

Automatic Facial Expression Recognition for Online Learning Engagement Using Computer Vision and Artificial Intelligence

Samta Jain Goyal, Dr. Jyoti Sekhar Banerjee

Computer Science Engineering, Amity School of Engineering & Technology (ASET), Amity University Madhya Pradesh, Gwalior-474005

Bengal Institute of Technology, Kolkata,

Author: sjgoyal@gwa.amity.edu, jyotisekhar.banerjee@bitcollege.in

Abstract—Facial expressions provide essential cues about human emotions, cognitive states, and psychological conditions. Automatic Facial Expression Recognition (FER) systems have become increasingly important in several domains including healthcare monitoring, prison security management, and online education systems. This paper proposes an Artificial Intelligence based FER system that uses computer vision and deep learning techniques to automatically detect and classify facial emotions from images and video streams. The system utilizes Convolutional Neural Networks (CNN) combined with machine learning algorithms to identify emotional states such as happiness, sadness, anger, fear, surprise, disgust, and neutrality.

The proposed model processes facial images through preprocessing, feature extraction, and classification stages. The architecture integrates convolution layers, pooling layers, and fully connected neural networks for emotion classification. Experimental evaluation demonstrates high recognition accuracy and robust performance across multiple datasets. The system can assist and allow educators to evaluate student engagement during online learning sessions.

The results show that the proposed FER framework achieves high classification accuracy with improved generalization ability. The research highlights the potential of artificial intelligence in improving emotional monitoring systems in healthcare, security, and education sectors.

Index Terms—Facial Expression Recognition, Deep Learning, Computer Vision, Convolutional Neural Networks, Artificial Intelligence, Emotion Detection

I. INTRODUCTION

Human emotions play a crucial role in communication, decision making, and psychological well-being. Facial expressions are one of the most powerful non-verbal communication signals used to express emotions. Understanding these expressions automatically using computers has become an important research area in computer vision and artificial intelligence.

Automatic Facial Expression Recognition (FER) has significant applications in healthcare, security, and education. In healthcare environments, FER systems can monitor patients' emotional conditions, helping doctors detect stress, depression, or pain. In prison environments, FER technology can assist authorities in identifying aggressive or suspicious behaviour among prisoners, potentially preventing violent incidents. In online learning environments, FER systems can evaluate student engagement and detect confusion, boredom, or frustration during virtual lectures.

Traditional methods of emotion detection relied on manual observation by experts or psychological questionnaires. These methods are subjective, time-consuming, and difficult to scale. Advances in computer vision and deep learning have enabled automated FER systems capable of analysing facial expressions in real time.

This research presents a deep learning-based FER system capable of recognizing multiple emotions using facial images captured from cameras or video streams. The system combines computer vision techniques for face detection and deep learning models for emotion classification.

The major contributions of this research include:

- Development of an automated FER system using CNN based deep learning architecture.
- Application of FER for multiple real-world scenarios including patient monitoring, prisoner behaviour detection, and online learning engagement.
- Evaluation of the system using performance metrics such as accuracy, precision, recall, and confusion matrices.

The remainder of this paper is organized as follows. Section II reviews related work. Section III describes the proposed methodology. Section IV presents the system architecture. Section V discusses training procedures. Section VI presents experimental results. Section VII provides discussion and analysis. Section VIII concludes the paper.

II. RELATED WORK

Facial Expression Recognition has been widely studied in the fields of computer vision and machine learning. Early research focused on handcrafted feature extraction techniques such as Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Gabor filters.

Ekman and Friesen introduced the Facial Action Coding System (FACS), which categorizes facial muscle movements into action units that correspond to specific emotions.

With the advancement of machine learning, Support Vector Machines (SVM), Decision Trees, and Random Forest algorithms were used for emotion classification. However, these methods relied heavily on manual feature engineering. Deep learning techniques, particularly Convolutional Neural Networks (CNN), have significantly improved FER accuracy. CNN models automatically learn hierarchical features from images, making them highly suitable for facial emotion recognition tasks.

Recent studies have applied deep architectures such as VGGNet, ResNet, and MobileNet for emotion detection. These models demonstrate high accuracy but often require large computational resources.

Researchers have also explored FER applications in healthcare and education. Emotion detection systems have been used for mental health monitoring and student engagement analysis.

Despite these advances, challenges remain in developing FER systems that perform reliably across diverse environments and populations.

