



# Facial Expression Recognition with Gender Identification

Shubham Gupta, Aarushi Dhawan, Arpit Gupta, A. Kumar Dubey\*

Bharati Vidyapeeth's College of Engineering, INDIA

Emails: [shubhamgupta5111997@gmail.com](mailto:shubhamgupta5111997@gmail.com); [aarushidhawan99@gmail.com](mailto:aarushidhawan99@gmail.com); [guptaarpit09@yahoo.com](mailto:guptaarpit09@yahoo.com);  
[arudubey@gmail.com](mailto:arudubey@gmail.com)

\* Correspondence: [arudubey@gmail.com](mailto:arudubey@gmail.com)

## Abstract

Human facial emotion recognition has attracted interest in the field of Artificial Intelligence. The emotions on a human face depict what's going on inside the mind. Facial expression recognition is the part of Facial recognition which is gaining more importance, and the need for it has increased tremendously. Though there are methods to identify expressions using machine learning and Artificial Intelligence techniques, this work attempts to use convolution neural networks to recognize expressions and classify the expressions into 6 emotional categories. Various datasets are investigated and explored for training expression recognition models are explained in this paper, and the models which are used in this paper are VGG 19 and RESNET 18. We included facial emotional recognition with gender identification also. In this project, we have used the fer2013 and ck+ datasets and ultimately achieved 73% and 94% around accuracies, respectively.

**Keywords:** Facial Expression Recognition; Gender Identification

## 1. Introduction

For the past years, human facial attributes like age, gender, and emotions are always in the list of researchers. The reason being is that there are many applications of this challenging problem ranging from security, the interaction between humans and computers, identification of human emotions, etc. Also, as a part of communication, humans show their emotions using facial expressions. It is very easy for humans to recognize emotions and facial expressions, but emotion recognition is challenging for machines. In recent years many techniques have been used for emotion classification but developing an automated system to accomplish this task is difficult. In this project, we have developed a Convolutional Neural Network to accomplish the task of emotion recognition. Image is given as input to our network, which then predicts the emotion out of the mentioned categories. The emotions are classified into seven classes: anger, happiness, fear, sadness, disgust, surprise, and neutral. As we all know, in today's world, one of the most important things is customer satisfaction, so we developed a system that can recognize facial expressions and can be used for customer feedback. Facial expression recognition is a technology that uses a confusion matrix to detect emotions on human faces. More precisely, this system is a sentiment analysis tool and is able to detect basic & universal expressions like happy, sad, fear, anger, etc. Facial expression recognition systems are important because of the ability to understand human expressions. Facial expressions and other gestures convey nonverbal types of communication that play an important role in interpersonal relationships. This complements speech by helping the listener interpret the words' meaning.

Therefore, facial expression recognition, because of its extraction and analysis nature of an image or video, is able to deliver unfiltered, unbiased emotional responses to data.

We have taken a dataset, modified it, and trained it in this system. When we put an image in front of the camera, the image is captured on a grey scale. The image is saved in the resulting library. We have used a confusion matrix to recognize the real-time frames. We have added seven types of expressions angry, disgusted, fear, happy, neutral, sad, and surprised, which can be identified. Also, we have added the option for identifying the age and gender of the person. The system identifies the expression by observing the classification score and expression category ratio. When the classification score is high for a specific expression, the system recognizes it. The objective is to construct a system that can recognize these expressions and differentiate between them so that it can be used for feedback and other applications. The scientific study of facial expression recognition tells that it represents emotions and intention communication. In the first category, a filter bank or an image filter processes the face images. This processing of the image is done either on some regions of the face or on the whole part of the image. The motive behind the processing is to extract any changes in the facial appearance. Typically, some filters convolve specific regions of the face or the whole face. The entire face is divided into different facial filters. During the extraction of geometric features gained from the facial appearance processing, a feature vector can be formed by the coordinates of the fiducial points or the angles, shapes, and distances. Under universal emotions classifications, various emotions are categorized on the basis of movements of the face.

Emotions	Details	Features
Anger Emotion	It's a very harmful emotion that can destroy many people and their relationships with others.	Widely opened eyes highly raised eyebrows.
Fear	It is an emotion of panic, horror, or danger from surroundings.	Open mouth, shaking of face
Sad	This emotion depicts the sorrow in one's life.	Lips pulled down, water in the eye
Happy	This is an emotion of joy. It depicts hope in one's eyes. The person is always calm and cheerful.	Cheeks pulled over, smiles on one's face.
Disgust	It is a feeling when a person is irritated and not happy. The reason may be anything.	-

Figure 1: Universal Expressions Table with details and features.

