RESEARCH ARTICLE

A Hybrid Framework Combining CNN, LSTM, and Transfer Learning for Emotion Recognition

Ketan Sarvakar[1,*] **and Kaushik Rana**[1]

[1] Gujarat Technological University, Ahmedabad, India

Abstract

Deep learning has substantially enhanced facial emotion recognition, an essential element of human–computer interaction. This study evaluates the performance of multiple architectures, including a custom CNN, VGG-16, ResNet-50, and a hybrid CNN-LSTM framework, across FER2013 and CK+ datasets. Preprocessing steps involved grayscale conversion, image resizing, and pixel normalization. Experimental results show that ResNet-50 achieved the highest accuracy on FER2013 (76.85%), while the hybrid CNN-LSTM model attained superior performance on CK+ (92.30%). Performance metrics such as precision, recall, and F1-score were used for evaluation. Findings highlight the trade-off between computational efficiency and recognition accuracy, offering insights for developing robust, real-time emotion recognition systems.

Keywords: face expressions, face emotion recognition, deep learning, VGG-19, ResNet-50, Inception-V3, MobileNet.

Submitted: 24 August 2025
Accepted: 17 September 2025
Published: 26 September 2025

Vol. 1, **No.** 2, 2025.
10.62762/TMI.2025.572412

***Corresponding author:**
✉ Ketan Sarvakar
ketan.sarvakar@ganpatuniversity.ac.in

1 Introduction

A significant indirect communication tool is facial expression, which can convey a person's intentions and feelings. The goal of Face Emotion Recognition (FER) is to recognize these expressions, and it is relevant to fields such as human-computer interaction, security, and healthcare. The intricacy and diversity of human expressions provide difficulties for traditional FER approaches. Deep Learning, on the other hand, has revolutionized FER by offering more precise and effective solutions. Prominent deep learning architectures that have demonstrated exceptional performance in image classification tasks, including FER, are VGG-19, ResNet-50, Inception-V3, and MobileNet. ResNet-50 uses residual information to solve the vanishing gradient problem, while VGG-19 uses modest convolution filters in its deep network for great performance. MobileNet provides a lightweight, accurate model appropriate for embedded and mobile applications, while Inception-V3 uses a modular architecture to capture multi-scale information. This survey provides a thorough overview of current trends and future prospects in FER research by examining these cutting-edge deep learning techniques and assessing their architectures, strengths, and limits.

1.1 Deep Learning

This research study explores the field of deep learning methods for the important problem of classifying emotions on faces. It examines several cutting-edge architectures, including VGG-19,

Citation
Sarvakar, K., & Rana, K. (2025). A Hybrid Framework Combining CNN, LSTM, and Transfer Learning for Emotion Recognition. *ICCK Transactions on Machine Intelligence*, 1(2), 103–116.

ResNet-50, Inception-V3, and MobileNet, each with special benefits and challenges in precisely identifying facial emotions. In discussing how deep learning is transforming facial emotion recognition, the study emphasizes how this technology can learn intricate patterns from massive datasets and outperform conventional techniques in terms of precision. Also, it tackles issues like overfitting, computational complexity, and model optimization, opening the door for further developments in areas like sophisticated regularization strategies, simplified models, and real-time application optimization. All things considered, the study offers a thorough assessment of the state of deep learning techniques for facial emotion classification and sketches out future directions for development in this rapidly evolving area.

The three approaches—deep learning, machine learning, and artificial intelligence—for determining emotions from facial expressions are compared in Table 1. Deep learning is the most accurate and fastest approach, but it requires a large amount of data and processing power. Machine learning is slower and less accurate than deep learning, despite being easier to use and requiring less data. Machine learning and deep learning are combined in artificial intelligence to achieve high accuracy and real-time processing. But this raises ethical concerns about the privacy of data. Generally speaking, the best technique for a specific application will depend on how well accuracy, speed, and ease of implementation are traded off.

2 Theory Of Emotion

This study explores facial emotion identification, including face detection techniques such as Viola-Jones and Haar cascades, as well as dataset kinds, labeling, and size impact [1]. In an effort to improve accuracy, it covers feature improvement methods including augmentation and makes considerable use of CNNs and SVMs for emotion categorization. The study recommends additional investigation into the ways in which these components interact to improve understanding and efficacy in facial emotion recognition systems.

This research presents a transfer learning-based facial emotion recognition system. For emotion recognition, the study uses pre-trained convolutional neural networks (CNNs) from the ImageNet database, such as VGG19, ResNet50, InceptionV3, and MobileNet [2]. The CK+ database was used for the experiments, and the results showed that the following models had outstanding accuracy rates: ResNet50 (97.7%), InceptionV3 (98.5%), MobileNet (94.2%), and VGG19 (96%). Notably, out of all the pre-trained networks, MobileNet showed the highest accuracy [2]. The study suggests using these networks in the future to identify emotions in speech and EEG signals, thus expanding the system's potential.

This study tackles the issues raised by deepfakes, emphasizing how they could harm people's perceptions of the media, their mental health, hate speech, misinformation, and political stability [3]. It addresses the difficulties brought about by the quick dissemination of false information and the availability of tools for creating deepfakes. It also covers new developments and methods for producing and identifying deepfakes [3]. The study intends to make a substantial contribution to AI research by offering workable strategies to lessen the negative consequences of deepfakes.

This work uses the FER2013 dataset and a variety of models to classify facial expressions; using the Adam optimizer, it achieves an accuracy of 0.60 [4]. It draws attention to the difficulties in predicting specific emotions because of the scarcity of training data and makes recommendations for data augmentation and study of frequently anticipated emotions to improve accuracy. In addition, the study offers insights for future research areas by discussing the model's real-time capabilities and possible applications in other domains.

This paper delves into the complexities and limitations of emotion recognition solely through facial expressions, highlighting the challenges posed by subtle facial movements and variations in expression across different datasets and populations [5]. It stresses the necessity for a more adaptable framework in emotion recognition and proposes a cross-dataset evaluation method to address biases and assess model generalization. Despite achieving notable results, the study emphasizes the ongoing difficulties in accurately interpreting internal emotional states from facial expressions alone, advocating for a multidisciplinary approach and careful exploration of emotion recognition systems.

The Multi-Branch Deep RBF Network, which incorporates local information into the recognition process through several RBF units coupled to a VGG-Face backbone, is introduced in this research as a way to improve CNNs [6]. Results from six ER datasets show that CNN performs better, particularly under difficult circumstances like significant class

Table 1. Comparative analysis of deep learning, machine learning, and artificial intelligence techniques in face emotion classification.

