A Deep Convolutional Neural Network Approach for Facial Expression Recognition Using FER-2013 Dataset

[1]Mr.Sahil A. Mulani,[2]Dr.Seema S.Patil

[1]PG Student,
[2]Associate Professor,
[1]Electronics and Telecommunication Engineering,
[1] Ashokrao Mane Group of Institution, Vathar
[1]sahilmulani5554@gmail.com [2]ssp@amgoi.edu.in

Abstract

Recognizing facial expressions (FER) is important for areas such as human–computer interaction, surveillance applications, health tracking, and affective computing. In this work, a method built on Convolutional Neural Networks (CNNs) is introduced to identify facial expressions from the FER-2013 dataset. The suggested system assigns grayscale face images to seven emotion classes: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. Its CNN design uses three convolution stages, includes max-pooling steps, adds a dense fully connected layer, and applies dropout as a regularizer. The experiments indicate roughly 84% accuracy on the training set and about 58% on the validation set. From the learning plots and confusion-matrix evaluation, overfitting and uneven class distribution are evident, with smaller groups like Disgust affected most. The findings emphasize that CNNs work well for expression recognition and point to possible gains via augmentation, transfer learning, and more sophisticated network designs.

Keywords: Facial Expression Recognition, CNN, Deep Learning, FER-2013, Emotion Classification, Computer Vision.

I. Introduction

Facial movements in people are one of the most instinctive and powerful ways to communicate feelings. With Automatic Facial Expression Recognition (FER), computers can infer emotional conditions, which strengthen interaction between users and machines. Use cases span smart tutoring platforms, monitoring for drivers, medical assessment, surveillance for safety, and robots designed for social settings. Earlier FER methods depended on manually designed descriptors like Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG), paired with classifiers such as Support Vector Machines (SVMs). In contrast, modern progress in deep learning—especially Convolutional Neural Networks (CNNs)—has markedly raised accuracy by letting models learn layered feature patterns directly from image data. The present study aims to build a CNN-driven FER model with the FER-2013 standard dataset and to assess how well it performs over seven distinct emotion classes.

II. Literature Review

A number of papers have investigated how deep learning methods can be applied to recognizing facial expressions.

LeCun and colleagues presented convolutional neural network models, bringing major change to image classification work.
Goodfellow and coauthors introduced the FER-2013 dataset, and it went on to serve as a common baseline in emotion-recognition studies.
Mollahosseini and collaborators showed that using more layers in CNN designs can markedly raise FER performance.
More recent research applies transfer learning with models like VGG16, ResNet50, and EfficientNet, reaching leading results in the field.

Even with these developments, factors including uneven class distributions, changes in lighting, occlusion effects, and faint emotional cues still reduce recognition precision.
Several studies have explored deep learning techniques for facial expression recognition.

III. Dataset Description

The FER-2013 dataset is widely used for facial expression recognition research.

FER-2013 is a common standard for evaluating methods in facial expression recognition. The collection includes 48 × 48 pixel grayscale face pictures, sorted into seven emotion categories. It provides 28,709 images for training and 3,589 images for validation/testing.

Every picture corresponds to one of these emotions: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. Because the images are low-detail and expressions vary widely, the dataset poses a practical, real-world difficulty. Moreover, class counts are imbalanced across categories, and this affects how the model learns.

Dataset Characteristics

Parameter	Value
Image Size	48 × 48 pixels
Color Mode	Grayscale
Number of Classes	7
Training Images	28,709
Validation/Test Images	3,589

IV. Methodology

The proposed approach utilizes a Convolutional Neural Network designed to extract meaningful features from facial images and classify them into predefined emotion categories.

During the pre-processing stage, all images are resized to 48 × 48 pixels and normalized to a range of 0 to 1 to ensure stable training. The labels are converted into one-hot encoded vectors to facilitate multi-class classification. The dataset is then divided into training and validation sets.

The CNN architecture consists of three convolutional layers with increasing filter sizes, allowing the network to capture low-level as well as high-level features. Each convolutional layer is followed by a max-pooling operation, which reduces spatial dimensions and enhances feature robustness. The extracted features are flattened and passed through a fully connected dense layer with ReLU activation. A dropout layer with a rate of 0.5 is incorporated to prevent over-reliance on specific neurons and to improve generalization. The final output layer uses a softmax activation function to classify the input into one of the seven emotion categories.

