

Comprehensive Evaluation of Preprocessing Techniques for Facial Emotion Recognition

**Team Members**

* Mahmoud Adel – 233561
* Youssef Omran – 235049

# Contents

[Contents 1](#_Toc197725129)

[1. Introduction 2](#_Toc197725130)

[2. Methodology 2](#_Toc197725131)

[2.1 Dataset 2](#_Toc197725132)

[2.2 Pipeline Overview 2](#_Toc197725133)

[2.3 CNN Model 3](#_Toc197725134)

[3. Preprocessing Techniques Evaluated 3](#_Toc197725135)

[3.1 Techniques That Reduced Accuracy 3](#_Toc197725136)

[3.2 Denoising Techniques 4](#_Toc197725137)

[3.3 Best Performing Techniques (Applied Together) 5](#_Toc197725138)

[4. Face Detection Integration: YOLOv8 6](#_Toc197725139)

[4.1 Web Interface Deployment Using Gradio 7](#_Toc197725140)

[5. Experimental Results 8](#_Toc197725141)

[6. Challenges 9](#_Toc197725142)

[7. Conclusion 11](#_Toc197725143)

# 1. Introduction

Facial Emotion Recognition (FER) is a crucial subfield in affective computing and human-computer interaction, enabling systems to interpret human emotions from visual cues. Despite significant advances through deep learning models such as Convolutional Neural Networks (CNNs), performance remains sensitive to image quality, especially in datasets like FER-2013, which contain low-resolution, grayscale images with noise and poor contrast.

This project explores a variety of preprocessing techniques to determine which approaches enhance emotion recognition performance. We also integrated a modern face detection framework, YOLOv8, to enable automated face region extraction from raw images. The processed face regions were then fed into a CNN trained on FER-2013 to classify emotions. The study reveals that certain commonly-used methods like CLAHE or gamma correction degrade performance, while others—when applied cumulatively—significantly improve recognition accuracy.

# 2. Methodology

## **2.1 Dataset**

The FER-2013 dataset consists of approximately 35,000 grayscale facial images, each 48×48 pixels, labeled into seven emotion categories:  
**Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral.**  
The dataset is split into 80% for training and 20% for testing.

## **2.2 Pipeline Overview**

Our emotion recognition pipeline includes the following stages:

* **Face Detection**: YOLOv8 is used to detect faces in the original input images. It localizes and crops face regions effectively, even in challenging conditions such as complex backgrounds or multiple subjects.
* **Preprocessing**: Detected facial regions undergo a series of preprocessing steps aimed at enhancing visual clarity while preserving key emotional features.
* **Emotion Classification**: A Convolutional Neural Network (CNN) classifies the preprocessed face into one of the seven emotion categories.

## **2.3 CNN Model**

The CNN model receives 48x48 grayscale images and includes:

* 3 convolutional layers with ReLU activation and max pooling
* Dropout for regularization
* Fully connected dense layer with SoftMax output

Training was performed using the Adam optimizer for 50 epochs with a batch size of 64 and categorical cross-entropy loss.

# 3. Preprocessing Techniques Evaluated

We experimented with 13 preprocessing methods, evaluating each one both visually and in terms of its effect on classification accuracy. All techniques were applied to the same test images to ensure fair comparison.

## 3.1 Techniques That Reduced Accuracy

|  |  |  |
| --- | --- | --- |
| **Technique** | **Visual Effect** | **Accuracy Impact** |
| **Gamma Correction** | Overexposed midtones, facial features became flat and washed out. | ↓ (22.4%) |
| **Laplacian Filtering** | Enhanced edges excessively, introduced noise halos. | ↓ (28.6%) |
| **Thresholding** | Eliminated gradients resulted in binary blobs. | ↓ (30.3%) |
| **Smoothing (Average)** | Over-blurred; facial features like eyes and mouth lost sharpness. | ↓ (40.1%) |
| **CLAHE** | Uneven local contrast; introduced artifacts in dark areas. | ↓ (39.0%) |



**clahe**



## 3.2 Denoising Techniques

These techniques were used to reduce noise, with varying success:

|  |  |  |
| --- | --- | --- |
| **Technique** | **Visual Effect** | **Accuracy Impact** |
| **Median Filtering** | Smoothed noise but blurred edges. | 48.9% |
| **Gaussian Filtering** | Gentle blur, but still reduced detail. | 51.0% |
| **Bilateral Filtering** | Preserved edges better than Gaussian; moderate success. | 52.7% |
| **Wavelet Denoising** | Reduced background noise but removed fine details. | 54.5% |
| **Non-Local Means Denoising ✅** | Preserved facial texture and edge clarity effectively. | **62.9%** |



A person with glasses and beard

AI-generated content may be incorrect.

## 3.3 Best Performing Techniques (Applied Together)

Four techniques were found to work best when **applied sequentially**, leading to optimal emotion recognition performance:

1. **Histogram Equalization** – Increased global contrast across the image.
2. **Normalization** – Scaled pixel values to range [0, 1] for better convergence.
3. **Sharpening (custom kernel)** – Highlighted key facial contours like eyes and mouth.
4. **Non-Local Means Denoising** – Retained texture while reducing background interference.

These four were used in the final version of the pipeline, **cumulatively**, with the highest recorded accuracy of **62.9%**.



# 4. Face Detection Integration: YOLOv8

To emulate real-world use cases and test the model's reliability in dynamic settings, we replaced static cropped inputs with YOLOv8-based face detection.