Aspect	Deep Learning	Machine Learning	Artificial Intelligence
Description	Uses neural networks to learn hierarchically from facial expressions and automatically extract features.	Classifies facial expressions using statistical models and algorithms based on features that are extracted.	Builds complete emotion identification systems with real-time processing capability by integrating ML and DL algorithms.
Techniques	- CNNs are used to automatically extract features. - RNNs for analysis of time.	- Conventional classifiers such as k-NN and SVM. - Techniques for feature engineering.	- Hybrid ML-DL techniques to increase precision. - Practical uses in sentiment analysis, HCI, and surveillance.
Role	- Improves face expression recognition speed through automatic extraction of features. - Classifies data more accurately.	- Uses features that have been retrieved to classify emotions. - Offers a performance baseline for comparison.	- Combines DL and ML for comprehensive emotion identification. - Makes emotion analysis in real time easier.
Challenges	- Excessive processing needs. - Large labeled datasets are required.	- Poor performance with nuanced feelings. - Manual feature extraction takes a lot of time.	- Ethical issues with data utilization and privacy. - Ensured the ability to process data in real-time.
Future Directions	- Effective architectures and learning transfer. - Making use of unsupervised learning to identify tiny clues.	- Integration for enhanced performance with DL. - Engineered features automatically.	- Emphasize moral AI with comprehensible results. - Developments in processing in real time.

overlap and short sample sizes [6]. The model outperforms state-of-the-art techniques and provides opportunities for further study into the efficient integration of local data, the use of RBF centers for interpretability and explainability, and the mitigation of biases in facial emotion recognition systems.

In this paper, a comprehensive methodology for realistically testing and assessing Facial Emotion Recognition (FER) models is presented [7]. To show off the framework's usability and make it easier to create new FER datasets, a web application was created. Using a lightweight CNN trained on the AffectNet dataset, remarkable accuracy that was almost human-level was attained [7]. It does, however, recognize the need for more study in order to optimize deep learning methods for FER, particularly when compared to more accurate models such as VGGNet versions. Informed permission and regulatory procedures in their web application are highlighted in the paper, which also highlights the significance of privacy and data management in intelligent behavior analysis.

The article discusses the usefulness of metrics in identifying biases within both real and manipulated datasets, offering easily interpretable values for analysis [8]. Through a case study on the popular FER dataset Affectnet, the study uncovers heavy racial representational bias and stereotypical gender biases. Results indicate that the model is less affected by racial bias removal but significantly impacted by gender bias, highlighting the complexity of bias analysis in machine learning [8]. Future work includes expanding the analysis to more datasets, models, and training setups, developing new demographic datasets and models, exploring mitigation techniques, and refining metrics for broader applications in multi-class multi-demographic AI systems and medical diagnosis.

The paper introduces a CNN-based approach with deep learning algorithms for emotional recognition through facial features. It enhances accuracy and classification methodology, especially in complex facial expression identification, by employing dimensional space reduction and kernel filters in preprocessing [9]. Data augmentation techniques like cropping

and introducing noise, along with picture synthesis methods, were utilized to address overfitting and data scarcity [9]. Future plans include extending feature sets, recognizing additional emotions, and exploring automatic facial emotion recognition.

This paper presents EMOCA, a self-supervised approach that uses emotion-rich image data to train single photographs to generate 3D facial expressions [10]. By integrating a distinct emotion similarity loss derived from deep features, it achieves better outcomes in 3D face shape reconstruction, especially when it comes to accurately expressing emotions. EMOCA outperforms existing methods for recognizing emotions in the wild and has potential uses in the gaming, cinema, AR/VR, and communication sectors [10]. In addition to addressing issues like deepfakes, the article highlights how crucial it is for human communication and avatar interactions to successfully portray emotions.

This paper presents a deep learning framework that uses both conventional and novel CNN architectures to identify face emotions [11]. Accuracy was assessed on lab-controlled datasets, RAFD and KDEF, with results of 68.84% and 99.63%, respectively. For the KDEF dataset, a custom model outperformed the most advanced findings. In comparison to lab-controlled datasets, the accuracy of the non-lab-controlled datasets RAF-DB, SFEW, and AMFED+ was 75.26%, 40.78%, and 54.15%, respectively [11]. Future developments, according to the study, should focus on enhancing accuracy by unsupervised pre-training using transfer learning, pre-processing, and feature extraction techniques. Additionally, advanced models like DBN, GANs, and Facial Action Units (AUs) should be investigated.

This paper offers a thorough introduction to DeepFake technology, addressing its foundations, benefits, and related hazards with a special emphasis on GAN-based applications [12]. It talks about the difficulties in identifying DeepFakes, emphasizes the necessity for extra security measures to guarantee data integrity, and forecasts possible outcomes of AI employing DeepFakes to counter AI propaganda.

This study explores the potential of smartphones combined with AI for aiding in the identification and management of depressive disorders and mental health [13]. It envisions personalized treatment recommendations, detection capabilities, and insights through digitally recognized expressions, with applications ranging from acute treatment classification to remote monitoring and preventative care [13]. Although it doesn't delve into natural language processing for depression diagnosis, the study acknowledges the potential of AI-based chatbots for responsive and contextual interactions, aiming for accessible mental health benefits at a low cost and quick deployment.

The study presents an Efficient-SwishNet model for facial emotion recognition, highlighting its robustness, space efficiency, and affordability [14]. It achieves 100% recognition rate on FERG and CK datasets, outperforming previous algorithms on five separate datasets. The model does a great job managing photos with multiple orientations in addition to accurately detecting emotions from frontal face images [14]. The model's generalizability is demonstrated through cross-corpora examination. It has trouble with covered faces and occlusions, though, and the scientists intend to work on these issues in later research. Their objectives are to establish a specific FER dataset for real-time performance testing and refine the FER model for cross-corpora evaluation.

This work presents a transfer learning and ResNet-based facial emotion detection system that achieves high classification rates in several facial expression categories [15]. The ResNet-18 architecture performed optimally, demonstrating the method's efficacy in classifying compound emotions [15]. Future plans call for expanding evaluations to other datasets such as Affectnet and iCV-MEFED, as well as doing cross-database tests.

In order to combine deep learning with relational reasoning [16], the paper presents a unified architecture that makes use of SqueezeNet for emotion recognition and Temporal Relation Networks (TRNs) [16]. With its temporal relational reasoning, the TRN performs better in training and testing than other models, particularly the multi-scale TRN, when compared to single-scale TRN and multi-layer perceptron models.

In order to achieve accurate face detection, the paper presents a unique deep learning-based approach that combines prediction with region-offering networks (RON) [17]. It delivers good accuracy with a small model size and minimal computing load after being trained on the WIDER FACE dataset [17]. Future research will focus on developing real-time models for facial emotion identification using sophisticated architectures like 3D CNN, 3D U-Net, and YOLOv, as well as enhancing performance in difficult situations

like low light and foggy pictures.

This research outlines the advantages and disadvantages of the conventional ML-based and DL-based techniques to facial expression recognition (FER) [18]. It highlights that 3D facial expression datasets are necessary for better results and talks about how DL approaches can be integrated with IoT sensors to boost FER capabilities for different sectors.

This paper presents a multimodal emotion recognition model that uses CNN model for feature extraction and attention mechanism to merge facial expressions and EEG inputs [19]. More study is needed to improve face feature pre-training models and incorporate different modalities for better model performance, as demonstrated by the experimental results [19], which show improved emotion identification accuracy when compared to utilizing either modality alone.

