnering increasing interest from researchers. Deep learning methodologies are becoming significant in the realm of categorization [63]. A facial emotion identification method for masked facial images employs convolutional neural network feature analysis

of key facial features and low-light image augmentation [64]. Facial emotions and expressions Real-time recognition has several applications, including security, emotion analysis, healthcare, and safety assessments [65]. A study developed an emotion detector system capable of identifying individuals with autism in real time [66]. The study [67] presents a novel approach for constructing a real-time engagement detection system based on deep learning models. Evaluating the learner's facial expressions during online education allows an automatic measurement of their level of engagement. The FER model employs an expressional vector to classify typical facial expressions into five distinct categories. A convolutional neural network (CNN) was used to identify significant face landmarks for the purpose of exhibiting the emotional attributes extracted from the photos [68]. The research article [69] proposes a technique for emotion detection via feature extraction and convolutional neural networks.

The need for enhanced model recognition rates and stability is increasing as an increasing number of people pursue deep learning for innovative biomedical applications [70]. The domain of facial expression analysis has undergone much investigation. Facial expressions are complex and dynamic, complicating the mastery of their identification despite much work in this domain [71]. Authors can improve the accuracy of FER's classification in challenging scenarios by utilizing a MobileNetV1-based method [72]. The objective of the project is to develop a system for facial expression recognition using convolutional neural networks and other data [73]. This method can categorize seven basic emotions using visual data. The study [74] demonstrates the efficacy of a hierarchical attention network that integrates progressive feature fusion for practical facial emotion identification. The varied feature extraction module integrates several complementing properties. The qualities include both low and high levels, features that are gradients and hence stay invariant under varying lighting circumstances and features that include context on both local and global scales. Proposed an alternate approach for drone-based face detection and identification that would enhance the accuracy of facial recognition in low-light or airborne conditions [75]. The integration of CNN and ViT architectures with a CoT module enables precise detection of facial expressions in intricate situations. By instructing on intricate relationships among attributes in specific areas, this facilitates the recognition of small variations [76]. This study presents an innovative method for face emotion identification via a selective kernel network [77]. The Cross-Centroid Ripple Pattern (CRIP) is an innovative feature descriptor for computer-generated face expression identification [78].

Overview abstract of deep learning approaches for the FER system, utilizing preprocessing, augmentation, CNN models with transfer learning weights, training and tuning, and recognition for more than 5 emotions is presented in Fig. 5.

Table 2 shows the critical analysis of deep learning for FER system. The existing models for facial expression recognition are inadequate for the complex situations encountered in daily life. These concerns include face occlusion, fluctuations in illumination, image noise, and other related problems. The answer to these challenges is the incorporation of the CoT function into the CNN and ViT systems. It enhances the model's sensitivity to diverse lighting conditions, augments its capacity to discern nuanced correlations among local attributes, and elevates the probability that those features will align with the global representation. The authors refine models derived from the ConvNet architecture, such as ConvNeXt, capable of recognizing and categorizing faces exhibiting a range of emotions with the IMED Dataset. These iterations include Swin Transformer v2, Vision Transformer, and other variants. Each model undergoes testing with several configurations using improved hyperparameters [79].

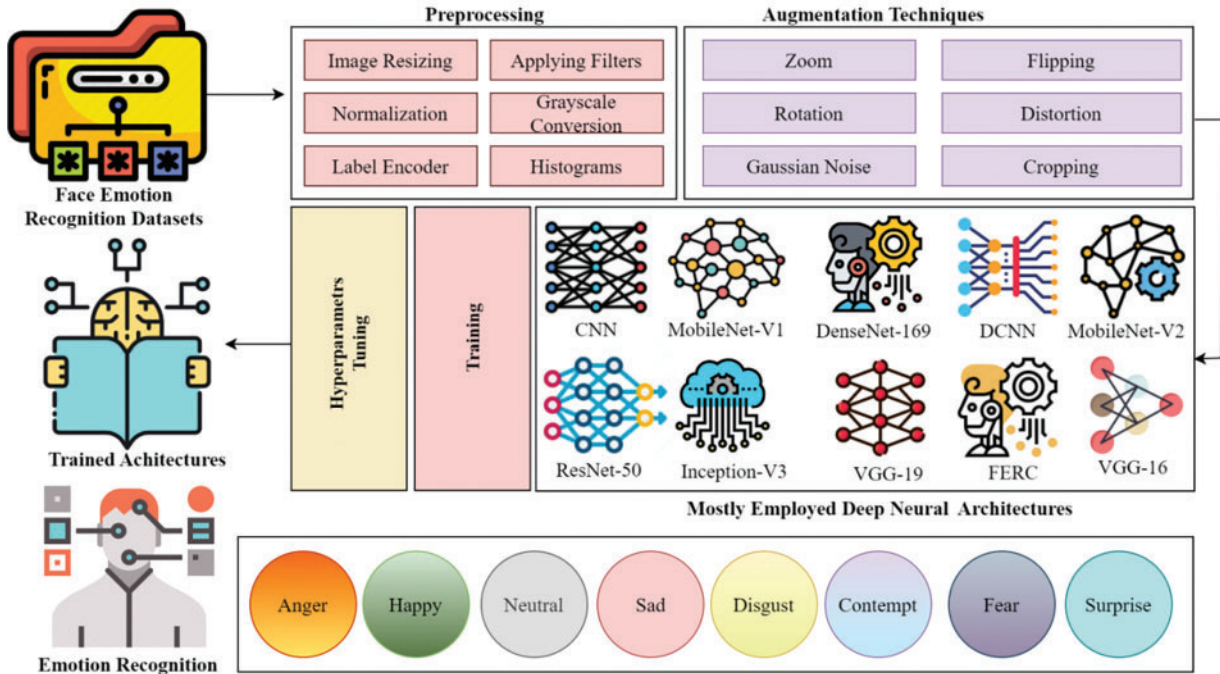


Figure 5: Overview abstract of deep learning approaches for the FER system, utilizing preprocessing, augmentation, CNN models with transfer learning weights, training and tuning, and recognition for more than 5 emotions

Table 2: Critical analysis of deep learning for FER system

Ref.	Datasets	Methods	Emotions
[49]	CK+, JAFFE and FER2013	MobileNet V2	7
[50]	FERPlus and RAF-DB	MobileNet	7
[51]	JAFFE and CK	CNN	7
[52]	FER-2013, Random dataset	MobileNet-V1	7
[53]	FER2013 dataset and CK+	VGG16	7
[54]	FER-2013, CK+ and RAF-DB	ResNet-50	7
[55]	KDEF	VGG16	6
[56]	CK	ResNet-50	7
[57]	FER2013 and private VEMO	CNN	7
[58]	FER2013, JAFFE, CK+, KDEF, RAF	DCNN	7
[59]	Caltech faces, CMU and NIST	FERC	7
[60]	KDEF and JAFFE	DenseNet-161	7
[61]	FER-2013, CK+, FERG, and JAFFE	CNN	7
[62]	Private	DCNN	5
[63]	FERC-2013, JAFFE	CNN	7

4 Hybrid Learning based Approaches

Extensive empirical studies have shown that the act of concealing mouths significantly complicates the interpretation of facial emotions. Recent research has shown that the majority of emotional facial expressions are attributed to the concealment of the lips, perhaps elucidating the reason for the limited assurance in reaching this conclusion. VGG-16 and AlexNet are the two predominant deep learning models used for feature extraction. The study used many classifiers for FC6, FC7, and FC8 after feature extraction in order to enhance the predictive accuracy of concealed facial emotions [80]. This study shows the first results of a system that uses both CNN-based facial geometry displacement classification and SVM-based facial motion classification to sort facial motion flows into groups. In terms of performance, the system is state-of-the-art. Compared to using individual classifiers, the hybrid model achieves much higher performance when it comes to facial expression categorization [81]. Researchers used facial recognition software to extract network difference features (NDF) from the relationships between nearby locations. In preparation for testing, the researchers created a random forest classifier that could classify facial expressions into seven separate groups [82].

Cross-database studies provide encouraging results compared to state-of-the-art models, further demonstrating the enormous potential of integrating shallow and deep learning features [83]. This combined approach makes use of the energy-entropy-based dual-tree m-band wavelet transform (DTMBWT) method to achieve a high level of precision. Also, it uses a Gaussian mixture model (GMM) for classification purposes, which allows it to accurately identify facial emotions in database photographs. Using the DTMBWT dataset, one may extract many expression characteristics from levels 1 to 6. It is also feasible to get the homogeneity and contrast indices in addition to the GLCM characteristics. In the end, the GMM classifier makes it easier to classify and identify certain traits [84]. This research proposes a state-of-the-art deep learning method for emotion prediction by analysing photographed faces using CNN. CNN for analysing the main emotion (happy or sad) and another for predicting the secondary emotion make up the model of this study [85]. The paper presents a technique for emotion recognition using lighting, occlusion, and posture in a range of face photos. Facial expression identification using a hybrid approach has not been the subject of any previous studies. In spite of training on a dataset of static head poses and illuminations, their model can adjust to variations in head postures, color, contrast, and light [86]. Overview of Hybrid learning approaches for the FER system, utilizing preprocessing, augmentation, feature extraction, classification, training and tuning, and recognition is depicted in Fig. 6.

