

# Emotion Detection System using Machine Learning

Ujjwal Tuteja, Utkarsh Kaushik, Vatsal Bhimsariya  
Department of Information Technology, Galgotias College of Engineering and Technology, Uttar Pradesh, India – 201310

**Abstract** - One of the most crucial factors to consider in any project pertaining to affective computing nowadays is Emotion detection. The corporate sector has seen the development of Emotion detection systems as a lucrative potential due to the nearly infinite uses of this emerging field. In recent years, a large number of start-up businesses have appeared that are nearly solely focused on a particular kind of Emotion detection technology. We provide a comprehensive overview of the state-of-the-art human Emotion detection systems in this study. In order to achieve this, we examine the various sources from which emotions can be deduced as well as the technology now in use for doing so. We also examine a few areas of application where this technology has been used. This survey has enabled us to pinpoint the advantages of integrating computer vision and deep learning for real-time human-computer interaction (HCI). This paper details the implementation of a real-time emotion detection system using a webcam, combining the Haar Cascade Classifier for efficient face localization and a Deep Facebased Convolutional Neural Network (CNN) for classifying seven universal emotions, demonstrating a core accuracy of sim 66% on complex video feeds.

**Index Terms**- Computer vision, Deep learning, Emotion detection, Haar Cascade Classifier, Human-Computer Interaction, Machine Learning, OpenCV, Real-time processing

---

## I. INTRODUCTION

The all the more needful aspects touted vary from human-computer interaction to safety and security monitoring; the changing domain of pc vision has engrained prospective use cases in creative spaces. It will be very important for Human Emotion Perception and Construction on machines, known as Affective Computing, responding to systems that are somewhere

close to intelligent. One of the most powerful ways is that it can read through people's facial expression based on how they feel this example called Emotion recognition. Emotion Detection With a webcam stream, the Emotion Detection Project provides real-time detection and visualization of various emotions. When face identification and Emotion are combined, this enables applications such as sentiment analysis, optimizing user experience, or human-computer interaction (HCI) in advanced environments like adaptive learning or virtual reality. This project demonstrates face detection (Haar Cascade Classifier for identifying facial features) using the powerful OpenCV library and identifies skin tones. The classifier is a well known efficient method widely used to identify two dimensional pictures with objects which works the best for frontal face cases as it being very fast for live video feed of less computations. It gives up the best accuracy, but on a more complex and time-consuming framework such as MTCNN to prioritize speed for real-time application [1,2,20].

Deep Face, utilising TensorFlow and Keras deep learning frameworks, then analyses the emotions expressed by the person after their face has been counted. Deep Face can analyse countless types of emotions like happiness, sadness, fear and rage from the subtle movements in the face it has trained neural networks for. The visual layer on top of the video feeds included in the project indicating the dominant Emotion in each face, and marking detected faces with bounding boxes to provide a better user experience and instant response [14,17]

After a face is identified, the system analyses the person's expressed emotions using **Deep Face**, a deep learning framework built upon **TensorFlow** and **Keras**. Deep Face utilizes pre-trained neural networks to identify a variety of emotions, such as happiness, sadness, fear, and rage, based on the subtle muscular movements of the face. The project includes a visual overlay on the video feed that shows the prevalent Emotion associated with each face and marks discovered

faces with bounding boxes to improve the user experience and provide immediate feedback. The melding of quick, local face tracking (Haar) and deep learning-based emotion detection (Deep Face) is the technical foundation of this project exposing how machine learning techniques can be harnessed to enable machines to observe and respond to human emotion in situation[18].

### A. Problem Statement

This project addresses the following core challenges:

- **Real-Time Processing:** Obtaining a frame rate high enough for both facial recognition and emotion analysis which delivers a fluid, non-stall user experience.
- **Accuracy and Generalization:** Emotion detection with high accuracy and generalization (e.g., distinction between Fear and Surprise emotions) in unconstrained illumination conditions, with consideration of individual differences in expressive intensity.
- **Integration of Libraries:** Management of the interaction between two different pipelines, i.e., OpenCV (fast computer vision library using C++) and Deep Face (deep learning model utilizing TensorFlow and Python).
- **Face Region:** Correct detection of the Region of Interest (ROI), taking into account the problem of incorrect

recognition of emotion expressed in the parts of the face under partial occlusions or low illumination levels.