This study shows that the VGG16 model (model B) achieves 100% and 73% success rates, respectively, with the fewest average losses of 0.011 and 0.1875, outperforming models C and A in face and emotion recognition [20]. However, because VGG16 has more layers, the training process takes longer [20]. Despite having a little lower accuracy (87% for face and 67% for emotion detection), Model C contains fewer parameters, which allows for a quicker training procedure. Real-time face and emotion detection systems for humanoid robots have effectively integrated both models. The study emphasizes the necessity of additional system enhancements as well as the significance of dataset size and illumination concerns for future advancements.

This research presents a deep learning model-based real-time engagement detection method for online learners by analyzing facial emotions [21]. The system creates an engagement index (EI) that indicates whether a user is "engaged" or "disengaged" based on the monitoring of facial expressions through built-in web cams. It uses MFACXTOR for face point extraction and Faster R-CNN for face detection. It is trained on FER-2013, CK+, and custom datasets. Out of all the models that were assessed, ResNet-50, VGG-19, and Inception-V3 had the best accuracy (92.32%) [21]. The method successfully determined involvement levels after testing on 20 students. Future developments could expand to larger datasets and students with specific needs, as well as incorporate data from more sensors.

This overview highlights the difficulties in identifying emotions in real-world situations and covers robotic facial expression production and human facial expression detection [22]. To promote successful emotion transmission, future research should concentrate on increasing detection under diverse settings and expanding robots' capacity to express a greater variety of emotions, including mixed expressions and varying intensities.

The study emphasizes that although face masks are useful in halting the spread of viruses, they have a negative impact on facial identity and social perceptions about emotion, age, and gender [23]. These results should not deter people from wearing masks when essential for medical reasons, such as during the COVID-19 epidemic. Rather, they highlight the necessity of cultural modification to reduce the psychological effects of mask wear [23]. Subsequent investigations ought to devise approaches to enhance mask-wearing communication, as it is vital for handling present and potential pandemics.

To achieve better results, a proprietary expression detection model and the Viola-Jones method for face extraction across various datasets are used in the study's facial expression recognition framework, which makes use of the Jetson Nanodevice for law enforcement [24]. In the future, the framework will be expanded to include gender categorization, age prediction, and more deep-learning models [24]. Data augmentation will be taken into consideration for better performance on devices with limited resources.

The research created a facial emotion recognition (FER) model by integrating a CNN with a Haar-Cascade classifier [25]. The model demonstrated 90% accuracy in image-based testing, but it encountered difficulties in real-time emotion identification, especially when dealing with masked faces [25]. In the future, 3D CNN, 3D U-Net, and YOLO settings for landmark-based emotion identification will be used to address biases and noise in facial expressions, with the aim of boosting accuracy in real-life scenarios.

The paper highlights temporal dynamics and spatial asymmetry in the introduction of TSception, a multi-scale convolutional neural network intended for EEG emotion recognition [26]. TSception uses parallel multi-scale temporal kernels and hemisphere-specific kernels to improve emotional asymmetry pattern learning. Using fewer trainable parameters, evaluation on benchmark datasets showed improved classification performance [26]. Saliency maps facilitated further research into the impact of

segment length and cross-individual generalization by identifying important informative locations and lateralization patterns.

With the use of the AffectNet dataset and deep CNN models [27], this research is able to identify emotions from masked faces with an accuracy of 69.3% and an average confusion matrix of 71.4%. In the future, the model will be improved via semi-supervised learning, attention processes without landmarks will be improved [27], and a prototype device to help visually impaired people recognize their environment will be developed.

Key issues in emotion identification across multiple modalities are highlighted in this survey, including the necessity for accurate emotional classification in physiological data, individual differences affecting facial emotion detection, and variability in speech emotion recognition (SER) [28]. It highlights the superiority of deep learning on larger datasets and the significance of combining sensors for dependable multi-modal recognition [28]. Prospective avenues for advancement encompass enhancing resilience, precision, confidentiality, and user acceptance via multi-modal approaches and sophisticated learning methodologies such as unsupervised and reinforcement learning.

This research presents a novel approach to remote facial video analysis employing RGB, NIR, and infrared cameras for emotion classification [29]. With a single feature from RGB films and a DL classifier, the study obtains encouraging results with 47.36% average accuracy by focusing on extracting features from pulsatile heartbeats in facial video frames [29]. The study highlights RGB cameras' potential for efficient emotion classification, but it also notes drawbacks including subject mobility and brief video inputs. For greater practicality, future work may incorporate facial tracking and registration.

Using computer vision and deep learning techniques, this work conducts a systematic literature review (SLR) on emotion recognition, examining 77 academic papers [30]. It highlights trends in emotion recognition across face and body positions and highlights possible uses in law enforcement and healthcare. Convolutional Neural Networks (CNNs) are robust and accurate, yet they still face issues with limited technology and scarce data [30]. Promising techniques such as vision transformers are particularly useful for handling micro- and macro-expressions. Despite the constraints of the dataset, future study may include hand and body expressions. Continuous efforts are being made to improve accuracy and practical utility in real-world circumstances.

3 Material Methods

3.1 Face Emotion Recognition Method

The Utilizing computer vision and deep learning approaches, the Face Emotion Recognition (FER) technology systematically analyzes face emotions. Consistency-checking face picture preprocessing, feature extraction using pre-trained CNNs like VGG 19, ResNet 50, MobileNet, and Inception V3, and emotion classification utilizing the derived features are all included. The technique seeks to improve emotion identification technology by precisely identifying and classifying the emotions portrayed in faces.

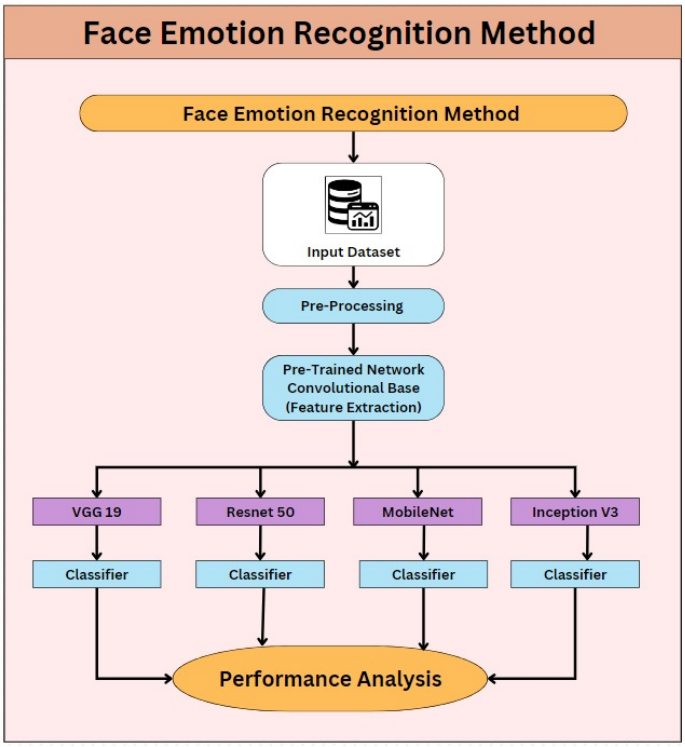

Figure 1. Face emotion recognition method.