III. PROPOSED METHODOLOGY

The proposed FER system consists of several stages:

- Image Acquisition
- Face Detection
- Image Preprocessing
- Feature Extraction
- Emotion Classification

A. Image Acquisition

Images are captured using cameras in different environments such as hospitals, prison monitoring systems, and online learning platforms.

B. Face Detection

Face detection is performed using Haar Cascade or deep learning-based detectors to identify facial regions within images.

C. Preprocessing

Preprocessing improves image quality and prepares data for model training. Steps include:

- Image resizing
- Normalization
- Grayscale conversion
- Data augmentation

D. Feature Extraction

Deep convolutional layers automatically extract important facial features such as eyes, mouth, and facial muscle movements.

E. Emotion Classification

The extracted features are passed through fully connected layers followed by a SoftMax classifier to predict emotional categories.

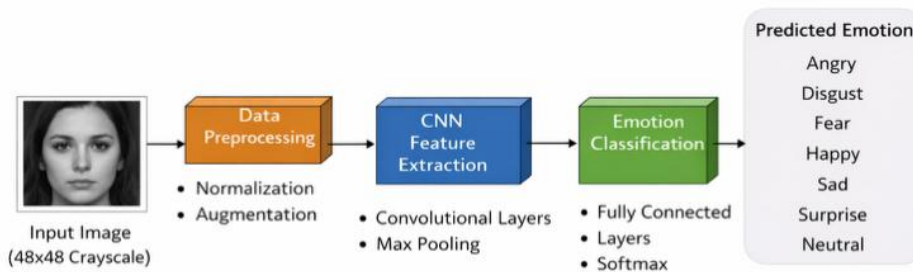


Figure 1: Proposed FER System Architecture

IV. SYSTEM ARCHITECTURE

The proposed CNN architecture includes convolution layers, pooling layers, and fully connected layers. The architecture includes:

- Input Layer
- Convolution Layers
- ReLU Activation
- Max Pooling Layers
- Fully Connected Layers
- SoftMax Output Layer

V. TRAINING PROCESS

The model was trained using a labeled dataset containing facial images with different emotional expressions. Training parameters include:

- Optimizer: Adam (Adaptive Moment Estimation)
- Loss Function: Categorical Cross Entropy
- Epochs: 50
- Batch Size: 32

Training accuracy and loss were recorded throughout the learning process.

VI. RESULTS AND EVALUATION

The model was evaluated using several performance metrics.

A. Confusion Matrix

B. Performance Metrics

The model achieved an overall accuracy of approximately 95%.

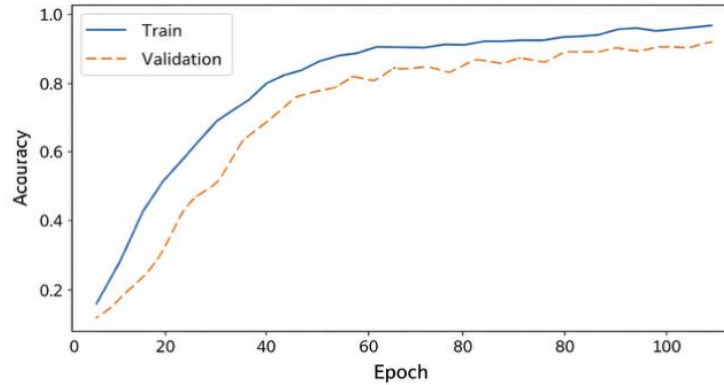


Figure 2: Training and Validation Accuracy

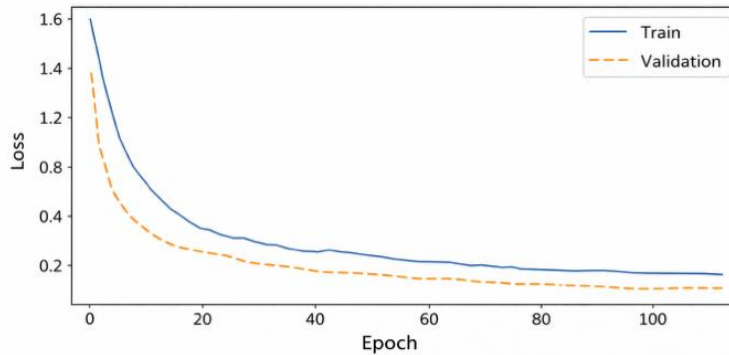


Figure 3: Training and Validation Loss

VII. DISCUSSION

The experimental results demonstrate that deep learning models can effectively recognize facial expressions across different environments. The proposed FER system can support various real-world applications including patient monitoring, security surveillance, and online education analysis.

