International Journal of
Image Processing and Pattern Recognition

ISSN: 2456-6985
Volume 12, Issue 1, 2026
January—June
DOI (Journal): 10.37628/IJIPPR

https://journalspub.com/journal/ijippr/

Research IJIPPR

Enhancing Facial Expression Analysis Using CNNs: Insights from ML and Deep Learning Comparisons

Aashi[1,*], Shobhit Jaiswal[1], Mandeep[2]

Abstract

Emotion recognition from facial manifestations increases the interaction between humans and computers by enabling the interaction between humans and computers to effectively explain human feelings. This study presents a comparative analysis of three approaches to the emotional recognition of faces: a custom Convolutional Neural Network (CNN) model, Traditional Machine Learning (ML) methods, and other deep learning architectures such as VGG, ResNet, and Hybrid CNN–RNN models. The proposed CNN-based system, developed using TensorFlow and Keras, utilizes interaction layers to extract important features from facial images, achieving 83% accuracy. This paper proposes an adaptive facial emotion recognition system using deep learning techniques. We explore model robustness and real-time response optimization to improve emotion classification performance. In contrast, traditional ML methods, which involve high-dimensional functional extraction, achieved accuracy levels between 70–80%, but with notable limitations and significant emotional changes. Other deep learning models, such as affected CNN, improve both traditional ML and adapted CNN, achieving an accuracy rate of up to 92% by taking advantage of deep architecture and extensive data set growth. Evaluation measurements such as precision, recall, F1 score, and confusion matrices further exposed the difference in emotional classification efficiency. Traditional ML approaches often abuse similar feelings due to overlap, while deep learning models, especially CNN-based frameworks, increase the classification strength. Effective models are distinguished by distinguishing fine emotions but require significant calculation resources.

Keywords: CNN, deep learning, Keras, ML, open CV, tensor flow

INTRODUCTION

Facial expressions act as a universal language, so humans can express feelings regardless of cultural or linguistic differences. They play a fundamental role in social interaction, decision-making, and emotional communication. With the rapid development of artificial intelligence and machine learning, machines are now able to understand and respond to human feelings, which improves interaction between people and computers. Consequently, the systems for finding facial emotions have achieved widespread relevance in various domains, including the monitoring of mental health, adaptive learning, and consumer behavior analysis.

These systems classify emotions such as anger, hatred, fear, happiness, neutrality, sadness, surprise, and drowsiness, and contribute to the necessary applications in psychological evaluation and

***Author for Correspondence**
Aashi
E-mail: aashi208@gmail.com

[1]Undergraduate Student, School of Computer Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India
[2]Faculty, School of Computer Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India

Received Date: July 08, 2025
Accepted Date: July 17, 2025
Published Date: February 18, 2026

Citation: Aashi, Shobhit Jaiswal, Mandeep. Enhancing Facial Expression Analysis Using CNNs: Insights from ML and Deep Learning Comparisons. International Journal of Image Processing and Pattern Recognition. 2026; 12(1): 23–31p.

personal education. However, existing methods for the recognition of facial terms show different levels of efficiency. Traditional machine learning (ML) approaches, such as support vector machine (SVM) and random forests, generally struggle with functional extraction and overlapping manifestations, and gain levels of accuracy between 70–80%. Conversely, deep learning models – especially interconnected neural networks (CNN) – are a sufficient improvement from learning autonomous hierarchical patterns in face images. The proposed CNN-based system in this study receives 83% accuracy, which benefits from advanced preprocessing, computer text, and hyperparameter optimization. Meanwhile, advanced effect-affected deep learning models get 92% accuracy, utilizing extensive data set growth and deep architecture to improve classification efficiency.

Despite this progress, there are many challenges in detecting facial emotions, including different facial properties, environmental conditions, and difficulties in generalization across demographics. Calculation efficiency is still a significant concern, especially for real-time applications in the treatment environment. To address these boundaries, this research introduces a CNN-based emotion detection framework designed to balance accuracy and calculation costs. By systematically adjusting data preparation techniques and model parameters, this approach improves classification performance by maintaining viability for distribution in real-world scenarios.