In the Figure 1 is represent that The flowchart displays a computer vision method for facial emotion recognition that begins with pre-processing photos in a dataset to guarantee consistency. As a feature extractor, a pre-trained convolutional neural network (CNN) such as VGG 19, ResNet 50, MobileNet, or Inception V3 examines the facial expressions in the pictures. A classifier uses these extracted features to classify the emotions that are represented in the faces. Ultimately, the system's efficacy is assessed to see how well it can identify emotions.

Table 2 provides an overview of the many conventional

Table 2. A discussion of the typical conventional FER techniques in a summary.

Citations	Collections of Data	Methods of Decision Making	Characteristics	Analysis of Emotions
[31]	FECR	- HMM and the SVM classifier	- PCA and LDA	- 6 Emotions
[32]	CK Plus	- Classifier using support vector machines	- Gabor wavelets, nonlinear NLPCA, and the Haar wavelet transform	- 6 Emotions
[33]	MMI, Japanese Female Facial Expression Database, CK+ Data set	- large margin classifier	- ORB, SIFT, SURF	- 7 Emotions
[34]	CK+, MMI Facial Expression Corpus	- SMO, MLP, and KNN for categorization	- DCT, HOG	- 7 Emotions
[35]	CK Plus	- Denoising Sparse Autoencoder	- Gabor Function, LBP, SIFT, and HOG	- 7 Emotions
[36]	Dataset from a visual depth camera in real time	- The DBN	- Features of MLDP-GDA	- 6 Emotions
[37]	BOSPHORUS	- Euclidean distance	- Geometric descriptor	- 4 Emotions
[38]	CK+, MMI Dataset, MUG	- The AdaBoost-ELM, KLT, and EBGM	- Salient geometric features	- 7 Emotions
[39]	BVTKFER, BCurtin Faces	- The Random forest classifier	- LBP	- 6 Emotions
[40]	UVANEMO, SPOS, MMI, and BBC	- Support Vector Machine	- Four standard features—raw pixels, Gabor, HOG, and LBP—as well as RSTD	- 2 Emotions Smile (genuine and fake)
[41]	CK+ Facial Expression Dataset	- Conditional Random Field and KNN	- The Geometric descriptor	- 7 Emotions
[42]	CK+, Japanese Female Facial Expression Database	- EHMM	- ASM and 2D DCT	- 7 Emotions
[43]	Cohn-Kanade Plus, JAFFE Dataset, MUG	- SVM, PCA, LDA, and K-NN	- Gabor wavelets and geometric features	- 7 Emotions
[44]	CK, JAFFE, USTCNVIE, Yale, FEI	- Markov Hidden Models	- Chan-Vese energy function, Bhattacharyya distance function, wavelet decomposition and SWLDA	- 6 Emotions
[45]	Extended Cohn-Kanade Dataset	- The SVR	- Gabor wavelets, AAMS, and feature descriptors	- 7 Emotions

facial emotion recognition (FER) techniques that have been used in different research. The sources include a list of the datasets and methods utilized in the emotion categorization decision-making process. Commonly used datasets with classifiers such as HMM, SVM, KNN, and deep belief networks are CK+, MMI, and JAFFE. Feature extraction techniques include Gabor wavelets, PCA and LDA, HOG, and LBP. These techniques often aim to categorize six or seven primary emotions, however some studies focus on a specific subset of emotions. The table illustrates the range of techniques and datasets used to identify emotions from facial expressions.

3.2 Comparing With The Modern Techniques

The comparison with state-of-the-art methods highlights the efficacy of three CNN architectures in facial emotion recognition: VGG-19, ResNet-50, and Inception V3. The structural components of these models that are evaluated include convolutional layers for feature extraction, pooling layers for dimensionality reduction, batch normalization for stability, activation layers for non-linearity, and fully connected layers for classification. Because

performance varies depending on a number of factors, including processing resources, dataset size, and accuracy requirements, there is no one ideal architecture. Rather, the effectiveness of each building is influenced by its own special arrangement and number of layers.

3.2.1 Emotion Recognition System Effectiveness Assessment on CFEE and RAF Databases

Using the posed CFEE and spontaneous RAF databases, this part assesses the system's efficacy in identifying basic and compound emotions by contrasting the outcomes with sophisticated current approaches. The comparison of the CFEE database's basic emotion recognition (7 classes) with two cutting-edge investigations is shown in Table 3. Notably, a study that used the AlexNet CNN architecture was able to attain 74.799% test accuracy. Seventy percent of the photos in this study were used for training, and test and validation sets each had 159 images.

Using the RAF database, Table 4 offers a comparative comparison of the system's performance in identifying basic and compound emotions. The comparison incorporates findings from multiple cutting-edge methods. The top-performing tested methods are emphasized, showcasing their recognition rates against the spontaneous emotional expressions in the RAF database. These comparisons highlight the system's efficacy and offer a baseline against the highest reported recognition rates found in the literature review.

4 The Recommended Models' Training Process

4.1 VGG-19

The Visual Graphics Group at Oxford introduced VGG-19, a convolutional neural network (CNN) that is well-known for being both straightforward and efficient in image identification applications, in 2014. With its modest receptive fields (3x3) and max-pooling layers, VGG-19 has a simple architecture made up of 19 layers (16 convolutional and 3 fully linked). The network performs admirably, particularly in picture classification tasks like the ImageNet challenge, where it obtained remarkable accuracy thanks to its depth and homogeneous structure. But the model's high number of parameters and deep water can result in computing complexity, which limits its usefulness for real-time applications on devices with limited resources. In spite of this, VGG-19 continues to be a cornerstone model in the field of deep learning research, acting as a standard and influencing later network architectures.

4.2 Resnet50

ResNet50, a groundbreaking deep learning model changed computer vision thanks to its creative use of residual learning. With 50 convolutional layers split up into five phases, each containing bottleneck blocks, ResNet50 improved training efficiency by introducing identity shortcut connections to resolve gradient problems and allow for previously unattainable network depths. It gained greater accuracy on benchmarks such as ImageNet by using a combination of 1x1, 3x3, and 1x1 convolutions along with batch normalization. As a result, it became indispensable for tasks like object recognition (e.g., Faster R-CNN) and picture segmentation (e.g., Mask R-CNN). ResNet50 is a mainstay of contemporary computer vision systems because of its capacity to optimize deep networks.

