# Facial Expression Recognition using State of Art techniques

Cristian Pérez Díaz (1565487), Pau Carracelas Expósito (1569293), Gabriel Bardají Biescas (1568297)

**Abstract**— Facial Expression Recognition (FER) has gained significant attention in recent years due to its various applications in areas such as mental health, advertising, security, and entertainment. This article presents a review of the state of the art in FER and proposes a solution using both conventional and deep learning techniques. The conventional method involves image preprocessing, feature extraction using Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG), and classification using Random Forest. The deep learning approach utilizes a custom Convolutional Neural Network (CNN) architecture. The proposed solution is evaluated using the Fer2013 dataset and achieves promising results.

**Keywords**—facial expression recognition; feature extraction; classification; deep learning; fer2013;

## 1 Introduction

FACIAL Emotion Recognition (FER) has gained significant attention in recent years due to its various applications in areas such as mental health, advertising, security, and entertainment. In this project, we aim to tackle the challenge of recognizing facial emotions using computer vision techniques and algorithms.
The main motivation is the interest in learning and exploring the field of FER and applying different techniques to achieve the goal. Through this project, we hope to gain knowledge in image processing and facial emotion analysis, which can be useful for developing innovative solutions in the future. This article will present a review of the state of the art in FER as well as a proposed solution that will be evaluated in future experiments.

## 2 Objectives

In this project, our objective is to develop an automated facial emotion recognition solution that can potentially be functional and applicable for future projects, companies, etc. To achieve this, we will implement and compare different facial emotion recognition techniques, both conventional and deep learning-based, in order to evaluate their effectiveness and efficiency in terms of accuracy and speed. Additionally, we will investigate new deep learning-based facial emotion recognition techniques, to compare their performance with conventional techniques and determine their potential in practical applications. By achieving these objectives, we hope to significantly improve our knowledge in the field of facial emotion recognition and computer vision. Our main goal is to get into the field of facial and emotion recognition, as we believe its applications in larger-scale projects are of our interest and with high potential for the future.

## 3 State of art

In terms of conventional methods for FER, techniques such as Local Binary Patterns (LBP), Haar-like features, and Support Vector Machines (SVM) have been used for facial feature analysis. These methods rely on the intensity of the expression, position and shape of facial features, and temporal dynamics of facial expressions. LBP is a texture-based approach that captures local information about facial features, while Haar-like features are based on simple rectangular areas of the face [4]. SVM is a popular classification method used for FER, which works by finding a hyperplane that separates the facial expression feature space into different classes. However, these conventional methods heavily rely on manual feature engineering, which can be time-consuming and may not be suitable for complex facial expressions.

On the other hand, the use of deep learning in FER has shown promising results in recent years. Convolutional neural networks (CNNs) have been widely used to extract features from facial images, while recurrent neural networks (RNNs) have been used for temporal modeling of facial image sequences to improve FER accuracy [4]. Deep learning methods automatically learn features from raw data, eliminating the need for manual feature engineering, which makes them more suitable for complex tasks such as FER. CNNs have multiple layers that can capture features of different complexities, allowing them to learn representations of the face that are more robust to variations in facial expressions and lighting conditions. RNNs, on the other hand, can model temporal dependencies between frames, which is particularly useful for analyzing dynamic facial expressions.

## 4 Implementation Proposal

### 4.1 Conventional method

To implement FER, we will first use the conventional method, which involves the following steps:

1. **Image input**
2. **Preprocessing of the image:** We will perform noise reduction,histogram equalization and median filter to improve the quality of the facial images.
3. **Feature extraction:** After investigating various conventional feature extraction methods such as Gabor Feature Extraction [9], Local Binary Pattern (LBP) [3], ASM/AAM [2] [1], Optical Flow Method [7], Haar-like Feature [8], and Feature Point Tracking Extraction, we have ultimately chosen to utilize the LBP and HOG methods as our primary feature extraction techniques.
4. **Classification:** We have tested different classification algorithms such as KNN, SVM, and AdaBoost to determine the most suitable one for our FER system. After evaluating their performance, we have ultimately utilized the RandomForest algorithm. This step involves training the RandomForest classifier using the features extracted from the facial images and classifying the images into different emotion categories.