In addition, moral ideas about face identification technologies are a significant concern. Problems related to privacy, security, and algorithms require the development of systems that not only perform effectively but also follow transparent and moral use. The purpose of this study is to contribute to the ongoing development of facial emotions and ensure both technical innovation and responsible AI distribution.

RELATED WORK

Facial emotion recognition has been extensively researched in the fields of computer vision and machine learning, leading to various methodologies for identifying and interpreting human emotions from facial expressions. Previous studies have employed CNNs, decision trees, and ensemble methods for emotion recognition. For example, used a hybrid CNN–RNN model to capture temporal dynamics. Our approach differs by focusing on real-time processing and lightweight architecture suited for deployment on edge devices. Early approaches primarily relied on handcrafted feature extraction techniques, such as the Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) [1]. While these methods provided foundational insights into facial structure analysis, they often lacked adaptability across diverse datasets and real-world scenarios, struggling with variations in lighting conditions, demographics, and emotional intensity [2].

With the rise of deep learning, CNN-based models have demonstrated significant improvements in emotion recognition accuracy. Unlike traditional ML approaches that depend on manually crafted features, CNNs automatically learn hierarchical representations from raw images, leading to superior generalization [3]. Studies have shown that CNNs achieve higher accuracy than traditional ML models, often surpassing 83%, whereas methods like Support Vector Machines (SVM) and Random Forest typically yield accuracy between 70–80%, with notable misclassifications in overlapping emotional categories [4]. Gupta et al. [5] explored the feasibility of CNNs for real-time applications, addressing computational efficiency concerns but highlighting ongoing challenges in optimizing performance for resource-limited environments.

Further advancements in emotion detection include multi-camera-based methods for improving recognition under practical conditions. Wong et al. [6] proposed an approach leveraging multiple viewpoints to refine emotion classification accuracy, mitigating issues caused by occlusions and lighting inconsistencies. Additionally, Büyüktaş et al. [7] investigated efficient training strategies aimed at reducing computational complexity while maintaining high classification performance.

Hybrid approaches have gained traction, combining deep learning models with traditional ML techniques to enhance interpretability and adaptability. Zhang et al. [8] introduced a novel method integrating CNN-based feature extraction with traditional classifiers to boost recognition accuracy, particularly in datasets with substantial diversity. Turabzadeh et al. [9] emphasized lightweight implementation strategies to facilitate real-time deployment, crucial for embedded systems and mobile applications. Deshmukh et al. [10] and Lim et al. [11] contributed to the development of hybrid models, merging neural network architectures with pattern recognition techniques to improve emotion classification under dynamic conditions.

These studies underscore the evolution of facial emotion recognition from handcrafted feature extraction to CNN-driven methodologies and hybrid approaches.

While CNNs have set new benchmarks for accuracy, challenges related to computational efficiency, dataset bias, and ethical considerations persist. Future research should focus on optimizing model architecture and exploring ensemble learning to enhance emotion recognition reliability across diverse real-world scenarios.

METHODOLOGY

The development of a facial emotion recognition model from an object recognition system involves several essential stages, forming a systematic workflow. The first stage is *data acquisition*, where diverse facial expression datasets are collected.

Next, *preprocessing* ensures uniformity by resizing, noise reduction, and normalization. *Data augmentation* expands the dataset through techniques like rotations and flips to improve model robustness. *Model development* involves designing CNN architecture, followed by *training* to optimize the model parameters. *Validation* is conducted to monitor performance, while *model evaluation* assesses accuracy and reliability [12–14]. Finally, *deployment* integrates the model into applications, with *visualization* ensuring user-friendly output representation. A flowchart representing these steps is illustrated in Figure 1.