* **Model Used:** face\_yolov8n.pt
* YOLOv8 outputs bounding boxes around detected faces.
* The crops are converted to grayscale, resized to 48×48, and passed through the preprocessing pipeline.
* **The CNN then predicts the emotion label.**

This automated flow ensures that the emotion recognition system can process unconstrained inputs like single or group images.

A person smiling in front of a poster

AI-generated content may be incorrect.

A person and person holding shopping bags

AI-generated content may be incorrect.

## 4.1 Web Interface Deployment Using Gradio

To enhance accessibility and streamline testing, a lightweight web interface was developed using **Gradio**. This interface offers an intuitive way for users to interact with the trained emotion recognition model without writing code or running Jupyter cells.

**Key Characteristics:**  
• Allows image uploads directly through the browser.  
• Internally utilizes the YOLOv8 model to detect all visible faces in the uploaded image.  
• Each detected face is preprocessed using the optimized four-step pipeline: Histogram Equalization, Normalization, Sharpening, and Non-Local Means Denoising.  
• The processed faces are passed to the trained CNN for emotion prediction.  
• Results are visualized with bounding boxes and corresponding emotion labels overlaid on the original image.

**Benefits:**  
• Offers an accessible demo without requiring a coding environment.  
• Supports real-time experimentation with user-supplied images.  
• Optionally, the app can be shared via public URLs (using share=True) for remote testing or presentation purposes.

This interface adds practical value by bridging the model's backend with an intuitive frontend, demonstrating potential for deployment in applications like education, customer experience analysis, or research.

# 5. Experimental Results

Below is a summary of test accuracy across all evaluated preprocessing configurations:

|  |  |
| --- | --- |
| **Preprocessing Method** | **Accuracy (%)** |
| No Preprocessing | 42.6 |
| Gamma Correction | 22.4 |
| Laplacian Filtering | 28.6 |
| Thresholding | 30.3 |
| CLAHE | 39.0 |
| Smoothing | 40.1 |
| Median Filter | 48.9 |
| Gaussian Filter | 51.0 |
| Bilateral Filter | 52.7 |
| Wavelet Denoising | 54.5 |
| Histogram Equalization | 56.2 |
| Normalization | 57.5 |
| Sharpening | 60.4 |
| Non-Local Means Denoising. | 63.07 |
|  |
| **Combined (4 techniques)** | **63.07** |

# 6. Challenges

**• Image Resolution Constraints**

The FER-2013 dataset consists of grayscale facial images sized at 48×48 pixels, which poses a major limitation. At such a low resolution:

* Fine-grained emotion indicators, like subtle wrinkles, eyebrow shapes, and eye movements, are easily lost or distorted.
* Aggressive preprocessing techniques—particularly those involving filters or transformations—tend to degrade already minimal details.
* Any misstep in enhancement risks amplifying noise more than useful features.

**• Noise vs. Detail Tradeoff**

Denoising techniques such as Gaussian filtering, Wavelet denoising, and even Non-Local Means must carefully balance:

* Suppressing background or compression noise
* Preserving soft contours in expressions like fear or sadness

**In practice:**

* Many denoising filters smoothed out crucial emotion cues like fine lines around the mouth or furrowed brows.
* Wavelet denoising, for example, slightly improved clarity but sometimes oversimplified facial textures.
* Non-Local Means Denoising proved to be the best, retaining fine details while reducing background noise—though it was computationally expensive.

**• Overprocessing Effects**

Certain preprocessing steps negatively affected accuracy:

* Thresholding obliterated gradients entirely, leading to classification failure.
* CLAHE (Contrast Limited Adaptive Histogram Equalization) often over-amplified local contrasts, especially in already clear images, producing artificial artifacts and worsening model predictions.
* Gamma correction and Laplacian filtering added either too much brightness or exaggerated edges, confusing the CNN’s feature extraction layers.

These findings confirm that more preprocessing doesn't always mean better performance.

**• Real-Time Processing Limitations**

In the context of deploying the model in real-time environments like surveillance or retail:

* Techniques like Non-Local Means Denoising and multi-step sharpening pipelines are computationally intensive, causing delays in frame-by-frame processing.
* Face detection using YOLOv8 is fast, but the cumulative effect of multi-stage preprocessing slows down full pipeline execution.
* Optimization or GPU acceleration would be essential for real-time use cases.

**• Dataset Alignment and Variability**

The FER-2013 dataset is known to have inconsistencies in face alignment, including:

* Faces looking in different directions
* Off-center compositions
* Varying degrees of zoom and head tilt

These inconsistencies introduce noise that preprocessing alone cannot correct. While YOLOv8 detects faces effectively, the lack of precise alignment or facial landmark standardization further complicates learning consistent emotion features.

# 7. Conclusion

This study demonstrates that carefully selected and sequentially applied preprocessing techniques significantly enhance FER performance. While some widely-used methods (e.g., CLAHE, gamma correction) surprisingly degrade accuracy on low-resolution facial images, others—especially Histogram Equalization, Normalization, Sharpening, and Non-Local Means Denoising—collectively lead to a marked improvement in test accuracy from 42.6% to 63.07%.

Integrating YOLOv8 for automated face detection allowed for a modular, scalable pipeline adaptable to real-world applications. Future work could explore learning-based or adaptive preprocessing methods and attention-based CNN architectures for better emotional localization.