### 4.2 Deep Learning

In addition to conventional feature extraction and classification methods, we also explored the power of deep learning techniques

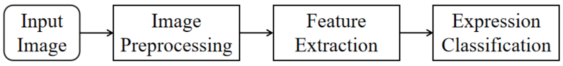


Fig. 1: Conventional method design [4]

for our FER system. We implemented a Convolutional Neural Network (CNN) using the Keras framework, allowing us to design and train our own network architecture specifically tailored to the task of facial emotion recognition.
Our custom CNN architecture, consists of multiple layers designed to extract meaningful features from facial images and classify them into different emotion categories. The architecture includes convolutional layers, which detect various visual patterns and features in the input images. These convolutional layers are followed by pooling layers, which reduce the spatial dimensions of the extracted features, aiding in capturing the most important information. Additionally, we have incorporated dropout layers to prevent overfitting.
The output of the convolutional and pooling layers is then flattened and passed through fully connected layers. These layers perform feature extraction and make predictions by mapping the learned features to the corresponding emotion categories. The final layer utilizes the softmax activation function to produce probability scores for each emotion category. For a more detailed description of the deep learning model, please refer to Appendix.
In addition to developing our own custom CNN for facial emotion recognition, we aimed to compare its performance with a pre-existing CNN architecture. This approach enabled us to gain valuable insights into the effectiveness of different CNN models specifically tailored to our task.

## 5 Data

In this work, we will use the Fer2013 dataset [5], which contains grayscale images of faces with a size of 48x48 pixels. The faces have been preprocessed and registered automatically so that they are more or less centered and occupy a similar amount of space in each image. The goal is to classify each face into one of seven categories based on the emotion shown in the facial expression: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. The training set of the Fer2013 dataset consists of 28,709 examples, and the public test set consists of 3,589 examples.

Fig. 2: Fear labeled picture [5]

Additionally, we will utilize our own set of images to evaluate the performance of our final trained FER model. These images will not have the face cropped initially. To extract the faces from these images, we will employ a pre-trained Haar Cascades classifier [6]. The Haar Cascades classifier will aid in detecting and localizing the faces in the images. Once the faces are cropped, we will apply our FER model for emotion classification.
Once our FER model is adequately trained and validated, we will employ it to classify the emotions in the faces obtained from our own set of images. This evaluation will provide insights into the performance and generalization capabilities of our FER model when applied to real-world images.
By utilizing our own set of images and leveraging the pre-trained Haar Cascades classifier, we aim to ensure the effectiveness and practicality of our FER system in real-world scenarios where the faces are not pre-cropped.

## 6 Experiments, results and analysis

In this section, we present the experiments, results, and analysis of our two proposed implementations. We explore both a conventional method and a deep learning approach for facial emotion recognition, aiming to provide a comprehensive evaluation of their performance.

### 6.1 Conventional Method

In our experiments, we utilized Local Binary Pattern (LBP) and Histogram of Oriented Gradients (HOG) as feature extractors for facial emotion recognition. To determine the most suitable classifier for our task, we conducted a quick test using the LazyPredict library. This test helped us identify that Random Forest achieved the best performance among the classifiers tested. Therefore, we employed Random Forest as our chosen classifier.
Furthermore, we investigated the combination of LBP and HOG features to assess their impact on the classification results. We computed the accuracy of the classification using LBP and HOG features separately and then combined them. This allowed us to evaluate the effectiveness of each feature extraction technique individually and their combined performance.

Taula 1: Accuracy of the Conventional Method

| Feature Extractor | Accuracy (Train) | Accuracy (Test) |
|---|---|---|
| LBP | 99.85 | 36.08 |
| HOG | 99.85 | 38.11 |
| Combined (LBP+HOG) | 99.85 | 38.40 |

It is worth noting that we observed signs of overfitting in our model. To address this issue, we experimented with reducing the depth of the decision trees in the Random Forest and increasing their width. Unfortunately, these adjustments did not completely solve the overfitting problem. However, we did manage to find some parameter settings that slightly improved the accuracy.

Taula 2: Accuracy of the Conventional Method with Adjusted Parameters

| Feature Extractor | Accuracy (Train) | Accuracy (Test) |
|---|---|---|
| LBP | 99.83 | 36.91 |
| HOG | 99.85 | 39.74 |
| Combined (LBP+HOG) | 99.85 | 39.64 |

Based on our experiments, we found that the best-performing model was the one using HOG features with the adjusted parameters. This model achieved an accuracy of 39.74 .

To provide a more comprehensive analysis of the model's performance, we present the confusion matrix in Figure 3.