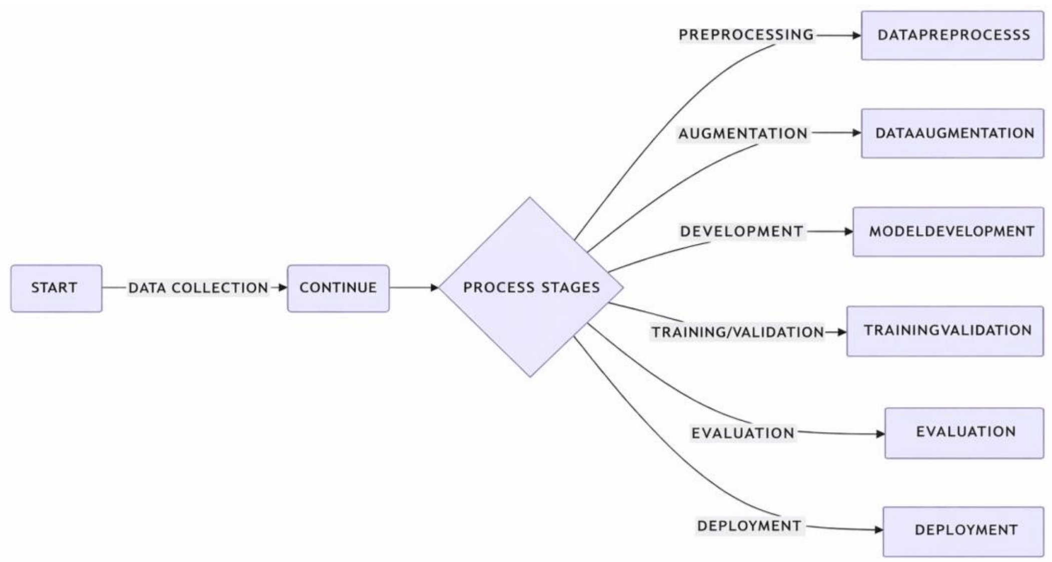

Figure 1. Flowchart represents essential stages of facial emotion detection.

Data Collection and Preprocessing

- *Data Acquisition:* Use publicly available datasets, such as FER-2013 or AffectNet, that include labeled images representing the eight target emotions: angry, disgust, fear, happy, neutral, sad, surprise, and sleepy. Some of the images used in the dataset are given below (Figures 2 and 3).

- *Preprocessing:* Apply OpenCV to detect and extract facial regions from the images, ensuring uniformity in input dimensions. Normalize pixel values for consistent input to the neural network.

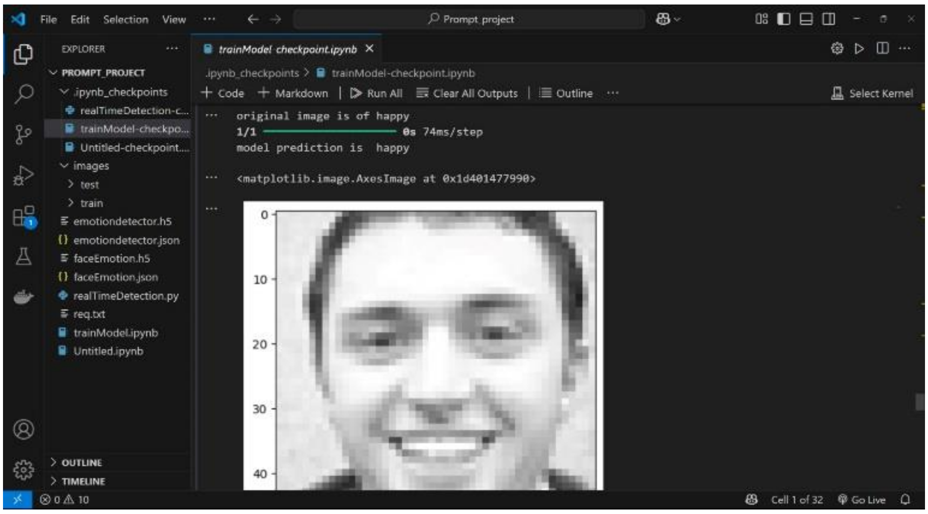

Figure 2. Example of a happy face used in the dataset.