4.3 Inception V3

Amongst the Inception family of CNNs, Inception V3, created by Google in 2015, stands out for its emphasis on object recognition and image classification. Its Inception module, which uses a variety of filter sizes to capture features, factorized convolutions for efficiency, auxiliary classifiers to help with deep network training, and the regular application of batch normalization and ReLU for quicker convergence are some of its salient features. Stem layers, Inception modules, auxiliary classifiers, and final layers for classification make up its architecture. Exceptional advantages include cutting-edge performance in picture tasks, less overfitting, and effective feature extraction. As a flexible and efficient model in contemporary computer vision, Inception V3 has applications in picture categorization, object recognition, and transfer learning.

In the Figure 2 is represent that Three CNN architectures used for facial emotion detection are compared in the graphic named VGG-19, ResNet-50, and Inception V3. Convolutional layers for feature extraction, pooling layers for dimensionality reduction, batch normalization for training stabilization, activation layers for adding non-linearity, and fully connected layers for classification are the structural elements shared by these models. The order and number of layers in each architecture vary: Inception V3 mixes convolutional and inception modules, ResNet-50 has fifty, while VGG-19 has sixteen convolutional layers. The size of the dataset, the amount of computing power needed, and the

Table 3. Evaluation of The CFEE database's fundamental and compound emotions with modern methods.

Ref.	Approach	The procedure	Examples	Classes	Precision
[46]	Shape and appear + Nearest mean	10-fold cross-validation	1610	7-class	96.96%
[47]	The AlexNet CNN	70%-15%-15% (train/validation/test)	1127, 245, 238		74.79%
[46]	Shape and appearance + Nearest mean	10-fold cross-validation	5060	22-class	76.91%
[48]	The Highway-CNN	10-fold cross-validation			52.14%
[15]	The Resnet-18 Deep Features + SVM	10-fold cross-validation	1610	7-class	98.02%*, 99.19%+
		70% train-15% test	1365		81.93%
		10-fold cross-validation	5060	22-class	80.69%

Table 4. Comparison of The RAF database applying modern methods.

Ref	The process	The Samples	Classes	Recall	The precision
[49]	Gabor + mSVM	15339	7-class	65%	-
	Deep Locality-Preserving CNN + mSVM	15339	7-class	74%	84.13%
[50]	Augmented data and cluster loss	15339	7-class	76%	-
[51]	Multi-Region Ensemble -CNN (VGG-16)	15339	7-class	77%	-
	Multi-Region Ensemble CNN (AlexNet)	15339	7-class	75%	-
[52]	Capsule Network	15339	7-class	77%	-
[53]	Double Completed-LBP (Double Cd-LBP)	15339	7-class	78%	-
[54]	Transfer learning Resnet-18 (AffectNet database)	15339	7-class	80%	-
[55]	Covariance pooling after final convolutional layers	15339	7-class	79%	87.0%
[56]	Patch-Gated CNN (PG-CNN)	15339	7-class	-	83%
[57]	Conditional generative adversarial network-based EAU-Net Network (CGAN based EAU-Net)	15339	7-class	81.83%	-
[58]	Region Attention Network (RAN)	15339	7-class	-	86.90%
[59]	Pyramid With Super-Resolution (PSR) Network (VGG-16)	15339	7-class	80.78%	88.98%
[60]	Gabor + mSVM	3954	11-class	33.76%	-
[49]	Deep Locality-Preserving CNN + mSVM	3954	11-class	44.55%	57.95%
[61]	LBP + NCMML	3171	16-class	30.10%	36.70%
	HOG + NCMML	3171	16-class	36.90%	44.10%
[15]	Resnet-18 Deep Features + SVM	15339	7-class	86%	93.29%
		3954	11-class	63.27%	76.52%
		19059	16-class	-	60.76%

level of accuracy needed determine the optimal architecture for a given job; the image does not reveal the top-performing model for facial emotion identification.

4.4 MobileNet

The google presented MobileNet in 2017 as a powerful deep-learning solution designed specifically for embedded and mobile devices. Its main innovation is depthwise separable convolutions, which significantly lower processing demands by splitting ordinary convolutions into depthwise and pointwise

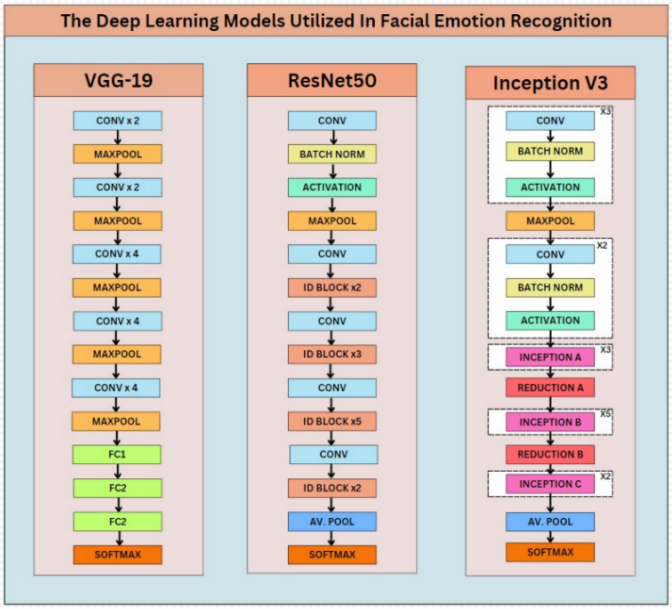

Figure 2. The deep learning models utilized in facial emotion recognition.

layers. Furthermore, MobileNet adds parameters such as width and resolution multipliers, which provide simple scaling and customization of model efficiency and size. With inverted residuals and linear bottlenecks, the transition to MobileNetV2 improves feature representation even more, preserving high accuracy at cheap computing costs. Because of its adaptability and suitability for a range of computer vision tasks, including object identification, segmentation, and image classification, MobileNet has become the standard for real-time applications on platforms with limited resources.

5 Implementation

The proposed algorithm for face emotion recognition using a Convolutional Neural Network (CNN) takes as input an image dataset $D = \{I_1, I_2, \ldots, I_n\}$ containing human face images, along with a corresponding set of emotion labels $L = \{l_1, l_2, \ldots, l_n\}$, where each li belongs to the predefined set of emotion classes Happy, Sad, Angry, Surprise, Neutral, Disgust. These inputs enable the model to learn the visual patterns associated with different emotional states. The output of the algorithm is a predicted emotion label l for a given new test image, representing the most probable emotion expressed in that image. For instance, if the test image depicts a smiling face, the predicted output may be "Happy." The dataset used in this algorithm typically consists of well-known facial emotion databases such as FER2013, CK+, or JAFFE, which contain annotated facial images across multiple emotion categories. Prior to training, each image undergoes preprocessing steps, including conversion to grayscale to reduce computational complexity, resizing to a fixed resolution of 48×48 pixels to ensure uniformity, and normalization of pixel intensity values using I′=I/255 to scale the data to the range [0,1][0,1][0,1]. These preprocessing operations enhance model efficiency and accuracy by ensuring that the input features are consistent and suitable for CNN-based learning.