## 2. Related Work

In the last few years, researchers have been successful in developing automatic emotion classifiers [1][2][3]. The emotions like happiness, sadness, and anger are recognized by some expression recognition systems, and other systems are able to recognize the movement of muscles on an individual's face. Facial Action Coding System (FACS) [4] is one of the best psychological frameworks which describe muscle movements that the face can produce. CNN shows better performance in Yu and Zhang's earlier work of Emotion 2015 [5]. They used a group of Convolutional Neural Networks with five convolutional layers in their paper. They achieved excellent performance in recognition and proposed data interference and voting methods which further improved the performance of recognition. In the work of Deepali Aneja [6], they proposed a novel approach to the transfer of expression from people to multiple stylized personalities. To independently recognize human expression and stylized characters. The first train two Convolutional Neural Networks and use a transfer learning technique to learn how to map human characters to create a common integration feature space. For gender identification, we have used a paper where gender is identified (Male or Female) using front images[9]. We have used a research paper for comparison purposes that use the fer2013 dataset Tumen, V., Soylemez, O. F., & Ergen, B. (2017). They used the convolutional neural network to process humans' facial expressions on a database [10]. They have achieved cnn accuracy of 57.1% on the fer2013 database. We received an accuracy of 73.112% on fer2013 which is approximately a 28% increase.

### 3. Proposed Methodology

#### 3.1 Facial Expression Recognition Systems

i) **Preprocessing** is a process that is used to improve the performance of the FER system, and it can be carried out before the feature extraction process. Image preprocessing includes different types of processes such as image clarity and scaling, contrast adjustment, and additional enhancement processes to improve the expression frames. Here in preprocessing, the input image, which is in RGB form, is converted into Grayscale, and detection is performed by cropping the image to the appropriate size. After that, normalizing the image is important.

ii) **Normalization**: Normalization is the preprocessing method that can be designed for the reduction of illumination and variations of the face images to achieve an improved face image. The normalization method is also used to extract eye positions, making it more robust to personality differences for the FER system and providing more clarity to the input images.

iii) **Feature Extraction**: Feature extraction is finding and depicting positive features of concern within an image for further processing. In image processing, computer vision feature extraction is a significant stage, whereas it spots the move from graphic to implicit data depiction. Then these data depictions can be used as input for the classification. The feature extraction methods are categorized into five types such as texture feature-based method, edge-based method, global and local feature-based method, geometric feature-based method, and patch-based method.

iv) **Classification**: Classification is the final stage of the FER system in which the classifier categorizes expressions such as smile, sad, surprise, anger, fear, disgust, and neutral.

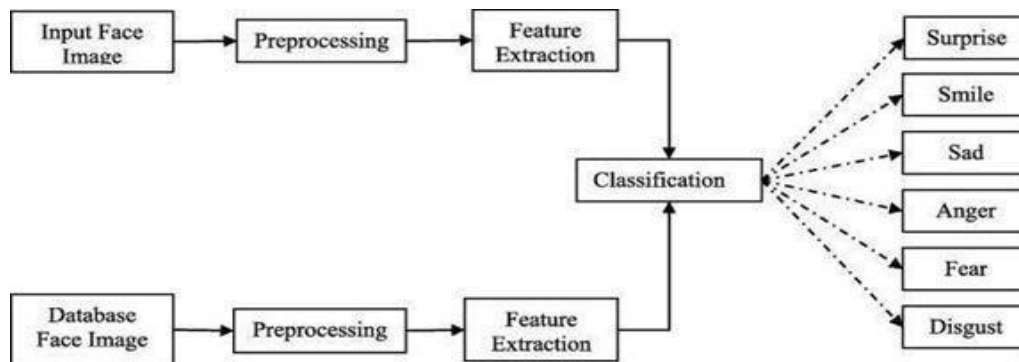


Figure 2: Architecture of Facial Expressions Recognition System.