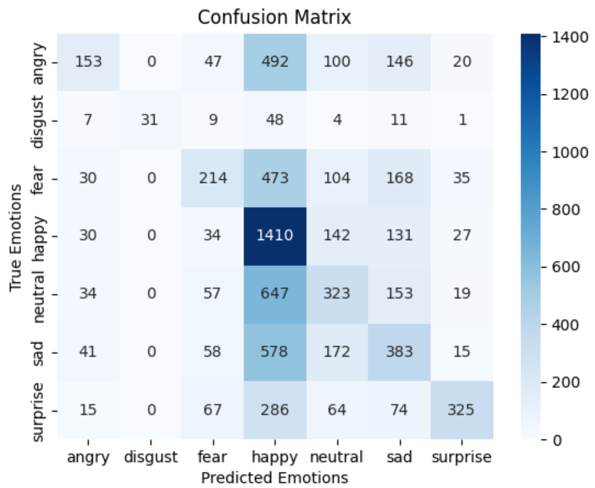


Fig. 3: Confusion Matrix for the Conventional Method

In addition to evaluating the performance of the conventional method using different feature extractors, we also explored the impact of reducing the number of emotions to predict. By limiting the emotion categories, we aimed to assess whether a smaller set of emotions could lead to improved classification results.
The following table summarizes the accuracy of the conventional method for different emotion combinations:

Taula 3: Accuracy of the Conventional Method with Reduced Emotion Categories

| Emotion Combination | Accuracy (Train) | Accuracy (Test) |
|---|---|---|
| Happy & Angry | 99.99 | 70.75 |
| Angry & Surprise | 99.97 | 78.25 |
| Angry, Surprise, Neutral | 99.95 | 61.68 |
| Surprise, Neutral | 99.98 | 80.95 |
| Surprise, Neutral, Fear | 99.88 | 60.78 |

We observed that reducing the number of emotions to predict yielded mixed results. In some cases, such as the combination of ['angry', 'surprise'], ['surprise', 'neutral'], and ['angry', 'happy'], the accuracy remained high for both the training and testing sets. However, for other combinations, such as ['surprise', 'neutral', 'fear'] and ['angry', 'surprise', 'neutral'], the accuracy decreased significantly.

### 6.2 Deep Learning

In the deep learning approach, we decided to employ our own CNN (Convolutional Neural Network) to tackle the emotion recognition problem. After training and evaluating this model, we achieved promising results in terms of accuracy.
The results obtained with our CNN were as follows:

- Train accuracy: 67.34
- Test accuracy: 65.46

These results demonstrate a significant improvement compared to the conventional methods discussed in the previous section. Our CNN was able to capture relevant features from the images and make more accurate predictions regarding the depicted emotions.

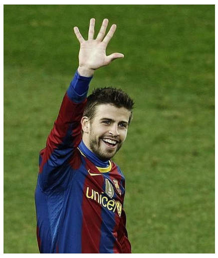

Fig. 4: Piqué celebrating a win

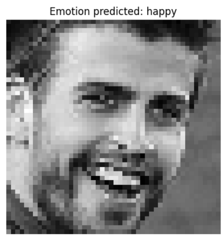


Fig. 5: Piqué labeled as Happy

Next, we evaluated the performance of another model, namely ResNet50. The results obtained with ResNet50 were as follows:

- Train accuracy: 66.10
- Test accuracy: 60.94

Comparing these results with those obtained by our CNN, we can conclude that our custom CNN outperforms ResNet50. Despite ResNet50 being a state-of-the-art and widely-used architecture, our customized approach managed to surpass it in terms of test accuracy.
These results support the choice of our CNN as the preferred model for emotion recognition in this specific context.