To characterise the relative geometric position interaction of face landmarks, a deep geometric feature descriptor is the appropriate choice. Basic geometric features may be acquired by categorising the landmarks into seven distinct groups. Adaptive relevance estimation of several landmark sites is achievable using the attention module in ALSTMs. The integration of ALSTMs and SACNNs enables the development of hybrid features for expression recognition [87]. In the domain of subjective evaluation, the GMM-DNN model yields results that are equivalent to those of human assessors. Furthermore, they conducted tests on the classifier in both quiet and noisy conversation environments, as well as on two distinct emotional datasets [88]. Given situations where facial expressions are inherently misleading, the provided paradigm demonstrates significant utility in identifying the accurate emotional condition. This work surpasses or compares favorably the accuracy of the reference subject-independent multimodal emotion recognition research in the available literature [89]. DBN averages segment-level characteristics of a video sequence to build a global video feature representation with a given duration. A linear SVM trained on global video feature representations can classify face expressions [90]. This study compares deep learning with transfer learning methods for categorizing facial emotions as happy or angry. These include CNN, LSTM, Inception, ResNet, VGG, Xception,

and InceptionResnet. Deep hybrid learning (DHL) uses DL and transfer learning to recognise face emotions [91]. Critical analysis of Hybrid learning for FER system is presented in Table 3.

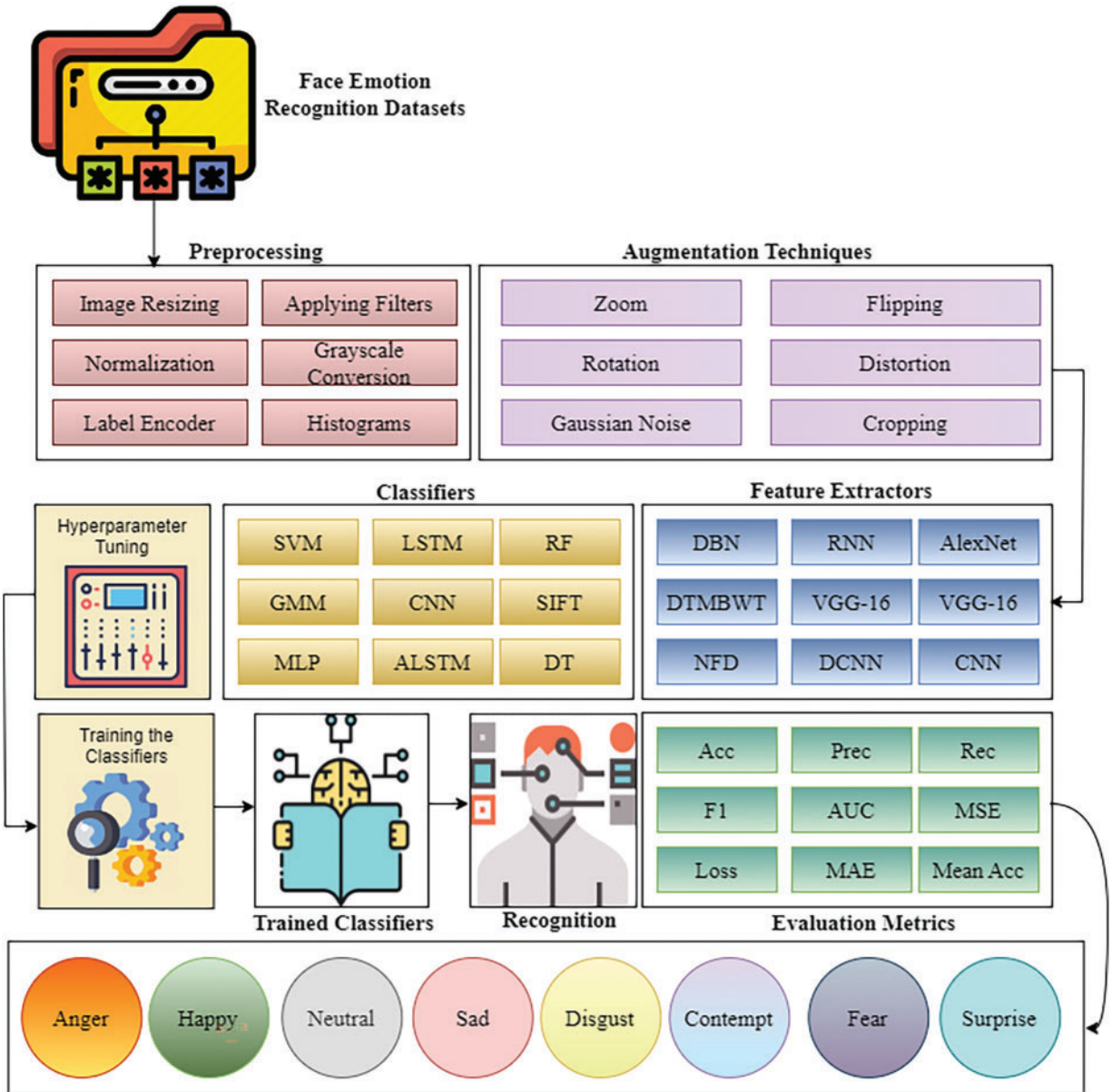


Figure 6: Overview of hybrid learning approaches for the FER system, utilizing preprocessing, augmentation, feature extraction, classification, training and tuning, and recognition

Table 3: Critical analysis of hybrid learning for FER system

Ref.	Datasets	Feature extraction	Classification	Emotions
[80]	M-LFW-FER	AlexNet and VGG-16,	SVM	7
[81]	CK+, BU4D	CNN	SVM	7
[82]	RAF, SFEW	Neighborhood difference features NFD	RF	7
[83]	CK+	CNN	Scale-invariant feature transform (SIFT)	7
[84]	JAFFE	Dual-tree m-band wavelet transform (DTMBWT)	Gaussian mixture model (GMM)	7
[85]	JAFFE, FER2013	CNN	CNN	–
[86]	EMOTIC, FER-13 and FERG	RNN+CNN	LSTM	7
[87]	FER2013, CK+, and JAFFE	Spatial attention convolutional neural network (SACNN)	Long short-term memory networks with attention mechanism (ALSTMs)	7
[88]	ESD	Gaussian mixture model and deep neural network (GMM-DNN)	SVMs and MLP	6
[89]	LUMED-2, DEAP	CNN	DT	3
[90]	BAUM-1s, RML, and MMI	Deep belief network (DBN)	SVM	6
[91]	Multisource	CNN	LSTM	6

5 Facial Emotion Recognition Datasets



The authors utilized various FER datasets for the classification and recognition of human emotion with different methods. A brief explanation and each dataset samples are illustrated in [Table 4](#).

6 Comparative Analysis of Different Learning Methods

Machine learning, deep learning, and hybrid learning are present distinctive advantages and challenges, particularly in the areas of data augmentation and feature extraction. The foundation of machine learning is manual feature extraction and simplified algorithms that are beneficial for tasks that necessitate profound knowledge, but have limited data and computing resources. Conversely, DL autonomously extracts features through the use of sophisticated neural networks. This approach is more advantageous for large and intricate data sets, despite the fact that it necessitates a greater amount



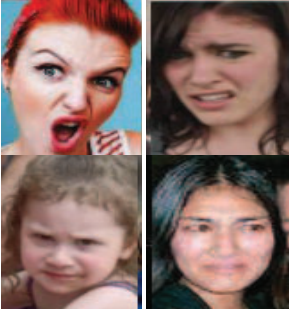

of computing capacity and may offer inferior accuracy. Fuzzy learning is a complex and resource-intensive technique that integrates machine learning and Hybrid learning techniques. It leverages their shared assets to solve more complex problems and increase efficiency. In all applications, data augmentation is advantageous due to its ability to enhance model flexibility, prevent overfitting, and generally enhance model accuracy by artificially increasing their training data. Therefore, comparative Analysis of different learning methods is shown in [Table 5](#).

