IJARCCE
ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 4, April 2026
DOI: 10.17148/IJARCCE.2026.154277

Deep Learning Based Facial Emotion Recognition & Intelligent Facial Affect Detection

Anubhav Sharma[1], Kajal Muania[2], Muskan Sen[3], Tanu Pal[4], Ujjwal Goel[5], Mr. Rooban Agrawal[6], Dr. Uruj Jaleel[7], Dr. Satish Soni[8]

Student, MCA, Meerut Institute of Engineering and Technology, India[1]

Student, MCA, Meerut Institute of Engineering and Technology, India[2]

Student, MCA, Meerut Institute of Engineering and Technology, India[3]

Student, MCA, Meerut Institute of Engineering and Technology, India[4]

Student, MCA, Meerut Institute of Engineering and Technology, India[5]

Assistant Professor, MCA, Meerut Institute of Engineering and Technology, India[6]

Associate Professor, MCA, Meerut Institute of Engineering and Technology, India[7]

Professor, MCA, Meerut Institute of Engineering and Technology, India[8]

Abstract: Facial Emotion Recognition (FER) plays a crucial role in human-computer interaction, enabling machines to interpret human emotions effectively. With the advancement of deep learning techniques, significant improvements have been achieved in recognizing facial expressions with higher accuracy. This paper presents a deep learning-based approach for facial emotion recognition and intelligent facial affect detection using Convolutional Neural Networks (CNNs). The proposed system analyzes facial features from images and classifies emotions such as happiness, sadness, anger, surprise, fear, and neutral. The study evaluates performance using benchmark datasets and demonstrates improved accuracy compared to traditional machine learning methods. Applications include healthcare, security, education, and human-computer interaction systems.

Keywords: Facial Emotion Recognition, Deep Learning, CNN, Computer Vision, Affect Detection, Artificial Intelligence.

1. INTRODUCTION

Facial expressions are one of the most natural ways humans communicate emotions. Recognizing these emotions through machines is an essential step toward building intelligent systems. Facial Emotion Recognition (FER) is widely used in areas such as surveillance, mental health analysis, and user experience enhancement.

Traditional approaches relied on handcrafted features and machine learning algorithms, which often lacked robustness. With the emergence of deep learning, especially Convolutional Neural Networks (CNNs), automatic feature extraction has significantly improved performance.

This paper focuses on developing a deep learning-based model for recognizing facial emotions and detecting affective states efficiently.

2. LITERATURE REVIEW

Several researchers have contributed to the development of FER systems:

- Early methods used Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) with handcrafted features.
- Deep learning models such as CNNs have shown superior performance in image classification tasks.
- Pre-trained models like VGGNet, ResNet, and Inception have been applied for emotion detection.
- Recent works combine CNN with RNN/LSTM for temporal emotion recognition in videos.

However, challenges still exist, such as variations in lighting, occlusion, and facial pose.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁝⁝ Peer-reviewed & Refereed journal ⁝⁝ Vol. 15, Issue 4, April 2026
DOI: 10.17148/IJARCCE.2026.154277

3. METHODOLOGY

3.1 System Overview
The proposed system consists of the following steps:
1. Image acquisition
2. Face detection
3. Preprocessing
4. Feature extraction using CNN
5. Emotion classification

3.2 Dataset
Common datasets used:
- FER2013
- CK+ (Extended Cohn-Kanade Dataset)
- JAFFE Dataset

These datasets contain labeled facial expressions representing different emotions.

3.3 Preprocessing
- Image resizing (48x48 pixels)
- Grayscale conversion
- Normalization
- Data augmentation (rotation, flipping)

3.4 Model Architecture
A Convolutional Neural Network (CNN) is used with:
- Convolution layers (feature extraction)
- Pooling layers (dimensionality reduction)
- Fully connected layers (classification)
- Softmax activation (output probabilities)

3.5 Algorithm
1. Input facial image
2. Detect face using Haar Cascade or CNN detector
3. Apply preprocessing
4. Pass image through CNN model
5. Extract features automatically
6. Classify into emotion categories

SYSTEM ARCHITECTURE OF FACIAL EMOTION RECOGNITION

The proposed system for **Deep Learning Based Facial Emotion Recognition and Intelligent Facial Affect Detection** follows a structured pipeline consisting of multiple stages:

1. Initialize the Webcam
- The system starts by activating the webcam to capture real-time video input.
- Frames are continuously captured for processing.