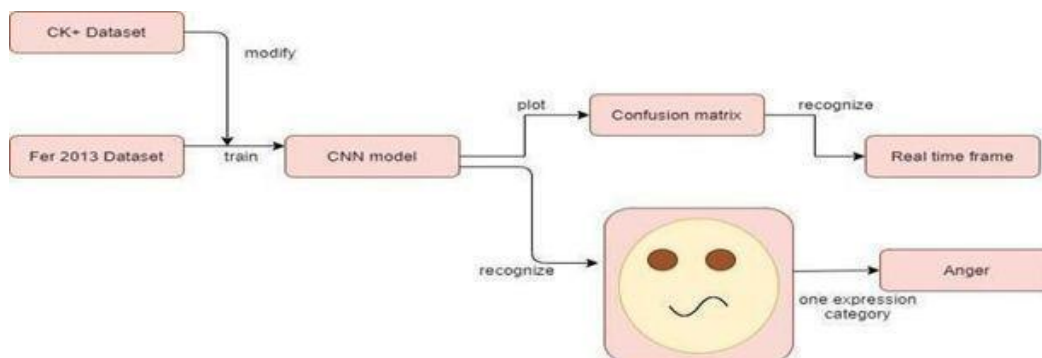


Figure 3: Methodology of Facial Expression Recognition (ROADMAP)

