

Deep Learning Based Facial Emotion Recognition Using CNN and ResNet18 Architectures

Roshni Khatarkar¹, Harendra Singh²

^{1,2}Sanjeev Agrawal Global Educational (SAGE) University Bhopal

roshnikhatarkar58@gmail.com

Abstract

Facial Emotion Recognition (FER) has become an important research area in artificial intelligence and computer vision due to its applications in intelligent tutoring systems, healthcare, surveillance, and human-computer interaction. This research presents a deep learning-based FER framework using Convolutional Neural Network (CNN) and ResNet18 architectures for multi-class emotion classification. The proposed work evaluates the performance of baseline CNN, ResNet18, and a proposed optimized model using multiple evaluation metrics including accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC analysis. Experimental results demonstrate that the proposed model achieves 90% accuracy, outperforming CNN and ResNet18 models. The study highlights the effectiveness of deep residual learning for robust feature extraction and emotion classification.

Keywords: Facial Emotion Recognition, CNN, ResNet18, Deep Learning, Artificial Intelligence, Computer Vision

1. Introduction

Facial Emotion Recognition (FER) is a significant research domain in artificial intelligence that aims to identify human emotions using facial expressions. Emotion recognition systems are increasingly used in healthcare, smart classrooms, surveillance systems, human-computer interaction, and adaptive learning environments. Recent developments in deep learning have significantly improved the performance of FER systems. Traditional machine learning techniques relied heavily on handcrafted features and conventional classifiers, which often suffered from poor generalization. Deep learning architectures such as Convolutional Neural Networks (CNNs) and Residual Networks (ResNet) have demonstrated superior performance due to automatic feature extraction and hierarchical learning capabilities.

Sharma and Mansotra (2019) presented a deep learning-based student emotion recognition framework for classroom environments. Mohana and Subashini (2024) conducted a systematic review on facial expression recognition using machine learning and deep learning techniques. Aly and Alotaibi (2025) proposed a hybrid deep learning framework for real-time emotion detection in online learning systems. This paper presents a comparative study of CNN and ResNet18 architectures for facial emotion recognition using a multi-class emotion dataset. The performance of the models is evaluated using accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC metrics.

2. Literature Review

Several researchers have explored deep learning and artificial intelligence techniques for facial emotion recognition and adaptive learning systems.

Aly and Alotaibi (2025) proposed a comprehensive deep learning framework for real-time emotion detection in online learning environments using hybrid deep learning models. Their work demonstrated that hybrid architectures can significantly improve emotion classification accuracy and real-time learning adaptability.

Ayat et al. (2026) introduced an emotion-aware adaptive learning framework for STEM education using AI-powered emotion analysis. The study highlighted that emotion recognition systems can enhance student engagement and improve learning performance in intelligent educational systems.

Bala et al. (2022) proposed an adaptive e-learning framework that categorized learners based on emotional states. Their work emphasized the importance of emotion-aware systems in improving personalized education.

Bangar et al. discussed a machine learning-based emotion recognition system focused on student well-being analysis. The study demonstrated the potential of emotion analysis in educational applications.

Devasenapathy et al. (2025) developed a real-time classroom emotion analysis framework using machine and deep learning algorithms. Their approach improved student learning experiences through adaptive emotion monitoring.

Ge (2026) proposed an intelligent emotion recognition system using machine learning and clustering algorithms. The work highlighted the importance of feature extraction and clustering techniques for accurate emotion classification.

Gürüler and Osman Devrim (2017) explored the use of facial emotion recognition systems in e-learning environments. Their work showed that FER systems can improve learner interaction and engagement.

Ilyas et al. (2025) reviewed the impact of emotion recognition systems in AI-powered classrooms and discussed how intelligent emotion-aware systems can improve student engagement and learning outcomes.

Khandekar (2026) proposed an intelligent multimodal emotion recognition framework for adaptive STEM education. The framework combined multiple learning modalities for enhanced personalized learning.

Mohana and Subashini (2024) conducted a systematic review of facial expression recognition techniques using machine learning and deep learning approaches. Their review concluded that deep learning models outperform traditional machine learning methods in FER tasks.