The model is trained using the Adam optimizer with a learning rate of 0.001. The categorical cross-entropy loss function is used to measure classification error. Training is performed for 50 epochs with a batch size of 64.

Data Pre-processing

To prepare for model training, a set of pre-processing operations was carried out to maintain uniformity and boost learning effectiveness. Each image input was scaled to 48 × 48 pixels so it aligns with the FER-2013 dataset's expected input size. Pixel values were rescaled into the 0–1 interval, aiding training stability and making convergence faster. Because the collection contains grayscale face images, chromatic details were unnecessary, so processing was done with that in mind. Emotion categories were converted into one-hot vectors so multi-class prediction could be handled through a softmax output layer. In the end, the data was split into training and validation groups to assess how well the model generalizes to unseen samples while training.

CNN Architecture-

The proposed CNN architecture consists of three convolutional blocks followed by fully connected layers.

Layer #	Layer Type	Output Shape	Parameters
1	Conv2D (32 filters, 3×3, ReLU)	46 × 46 × 32	320
2	MaxPooling2D (2×2)	23 × 23 × 32	0
3	Conv2D (64 filters, 3×3, ReLU)	21 × 21 × 64	18,496
4	MaxPooling2D (2×2)	10 × 10 × 64	0
5	Conv2D (128 filters, 3×3, ReLU)	8 × 8 × 128	73,856
6	MaxPooling2D (2×2)	4 × 4 × 128	0
7	Flatten	2048	0
8	Dense (128, ReLU)	128	262,272
9	Dropout (0.5)	128	0
10	Dense (7, Softmax)	7	903

The Softmax activation function is used in the output layer for multi-class classification

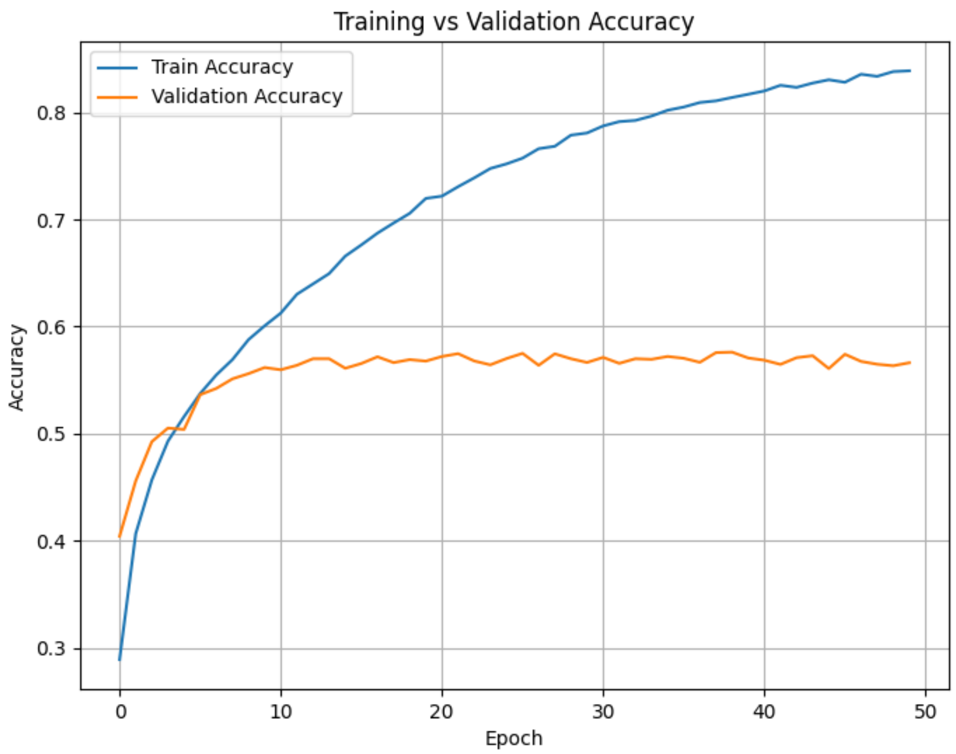
Hyper parameters

Hyper parameter	Value
Image Size	48 × 48
Batch Size	64
Epochs	50
Learning Rate	0.001
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Dropout Rate	0.50