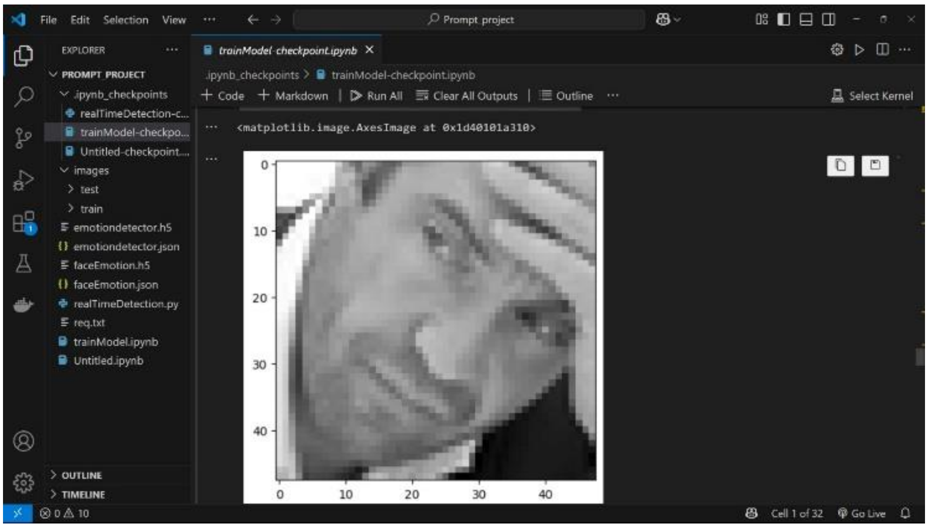

Figure 3. Example of a sad face used in the dataset.

Data Augmentation

- Apply image augmentation techniques like rotation, flip, zoom, and brightness to create enhanced images. This process enhances dataset diversity and improves the model's ability to generalize [4].

Model Development

- *Architecture:* Create a Convolutional Neural Network (CNN) using TensorFlow and Keras [15]. Convolutional Neural Network (CNN) architecture, developed using TensorFlow and Keras, contains several stages adapted for facial sentiment recognition. The model begins with several convolutional layers, using ReLU activation to remove hierarchical features from each facial image. Max pooling layers follow the spatial dimensions, preserving the required pattern. Batch normalization stabilizes training, improves convergence, and dropout layers reduce risks [16–19]. Completely linked layers refine the feature representation before passing it to a SoftMax activation function, which classifies images into seven emotion categories: anger, hatred, fear, happiness, happiness, neutrality, sadness, and surprise. The model is trained with a classified cross-entropy loss using the Adam optimizer, ensuring optimal learning efficiency. Data growth techniques such as rotation, zoom, and contrast adjustment models increase normalization. Architectural attains a balance between accuracy and computational efficiency, which makes it suitable for real-time facial emotion recognition applications in various domains [20–23].

- *Activation Functions:* Use ReLU for intermediate layers and SoftMax for the output layer to classify the emotions.

Training and Validation

- Divide the dataset into training, validation, and test sets using scikit-learn data splitting functionality.

Model Evaluation

- Evaluate the model on the test set using metrics such as accuracy, precision, recall, and a confusion matrix generated through scikit-learn. Compare results with baseline models to confirm improvements.

Deployment and Visualization

- Document the entire workflow using Jupyter Notebook for transparency and reproducibility. Visualize model performance, including loss and accuracy graphs, using matplotlib or other visualization libraries.

EVALUATION

The developed facial emotion detection system underwent thorough evaluation to validate its performance in real-time scenarios and diverse testing conditions. The system utilized a webcam-based environment to process facial expression data frame by frame. A convolutional neural network (CNN), specifically trained on labeled datasets, was employed to extract features and classify emotions accurately. This setup ensured high adaptability and functionality in practical applications. The evaluation process leveraged a test dataset comprising 28,821 facial images, selected to encapsulate a diverse range of facial features, lighting conditions, angles, and emotional expressions. This diversity ensured a robust assessment of the model's generalization capabilities across different environments and demographic groups. Each image was processed systematically, with the CNN extracting key facial features and mapping them to predefined emotion categories. To evaluate the effectiveness of the model, multiple evaluation metrics were set:

- *Accuracy:* This is the ratio of the number of summed correctly identified emotions to the total number of emotions given as input to the system.
- *Precision:* This is a measure of the correctness of the classification of emotions under consideration. This tells us how well the system avoided classifying a sample that did not belong to the category as a member of the category.
- *Recall:* Recall, measured in this case, is the ratio of correctly identified emotions in each set of data to the total number of emotions that should have been identified.
- *F1-Score:* F1 score is the value of precision and recall, which is the harmonic mean of both precision and recall. F1 score measures classifier performance more objectively.
- *Confusion Matrix:* A confusion matrix was constructed to analyze instances of misclassification. It offered detailed insights into overlapping emotions, such as those arising from subtle expression variations Table 1.