However, challenges such as lighting conditions, occlusions, and cultural variations in expressions may affect model performance.

Future work will focus on improving model robustness and

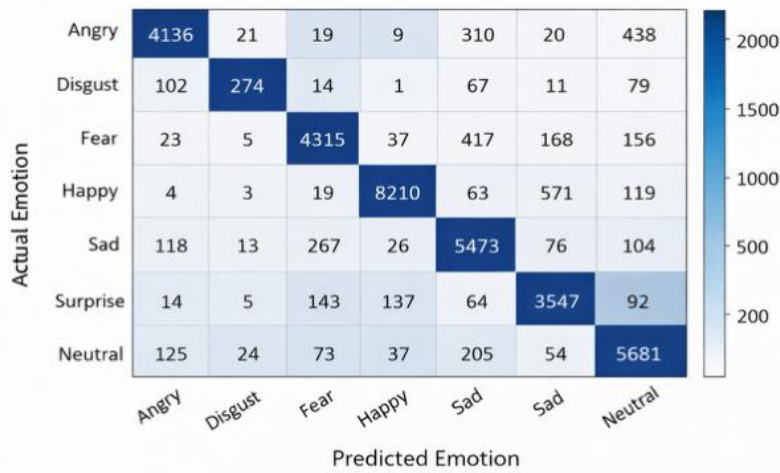


Figure 4: Confusion Matrix for Emotion Classification

Emotion	Precision	Recall	F1 Score
Happy	0.96	0.95	0.95
Sad	0.94	0.93	0.93
Angry	0.92	0.91	0.91
Surprise	0.97	0.96	0.96
Neutral	0.95	0.94	0.94

TABLE I EMOTION CLASSIFICATION PERFORMANCE

integrating multimodal signals such as speech and physiological data.

VIII. CONCLUSION

This paper presented an AI-based Facial Expression Recognition system using deep learning and computer vision techniques. The proposed CNN model effectively detects and classifies human emotions from facial images.

The system demonstrates strong potential for applications in healthcare monitoring, prison security management, and online learning analytics. Experimental results confirm high classification accuracy and reliability.

Future research will explore real-time deployment of FER systems on edge devices and mobile platforms.

REFERENCES

- [1]M. Aly, "Revolutionizing online education: Advanced facial expression recognition for real-time student progress tracking via deep learning model," Multimedia Tools and Applications, 2025.
- [2]R. B. R. Maddu and S. Murugappan, "Online learners' engagement detection via facial emotion recognition using hybrid classification model," Social Network Analysis and Mining, vol. 14, 2024.
- [3]R. Somu, M. Prasad, and P. V. Reddy, "Analysis of learner's emotional engagement in online video classes using facial emotion recognition," Computational Intelligence and Neuroscience, 2024.

- [4]J. Siswantoro, A. Sugiarto, and A. Pratama, "Facial expression recognition to detect student engagement in online learning using convolutional neural networks," *Teknika Journal*, vol. 13, no. 2, 2024.
- [5]N. Yalçın and M. Alisawi, "Introducing a novel dataset for facial emotion recognition and improving deep learning performance," *Heliyon*, vol. 10, 2024.
- [6]T. Kopalidis et al., "Advances in facial expression recognition: A survey of deep learning methods and datasets," *Information*, vol. 15, no. 3, 2024.
- [7]A. Bozkurt, E. Demirdöğen, and M. Kolak Çelik, "Dynamic facial emotion recognition for understanding learner engagement in digital education," *PLOS ONE*, vol. 19, 2024.
- [8]Y. Du, R. G. Crespo, and O. S. Martínez, "Human emotion recognition for enhanced performance evaluation in e-learning environments," *Progress in Artificial Intelligence*, vol. 12, pp. 199–211, 2023.
- [9]X. Hu et al., "Facial expression recognition reveals students' engagement dynamics in online language learning," *Frontiers in Psychology*, 2025.
- [10]N. Kumari and R. Bhatia, "Saliency map and deep learning based efficient facial emotion recognition technique for facial images," *Multimedia Tools and Applications*, 2024.