V. Results

Training and Validation Accuracy

In training, the suggested model reached about 84% accuracy, whereas the validation performance stayed around 57–58%. As epochs progressed, the training score rose steadily, suggesting the network was successfully capturing patterns within the training set. Conversely, validation performance increased mainly at the beginning and then leveled off at roughly the tenth epoch. The widening gap between training and validation outcomes indicates that the model is fitting fine-grained regularities tied to the training samples, while showing weaker generalization on previously unseen inputs. This pattern underscores the distinction between extracting broadly representative features and becoming tuned to properties unique to the dataset over the course of training.

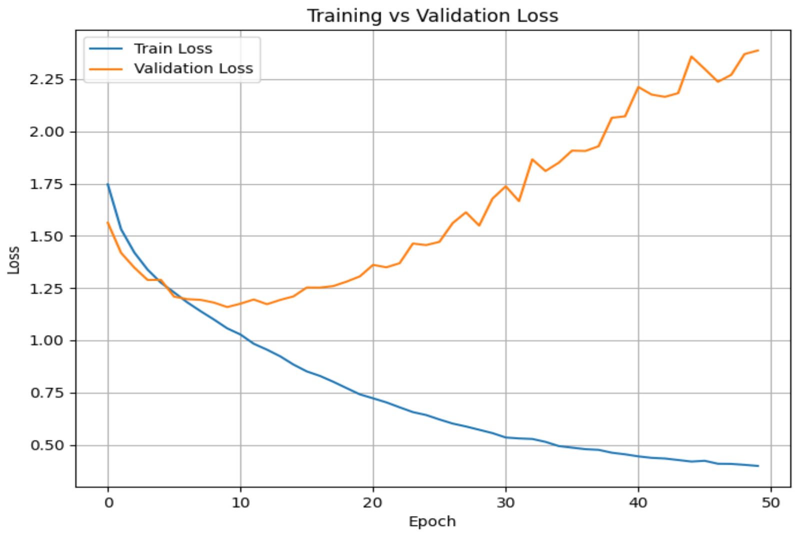


Graph 1: Training vs Validation Accuracy Curve

Training and Validation Loss

Training loss steadily decreased to approximately 0.4, while validation loss increased significantly after epoch 15 and reached approximately 2.35 by epoch 50.

This divergence further confirms over fitting.



Graph 2: Training vs Validation Loss Curve

VI. Confusion Matrix Analysis

The confusion matrix provides detailed insights into class-wise performance.

Emotion	Correct Predictions	Total Samples	Recall (%)
Angry	449	958	46.9
Disgust	45	111	40.5
Fear	422	1024	41.2
Happy	1375	1774	77.5
Neutral	611	1233	49.6
Sad	553	1247	44.3
Surprise	609	831	73.3

VII. Observations

Reviewing the confusion matrix indicates clear differences in how well the model handles the various emotion categories. Emotions like Happy and Surprise achieve the strongest recall, implying their facial cues are more pronounced and simpler to identify. By comparison, Disgust records weaker results, likely because this class has fewer training examples. Moreover, Fear, Sad, and Angry are misidentified more often, which points to shared facial patterns that make separating them challenging. Neutral faces are difficult as well, since they can resemble several emotional conditions, creating uncertainty in labelling. Altogether, these findings emphasize how hard it is to correctly detect subtle expressions that are closely related.