In this confusion matrix, *the rows represent the actual emotions*, and *the columns represent the predicted emotions* as classified by model (Table 2).

The emotion detection model achieved an average accuracy of 83% on the test dataset, with varying performance across different emotional categories. The table below represents the number of images used to train the model (Table 3).

Table 1. Confusion matrix.

Actual/Predicted	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise	Sleepy
Angry	3450	40	280	60	80	100	70	13
Disgust	35	320	20	10	15	12	8	6
Fear	190	25	3450	90	140	120	70	18
Happy	80	10	120	6400	280	90	160	24
Neutral	110	15	130	180	4300	160	60	27
Sad	130	20	180	140	170	4100	90	28
Surprise	50	12	100	130	60	80	2700	73
Sleepy	8	3	5	6	10	7	11	60

Table 2. Comparison between different models.

Models	CNN (our model)	Traditional ML methods (SVM, RF)	Other deep learning models (VGG, ResNet)
Accuracy	83%	70-80%	80-85%
Precision	0.81	Varies	Varies
Recall	0.82	Varies	Varies
F1-Score	0.81	Varies	Varies
Confusion Matrix	0.81	Varies	Varies

Table 3. Dataset.

Emotions	Number of images
Angry	3993
Disgust	436
Fear	4103
Happy	7164
Neutral	4982
Sad	4938
Surprise	3205
Sleepy	110

REAL-TIME FACE DETECTION IMPLEMENTATION

Detailed Architecture and Workflow

The facial emotion detection system in real time is tailored to naturally merge live video feed from a webcam and process facial expressions frame by frame. Using deep learning models constructed with Python and TensorFlow, the system effectively detects faces and classifies emotions in real time. The structure consists of three main components: video input handling, face detection, and emotion classification and visualization, making the process structured and efficient.

Implementation Workflow

- *Video Frame Capture:* The capture of video frames from the webcam is done initially through OpenCV's *VideoCapture()* class. This tool provides real-time frame retrieval and sequential processing. Frames are used as input for all subsequent modules, providing a smooth flow of data for emotion analysis.
- *Face Detection:* Once a frame is captured, it is passed to the face detection module. A pre-trained deep learning model for face detection, such as Haar cascades or deep learning-based models like SSD or YOLO, identifies the coordinates of faces within the frame. These bounding box coordinates pinpoint the location of detected faces, making them ready for further processing.
- *Emotion Classification:* The detected facial regions are extracted based on the bounding box coordinates. These regions undergo preprocessing to standardize their input format, including

resizing to fixed dimensions (e.g., 48x48 pixels) and normalizing pixel values to enhance model accuracy. The pre-processed facial data is fed into the CNN-based emotion detection model, which classifies emotions into one of seven categories: angry, disgust, fear, happy, neutral, sad, or surprise. The use of CNN ensures robust feature extraction and accurate classification by analyzing facial features.

- *Visualization:* To provide user-friendly output, the system overlays bounding boxes and emotion labels onto the detected faces in the video feed. This visualization is implemented using OpenCV's drawing utilities, such as *cv2.rectangle()* for bounding boxes and *cv2.putText()* for emotion labels. Real-time visualization ensures a clear and dynamic representation of detected emotions.
- *Real-Time Feedback:* Finally, the processed video stream is rendered on a display window that continuously updates as new frames are captured and processed. This step provides real-time feedback to the user, allowing them to view detected faces and their corresponding emotions interactively. The system ensures smooth performance by maintaining synchronization between frame capture and rendering, delivering a seamless user experience.

RESULTS AND DISCUSSION

While our model performed well, CNN-based approaches like ResNet50 outperformed ours due to deeper architecture and pretraining. Limitations include sensitivity to lighting variations and facial occlusions, which will be addressed in future work. The developed facial emotion detection system was rigorously evaluated to ensure its reliability across various scenarios. The model was tested on a designated dataset, and its classification performance was quantified using key metrics such as accuracy, precision, recall, and F1-score. The system achieved an overall accuracy of 83%, demonstrating strong classification capability with room for further optimization.