### 3.2 Architecture of the convolutional neural network used in this project:

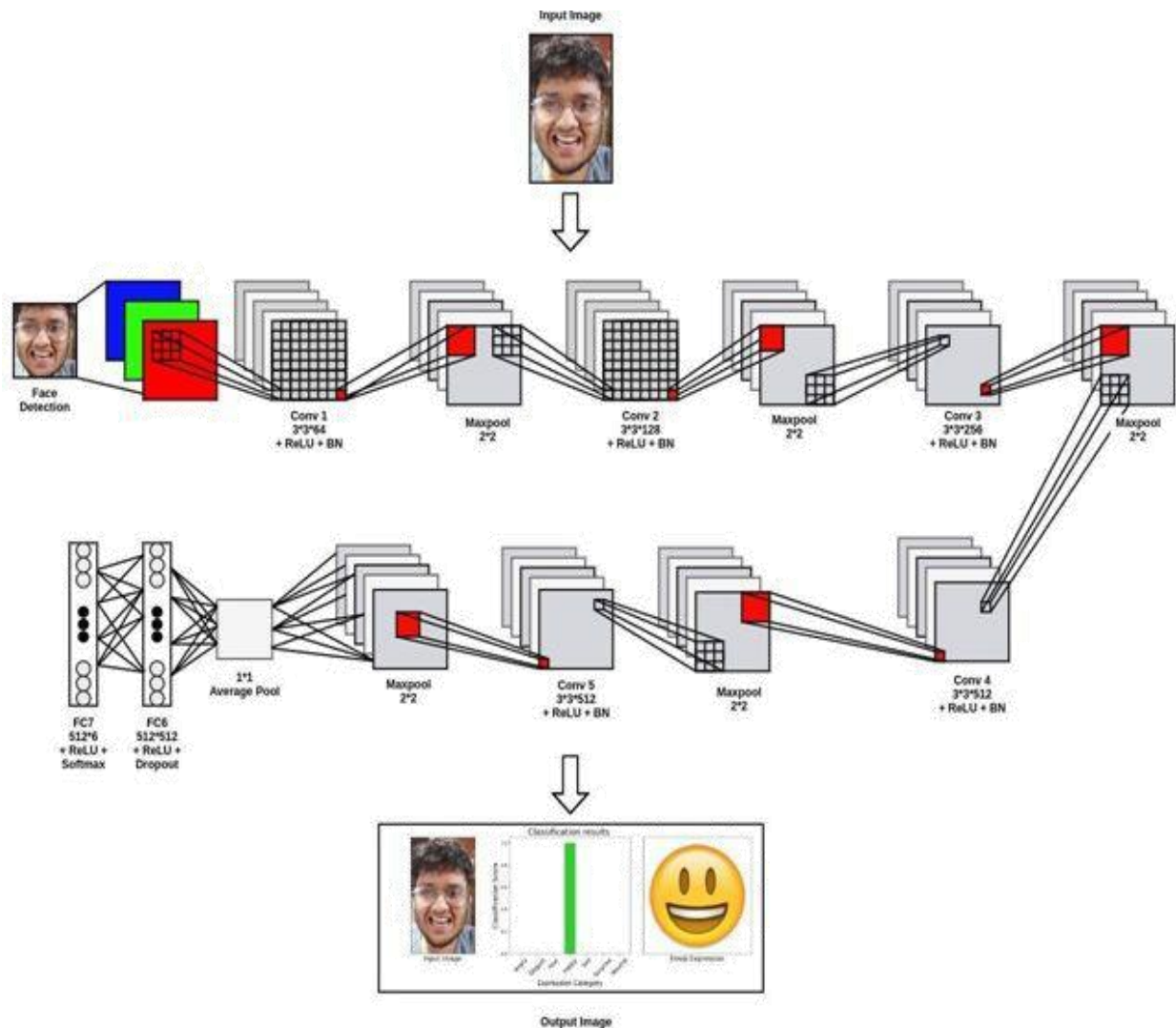


Figure 4: Block diagram of VGG Model and Layers used

Figure 4 shows our VGG Model and the Layers used in it. The input layer is where the input image is sent to RGB Layer, which converts to Grayscale. This is the preprocessing step. After that, 5 convolution layers are used along with max pooling, which comes under the feature extraction, and after feature extraction for classification, fully connected layers are used in which out of the 6 emotions, one emotion is depicted. In all the above layers, the value of dropout used is 0.5.

## 4. IMPLEMENTATION & EXPERIMENTAL RESULTS

**DATASETS USED:** The dataset used for this research is FER2013, introduced in ICML 2013 workshop's facial expression recognition challenge[8]. The dataset contains around 35,887 well-structured 48x48 pixel grayscale images. The dataset is the collection of seven different emotions labelled as follows 0=>Angry, 1=>Disgust, 2=>Fear, 3=>Happy, 4=>Sad, 5=>Surprise, 6=>Neutral. Basically, the whole dataset is in the form of .csv files. The

.csv contains columns as "emotion" and "pixels". The emotion column contains numeric codes ranging from 0 to 6, representing respective emotions. Whereas the "pixel" column contains 2304 fiducial points for each label. In terms of numbers, the dataset has 28,710 examples as a training set. The public and final test sets contain 3587 and 3590 examples, respectively. *Figure 1* depicts one example for each emotion category.



Figure 5: Statistical distribution of images in the dataset that contains 4953 images of anger, 547 images of disgust, 5121 images of fear, 8989 images of happiness, 6077 images of sadness, 4002 images of surprise, and 6198 images of neutral.



Figure 6: Sample of Images in Dataset

In this model, the image is sent to the model, and it gives the corresponding probability graph of emotions depicted from that image called the input image. We have also added emoticons for the expression, which scored the highest category score. The graph is made using the sklearn confusion matrices shown above.

We build Convolutional Neural Networks with different depths to assess the performance of these models for emotion recognition. A CNN architecture consists of all or some of the following layers: The first part of the network refers to the first kind of layers, which usually contains a Convolutional layer with a ReLU activation function and a Max-pooling layer. When we use our first layer for M times, then the network convolves to Fully-Connected layers with Rectified linear unit(ReLU) activation function, and it also includes dropout. In the end, the network is led to a softmax loss function, which is used to calculate the probability. This Convolutional Neural Network model provides us with good scalability and elasticity to personalize the number of Convolutional and Fully-Connected layers, as well as the existence of dropout methods. Furthermore, the number of filters, size of filters, strides, and zero-padding can be specified by users. We used TensorFlow and Keras to build a deeper

Convolution Neural Network. We have explored only limited layers of the model due to the limitation of our computer CPU. Figure 7 depicts the result for test images.

DATASETS	MODEL			
	LBP TOP	3DFCN	VGG19	RESSNT 18
CK+	89.3%	92.4%	94.646%	94.040%
FER 2013	42.32%	55.49%	73.112%	72.973%

Table 1: .Comparison of accuracies between the previous research and our research