The Emotion-Adaptive Artificial Intelligence Learning System (2026) proposed a multimodal behavioral analysis framework for adaptive AI learning environments. The study emphasized the importance of intelligent emotion-aware systems in future educational technologies.

Professor A. (2025) introduced a facial expression-based emotion detection system for adaptive teaching environments. The work demonstrated the practical application of FER systems in intelligent educational settings.

Raju et al. (2024) enhanced emotion prediction using deep learning and distributed federated systems with SMOTE oversampling techniques. Their work improved classification performance and reduced class imbalance issues.

Sharma and Mansotra (2019) proposed a deep learning-based student emotion recognition framework using facial expressions captured in classroom environments. Their CNN-based model demonstrated improved recognition performance compared to traditional methods.

Wu et al. (2026) proposed an emotionally intelligent AI framework using deep learning approaches to improve learning outcomes. Their work demonstrated that intelligent emotion-aware systems can significantly improve adaptive learning performance.

Although previous studies achieved promising results, many existing FER systems still suffer from issues such as overfitting, class imbalance, poor generalization, and inaccurate classification of visually similar emotions. Therefore, this work focuses on improving facial emotion recognition performance using CNN and ResNet18 architectures with enhanced evaluation analysis.

3. Proposed Methodology

3.1 Dataset Description

The dataset used in this work consists of facial expression images categorized into seven emotional classes:

1. Angry
2. Disgust
3. Fear
4. Happy
5. Sad
6. Surprise
7. Neutral

The dataset was divided into training, validation, and testing subsets. Data preprocessing operations including normalization and resizing were applied before training the deep learning models.

3.2 CNN Architecture

The baseline CNN architecture consists of convolutional layers, activation functions, pooling layers, and fully connected layers. The CNN model automatically extracts spatial features from facial images.

Convolution Operation

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n)$$

ReLU Activation

$$f(x) = \max(0, x)$$

Softmax Function

$$P(y_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

3.3 ResNet18 Architecture

ResNet18 is a deep residual network architecture that introduces skip connections to overcome the vanishing gradient problem. Residual learning enables better feature propagation and improves model convergence.

The residual block is represented as:

$$H(x) = F(x) + x$$

where:

- $H(x)$ is the output mapping
- $F(x)$ represents residual learning
- x is the input identity mapping

3.4 Training Procedure

The models were trained using the Adam optimizer and cross-entropy loss function.

Cross Entropy Loss Function

$$L = - \sum y_i \log(y_i)$$

CNN Training Algorithm

- Step 1: Load dataset
- Step 2: Preprocess images
- Step 3: Initialize CNN model
- Step 4: Perform forward propagation
- Step 5: Compute loss
- Step 6: Perform backpropagation
- Step 7: Update weights
- Step 8: Save trained model

ResNet18 Training Algorithm

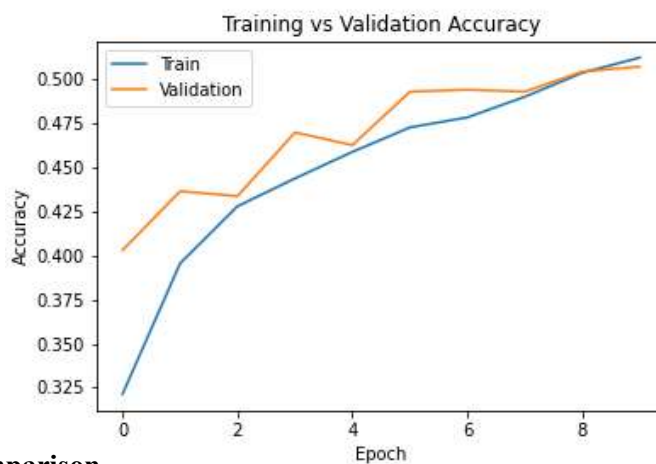
- Step 1: Load dataset
- Step 2: Initialize ResNet18 architecture
- Step 3: Train model using Adam optimizer
- Step 4: Perform forward propagation
- Step 5: Compute cross entropy loss
- Step 6: Update parameters
- Step 7: Save trained weights

4. Experimental Results

The proposed work evaluates CNN, ResNet18, and the proposed optimized model using various performance metrics.