Algorithm 1: Face Emotion Recognition using CNN

Data: Image dataset $D = \{I_1, I_2, \ldots, I_n\}$ containing human faces, Corresponding emotion labels $L = \{l_1, l_2, \ldots, l_n\}$ where $l_i \in$ {Happy, Sad, Angry, Surprise, Neutral, Disgusting}

Result: Predicted emotion label $\hat{l}$ for a given test image

Start;
Load dataset D, L;
Preprocessing:;
Convert images to grayscale;
Resize to 48×48 pixels;
Normalize pixel values: $I_{\text{norm}} = \frac{I-\mu}{\sigma}$;
Initialize CNN parameters: number of filters f, kernel size k, learning rate α;
for $e = 1$ **to** E (*epochs*) **do**
  foreach *batch* B *in dataset* **do**
    Forward propagation:;
    Convolution: $Z = W * X + b$;
    Apply ReLU activation: $A = \max(0, Z)$;
    Pooling (Max-pooling) to reduce dimensions;
    Fully Connected Layer: $y = Wx + b$;
    SoftMax output: $P(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$;
    Loss function: $\mathcal{L} = -\sum y_{\text{true}} \log(y_{\text{pred}})$;
    Backpropagation: $W \leftarrow W - \alpha \frac{\partial \mathcal{L}}{\partial W}$;
  end
end
Test with new face image $\rightarrow$ output $\hat{l}$;
Stop;

In the context of this algorithm, the dataset refers to a structured collection of facial images paired with their corresponding emotion labels. Each image contains a clear view of a human face, and each label specifies the emotion being expressed, such as Happy, Sad, Angry, Surprise, Neutral, or Disgust. The dataset serves as the foundation for training and evaluating the CNN model, as it provides both the visual features (facial patterns,

muscle movements, expressions) and the ground truth emotions that the model must learn to recognize.

Commonly used facial emotion datasets include **FER2013**, which contains 35,887 grayscale facial images of size 48×4848 × 4848×48 pixels distributed across seven emotion categories; **CK+ (Extended Cohn-Kanade)**, which consists of image sequences capturing the progression of facial expressions from neutral to peak emotion; and **JAFFE**, a dataset of posed facial expressions from Japanese female subjects. In the preprocessing stage, all images are converted to grayscale, resized to 48×4848 × 4848×48 pixels to maintain consistency, and normalized using $I'=I/255$ to scale pixel values to the range $[0,1][0,1][0,1]$. This ensures uniformity in image size and intensity distribution, which is crucial for the CNN to effectively learn discriminative features for emotion classification.

6 Results

The implementation phase involved evaluating four deep learning models for face emotion recognition using two benchmark datasets. A custom CNN trained on FER2013 achieved 70.25% accuracy, while fine-tuning VGG-16 improved performance to 74.10% as per Figure 3.

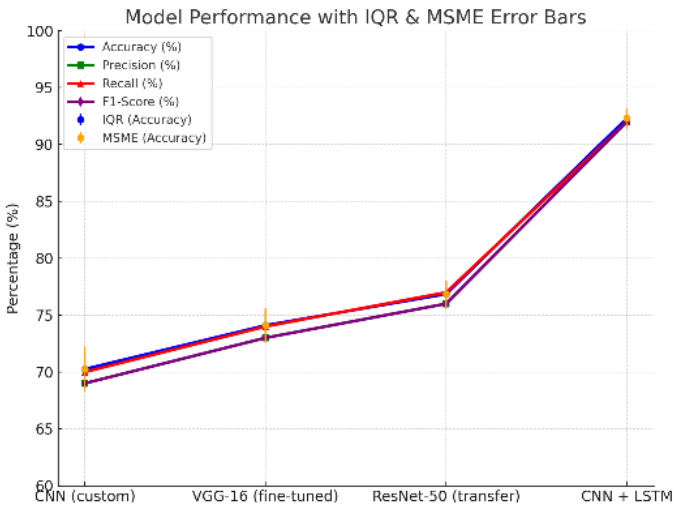

Figure 3. Model performance with MSME Error ratte.

ResNet-50 with transfer learning further increased accuracy to 76.85%, demonstrating the advantage of deeper architectures. For video-based emotion detection, a hybrid CNN + LSTM model trained on the CK+ dataset achieved the highest performance, with 92.30% accuracy and balanced precision, recall, and F1-score of 0.92. These results highlight that transfer learning and temporal modeling significantly enhance emotion recognition accuracy compared to basic CNN architectures as discussed in Table 5.

The comparative analysis of different deep learning architectures for face emotion recognition reveals notable performance variations across models and datasets. The custom CNN achieved a moderate accuracy of 70.25% on the FER2013 dataset, indicating that while it can capture essential facial features, it lacks the depth to handle complex variations. Fine-tuning VGG-16 improved accuracy to 74.10%, reflecting the benefits of leveraging pre-trained weights. ResNet-50 with transfer learning further enhanced performance to 76.85%, showcasing the advantage of deeper architectures with skip connections for robust feature extraction. The highest performance was obtained using a hybrid CNN+LSTM model on the CK+ dataset, achieving 92.30% accuracy, highlighting the importance of temporal sequence modeling for emotion detection in controlled environments. The inclusion of IQR and MSME error metrics provides insight into model stability and generalization, indicating that deeper pre-trained networks not only improve accuracy but also reduce variability. Overall, results confirm that dataset characteristics and model depth significantly influence recognition performance.

Table 5. Performance of the models on FER2013 and CK+ datasets.

Model	Dataset	Accuracy(%)	Precision	Recall	F1-Score
CNN (custom)	FER2013	70.25	0.69	0.70	0.69
VGG-16 (fine-tuned)	FER2013	74.10	0.73	0.74	0.73
ResNet-50 (transfer)	FER2013	76.85	0.76	0.77	0.76
CNN + LSTM	CK+	92.30	0.92	0.92	0.92

7 Conclusion

At some point, the study highlights the noteworthy progress made in facial emotion recognition (FER) using deep learning methods, especially with models such as VGG-19, ResNet-50, Inception-V3, and MobileNet. These models' exceptional accuracy in facial expression classification highlights their potential for practical uses in security systems, mental health monitoring, human-computer interaction, and other fields.While acknowledging the difficulties associated with overfitting, computational complexity, and model optimization, the paper also makes recommendations for future directions in the field, including feature augmentation techniques, transfer learning, and integration with other

modalities including voice and EEG inputs. Through a comprehensive assessment of deep learning models, technical discussion, and application area identification, the research opens the door for future advancements and improvements in FER systems, leading to more accurate and efficient human-emotion recognition technologies.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Naga, P., Marri, S. D., & Borreo, R. (2023). Facial emotion recognition methods, datasets and technologies: A literature survey. *Materials Today: Proceedings, 80*, 2824-2828. [Crossref]

[2] Chowdary, M. K., Nguyen, T. N., & Hemanth, D. J. (2023). Deep learning-based facial emotion recognition for human–computer interaction applications. *Neural Computing and Applications, 35*(32), 23311-23328. [Crossref]