Table 4: A brief explanation of each dataset used for the experimentation from different researchers, along with a few samples for understanding, is provided

Datasets	Sample images	Description
CK+ [92]		To capture the facial behavior of 210 individuals, authors used two Panasonic AG-7500 cameras that were synced using hardware. Out of the individuals aged 18 to 50, 69% were female, 81% were of Euro-American descent, 13% were Afro-American, and 6% were from another ethnic background. An observer administered a battery of 23 facial expressions to participants, each using a unique action unit or a combination of units. Unless specified differently, each show started and concluded with a neutral countenance. Photographic sequences captured from both 30-degree and frontal viewpoints were used to construct digital arrays of 640×490 or 640×480 pixels, each with 8-bit rescale or 24-bit color values.
JAFFE [93]		In order to streamline experimentation, the JAFFE database consolidates all essential information. Michael J. Lyons and colleagues (1997) established the term JAFFE to refer to the Japanese Female Facial Expression Database. The dataset consists of 213 photographs of Japanese women in seven distinct emotional states, including 10 different facial types (6 basic and 1 neutral). Six emotional descriptors guided the evaluation of each image by sixty Japanese volunteers.

(Continued)

Table 4 (continued)

Datasets	Sample images	Description
Fer2013 [94]		The FER+ dataset categories pictures accordingly as neutral, shocked, sorrowful, angry, disgusted, afraid, or contemptuous. This dataset is an expansion of the original FER one. To the normal Emotion FER dataset, the FER+ annotations provide an extra set of labels. Averaging the annotations of 10 crowd-sourced taggers, the ground truth for emotion in still photographs is superior in FER+ compared to the original FER labels. Employing 10 taggers each photo enables researchers to establish a probability distribution for the emotional state of each individual face.
KDEF [95]		The image collection included color photographs of human faces devoid of hairlines, sourced from the Karolinska Emotional Directed Faces database's a series. In all, these 490 photographs portray 70 people, with an equal distribution of 35 males and 35 females. Based on the findings of Lundqvist et al. (1998), the facial expressions of the people range from neutral to intense, fearful, repulsed, happy, sorrowful, and surprised. Everything converges in the frontal visual perspective.
RAF-DB [96]		The Real-world Affective Faces Database, also referred to as RAF-DB , is a database that captures face emotions. 40 distinct taggers have annotated 39,672 facial photos with simple or complex expressions. The ages, genders, and races of the individuals exhibit significant variation, and these photographs also include a diverse range of lighting arrangements, occlusions (such as spectacles, facial hair, or self-occlusion), and post-processing methods (such as the application of different filters and special effects).
MUG [97]		The database has information on eighty-six individuals of Caucasian descent, aged between twenty and thirty-five. The group consists of 35 females and 51 males, a portion of them do not possess beards. With the exception of seven individuals included in the second half of the database, none of the topics pertain to sports activities. Moreover, the only hindrances are misplaced facial hairs. Within some photographs only located in the second section of the database, a hand may obscure the face.

(Continued)

Table 4 (continued)

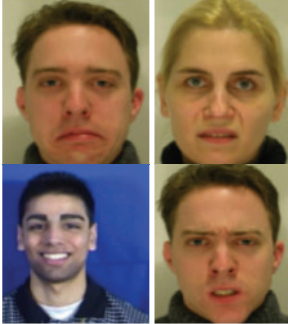
Datasets	Sample images	Description
MMI [98]		<p>This collection has over 2900 videos and high-resolution still photos that depict 75 different themes. The system employs event coding to indicate the appearance of autonomous units (AUs) in movies and partial frame coding to indicate whether an AU is in the offset, neutral, onset, or peak phase at a certain frame. Minor annotations were made for audio-visual amusement. The database is openly available to the scientist community without any cost.</p>

Table 5: Comparative analysis of different learning methods

Ref.	Datasets	Classification	Performance	Comments
[32]	Private	SVM-SMOTE	86	Subsequent presentations of facial expression recognition confirmed the results and preliminary experiment validated machine learning dataset gathering. The experiment showed that their technique could properly identify end-user happiness.
[33]	CK+ and JAFFE	CART	89.2	To calculate the face distance, one may use the Minkowski, Manhattan, or Euclidean distance metrics. To complete the examination of the three lengths, authors employed a decision tree. The JAFFE database's "Minkowski" distance indicated that the initial answer was better.
[34]	JAFFE	KNN	85.71	The method was implemented using MATLAB script codes, and the algorithm's efficacy was evaluated based on accuracy, incorrect classification, and accurate classification. The results indicated that 60 photographs were correctly classified, 10 were incorrectly classified, and 85.71% of the images were accurate.

(Continued)

Table 5 (continued)

Ref.	Datasets	Classification	Performance	Comments
[35]	Private	KNN	98.03	The Lucas-Kande optical flow method is used for the continuous monitoring of marker placements. By using one-way analysis of variance, they examine the relevance of the distance features obtained from six emotions.
[36]	CK+	SVM	93.53	Principal component analysis (PCA) is a method used to reduce the levels of dimensionality in characteristics. This paper presents an algorithm capable of identifying all eight archetypal face emotions. They used a 10-fold validation approach to train and evaluate the classifiers. They used three classifiers to do facial expression categorization.
[37]	CK+	SVM	91.5	The study's findings unequivocally demonstrate the value of the suggested method for extracting facial expressions from video feeds.
[49]	CK+, JAFFE and FER2013	MobileNetV2	97.98	This method can achieve state-of-the-art accuracy on the JAFFE and CK+ datasets thanks to MobileNetv2's benefits. It can also identify data much faster than standard classifiers. Others have done better study than us, but their methods don't work in real time or cost a lot more to run than ours.
[50]	FERPlus and RAF-DB	MobileNet	88.11	According on their empirical results on the FERPlus and RAFDB datasets, the improved model outperforms even the most advanced methods, such as the lightweight MoibleNet series models. On the scale of recognition accuracy, RAF-DB achieves a score of 84.49% whereas FERPlus achieves 88.11%.

(Continued)

Table 5 (continued)

Ref.	Datasets	Classification	Performance	Comments
[52]	FER-2013, Random dataset	MobileNet-V1	97.9	The precise and effective extraction of seven unique behaviours is a significant contribution that this work makes to the field of research. Utilising photographs or photographs, the suggested method was able to effectively accomplish the main purpose of constructing an emotion recognition system that is based on seven classes.
[53]	FER2013 dataset and CK+	VGG16	57.34	To reduce network parameters, adjust the pooling strategy, and include the pre-trained VGG16 model on ImageNet, they needed to substitute the two fully connected layers with convolutional layers.
[54]	FER-2013, CK+ and RAF-DB	ResNet-50	92.32	Twenty students took part in an online learning scenario to evaluate the proposed approach. Using automated facial expression recognition, the approach accurately distinguished between “engaged” and “disengaged” states.
[55]	KDEF	VGG16	88	The CNN model obtains 88% accuracy when using the performance measures as validation. Nevertheless, the results demonstrate how much superior the developed network design is over the earlier approaches.
[80]	M-LFW-FER	SVM	65.5	The mask, which obscures crucial facial features such as the lips and eyes, is only one factor contributing to the diminished realism. Additionally, the mask has the potential to alter the facial characteristics, hence increasing the difficulty for the algorithm to detect and identify them. One potential factor that might impact the accuracy of machines is their inability to comprehend the whole of human emotions.

(Continued)

Table 5 (continued)

Ref.	Datasets	Classification	Performance	Comments
[81]	CK+, BU4D	SVM	99.69%, 94.69%	When working with large datasets, the techniques outlined in may be useful, however they generate an overwhelming quantity of interest points. Currently, this system only relies on western databases to detect facial expressions. Nevertheless, the ultimate objective is to include an eastern dataset.
[82]	RAF, SFEW	RF	57.7, 59	Because they employed deep models, the previous approaches required plenty of computing power. Creators of these strategies concentrated on particular emotional states, making them difficult to apply to others. The suggested technology is adaptable and can recognise complicated face expressions. Training makes the difference, regardless of mood.
[83]	CK+,	Scale-invariant feature transform (SIFT)	94.82	The authors gave an example of how to extract characteristics from a single image frame to identify different facial emotions. Combining SIFT with CNN model deep learning features results in composite features that employ varying degrees of deep learning. For expression analysis, SVMs consider all of the attributes together.
[84]	JAFFE	Gaussian mixture model (GMM)	99.53	This research proposes a method for improving the high-performance computation facial expression recognition accuracy. A hybrid method using the DTMBWT algorithm—which depends on energy, a state of and the gray-level co-occurrence matrix (GLCM)—is recommended in order to achieve the required level of accuracy.