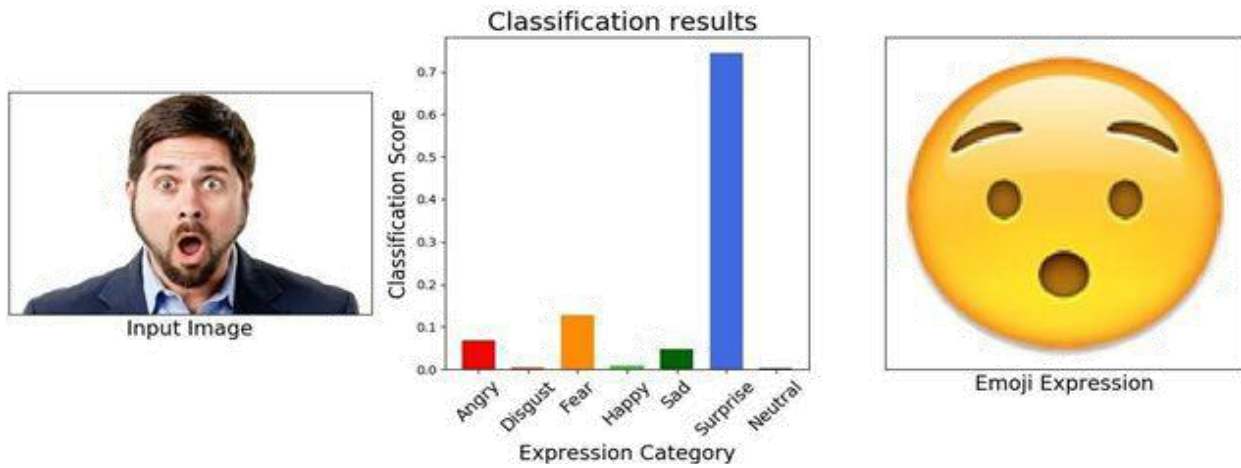


Figure 7: Classification results

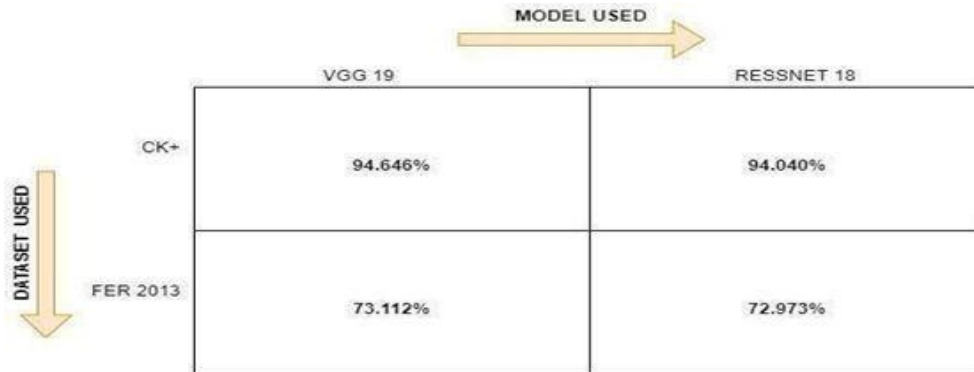


Figure 8: Comparison of Different Models

## 5. CONCLUSION AND FUTURE SCOPE

We have developed a smart vision for the device that can detect a human face and recognize its emotion as well as gender in a single integrated module. This offers human-like abilities to the device using our modern deep convolutional network. Our proposed model is capable of performing the classification of seven different emotions irrespective of the number of faces at a time. We have 2 models, shown below in figure 9. The datasets used are CK+ and FER 2013, and the accuracy achieved is shown in the table. These are the testing accuracy, and also we have implemented gender identification in our code. In the future, we shall tweak the dataset and make it more suitable for recognizing disgust emotion. Given more time, we would also like to use high-performance GPU and combat overfitting, and achieve state-of-art accuracies higher.

## References

- [1] Ververidis, Dimitrios, Constantine Kotropoulos, and Ioannis Pitas. "Automatic emotional speech classification." *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 1. IEEE, 2004.
- [2] Dubey, Arun Kumar, and Vanita Jain. "A review of face recognition methods using deep learning network." *Journal of Information and Optimization Sciences* 40.2 (2019): 547-558.
- [3] Casale, Salvatore, et al. "Speech emotion classification using machine learning algorithms." *2008 IEEE international conference on semantic computing*. IEEE, 2008.
- [4] Clark, Elizabeth A., et al. "The Facial Action Coding System for Characterization of Human Affective Response to Consumer Product-Based Stimuli: A Systematic Review." *Frontiers in Psychology* 11 (2020).
- [5] Yu, Zhiding, and Cha Zhang. "Image-based static facial expression recognition with multiple deep network learning." *Proceedings of the 2015 ACM on international conference on multimodal interaction*. 2015.
- [6] Aneja, Deepali, et al. "Modeling stylized character expressions via deep learning." *Asian conference on computer vision*. Springer, Cham, 2016.
- [7] Ramachandran, Prajit, Barret Zoph, and Quoc V. Le. "Searching for activation functions." *arXiv preprint arXiv:1710.05941* (2017).
- [8] Goodfellow, Ian J., et al. "Challenges in representation learning: A report on three machine learning contests." *International conference on neural information processing*. Springer, Berlin, Heidelberg, 2013..
- [9] Jain, Amit, Jeffrey Huang, and Shiao-fen Fang. "Gender identification using frontal facial images." *2005 IEEE International Conference on Multimedia and Expo*. IEEE, 2005.
- [10] Tümen, Vedat, Ömer Faruk Söylemez, and Burhan Ergen. "Facial emotion recognition on a dataset using convolutional neural network." *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*. IEEE, 2017.