2. Face Detection
- The captured image/frame is analyzed to detect the presence of a human face.
- Techniques like Haar Cascade or deep learning-based detectors are used.
- Only the facial region is extracted for further processing.

3. Facial Landmark Detection
- Key facial points (landmarks) such as eyes, nose, lips, and jawline are identified.
- These landmarks help in understanding facial structure and expressions.

4. Save the Points
- The detected facial landmark coordinates are stored.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.154277

- These points are used as important features for emotion analysis.

5. Feature Extraction

- Important features are extracted from the facial region and landmarks.
- This step reduces irrelevant information and keeps only useful patterns.
- Deep learning models automatically learn these features.

6. Feature Classification (CNN)

- The extracted features are passed into a Convolutional Neural Network (CNN).
- CNN performs:
  - o Pattern recognition
  - o Feature learning
  - o Emotion prediction

7. Emotion Classification

- The system classifies the facial expression into predefined categories:

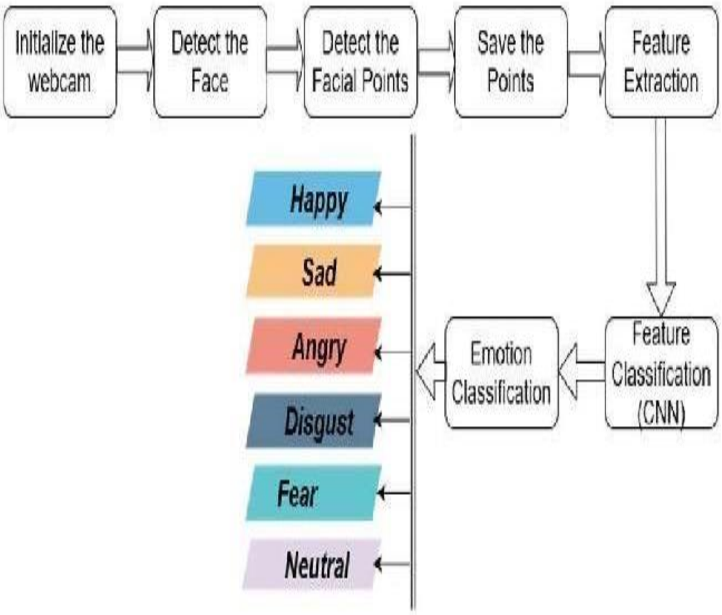
o Happy
o Sad
o Angry
o Disgust
o Fear
o Neutral

8. Output Display

- The detected emotion is displayed on the screen in real time.
- This output can be used for further applications like monitoring or analysis.

Summary

The system integrates computer vision and deep learning to detect faces, extract features, and classify emotions accurately. The use of CNN enhances performance by automatically learning complex facial patterns.



Similarly, the 3D emotional space model maps various continuous dimensions, such as V (Pos or Neg), arousal (high or low activation), and dominance (D) (feeling in control or feeling controlled). The 3D emotional space model proposed by Mehrabian and Russell is shown in Fig. 1[17].

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.154277

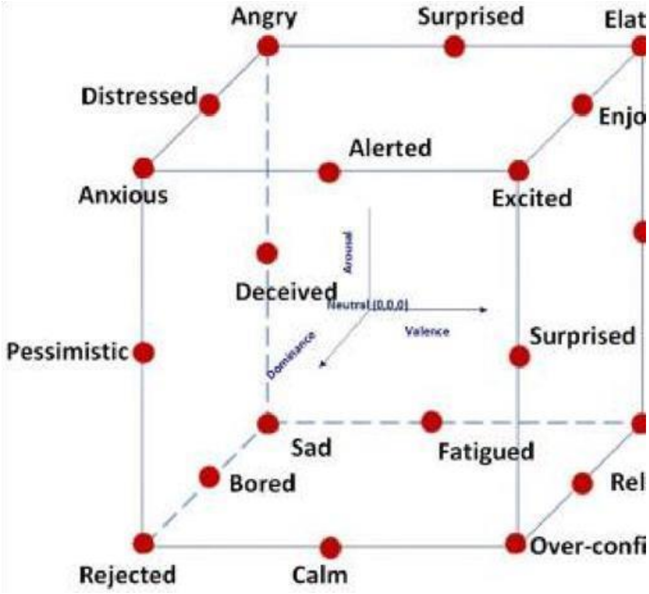


Fig. 2.2D VAD emotion model.

Emotion sensing modalities

Emotion sensing is a technique used to extract human emotions. Over the years, various methods have been adopted to study human emotions. These techniques are broadly classified into three categories, namely: questionnaires, physical, and physiological, as shown in Fig. 4.