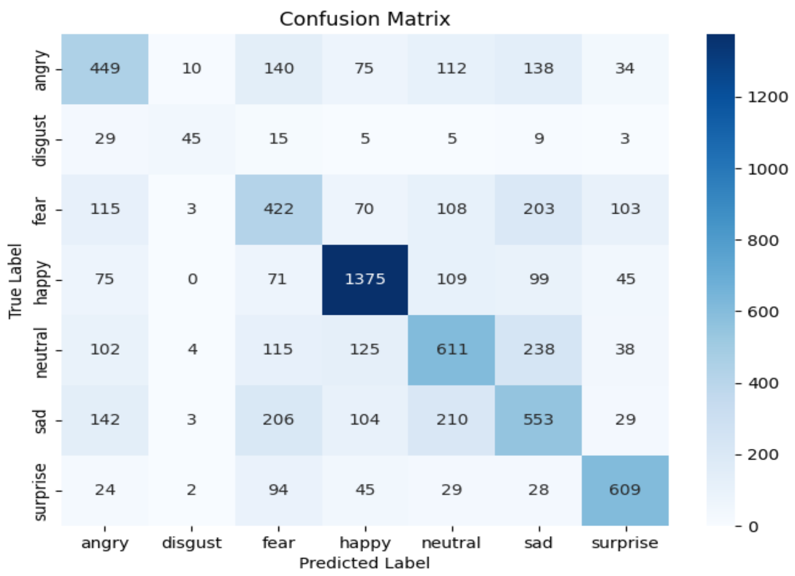


Figure 3: Confusion Matrix of CNN Model

Emotion	Correctly Classified	Total Samples	Approx. Recall
Angry	449	958	46.9%
Disgust	45	111	40.5%
Fear	422	1024	41.2%
Happy	1375	1774	77.5%
Neutral	611	1233	49.6%
Sad	553	1247	44.3%
Surprise	609	831	73.3%

VIII. Conclusion

In this work, a facial expression recognition method built on a Convolutional Neural Network was introduced and evaluated with the FER-2013 dataset. The suggested network proved it can learn facial patterns and sort them into seven separate emotion classes, reaching close to 84% accuracy on training data and roughly 58% on validation data. The findings from the experiments emphasize that CNN-based designs can derive useful feature representations from face images and handle emotion recognition as a multi-class problem. The training and validation patterns also offer additional understanding of how the model learns and how well it generalizes. In summary, the results support the value of deep learning for facial expression recognition and provide a solid base for future refinements. In upcoming efforts, we will investigate more sophisticated networks, transfer-learning methods, and tuning strategies to improve accuracy and strengthen robustness for practical real-world use.

IX.References

[1] H. Li, N. Wang, X. Ding, and J. Han, “Deep Learning for Facial Expression Recognition: A Survey,” *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195–1215, Jul.–Sept. 2022.

[2] S. Zhang, Z. Zhao, and Q. Ji, “Facial Expression Recognition With Deep Learning: A Comprehensive Review,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 4, pp. 1233–1248, Apr. 2022.

[3] Y. Wang, J. Li, and X. Liu, “Facial Expression Recognition Based on CNN and Attention Mechanism,” *IEEE Access*, vol. 10, pp. 45876–45885, 2022.

[4] M. Kim and S. Lee, “Hybrid CNN-Transformer Architecture for Facial Expression Recognition,” *IEEE Access*, vol. 11, pp. 23456–23467, 2023.

[5] A. Kumar, R. Sharma, and P. Singh, "Facial Emotion Recognition Using Transfer Learning With EfficientNet," *IEEE Access*, vol. 11, pp. 98765–98775, 2023.

[6] L. Chen, X. Zhao, and Y. Zhou, "Multi-Scale Feature Fusion Network for Facial Expression Recognition," *IEEE Transactions on Multimedia*, vol. 26, pp. 1123–1135, 2024.

[7] J. Park and K. Sohn, "Lightweight Deep Neural Network for Real-Time Facial Expression Recognition," *IEEE Access*, vol. 12, pp. 33445–33456, 2024.

[8] R. Gupta and S. K. Singh, "Facial Expression Recognition Using Vision Transformers and CNN Hybrid Model," *IEEE Access*, vol. 13, pp. 55678–55689, 2025.

[9] T. Nguyen, H. Pham, and D. Tran, "Robust Facial Expression Recognition Using Self-Supervised Learning," *IEEE Transactions on Affective Computing*, early access, 2025.

[10] P. Verma and A. Joshi, "Advanced Deep Learning Techniques for Facial Emotion Recognition: A Comparative Study," *IEEE Access*, vol. 14, pp. 11234–11250, 2026.