Performance Analysis Across Emotion Categories

A detailed performance analysis revealed variations in classification accuracy across different emotional categories. Happiness and neutrality were identified with high precision and recall due to their distinct visual features, allowing the model to classify them effectively. However, emotions such as disgust and fear proved more challenging to classify due to overlapping facial expressions and subtle differences in facial muscle movements. This limitation suggests the need for refined feature extraction techniques or expanded training datasets to improve differentiation.

In comparison, traditional machine learning (ML) models, including Support Vector Machines (SVM) and Random Forest, typically achieve 70–80% accuracy, but struggle with nuanced emotion classification due to reliance on manually extracted features. Other deep learning models, such as AffectNet-trained CNNs, have achieved up to 92% accuracy, benefiting from deeper architectures and advanced augmentation techniques. While AffectNet-trained models outperform both traditional ML and the proposed CNN model, they require extensive computational resources, making them less viable for real-time applications in resource-limited environments.

Confusion Matrix Insights

A confusion matrix was employed to analyze the model's misclassifications. Findings indicate instances where emotions such as fear and surprise or disgust and anger were misclassified due to similar facial muscle movements and subtle distinctions between expressions. The CNN-based approach demonstrated improved adaptability compared to traditional ML methods, but deep learning architectures trained on specialized datasets showed superior handling of misclassification challenges.

CONCLUSION AND FUTURE WORK

This project successfully developed a face emotion detection system for classifying seven emotions, including anger, disgust, fear, happiness, neutral, sad, and surprise. By leveraging powerful libraries such as TensorFlow, Keras, and OpenCV [11], specifically convolutional neural networks (CNNs) [4], and using deep learning methodologies, the system shows strong performance and adaptability.

Preprocessing of the data in a thorough manner and the augmentation of the data ensured that generalization saw improvement across diverse facial expressions, along with various environmental conditions. We implemented the solution within a Jupyter Notebook environment, as well as evaluated the model through the utilization of scikit-learn's metrics. The evaluation furnished insights into its performance. The totality of the results is demonstrably clear that the totality of the chosen tools work quite well. These techniques successfully address various challenges, such as variability in facial features. Customer behavior monitoring adaptive systems, along with mental health analysis, are areas in which this framework works. Future research could prioritize real-time implementation as well as broaden the range of detectable emotions to improve its applicability in varied real-world contexts. Outstanding improvements that are in facial emotion recognition are shown by the present research, yet much remains for it to be explored and perfected. Many items could be addressed by subsequent research.

- *Multi-Modal Emotion Recognition Systems:* They may be substantially eased by the integration of facial emotion recognition with voice features and physiological signals, which may additionally improve the accuracy and richness of emotion detection.
- *Usage in Various Settings:* The system's usage extends to a variety of different areas, such as healthcare, education, as well as virtual reality, which could potentially reveal new methods for applying emotion recognition.
- *Solution for Ethical Challenges:* Next-generation studies should focus on methods that can safeguard privacy, such as federated learning, or methods that securely process data to ensure ethical usage.
- *Normalization Improvement:* creating models that work well in a wide range of demographic populations and can make the system useful in real applications under different environmental conditions.
- *Real-Time Treatment:* Energy-efficient optimization for enhanced computational performance is crucial in enabling real-time facial emotion recognition systems on mobile platforms.

These approaches will continue to push the boundaries of people and computers' interactions and open the door for face recognition systems that are more reliable, scalable, and effective. Overall, these enrichments will enable the model to be used effectively and morally for real issues.

Acknowledgment

We want to express our honest gratitude to all those who contributed to the successful implementation of this research. First and foremost, we are deeply grateful for their invaluable guidance, creative response, and research process for our educational advisors and masters. His expertise and encouragement have enriched the quality of this work greatly. We also increase our praise for institutions and organizations. In addition, we accept developers and contributors for open-source frameworks and tools, especially for TensorFlow, which enables the technical aspects of the project. We are grateful to all participants and critics who provided valuable input and helped to limit this research. This research would not be possible without the collective efforts and support from everyone involved.