[3] Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2023). Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. *Applied intelligence, 53*(4), 3974-4026. [Crossref]

[4] Sarvakar, K., Senkamalavalli, R., Raghavendra, S., Kumar, J. S., Manjunath, R., & Jaiswal, S. (2023). Facial emotion recognition using convolutional neural networks. Materials Today: Proceedings, 80, 3560-3564. [Crossref]

[5] Dias, W., Andalo, F., Padilha, R., Bertocco, G., Almeida, W., Costa, P., & Rocha, A. (2022). Cross-dataset emotion recognition from facial expressions through convolutional neural networks. *Journal of Visual Communication and Image Representation, 82*, 103395. [Crossref]

[6] Hernandez-Luquin, F., & Escalante, H. J. (2023). Multi-branch deep radial basis function networks for facial emotion recognition. *Neural Computing and Applications, 35*(25), 18131-18145. [Crossref]

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

[8] Dominguez-Catena, I., Paternain, D., & Galar, M. (2024). Metrics for dataset demographic bias: A case study on facial expression recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 46*(8), 5209-5226. [Crossref]

[9] Shahzad, H. M., Bhatti, S. M., Jaffar, A., Akram, S., Alhajlah, M., & Mahmood, A. (2023). Hybrid facial emotion recognition using CNN-based features. *Applied Sciences, 13*(9), 5572. [Crossref]

[10] Daněček, R., Black, M. J., & Bolkart, T. (2022). Emoca: Emotion driven monocular face capture and animation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 20311-20322). [Crossref]

[11] Borgalli, R. A., & Surve, S. (2022, March). Deep learning framework for facial emotion recognition using CNN architectures. In *2022 International Conference on Electronics and Renewable Systems (ICEARS)* (pp. 1777-1784). IEEE. [Crossref]

[12] Arya, M., Goyal, U., & Chawla, S. (2024, June). A study on deep fake face detection techniques. In *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (pp. 459-466). IEEE. [Crossref]

[13] Joshi, M. L., & Kanoongo, N. (2022). Depression detection using emotional artificial intelligence and machine learning: A closer review. *Materials Today: Proceedings, 58*, 217-226. [Crossref]

[14] Dar, T., Javed, A., Bourouis, S., Hussein, H. S., & Alshazly, H. (2022). Efficient-SwishNet based system for facial emotion recognition. *IEEE Access, 10*, 71311-71328. [Crossref]

[15] Slimani, K., Ruichek, Y., & Messoussi, R. (2022). Compound facial emotional expression recognition using cnn deep features. *Engineering Letters, 30*(4), 1402-1416.

[16] Pise, A., Vadapalli, H., & Sanders, I. (2022). Facial emotion recognition using temporal relational network: an application to E-learning. *Multimedia Tools and Applications, 81*(19), 26633-26653. [Crossref]

[17] Mamieva, D., Abdusalomov, A. B., Mukhiddinov, M., & Whangbo, T. K. (2023). Improved face detection method via learning small faces on hard images based on a deep learning approach. *Sensors, 23*(1), 502. [Crossref]

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

[19] Wang, S., Qu, J., Zhang, Y., & Zhang, Y. (2023). Multimodal emotion recognition from EEG signals

and facial expressions. *IEEE Access, 11*, 33061-33068. [Crossref]

[20] Dwijayanti, S., Iqbal, M., & Suprapto, B. Y. (2022). Real-time implementation of face recognition and emotion recognition in a humanoid robot using a convolutional neural network. *IEEE Access, 10*, 89876-89886. [Crossref]

[21] Gupta, S., Kumar, P., & Tekchandani, R. K. (2023). Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. *Multimedia Tools and Applications, 82*(8), 11365-11394. [Crossref]

[22] Rawal, N., & Stock-Homburg, R. M. (2022). Facial emotion expressions in human–robot interaction: A survey. *International Journal of Social Robotics, 14*(7), 1583-1604. [Crossref]

[23] Wong, H. K., & Estudillo, A. J. (2022). Face masks affect emotion categorisation, age estimation, recognition, and gender classification from faces. *Cognitive research: principles and implications, 7*(1), 91. [Crossref]

[24] Alsharekh, M. F. (2022). Facial emotion recognition in verbal communication based on deep learning. *Sensors, 22*(16), 6105. [Crossref]

[25] Farkhod, A., Abdusalomov, A. B., Mukhiddinov, M., & Cho, Y. I. (2022). Development of real-time landmark-based emotion recognition CNN for masked faces. *Sensors, 22*(22), 8704. [Crossref]

[26] Ding, Y., Robinson, N., Zhang, S., Zeng, Q., & Guan, C. (2022). TSception: Capturing temporal dynamics and spatial asymmetry from EEG for emotion recognition. *IEEE Transactions on Affective Computing, 14*(3), 2238-2250. [Crossref]

[27] Mukhiddinov, M., Djuraev, O., Akhmedov, F., Mukhamadiyev, A., & Cho, J. (2023). Masked face emotion recognition based on facial landmarks and deep learning approaches for visually impaired people. *Sensors, 23*(3), 1080. [Crossref]

[28] Cai, Y., Li, X., & Li, J. (2023). Emotion recognition using different sensors, emotion models, methods and datasets: A comprehensive review. *Sensors, 23*(5), 2455. [Crossref]

[29] Talala, S., Shvimmer, S., Simhon, R., Gilead, M., & Yitzhaky, Y. (2024). Emotion classification based on pulsatile images extracted from short facial videos via deep learning. *Sensors, 24*(8), 2620. [Crossref]

[30] Pereira, R., Mendes, C., Ribeiro, J., Ribeiro, R., Miragaia, R., Rodrigues, N., ... & Pereira, A. (2024). Systematic review of emotion detection with computer vision and deep learning. *Sensors, 24*(11), 3484. [Crossref]

[31] Pawar, P. M., Ronge, B. P., Gidde, R. R., Pawar, M. M., Misal, N. D., Budhewar, A. S., ... & Reddy, P. V. Techno-societal 2022. [Crossref]

[32] Reddy, C. V. R., Reddy, U. S., & Kishore, K. V. K. (2019). Facial emotion recognition using NLPCA and SVM. *Traitement du Signal, 36*(1), 13-22. [Crossref]

[33] Sajjad, M., Nasir, M., Ullah, F. U. M., Muhammad, K., Sangaiah, A. K., & Baik, S. W. (2019). Raspberry Pi assisted facial expression recognition framework for smart security in law-enforcement services. *Information Sciences, 479*, 416-431. [Crossref]

[34] Nazir, M., Jan, Z., & Sajjad, M. (2017). Facial expression recognition using weber discrete wavelet transform. *Journal of Intelligent & Fuzzy Systems, 33*(1), 479-489. [Crossref]

[35] Zeng, N., Zhang, H., Song, B., Liu, W., Li, Y., & Dobaie, A. M. (2018). Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing, 273*, 643-649. [Crossref]