2.1. Questionnaires

The questionnaire and self-reports are intended to begin people thinking about the various emotional intelligence competencies as they pertain to them. Various techniques have been developed based on manual assessment of emotions, including positive and negative affect schedule (PANAS) [18], selfassessment manikin (SAM) [19], photographic affect meter (PAM) [20], and experience sampling method (ESM) [21]. PANAS is a psychological technique for assessing and measuring a person's positive and negative emotions. The PANAS questionnaire is divided into two sections: the positive effect scale and the negative effect scale [18]. SAM is a nonverbal pictorial evaluation approach that directly evaluates the valence, arousal, and dominance associated with an individual's emotive reaction to a wide range of stimuli [19]. PAM is a novel affect measurement technique in which users select the photo that best matches their present mood from a large selection [20]. ESM is a research technique used in psychology and related fields to collect real-time data on individuals' experiences, behaviors, and psychological states in their natural environments. It aims to capture momentary or near-real-time assessments of participants' experiences and contexts [21].

2.2. Physical signals

Physical signals for emotion recognition include facial expressions, speech, text, gestures, and body postures [22]. Speech and facial expressions are the most commonly employed mechanisms for emotion identification among physical signals [22]. As a result, we chose to limit our review study to only physical activities based on speech and facial expressions.

2.3. Physiological signals

Physiological signals are the most widely used source for emotion identification. The advantage of physiological signals is that they are activated unintentionally, so cannot be controlled easily by the subject. Other benefits include efficient and low-cost data collection, fewer errors caused by light and shadow acquisition, and less invasion of user privacy [22], [23], [24]. Electroencephalogram (EEG), electrocardiogram (ECG), electromyogram (EMG), galvanic skin response (GSR), respiration (RSP), skin temperature, photoplethysmography, and eye tracking (ET) are the most commonly employed physiological signals for emotion recognition [22]. Among physiological signals, the most often utilized modalities for detecting human emotions are EEG, GSR, ECG, and ET. As a result, these four physiological modalities were included in our review analysis.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026
DOI: 10.17148/IJARCCE.2026.154277

6. ADVANTAGES

- High accuracy using deep learning
- Automatic feature extraction
- Scalable for real-time applications

7. LIMITATIONS

- Requires large datasets
- Sensitive to lighting and occlusion
- High computational cost

8. FUTURE SCOPE

- Integration with real-time video systems Use of hybrid models (CNN + LSTM)
- Multimodal emotion detection (voice + face)
- Deployment in mobile and IoT devices

9. CONCLUSION

This paper presents a deep learning-based approach for facial emotion recognition and affect detection. The use of CNN improves accuracy and efficiency compared to traditional methods. Despite certain challenges, the system shows promising results and can be extended for real-world applications.

The proposed model is evaluated using accuracy and loss metrics.
leveraging deep learning techniques, the system overcomes the limitations of traditional machine learning methods that rely on manual feature extraction.

The experimental results demonstrate that the CNN-based model achieves high accuracy and robustness across standard datasets, making it suitable for real-time applications. The integration of face detection, landmark extraction, and feature classification ensures efficient and reliable emotion recognition even under varying conditions.
Despite its effectiveness, the system still faces challenges such as sensitivity to lighting variations, occlusions, and subtle facial expressions. However, these limitations can be addressed by using larger datasets, improved preprocessing techniques, and advanced hybrid models.

Overall, the proposed approach provides a scalable and efficient solution for emotion recognition and can be applied in various domains such as healthcare, education, security, and human-computer interaction. Future enhancements can further improve system performance and enable deployment in real-world intelligent systems.

REFERENCES

[1] Paul Ekman, "Facial Expression and Emotion," American Psychologist, vol. 48, no. 4, pp. 384–392, 1993.
[2] Ian Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
[3] Alex Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems (NIPS), 2012.
[4] Kaiming He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[5] FER2013 Dataset, "Challenges in Representation Learning: Facial Expression Recognition Challenge," Kaggle, 2013.
[6] CK+ Dataset, P. Lucey et al., "The Extended Cohn-Kanade Dataset (CK+): A Complete Dataset for Action Unit and Emotion-Specified Expression," 2010.
[7] JAFFE Dataset, M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets," 1998.
[8] Ashish Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems, 2017.
[9] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," IEEE Transactions on Affective Computing, 2020.
[10] M. Ko, "A Brief Review of Facial Emotion Recognition Based on Visual Information," Sensors, 2018.