## 7 Conclusions

Facial Expression Recognition (FER) is a challenging task that has gained significant attention in recent years due to its numerous applications in various fields. In this project, we aimed to develop a facial emotion recognition system using state-of-the-art techniques and algorithms.
We began by reviewing the state of the art in FER, exploring both conventional methods and deep learning techniques. Conventional methods, such as Local Binary Patterns (LBP) and Haar-like features, rely on manual feature engineering and classification algorithms like Support Vector Machines (SVM). On the other hand, deep learning approaches, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown promising results by automatically learning features from raw data.
For our implementation, we proposed a two-fold approach. First, we implemented a conventional method using LBP and Histogram of Oriented Gradients (HOG) as feature extractors, and Random Forest as the classifier. We evaluated the performance of different feature extractors and observed signs of overfitting, which we tried to address by adjusting parameters. Our best-performing model achieved an accuracy of 39.74
In addition to the conventional method, we developed a custom CNN architecture using the Keras framework for deep learning-based FER. The CNN consisted of convolutional layers for feature extraction and fully connected layers for classification. We compared the performance of our custom CNN with a pre-existing CNN architecture, aiming to gain insights into the effectiveness of different models tailored to FER.
In conclusion, this project has provided us with valuable knowledge and experience in the field of FER. We have explored and implemented both conventional and deep learning techniques, evaluated their performance, and gained insights into their strengths and limitations. Moving forward, we believe that further research and experimentation in FER will continue to ad-

vance the field and contribute to its applications in mental health, advertising, security, and entertainment.

## References


[1] Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6):681–685, 2001.

[2] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.

[3] Xiaoyi Feng, Matti Pietikäinen, and Abdenour Hadid. Facial expression recognition based on local binary patterns. *Pattern Recognition and Image Analysis*, 17:592–598, 2007.

[4] Yunxin Huang, Fei Chen, Shaohe Lv, and Xiaodong Wang. Facial expression recognition: A survey. *Symmetry*, 11(10), 2019.

[5] Aaron Courville Pierre-Luc Carrier. Challenges in representation learning: Facial expression recognition challenge. https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge, 2013.

[6] G. Solano. Deteccion de rostros con haar cascades python – opencv. https://omes-va.com/deteccion-de-rostros-con-haar-cascades-python-opencv/, 2020. Accessed: April 28, 2023.

[7] Yaser Yacoob and Larry S Davis. Recognizing human facial expressions from long image sequences using optical flow. *IEEE Transactions on pattern analysis and machine intelligence*, 18(6):636–642, 1996.

[8] Peng Yang, Qingshan Liu, and Dimitris N Metaxas. Boosting encoded dynamic features for facial expression recognition. *Pattern Recognition Letters*, 30(2):132–139, 2009.

[9] Shiqing Zhang, Lemin Li, and Zhijin Zhao. Facial expression recognition based on gabor wavelets and sparse representation. In *2012 IEEE 11th international conference on signal processing*, volume 2, pages 816–819. IEEE, 2012.


## Appendix

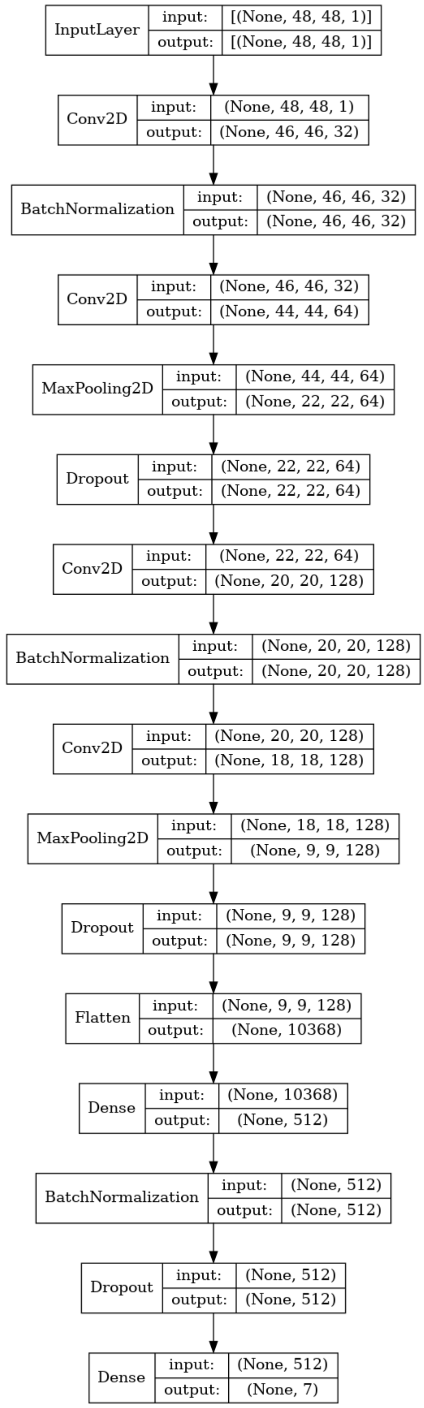


Fig. 6: Model of our own CNN