(Continued)

Table 5 (continued)

Ref.	Datasets	Classification	Performance	Comments
[85]	JAFFE, FER2013	CNN	94.12, 97.07	This study uses a CNN to determine whether a picture is mostly joyful or sad and to predict its secondary emotion. Training the proposed model on the FER2013 and JAFFE datasets showed that it predicts emotions from facial expressions better than current methods.
[86]	EMOTIC, FER-13 and FERG	LSTM	94.08	This model achieves better results on the FER13 dataset than on the FERG and EMOTIC datasets. Additional trials using other accessible datasets and deeper learning algorithms could help the researchers get better outcomes.
[87]	FER2013, CK+, and JAFFE	Long short-term memory networks with attention mechanism (ALSTMs)	74.31%, 95.15%, and 98.57%	The recommended technique achieves good performance on the datasets because to its ability to accurately detect frontal faces with little head movement. However, occlusions and variations in head orientations are common real-world situations that might directly lead to the failure of facial landmark recognition.
[88]	ESD	SVMs and MLP	83.97	The primary constraint was that the Emirati speech sample is synthetic and lacks naturalness, a consequence of the difficulty in accurately capturing genuine emotions on video. The majority of seized databases function in a similar fashion.
[89]	LUMED-2, DEAP	DT	74.2, 53.8	Although the subject-independent identification accuracy of the EEG modality is low due to its no stationary characteristics, this may be advantageous in cases with few output classes. To effectively ascertain the actual emotional state, it offered just two courses: high and low valence and arousal levels.

(Continued)

Table 5 (continued)

Ref.	Datasets	Classification	Performance	Comments
[90]	BAUM-1s, RML, and MMI	SVM	55.85, 73.73 and 71.43	After training the VGG16 model using ImageNet data, the authors modify the spatial and temporal CNN networks separately using video facial expression data. They employ a deep DBN model to simultaneously learn numerous spatiotemporal CNN features to effectively include them.
[91]	Multisource	LSTM	81.42	According to the experimental results, VGG19 has the lowest accuracy and remains consistent throughout the training process; whereas the InceptionResNet Model had the highest accuracy among the transfer learning approaches.

7 Discussions

Although current reviews have mostly concentrated on face emotion recognition in different situations, they have overlooked some groundbreaking ideas and neglected to include all types of models for classification. The main goal of this study is to analyse the deep model that can precisely recognize different facial emotions. The training, evaluation, and validation of the model necessitated the use of a manual image collecting protocol [52]. A separate research paper showcased the development of a deep learning framework that harnesses CNN to accurately identify emotions in photographs. This study provides a thorough analysis of several machine, deep, and hybrid fuzzy systems. Furthermore, this study examines a broad spectrum of applications and accompanying difficulties related to FER systems. The aim of this study is to do an extensive search for relevant papers published in the past and identify the prevailing classification and recognition architectures used for emotion recognition on datasets of facial expressions, using methods for processing face images. Fig. 7 presents the most employed models and its performance regarding different face emotion datasets.

The analysis began by using traditional machine learning approaches. The SVM, KNN, RF, and DT are the most often used models throughout the review process. By using these approaches with different hyper parameter ranges, the authors achieved divergent outcomes on separate datasets. RF achieved a statistical accuracy of 99.6% on the CK+ dataset and 57.1% on the RAF dataset when using the SVM model.

The work presents a succinct summary of many CNN architectures, focusing particularly on deep learning models. This research evaluates the performance of several CNN, VGG-16, MobileNet, and ResNet models, including both simple and complex architectures, in the FER domain. The task of comparing responses is complicated by the use of several databases for training and testing, some of which included sequences and others contained images. The range of images used for training and evaluating specific solutions was somewhat restricted. Fig. 8 presents the overall graphical abstract of different models and its performance regarding different face emotion datasets.

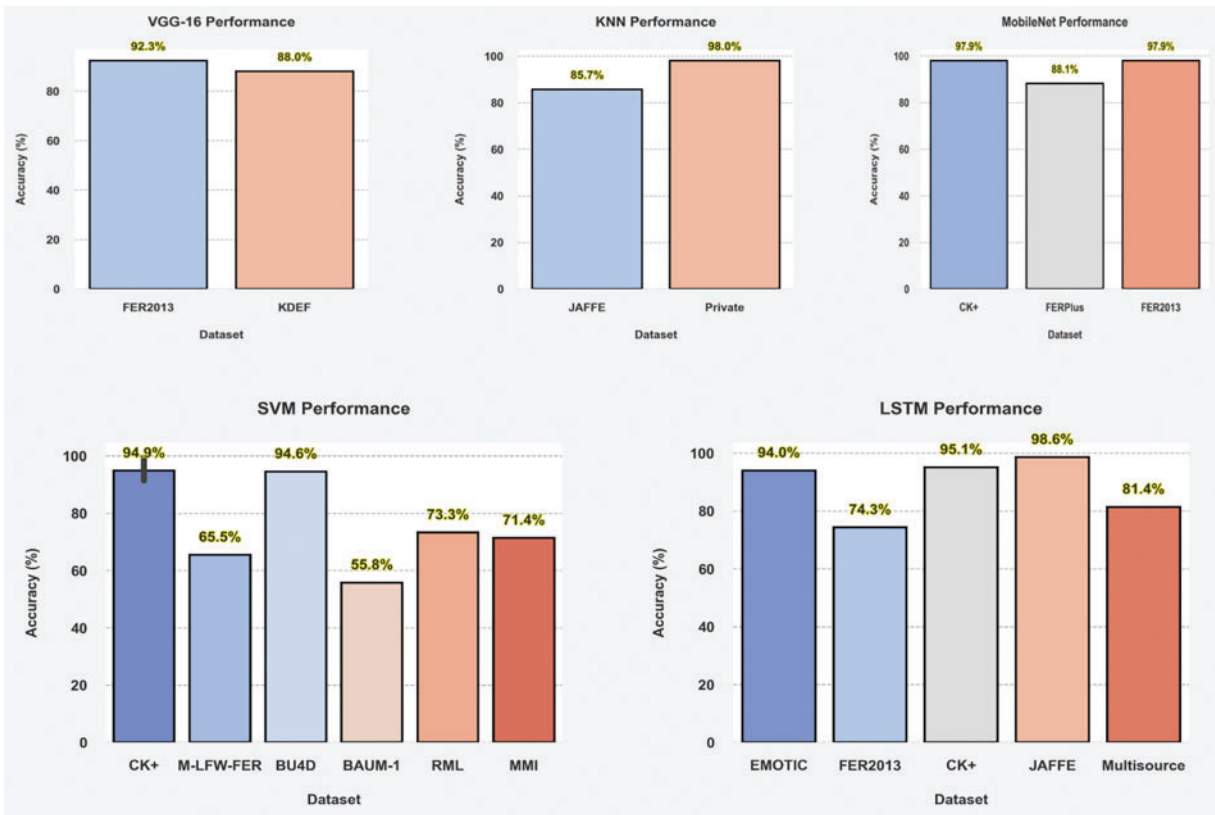


Figure 7: The most employed models and its performance regarding different face emotion datasets

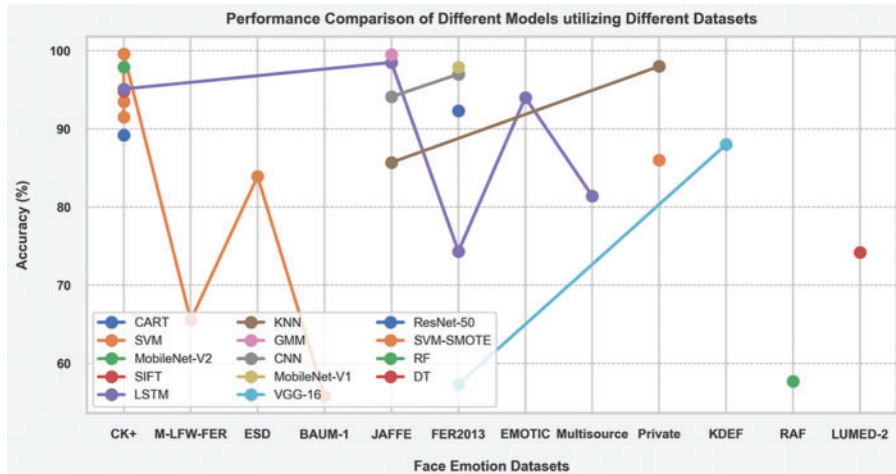


Figure 8: The overall graphical abstract of different models and its performance regarding different face emotion datasets

Experimental testing has demonstrated that deep and hybrid learning-based FER techniques are highly accurate. Nevertheless, there are a few lingering problems that require further investigation. As the framework’s complexity increases, data preparation necessitates a substantial dataset and

computational resources. Experiments and testing necessitate a substantial quantity of memory that consumes time, as well as a significant number of manually created and annotated datasets. The model capacity required by the techniques enumerated above, particularly those that employ deep learning, is substantial. Moreover, these methodologies aim to classify specific emotions, making them unsuitable for classifying other emotional states. Therefore, the development of a new framework that is generalizable to encompass the entire range of emotions would be highly relevant and potentially expandable to encompass the classification of intricate facial expressions.

7.1 Limitations

1. The FER system is compromised when portions of the face are concealed by hands or similar obstructions. In contrast to the actual world, it is not always feasible to visualize the whole spectrum of face traits.