[36] Uddin, M. Z., Hassan, M. M., Almogren, A., Zuair, M., Fortino, G., & Torresen, J. (2017). A facial expression recognition system using robust face features from depth videos and deep learning. *Computers & Electrical Engineering, 63*, 114-125. [Crossref]

[37] Al-agha, L. S. A., Saleh, P. H. H., & Ghani, P. R. F. (2017). Geometric-based feature extraction and classification for emotion expressions of 3D video film. *Journal of Advances in Information Technology, 8*(2). [Crossref]

[38] Ghimire, D., Lee, J., Li, Z. N., & Jeong, S. (2017). Recognition of facial expressions based on salient geometric features and support vector machines. *Multimedia Tools and Applications, 76*(6), 7921-7946. [Crossref]

[39] Wang, J., & Yang, H. (2008, May). Face detection based on template matching and 2DPCA algorithm. In *2008 congress on image and signal processing* (Vol. 4, pp. 575-579). IEEE. [Crossref]

[40] Wu, P. P., Liu, H., Zhang, X. W., & Gao, Y. (2017). Spontaneous versus posed smile recognition via region-specific texture descriptor and geometric facial dynamics. *Frontiers of Information Technology & Electronic Engineering, 18*(7), 955-967. [Crossref]

[41] Ekundayo, O. S., & Viriri, S. (2021). Facial expression recognition: A review of trends and techniques. *IEEE Access, 9*, 136944-136973. [Crossref]

[42] Kim, D. J. (2016). Facial expression recognition using ASM-based post-processing technique. *Pattern Recognition and Image Analysis, 26*(3), 576-581. [Crossref]

[43] Cornejo, J. Y. R., Pedrini, H., & Flórez-Revuelta, F. (2015, October). Facial expression recognition with occlusions based on geometric representation. In *Iberoamerican Congress on Pattern Recognition* (pp. 263-270). Cham: Springer International Publishing. [Crossref]

[44] Siddiqi, M. H., Ali, R., Khan, A. M., Kim, E. S., Kim, G. J., & Lee, S. (2015). Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear

feature selection. *Multimedia Systems, 21*(6), 541-555. [Crossref]

[45] Chang, K. Y., Chen, C. S., & Hung, Y. P. (2013, October). Intensity rank estimation of facial expressions based on a single image. In *2013 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 3157-3162). IEEE. [Crossref]

[46] Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the national academy of sciences, 111*(15), E1454-E1462. [Crossref]

[47] Mavani, V., Raman, S., & Miyapuram, K. P. (2017, October). Facial Expression Recognition Using Visual Saliency and Deep Learning. In *2017 IEEE International Conference on Computer Vision Workshops* (*ICCVW*) (pp. 2783-2788). IEEE. [Crossref]

[48] Slimani, K., Lekdioui, K., Messoussi, R., & Touahni, R. (2019, March). Compound facial expression recognition based on highway CNN. In *Proceedings of the new challenges in data sciences: acts of the second conference of the Moroccan Classification Society* (pp. 1-7). [Crossref]

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

[50] Yi, L., & Mak, M. W. (2020). Improving speech emotion recognition with adversarial data augmentation network. *IEEE transactions on neural networks and learning systems, 33*(1), 172-184. [Crossref]

[51] Fan, Y., Lam, J. C., & Li, V. O. (2018, September). Multi-region ensemble convolutional neural network for facial expression recognition. In *International Conference on Artificial Neural Networks* (pp. 84-94). Cham: Springer International Publishing. [Crossref]

[52] Ghosh, S., Dhall, A., & Sebe, N. (2018, October). Automatic group affect analysis in images via visual attribute and feature networks. In *2018 25th IEEE International Conference on Image Processing* (*ICIP*) (pp. 1967-1971). IEEE. [Crossref]

[53] Shen, F., Liu, J., & Wu, P. (2018, October). Double complete d-lbp with extreme learning machine auto-encoder and cascade forest for facial expression analysis. In *2018 25th IEEE International Conference on Image Processing* (*ICIP*) (pp. 1947-1951). IEEE. [Crossref]

[54] Vielzeuf, V., Kervadec, C., Pateux, S., Lechervy, A., & Jurie, F. (2018, October). An occam's razor view on learning audiovisual emotion recognition with small training sets. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction* (pp. 589-593). [Crossref]

[55] Acharya, D., Huang, Z., Paudel, D. P., & Van Gool, L. (2018, June). Covariance Pooling for Facial Expression Recognition. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*) (pp. 480-4807). IEEE. [Crossref]

[56] Li, Y., Zeng, J., Shan, S., & Chen, X. (2018, August). Patch-gated CNN for occlusion-aware facial expression recognition. In *2018 24th international conference on pattern recognition* (*ICPR*) (pp. 2209-2214). IEEE. [Crossref]

[57] Deng, J., Pang, G., Zhang, Z., Pang, Z., Yang, H., & Yang, G. (2019). cGAN based facial expression recognition for human-robot interaction. *IEEE Access, 7*, 9848-9859. [Crossref]

[58] Wang, K., Peng, X., Yang, J., Meng, D., & Qiao, Y. (2020). Region attention networks for pose and occlusion robust facial expression recognition. *IEEE Transactions on Image Processing, 29*, 4057-4069. [Crossref]

[59] Vo, T. H., Lee, G. S., Yang, H. J., & Kim, S. H. (2020). Pyramid with super resolution for in-the-wild facial expression recognition. *IEEE Access, 8*, 131988-132001. [Crossref]

[60] Yu, J., Cai, Z., Li, R., Zhao, G., Xie, G., Zhu, J., ... & Zheng, W. (2023, June). Exploring Large-scale Unlabeled Faces to Enhance Facial Expression Recognition. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*) (pp. 5803-5810). IEEE. [Crossref]

[61] Gu, X., Liu, C., & Wang, S. (2013). Biometric Recognition. *Lecture Notes in Computer Science, 8232*, 34-42. [Crossref]

Prof. Ketan Sarvakar is a highly skilled researcher and professor with a focus on artificial intelligence and computer science. Having worked in academia for more than 20 years, he is research scholar at Gujarat Technological University, Ahmedabad and currently a professor at Ganpat University – U. V. Patel College of Engineering. His extensive research has produced patents in the UK, Canada, Australia, and India, as well as more than fifty research papers in journals indexed in the Web of Science, Scopus, IEEE, and other databases covering facial emotion recognition, sentiment analysis, and machine learning. (Email: ksarvakar@gmail.com)

Dr. Kaushik K. Rana is an Associate Professor in the Department of Computer Engineering, L. D. College of Engineering, Ahmedabad, Gujarat Technological University, Ahmedabad. He received his Ph.D. in Computer Engineering from Gujarat Technological University, Ahmedabad, with research focused on SOA-based Software Slicing and Testing. His areas of interest include Software Engineering, Service-Oriented Architecture, Software Testing, and Emerging Computing Technologies. He has published research papers in reputed journals and conferences and actively guides students in advanced computing research. (Email: kkr@vgecg.ac.in)