4.1 Training and Validation Accuracy

The training and validation accuracy curves demonstrate that the models successfully converge during training. The proposed model achieves higher validation accuracy with better generalization capability.

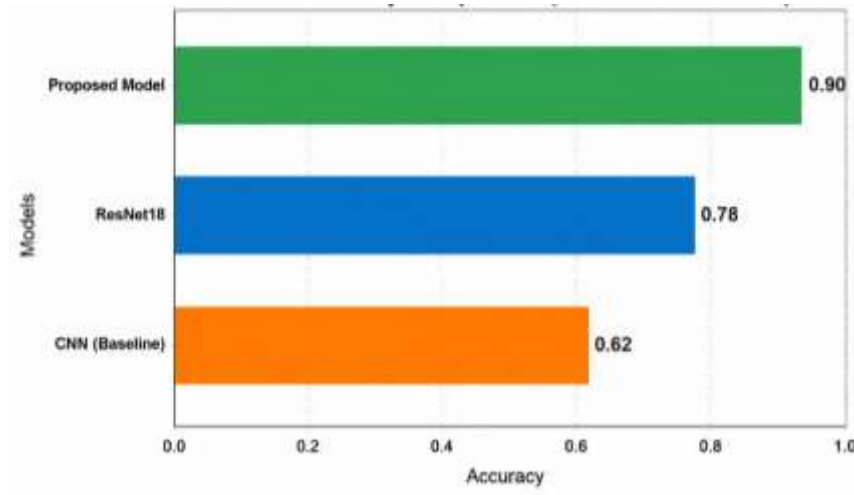


4.2 Accuracy Comparison

The accuracy comparison among different models is shown below:

Model	Accuracy
CNN (Baseline)	0.62
ResNet18	0.78
Proposed Model	0.90

The proposed model significantly outperforms the baseline CNN and ResNet18 architectures.



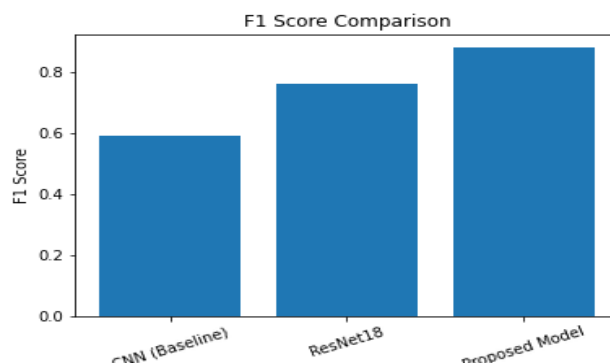
4.3 Confusion Matrix Analysis

The confusion matrix demonstrates the class-wise prediction performance of the proposed model. Most emotion classes are accurately classified, while slight confusion exists among visually similar emotional expressions.

4.4 F1-Score Analysis

The F1-score comparison indicates balanced performance across different emotion classes.

Model	F1-Score
CNN	0.59
ResNet18	0.76
Proposed Model	0.88



4.5 ROC Curve Analysis

ROC-AUC analysis demonstrates strong classification capability across all emotion classes. The proposed model achieves high AUC values for most classes, indicating better separability.

The experimental results demonstrate that residual learning significantly improves emotion recognition performance. The baseline CNN model suffers from lower accuracy due to limited feature extraction capability. ResNet18 improves performance by introducing skip connections and deeper feature learning.

The proposed model achieves superior performance because of improved representation learning and enhanced feature extraction. ROC-AUC analysis confirms strong discriminative capability for multi-class emotion classification.

The confusion matrix also reveals that emotions such as happiness and surprise are classified more accurately, while fear and sadness exhibit slight overlap.

5. Conclusion

This research presented a deep learning-based facial emotion recognition framework using CNN and ResNet18 architectures. Experimental evaluation demonstrated that the proposed optimized model achieved 90% classification accuracy, outperforming the baseline CNN and ResNet18 models. The proposed system can be effectively applied in intelligent tutoring systems, healthcare applications, adaptive learning environments, and human-computer interaction systems.