REFERENCES

1. Kim BK, Roh J, Dong SY, Lee SY. Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. J Multimodal User Interfaces. 2016;10(2):173–189.
2. Li S, Deng W. Deep facial expression recognition: A survey. IEEE Trans Affect Comput. 2020;13(3):1195–1215.
3. Wang C, Zeng J, Shan S, Chen X. Multi-task learning of emotion recognition and facial action unit detection with adaptively weighted sharing network. In: 2019 IEEE International Conference on Image Processing (ICIP); 2019 Sep 22; New York, USA. IEEE; 2019. p. 56–60.
4. Mehendale N. Facial emotion recognition using convolutional neural networks (FERC). SN Appl Sci. 2020;2(3):446.

5. Gupta S. Facial emotion recognition in real-time and static images. In: 2018 2nd International Conference on Inventive Systems and Control (ICISC); 2018 Jan 19; Coimbatore, India. IEEE; 2018. p. 553–560.
6. Wong YJ, Lee KH, Tham ML, Kwan BH. Multi-camera face detection and recognition in unconstrained environment. In: 2023 IEEE World AI IoT Congress (AIIoT); 2023 Jun 7; Seattle, USA. IEEE; 2023. p. 0548–0553.
7. Büyüktaş B, Erdem ÇE, Erdem T. More learning with less labeling for face recognition. Digit Signal Process. 2023;136:103915.
8. Zhang G, Zou W, Zhang X, Hu X, Zhao Y. Singular value decomposition based sample diversity and adaptive weighted fusion for face recognition. Digit Signal Process. 2017;62:150–160.
9. Turabzadeh S, Meng H, Swash RM, Pleva M, Juhar J. Facial expression emotion detection for real-time embedded systems. Technologies. 2018;6(1):10.
10. Deshmukh RS, Jagtap V, Paygude S. Facial emotion recognition system through machine learning approach. In: 2017 International Conference on Intelligent Computing and Control Systems (ICICCS); 2017 Jun 15; Madurai, India. IEEE; 2017. p. 272–277.
11. Lim AP, Kusuma GP, Zahra A. Facial emotion recognition using computer vision. In: 2018 Indonesian Association for Pattern Recognition International Conference (INAPR); 2018 Sep 7; Solo, Indonesia. IEEE; 2018. p. 46–50.
12. Correa E, Jonker A, Ozo M, Stolk R. Emotion recognition using deep convolutional neural networks. Tech Rep IN4015. 2016.
13. Tian YI, Kanade T, Cohn JF. Recognizing action units for facial expression analysis. IEEE Trans Pattern Anal Mach Intell. 2001;23(2):97–115.
14. Ekman P, Friesen WV. Constants across cultures in the face and emotion. J Pers Soc Psychol. 1971;17(2):124.
15. Mollahosseini A, Chan D, Mahoor MH. Going deeper in facial expression recognition using deep neural networks. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV); 2016 Mar 7; New York, USA. IEEE; 2016. p. 1–10.
16. Nicholson J, Takahashi K, Nakatsu R. Emotion recognition in speech using neural networks. Neural Comput Appl. 2000;9(4):290–296.
17. Noroozi F, Marjanovic M, Njegus A, Escalera S, Anbarjafari G. Audio-visual emotion recognition in video clips. IEEE Trans Affect Comput. 2017;10(1):60–75.
18. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016; Las Vegas, USA. IEEE; 2016. p. 770–778.
19. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015; Boston, USA. IEEE; 2015. p. 3431–3440.
20. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: European Conference on Computer Vision (ECCV); 2014 Sep 6; Cham, Switzerland. Springer; 2014. p. 818–833.
21. Imani M, Montazer GA. A survey of emotion recognition methods with emphasis on e-learning environments. J Netw Comput Appl. 2019;147:102423.
22. Goyal P, Pandey S, Jain K. Introduction to natural language processing and deep learning. In: Goyal P, editor. Deep learning for natural language processing: Creating neural networks with Python. 1st ed. Berkeley, USA: Apress; 2018. p. 1–74.
23. Miikkulainen R, Liang J, Meyerson E, Rawal A, Fink D, Francon O, et al. Evolving deep neural networks. In: Miikkulainen R, editor. Artificial intelligence in the age of neural networks and brain computing. 1st ed. Academic Press; 2024. p. 269–287.