2. Alterations in head position, such as leaning or lateral gazing, induce tension on face nerves, complicating the process. Most facial emotion recognition systems operate well when the occupant is seated in the front passenger seat.

3. Variations in brightness may influence the precision of facial tracking and detection. Insufficient or excessive brightness may conceal significant details and become distracting.

4. Numerous FER systems exhibit culturally unique activation of emotional expressions that was implementing at home. This results in diversity in emotion perception across many cultural situations. Due to substantial variations in face expressions throughout various geographies.

5. Although conventional manual and automated procedures are successful in execution, they are not straightforward to deploy. In complicated or unstructured situations, they often provide varying predictions depending on certain characteristics or datasets.

7.2 Future Work

Future investigations into FER must address current challenges, mitigate biases, and improve the model's applicability to real-world contexts. The fundamental objective is to provide more inclusive and diverse datasets that include a range of facial expressions from many locations, cultures, and ethnicities. Moreover, multimodal approaches improve the precision of emotion recognition by integrating facial expressions with other cues such as vocal tone and body language. Real-time FER systems are crucial for advanced applications requiring storage precision and optimisation in unstructured environments. Mitigating bias problems necessitates the formulation of strategies to reduce population bias in sample performance and provide objective results. The understanding and reliability of FER systems may be improved by the use of explainable artificial intelligence (AI) methodologies, especially in resource-constrained areas. Moreover, developments in practical applications may be effectively represented by improving temporal dynamics in forthcoming models like as Transformer or LSTM. Finally, advanced research indicates that FER techniques are secure in certain applications, whereas unsupervised and self-supervised learning methods reduce the need on large data sets.

8 Conclusion and Challenges

This work provided a brief review of the facial emotion recognition framework, leveraging a combination of behaviors from user evaluation. This approach has the ability to differentiate between seven unique emotional states: joy, anger, neutral, disgust, fear, surprise, and sad. Artificial intelligence

methods have a wide range of practical applications across many sectors. Notably, automated machines are already capable of understanding and interpreting facial emotions. This review analyzes and compares datasets collected from various sources, preprocessing, augmentation approaches, feature extraction, machine learning, and deep and hybrid learning frameworks. It is much less necessary for deep-learning-based face-entity recognition systems to use face-physics-based models and other preprocessing methods because they can learn directly from the photos that are fed into them. RF, SVM, KNN, and DT are among the classification algorithms used in traditional FER. CNN, a subtype of deep learning, uses graphic representations of input images to facilitate the understanding of models developed on various FER datasets. This showcases the effectiveness of networks trained on emotion detection across datasets and other FER-related tasks. Using hybrid methodologies could potentially overcome the issue of CNN-based FER methods not effectively capturing the temporal variations in face components.

Researchers mostly used LSTM and SVM to handle the temporal characteristics of successive frames and a CNN to handle the spatial features of individual frames. Compared to previous traditional methods that calculated the average of data across time, the use of hybrid approaches has demonstrated encouraging outcomes. Regrettably, deep learning-based FER methodology still has several drawbacks. Specifically, these techniques need extensive datasets, substantial memory and processing power, and a considerable duration for both training and testing. Despite the enhanced performance of hybrid architecture, the task of managing facial expressions, which refer to modest and involuntary facial movements that happen spontaneously, remains challenging? Researchers have established the criteria of identification accuracy and recall as the benchmarks for evaluation after extensive testing.

Recently, existing artificial intelligence techniques are unable to replicate intricate human emotions, such as empathy and the capability to understand context. Given that emotional intelligence facilitates the production of common human emotions, researchers are striving to suggest that systems with the ability to integrate it would attain much higher levels of success in the future. Researchers are currently conducting an ongoing investigation to develop continually improved techniques to address computational complexity, poor performance, and training difficulties with the goal of establishing an optimized architecture suitable for real-time applications.

Acknowledgement: The authors are thankful to the AIDA Lab, CCIS, Prince Sultan University, Riyadh, Saudi Arabia for support.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Conceptualization: Amjad Rehman, Muhammad Mujahid, Alex Elyassih, Methodology: Amjad Rehman, Bayan AlGhofaily and Saeed Ali Omer Bahaj; Software: Amjad Rehman, Muhammad Mujahid, Alex Elyassih; validation: Amjad Rehman, Muhammad Mujahid, Alex Elyassih; writing—original draft preparation: Amjad Rehman, Muhammad Mujahid, Saeed Ali Omer Bahaj; writing—review and editing: Amjad Rehman, Bayan AlGhofaily and Saeed Ali Omer Bahaj; visualization, Bayan AlGhofaily, Saeed Ali Omer Bahaj, Alex Elyassih; supervision: Amjad Rehman, Alex Elyassih, project administration: Bayan AlGhofaily, Alex Elyassih, Saeed Ali Omer Bahaj. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The authors confirm that the data supporting the findings of this study are available within the article in [Section 5](#).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

- [1] W. -S. Chu, F. De la Torre, and J. F. Cohn, "Selective Transfer Machine for personalized facial expression analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 529–545, Mar. 2017. doi: [10.1109/TPAMI.2016.2547397](https://doi.org/10.1109/TPAMI.2016.2547397).
- [2] A. J. A. AlBdairi *et al.*, "Face recognition based on deep learning and FPGA for ethnicity identification," *Appl. Sci.*, vol. 12, no. 5, 2022, Art. no. 2605. doi: [10.3390/app12052605](https://doi.org/10.3390/app12052605).
- [3] G. Simcock *et al.*, "Associations between facial emotion recognition and mental health in early adolescence," *Int. J. Environ. Res. Public Health*, vol. 17, no. 1, Jan. 2020, Art. no. 330. doi: [10.3390/ijerph17010330](https://doi.org/10.3390/ijerph17010330).
- [4] L. Li, X. Mu, S. Li, and H. Peng, "A review of face recognition technology," *IEEE Access*, vol. 8, pp. 139110–139120, 2020. doi: [10.1109/ACCESS.2020.3011028](https://doi.org/10.1109/ACCESS.2020.3011028).
- [5] H. Mliki, N. Fourati, S. Smaoui, and M. Hammami, "Automatic facial expression recognition system," in *2013 ACS Int. Conf. Comput. Syst. Appl. (AICCSA)*, Ifrane, Morocco, 2013. doi: [10.1109/AICCSA.2013.6616505](https://doi.org/10.1109/AICCSA.2013.6616505).
- [6] R. I. Bendjillali, M. Beladgham, K. Merit, and A. Taleb-Ahmed, "Improved facial expression recognition based on DWT feature for deep CNN," *Electronics*, vol. 8, no. 3, Mar. 2019, Art. no. 324. doi: [10.3390/electronics8030324](https://doi.org/10.3390/electronics8030324).
- [7] K. Shan, J. Guo, W. You, D. Lu, and R. Bie, "Automatic facial expression recognition based on a deep convolutional-neural-network structure," in *2017 IEEE 15th Int. Conf. Softw. Eng. Res., Manag. Appl. (SERA)*, London, UK, 2017.
- [8] E. Owusu, Y. Zhan, and Q. R. Mao, "A neural-AdaBoost based facial expression recognition system," *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3383–3390, Jun. 2014. doi: [10.1016/j.eswa.2013.11.041](https://doi.org/10.1016/j.eswa.2013.11.041).
- [9] A. Durmusoglu and Y. Kahraman, "Facial expression recognition using geometric features," in *2016 Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Bratislava, Slovakia, 2016.
- [10] Z. Yu, Y. Dong, J. Cheng, M. Sun, and F. Su, "Research on face recognition classification based on improved GoogLeNet," *Secur. Commun. Netw.*, vol. 2022, pp. 1–6, Jan. 2022. doi: [10.1155/2022/7192306](https://doi.org/10.1155/2022/7192306).
- [11] S. Kumar, V. Sagar, and D. Punetha, "A comparative study on facial expression recognition using local binary patterns, convolutional neural network and frequency neural network," *Multimed. Tools Appl.*, vol. 82, no. 16, pp. 24369–24385, 2023. doi: [10.1007/s11042-023-14753-y](https://doi.org/10.1007/s11042-023-14753-y).
- [12] H. Yang, U. Ciftci, and L. Yin, "Facial expression recognition by DE-expression residue learning," in *2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, 2018.
- [13] A. Majumder, L. Behera, and V. K. Subramanian, "Automatic facial expression recognition system using deep network-based data fusion," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 103–114, Jan. 2018. doi: [10.1109/TCYB.2016.2625419](https://doi.org/10.1109/TCYB.2016.2625419).
- [14] S. Deshmukh, M. Patwardhan, and A. Mahajan, "Survey on real-time facial expression recognition techniques," *IET Biom.*, vol. 5, no. 3, pp. 155–163, Sep. 2016. doi: [10.1049/iet-bmt.2014.0104](https://doi.org/10.1049/iet-bmt.2014.0104).
- [15] A. N. Ekweariri and K. Yurtkan, "Facial expression recognition using enhanced local binary patterns," in *2017 9th Int. Conf. Comput. Intell. Commun. Netw. (CICN)*, Girne, Northern Cyprus, 2017, pp. 43–47. doi: [10.1109/CICN.2017.8319353](https://doi.org/10.1109/CICN.2017.8319353).
- [16] S. Yasmin, R. K. Pathan, M. Biswas, M. U. Khandaker, and M. R. I. Faruque, "Development of a robust multi-scale featured local binary pattern for improved facial expression recognition," *Sensors*, vol. 20, no. 18, Sep. 2020, Art. no. 5391. doi: [10.3390/s20185391](https://doi.org/10.3390/s20185391).
- [17] C. Guo, J. Liang, G. Zhan, Z. Liu, M. Pietikäinen and L. Liu, "Extended local binary patterns for efficient and robust spontaneous facial micro-expression recognition," *IEEE Access*, vol. 7, pp. 174517–174530, 2019. doi: [10.1109/ACCESS.2019.2942358](https://doi.org/10.1109/ACCESS.2019.2942358).