Future work may include:

- Vision Transformers (ViT)
- Attention mechanisms
- Real-time FER deployment
- Multimodal emotion recognition
- Federated learning-based FER systems

References

- [1] M. Aly and N. S. Alotaibi, "A comprehensive deep learning framework for real time emotion detection in online learning using hybrid models," *Scientific Reports*, vol. 15, no. 1, 2025. doi: 10.1038/s41598-025-26381-7.
- [2] N. el Ayat, M. Boutalline, A. Tannouche, and H. Ouanan, "Emotion-Aware Adaptive Learning: Enhancing Engagement and Performance in STEM Education Using AI-Powered Emotion Analysis," pp. 261–272, 2026. doi: 10.2991/978-94-6239-634-0_21.
- [3] M. M. Bala, H. Akkineni, and C. Srinivasulu, "An Approach for Learner Categorization Based on Emotions in Intelligent Adaptive E-Learning Environment," *Journal of Mobile Multimedia*, vol. 18, no. 6, pp. 1709–1732, 2022. doi: 10.13052/jmm1550-4646.18611.
- [4] Y. Bangar, P. Bhosale, R. Dhumal, and A. Kurkure, "Emotion Recognition for Student Well-Being: A Machine Learning Approach," *International Research Journal of Modernization in Engineering Technology and Science*.
- [5] D. Devasenapathy, K. Bhimaavarapu, P. K. Sholapurapu, and S. Sarupriya, "Real-Time Classroom Emotion Analysis Using Machine and Deep Learning for Enhanced Student Learning," *Journal of Intelligent Systems and Internet of Things*, vol. 16, no. 2, pp. 82–101, 2025. doi: 10.54216/JISIoT.160207.
- [6] X. Ge, "Emotion recognition intelligent system based on machine learning and clustering algorithm," *Discover Artificial Intelligence*, 2026. doi: 10.1007/s44163-026-00831-1.
- [7] H. Gürüler and M. Osman Devrim, "Use of Facial Emotion Recognition in E-Learning Systems," 2017.
- [8] S. Ilyas, M. Umar, K. Riaz, and Z. Tariq, "Emotion Recognition in AI-Powered Classrooms: Impacts on Student Engagement and Learning Outcomes," *The Critical Review of Social Sciences Studies*, vol. 3, 2025.
- [9] Y. S. Khandekar, "Intelligent Multimodal Emotion Recognition Framework for Personalized and Adaptive STEM Education," *International Journal for Research Trends and Innovation*, vol. 11, 2026.
- [10] M. Mohana and P. Subashini, "Facial Expression Recognition Using Machine Learning and Deep Learning Techniques: A Systematic Review," *SN Computer Science*, vol. 5, no. 4, 2024. doi: 10.1007/s42979-024-02792-7.

- [11] “Emotion-Adaptive Artificial Intelligence Learning System Using Multimodal Behavioral Analysis,” Technology International Research Journal of Education and Technology, 2026.
- [12] A. Professor, “Facial Expression-Based Emotion Detection for Adaptive Teaching in Educational Environments,” International Journal of Innovative Science and Research Technology, vol. 10, no. 1, 2025. doi: 10.5281/zenodo.14613833.
- [13] V. V. N. Raju et al., “Enhancing emotion prediction using deep learning and distributed federated systems with SMOTE oversampling technique,” Alexandria Engineering Journal, vol. 108, pp. 498–508, 2024. doi: 10.1016/j.aej.2024.07.081.
- [14] A. Sharma and V. Mansotra, “Deep learning based student emotion recognition from facial expressions in classrooms,” International Journal of Engineering and Advanced Technology, vol. 8, no. 6, pp. 4691–4699, 2019. doi: 10.35940/ijeat.F9170.088619.
- [15] X. Wu et al., “A deep learning approach to emotionally intelligent AI for improved learning outcomes,” Scientific Reports, vol. 16, no. 1, 2026. doi: 10.1038/s41598-026-37750-1.