- [18] E. J. Cheng *et al.*, “Deep sparse representation classifier for facial recognition and detection system,” *Pattern Recognit. Lett.*, vol. 125, no. 4, pp. 71–77, Jul. 2019. doi: [10.1016/j.patrec.2019.03.006](https://doi.org/10.1016/j.patrec.2019.03.006).
- [19] P. A. Baffour, H. Nunoo-Mensah, E. Keelson, and B. Kommey, “A survey on deep learning algorithms in facial emotion detection and recognition,” *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 7, no. 1, pp. 24–32, Jun. 2022. doi: [10.25139/inform.v7i1.4282](https://doi.org/10.25139/inform.v7i1.4282).
- [20] H. Ge, Z. Zhu, Y. Dai, B. Wang, and X. Wu, “Facial expression recognition based on deep learning,” *Comput. Methods Programs Biomed.*, vol. 215, no. 5, 2022. doi: [10.1016/j.cmpb.2022.106621](https://doi.org/10.1016/j.cmpb.2022.106621).
- [21] S. K. Singh, R. K. Thakur, S. Kumar, and R. Anand, “Deep learning and machine learning based facial emotion detection using CNN,” in *2022 9th Int. Conf. Comput. Sustain. Glob. Dev. (INDIACom)*, New Delhi, India, 2022. doi: [10.23919/INDIACom54597.2022.9763165](https://doi.org/10.23919/INDIACom54597.2022.9763165).
- [22] A. Khattak, M. Z. Asghar, M. Ali, and U. Batool, “An efficient deep learning technique for facial emotion recognition,” *Multimed. Tools Appl.*, vol. 81, no. 2, pp. 1649–1683, Jan. 2022. [Online]. Accessed: Nov. 2, 2024. doi: [10.1007/s11042-021-11298-w](https://doi.org/10.1007/s11042-021-11298-w).
- [23] C. Bisogni, A. Castiglione, S. Hossain, F. Narducci, and S. Umer, “Impact of deep learning approaches on facial expression recognition in healthcare industries,” *IEEE Trans. Ind. Inform.*, vol. 18, no. 8, pp. 5619–5627, Aug. 2022. doi: [10.1109/TII.2022.3141400](https://doi.org/10.1109/TII.2022.3141400).
- [24] M. F. Alsharekh, “Facial emotion recognition in verbal communication based on deep learning,” *Sensors*, vol. 22, no. 16, Aug. 2022, Art. no. 6105. doi: [10.3390/s22166105](https://doi.org/10.3390/s22166105).
- [25] M. R. A. Borgalli and S. Surve, “Deep learning for facial emotion recognition using custom CNN architecture,” *J. Phys.: Conf. Ser.*, vol. 2236, 2022, Art. no. 012004.
- [26] A. A. Saad, M. Shah, M. R. Khurram Ehsan, A. Amirzada, and T. Mahmood, “Automated facial expression recognition framework using deep learning,” *J. Healthc. Eng.*, vol. 2022, no. 1, pp. 1–11, 2022. doi: [10.1155/2022/5707930](https://doi.org/10.1155/2022/5707930).
- [27] T. Kumar Arora *et al.*, “Optimal facial feature based emotional recognition using deep learning algorithm,” *Comput. Intell. Neurosci.*, vol. 2022, Sep. 2022, Art. no. 8379202.
- [28] D. Dukić and A. Sović Krzić, “Real-time facial expression recognition using deep learning with application in the active classroom environment,” *Electronics*, vol. 11, no. 8, Apr. 2022, Art. no. 1240. doi: [10.3390/electronics11081240](https://doi.org/10.3390/electronics11081240).
- [29] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, “Real time face detection and facial expression recognition: Development and applications to human computer interaction,” in *2003 Conf. Comput. Vis. Pattern Recognit. Workshop*, Madison, WI, USA, 2003. doi: [10.1109/CVPRW.2003.10057](https://doi.org/10.1109/CVPRW.2003.10057).
- [30] X. Zhao, J. Zhu, B. Luo, and Y. Gao, “Survey on facial expression recognition: History, applications, and challenges,” *IEEE Multimed.*, vol. 28, no. 4, pp. 38–44, Oct.–Dec. 1, 2021. doi: [10.1109/MMUL.2021.3107862](https://doi.org/10.1109/MMUL.2021.3107862).
- [31] A. A. M. Al-Modwahi, O. Sebetela, L. N. Batleng, B. Parhizkar, and A. H. Lashkari, “Facial expression recognition intelligent security system for real time surveillance,” in *World Congress Comput. Sci., Comput. Eng., Appl. Comput. (WORLDCOMP'12) World Congr. Comput. Sci., Comput. Eng., Appl. Comput. (WORLDCOMP'12)*, 2012.
- [32] K. Koonsanit and N. Nishiuchi, “Classification of user satisfaction using facial expression recognition and machine learning,” in *2020 IEEE Region 10 Conf. (TENCON)*, Osaka, Japan, 2020, pp. 561–566. doi: [10.1109/TENCON50793.2020.9293912](https://doi.org/10.1109/TENCON50793.2020.9293912).
- [33] F. Z. Salmam, A. Madani, and M. Kissi, “Facial expression recognition using decision trees,” in *2016 13th Int. Conf. Comput. Graph., Imag. Visualizat. (CGiV)*, Beni Mellal, Morocco, 2016, pp. 125–130. doi: [10.1109/CGiV.2016.33](https://doi.org/10.1109/CGiV.2016.33).
- [34] N. Iromini, B. T. Asi, C. O. Okuwa, and K. Gbadamosi, “Facial expression recognition system using KNN,” *Int. J. Sci., Engineer. Environment. Technolog. (IJOSEET)*, vol. 6, no. 4, pp. 31–38, 2021.
- [35] M. Murugappan *et al.*, “Facial expression classification using KNN and decision tree classifiers,” in *4th Int. Conf. Comput., Commun. Signal Process. (ICCCSP)*, Chennai, India, 2020, pp. 1–6. doi: [10.1109/ICCCSP49186.2020.9315234](https://doi.org/10.1109/ICCCSP49186.2020.9315234).

- [36] H. I. Dino and M. B. Abdulrazzaq, "Facial expression classification based on SVM, KNN and MLP classifiers," in *2019 Int. Conf. Adv. Sci. Eng. (ICOASE)*, Zakho-Duhok, Iraq, 2019, pp. 70–75. doi: [10.1109/ICOASE.2019.8723728](https://doi.org/10.1109/ICOASE.2019.8723728).
- [37] M. Priya, B. Ranjani, K. Rajammal, N. Murali, M. Tamilselvi and M. Kavitha, "Automatic emotion detection using SVM-based optimal kernel function," in *2024 11th Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, New Delhi, India, 2024, pp. 1607–1611. doi: [10.23919/INDIACom61295.2024.10498982](https://doi.org/10.23919/INDIACom61295.2024.10498982).
- [38] B. Tamer Ghareeb, F. Tarek, and H. Said, "FER_ML: Facial emotion recognition using machine learning," *J. Comput. Commun.*, vol. 2, no. 1, pp. 40–49, 2023. doi: [10.21608/jocc.2023.282094](https://doi.org/10.21608/jocc.2023.282094).
- [39] A. A. Alhussan *et al.*, "Facial expression recognition model depending on optimized SVM," *Optim. SVM. Comput., Mater. Contin.*, vol. 76, no. 1, pp. 499–515, 2023. doi: [10.32604/cmc.2023.039368](https://doi.org/10.32604/cmc.2023.039368).
- [40] X. Liu, X. Cheng, and K. Lee, "GA-SVM-based facial emotion recognition using facial geometric features," *IEEE Sens. J.*, vol. 21, no. 10, pp. 11532–11542, May 2021. doi: [10.1109/JSEN.2020.3028075](https://doi.org/10.1109/JSEN.2020.3028075).
- [41] L. Yao, Y. Wan, H. Ni, and B. Xu, "Action unit classification for facial expression recognition using active learning and SVM," *Multimed. Tools Appl.*, vol. 80, no. 16, pp. 24287–24301, Jul. 2021. doi: [10.1007/s11042-021-10836-w](https://doi.org/10.1007/s11042-021-10836-w).
- [42] M. I. H. Alzawali *et al.*, "Facial emotion images recognition based on binarized genetic algorithm-random forest," *Baghdad Sci. J.*, vol. 21, no. 2(SI), 2024, Art. no. 0780. doi: [10.21123/bsj.2024.9698](https://doi.org/10.21123/bsj.2024.9698).
- [43] V. A. Saeed, "A framework for recognition of facial expression using HOG features," *Int. J. Math.*, vol. 2, pp. 1–8, 2024.
- [44] S. Varma, M. Shinde, and S. S. Chavan, "Analysis of PCA and LDA features for facial expression recognition using SVM and HMM classifiers," in *Techno-Societal 2018*, Cham, Switzerland: Springer International Publishing, 2020, pp. 109–119.
- [45] M. Nazir, Z. Jan, and M. Sajjad, "Facial expression recognition using histogram of oriented gradients based transformed features," *Clus. Comput.*, vol. 21, no. 1, pp. 539–548, Mar. 2018. doi: [10.1007/s10586-017-0921-5](https://doi.org/10.1007/s10586-017-0921-5).
- [46] P. P. Wu, H. Liu, X. W. Zhang, and Y. Gao, "Spontaneous versus posed smile recognition via region-specific texture descriptor and geometric facial dynamics," *Front. Inf. Technol. Electron. Eng.*, vol. 18, no. 7, pp. 955–967, 2017. doi: [10.1631/FITEE.1600041](https://doi.org/10.1631/FITEE.1600041).
- [47] C. V. R. Reddy, U. S. Reddy, and K. V. K. Kishore, "Facial emotion recognition using NLPCA and SVM," *Trait. Du Signal*, vol. 36, no. 1, pp. 13–22, Apr. 2019. doi: [10.18280/ts.360102](https://doi.org/10.18280/ts.360102).
- [48] L. Hu and Q. Ge, "Automatic facial expression recognition based on MobileNetV2 in Real-time," *J. Phys. Conf. Ser.*, vol. 1549, no. 2, Jun. 2020, Art. no. 022136. doi: [10.1088/1742-6596/1549/2/022136](https://doi.org/10.1088/1742-6596/1549/2/022136).
- [49] Y. Nan, J. Ju, Q. Hua, H. Zhang, and B. Wang, "A-MobileNet: An approach of facial expression recognition," *Alex. Eng. J.*, vol. 61, no. 6, pp. 4435–4444, Jun. 2022. doi: [10.1016/j.aej.2021.09.066](https://doi.org/10.1016/j.aej.2021.09.066).
- [50] G. K. Prajapat, B. P. Sharma, A. Agrawal, R. Soni, and S. K. Saini, "Universal model for facial expression detection using convolutional neural network," *AIP Conf. Proc.*, vol. 3111, 2024, Art. no. 030022.
- [51] H. B. U. Haq, W. Akram, M. N. Irshad, A. Kosar, and M. Abid, "Enhanced real-time facial expression recognition using deep learning," *Acadlore Trans. AI Mach. Learn.*, vol. 3, no. 1, pp. 24–35, Jan. 2024. doi: [10.56578/ataiml030103](https://doi.org/10.56578/ataiml030103).
- [52] C. Dong, R. Wang, and Y. Hang, "Facial expression recognition based on improved VGG convolutional neural network," *J. Phys. Conf. Ser.*, vol. 2083, no. 3, 2021, Art. no. 032030. doi: [10.1088/1742-6596/2083/3/032030](https://doi.org/10.1088/1742-6596/2083/3/032030).
- [53] S. Gupta, P. Kumar, and R. K. Tekchandani, "Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models," *Multimed. Tools Appl.*, vol. 82, no. 8, pp. 11365–11394, 2023. doi: [10.1007/s11042-022-13558-9](https://doi.org/10.1007/s11042-022-13558-9).
- [54] S. A. Hussain and A. Salim Abdallah Al Balushi, "A real time face emotion classification and recognition using deep learning model," *J. Phys. Conf. Ser.*, vol. 1432, no. 1, Jan. 2020, Art. no. 012087. doi: [10.1088/1742-6596/1432/1/012087](https://doi.org/10.1088/1742-6596/1432/1/012087).
- [55] B. Li and D. Lima, "Facial expression recognition via ResNet-50," *Int. J. Cognit. Comput. Eng.*, vol. 2, pp. 57–64, Jun. 2021. doi: [10.1016/j.ijcce.2021.02.002](https://doi.org/10.1016/j.ijcce.2021.02.002).

- [56] L. Pham, T. Huynh Vu, and T. A. Tran, "Facial expression recognition using residual masking network," in *2020 25th Int. Conf. Pattern Recognit. (ICPR)*, Milan, Italy, 2021.
- [57] K. Mohan, A. Seal, O. Krejcar, and A. Yazidi, "Facial expression recognition using local gravitational force descriptor-based deep convolution neural networks," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021. doi: [10.1109/TIM.2020.3031835](https://doi.org/10.1109/TIM.2020.3031835).
- [58] N. Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)," *SN Appl. Sci.*, vol. 2, no. 3, Mar. 2020. doi: [10.1007/s42452-020-2234-1](https://doi.org/10.1007/s42452-020-2234-1).
- [59] M. A. H. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, "Facial emotion recognition using Transfer Learning in the Deep CNN," *Electronics*, vol. 10, no. 9, Apr. 2021, Art. no. 1036. doi: [10.3390/electronics10091036](https://doi.org/10.3390/electronics10091036).
- [60] S. Minaee, M. Minaei, and A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," *Sensors*, vol. 21, no. 9, Apr. 2021, Art. no. 3046. doi: [10.3390/s21093046](https://doi.org/10.3390/s21093046).
- [61] E. Pranav, S. Kamal, C. Satheesh Chandran, and M. H. Supriya, "Facial emotion recognition using deep convolutional neural network," in *2020 6th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Coimbatore, India, 2020.
- [62] A. Jaiswal, A. Krishnama Raju, and S. Deb, "Facial emotion detection using deep learning," in *2020 Int. Conf. Emerg. Technol. (INCET)*, Belgaum, India, 2020.
- [63] M. K. Chowdary, T. N. Nguyen, and D. J. Hemanth, "Deep learning-based facial emotion recognition for human-computer interaction applications," *Neural Comput. Appl.*, vol. 35, no. 32, pp. 23311–23328, Apr. 2023. doi: [10.1007/s00521-021-06012-8](https://doi.org/10.1007/s00521-021-06012-8).
- [64] M. Mukhiddinov, O. Djuraev, F. Akhmedov, A. Mukhamadiyev, and J. Cho, "Masked face emotion recognition based on facial landmarks and deep learning approaches for visually impaired people," *Sensors*, vol. 23, no. 3, Jan. 2023, Art. no. 1080. doi: [10.3390/s23031080](https://doi.org/10.3390/s23031080).
- [65] M. Sajjad *et al.*, "A comprehensive survey on deep facial expression recognition: Challenges, applications, and future guidelines," *Alex. Eng. J.*, vol. 68, no. 6, pp. 817–840, Apr. 2023. doi: [10.1016/j.aej.2023.01.017](https://doi.org/10.1016/j.aej.2023.01.017).
- [66] F. M. Talaat, "Real-time facial emotion recognition system among children with autism based on deep learning and IoT," *Neural Comput. Appl.*, vol. 35, no. 17, pp. 12717–12728, Jun. 2023. doi: [10.1007/s00521-023-08372-9](https://doi.org/10.1007/s00521-023-08372-9).
- [67] K. Sarvakar, R. Senkamalavalli, S. Raghavendra, J. Santosh Kumar, R. Manjunath and S. Jaiswal, "Facial emotion recognition using convolutional neural networks," *Mater. Today*, vol. 80, no. 2, pp. 3560–3564, 2023. doi: [10.1016/j.matpr.2021.07.297](https://doi.org/10.1016/j.matpr.2021.07.297).
- [68] Z. -Y. Huang *et al.*, "A study on computer vision for facial emotion recognition," *Sci. Rep.*, vol. 13, no. 1, May 2023, Art. no. 8425. doi: [10.1038/s41598-023-35446-4](https://doi.org/10.1038/s41598-023-35446-4).
- [69] C. Gautam and K. R. Seeja, "Facial emotion recognition using Handcrafted features and CNN," *Procedia Comput. Sci.*, vol. 218, no. 8, pp. 1295–1303, 2023. doi: [10.1016/j.procs.2023.01.108](https://doi.org/10.1016/j.procs.2023.01.108).
- [70] W. Gong, Y. Qian, W. Zhou, and H. Leng, "Enhanced spatial-temporal learning network for dynamic facial expression recognition," *Biomed. Signal Process. Control.*, vol. 88, no. 8, Feb. 2024, Art. no. 105316. doi: [10.1016/j.bspc.2023.105316](https://doi.org/10.1016/j.bspc.2023.105316).
- [71] M. J. A. I. Dujaili, "Survey on facial expressions recognition: Databases, features and classification schemes," *Multimed. Tools Appl.*, vol. 83, no. 3, pp. 7457–7478, Jan. 2024. doi: [10.1007/s11042-023-15139-w](https://doi.org/10.1007/s11042-023-15139-w).
- [72] J. L. Ngwe, K. M. Lim, C. P. Lee, T. S. Ong, and A. Alqahtani, "PAtt-Lite: Lightweight patch and attention MobileNet for challenging facial expression recognition," *IEEE Access*, vol. 12, pp. 79327–79341, 2024. doi: [10.1109/ACCESS.2024.3407108](https://doi.org/10.1109/ACCESS.2024.3407108).
- [73] R. Gupta, A. S. Shekhawat, M. S. Anand, and P. Selvaraj, "Face expression recognition master using convolutional neural network," in *AIP Conf. Proc.*, Kattankalathur, India, 2024, vol. 3075, Art. no. 020069.
- [74] H. Tao and Q. Duan, "Hierarchical attention network with progressive feature fusion for facial expression recognition," *Neural Netw.*, vol. 170, no. 8, pp. 337–348, Feb. 2024. doi: [10.1016/j.neunet.2023.11.033](https://doi.org/10.1016/j.neunet.2023.11.033).
- [75] M. Rostami, A. Farajollahi, and H. Parvin, "Deep learning-based face detection and recognition on drones," *J. Ambient Intell. Humaniz. Comput.*, vol. 15, no. 1, pp. 373–387, Jan. 2024. doi: [10.1007/s12652-022-03897-8](https://doi.org/10.1007/s12652-022-03897-8).

- [76] L. Xiong, J. Zhang, X. Zheng, and Y. Wang, "Context transformer and adaptive method with visual transformer for robust facial expression recognition," *Appl. Sci.*, vol. 14, no. 4, Feb. 2024, Art. no. 1535. doi: [10.3390/app14041535](https://doi.org/10.3390/app14041535).
- [77] Z. Gao, H. Gao, and Y. Xiang, "Facial emotion recognition based on selective kernel network," *J. Flow Vis. Image Process*, vol. 31, no. 1, pp. 33–52, 2024. doi: [10.1615/JFlowVisImageProc.2023048881](https://doi.org/10.1615/JFlowVisImageProc.2023048881).
- [78] M. Verma and S. K. Vipparthi, "Cross-centroid ripple pattern for facial expression recognition," *Multimed. Tools Appl.*, vol. 48, no. 4, May 2024, Art. no. 384. doi: [10.1007/s11042-024-19364-9](https://doi.org/10.1007/s11042-024-19364-9).
- [79] L. J. Akeh and G. P. Kusuma, "Mixed emotion recognition through facial expression using transformer-based model," *Stat., Optim. Inf. Comput.*, 2024. doi: [10.19139/soic-2310-5070-2103](https://doi.org/10.19139/soic-2310-5070-2103).
- [80] H. M. Shahzad, S. M. Bhatti, A. Jaffar, S. Akram, M. Alhajlah and A. Mahmood, "Hybrid facial emotion recognition using CNN-based features," *Appl. Sci.*, vol. 13, no. 9, 2023. doi: [10.3390/app13095572](https://doi.org/10.3390/app13095572).
- [81] J. C. Kim, M. -H. Kim, H. -E. Suh, M. T. Naseem, and C. -S. Lee, "Hybrid approach for facial expression recognition using convolutional neural networks and SVM," *Appl. Sci.*, vol. 12, no. 11, May 2022, Art. no. 5493. doi: [10.3390/app12115493](https://doi.org/10.3390/app12115493).
- [82] A. Alreshidi and M. Ullah, "Facial emotion recognition using hybrid features," *Informatics*, vol. 7, no. 1, Feb. 2020, Art. no. 6. doi: [10.3390/informatics7010006](https://doi.org/10.3390/informatics7010006).
- [83] X. Sun and M. Lv, "Facial expression recognition based on a hybrid model combining deep and shallow features," *Cognit. Comput.*, vol. 11, no. 4, pp. 587–597, Aug. 2019. doi: [10.1007/s12559-019-09654-y](https://doi.org/10.1007/s12559-019-09654-y).
- [84] J. Kommineni, S. Mandala, M. S. Sunar, and P. M. Chakravarthy, "Accurate computing of facial expression recognition using a hybrid feature extraction technique," *J. Supercomput.*, vol. 77, no. 5, pp. 5019–5044, May 2021. doi: [10.1007/s11227-020-03468-8](https://doi.org/10.1007/s11227-020-03468-8).
- [85] G. Verma and H. Verma, "Hybrid-deep learning model for emotion recognition using facial expressions," *Rev. Socionetwork Strat.*, vol. 14, no. 2, pp. 171–180, Oct. 2020. doi: [10.1007/s12626-020-00061-6](https://doi.org/10.1007/s12626-020-00061-6).
- [86] M. Rahul, N. Tiwari, R. Shukla, D. Tyagi, and V. Yadav, "A new hybrid approach for efficient emotion recognition using deep learning," *Int. J. Electr. Electron. Res.*, vol. 10, no. 1, pp. 18–22, 2022. doi: [10.37391/IJEER](https://doi.org/10.37391/IJEER).
- [87] C. Liu, K. Hirota, J. Ma, Z. Jia, and Y. Dai, "Facial expression recognition using hybrid features of pixel and geometry," *IEEE Access*, vol. 9, pp. 18876–18889, 2021. doi: [10.1109/ACCESS.2021.3054332](https://doi.org/10.1109/ACCESS.2021.3054332).
- [88] I. Shahin, A. B. Nassif, and S. Hamsa, "Emotion recognition using hybrid Gaussian mixture model and deep neural network," *IEEE Access*, vol. 7, pp. 26777–26787, 2019. doi: [10.1109/ACCESS.2019.2901352](https://doi.org/10.1109/ACCESS.2019.2901352).
- [89] Y. Cimtay, E. Ekmekcioglu, and S. Caglar-Ozhan, "Cross-subject multimodal emotion recognition based on hybrid fusion," *IEEE Access*, vol. 8, pp. 168865–168878, 2020. doi: [10.1109/ACCESS.2020.3023871](https://doi.org/10.1109/ACCESS.2020.3023871).
- [90] S. Zhang, X. Pan, Y. Cui, X. Zhao, and L. Liu, "Learning affective video features for facial expression recognition via hybrid deep learning," *IEEE Access*, vol. 7, pp. 32297–32304, 2019. doi: [10.1109/ACCESS.2019.2901521](https://doi.org/10.1109/ACCESS.2019.2901521).
- [91] R. K. Mishra, S. Urolagin, J. A. Arul Jothi, and P. Gaur, "Deep hybrid learning for facial expression binary classifications and predictions," *Image Vis. Comput.*, vol. 128, no. 10, Dec. 2022, Art. no. 104573. doi: [10.1016/j.imavis.2022.104573](https://doi.org/10.1016/j.imavis.2022.104573).
- [92] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The extended cohn-kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.-Workshops*, San Francisco, CA, USA, 2010, pp. 94–101. doi: [10.1109/CVPRW.2010.5543262](https://doi.org/10.1109/CVPRW.2010.5543262).
- [93] M. Lyons, M. Kamachi, and J. Gyoba, "The Japanese female facial expression (JAFFE) dataset," *Zenodo*, 1998. doi: [10.5281/zenodo.3451524](https://doi.org/10.5281/zenodo.3451524).
- [94] M. Sambare, "FER-2013: Facial expression recognition dataset. Kaggle," 2013. Accessed: Nov. 25, 2024. [Online]. Available: <https://www.kaggle.com/datasets/msambare/fer2013>
- [95] D. Lundqvist, A. Flykt, and A. Öhman, "Karolinska directed emotional faces," *APA PsycTests*, 2015. doi: [10.1037/t27732-000](https://doi.org/10.1037/t27732-000).
- [96] S. Li, W. Deng, and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2852–2861.

- [97] N. Aifanti, C. Papachristou, and A. Delopoulos, "The MUG facial expression database," in *11th Int. Workshop Image Anal. Multimed. Interact. Serv., WIAMIS 10*, IEEE, 2010, pp. 1–4.
- [98] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *2005 IEEE Int. Conf. Multimed. Expo.*, Amsterdam, The Netherlands, 2005.