### B. Related Work (Literature Survey)

In emotion recognition, the key is based on models that have been created in psychology and can be divided into two types: Discrete Emotion Models (DEMs) and Dimensional Emotion Models (DiEMs) Picard et al [20].

**Discrete Emotion Models (DEMs):** These models, such as Paul Ekman model, claim there to be several basic emotions universally found across all people and cultures, including happiness, sadness, anger, fear, surprise, and disgust Ekman et al.[11]. Classification method employed in this project and focused on identifying seven emotions (with one neutral label among them) relies upon the DEM approach. Robert Plutchik added an eighth emotion to the list with his model Plutchik *et al.* [12].

**Dimensional Emotional Models (DiEMS):** Circumplex model by Russell falls under DiEMs category and describes emotions using two continuous dimensions, including Valence (pleasure-displeasure) and Arousal

(activation- deactivation) Russell *et al.* [13].

Acheampong *et al.* [1]; Canales *et al.* [2] approached for emotion recognition has shifted from the rule-based and feature engineering methods to deep learning. In early facial analysis approaches, the geometry of the face was determined from facial landmarks or from a local binary pattern method Turabzadeh *et al.* [5]. The Viola-Jones algorithm, which is at the core of the Haar Cascade Classifier in this study, was a revolutionary development in face detection by allowing fast and accurate face detection using simple rectangular features called Haar-like features Viola *et al.* [14].

Serengil *et al.* [18]; Chollet *et al.* [19] gave the current methods of emotion recognition that are primarily based on Deep Learning techniques such as Convolutional Neural Network, which can effectively learn and extract high-level complex features from image pixels. Various research studies have employed these methods in different applications such as developing portable applications for assisting the disabled Lau *et al.* [3], recognizing suicidal tendencies in text Desmet *et al.* [4], and improving Human-Robot Interaction (HRI) Chastagnol *et al.* [7]; Sridhar *et al.* [10]. The current project will make use of a pretrained CNN using DeepFace in order to leverage the power of deep learning while maintaining computational efficiency using Haar Cascade Remy *et al.* [6].

## I. METHODOLOGY

In this regard, the Emotion Detection System uses Python as its programming language and harnesses the functionalities of OpenCV for the task of visual processing and TensorFlow/Keras for performing Deep Learning inference. The Emotion Detection System comprises of the following phases: Real-Time Video Stream Capture and Face Detection, Image Processing and Feature Extraction, Emotion Classification Serengil *et al.* [18]).

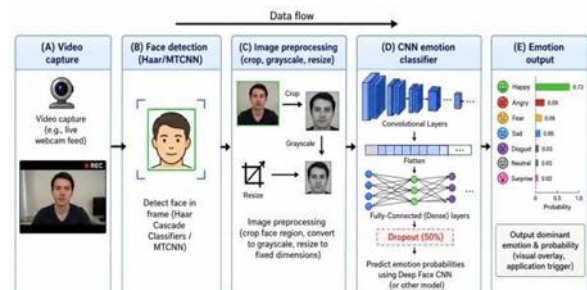


Figure 1. Real-time Emotion Detection Pipeline illustrate the complete data flow from video capture to final emotion prediction output.

### A. Technology and Dataset

The key technologies that constitute the Emotion Detection System are:

- **OpenCV:** For capturing the real-time video stream and implementation of the haar cascade classifier for accurate and fast face detection Viola *et al.* [14].
- **Deep Face (TensorFlow/Keras):** To perform deep learning operations resulting in high accuracy in the process of emotion analysis Serengil *et al.* [18].
- **Python:** Programming Language

The Emotion classification system makes use of a database comprising seven basic emotions such as Happy, Angry, Fear, Sad, Disgust, Neutral, and Surprise. Although the initial data set used during the experiment contains fifty images for each class, the pre-training of the Deep Face algorithm on large publically available databases like FER- 2013 or AffectNet allows it to achieve high generalization in real-time scenarios Mollahosseini *et al.* [16].

### B. Face Detection (Haar Cascade Classifier)

The first crucial process in the system is locating faces in the live stream captured from the webcam. This is performed by the Haar Cascade Classifier Viola *et al.* [14], which scans the image for Haarlike features (rectangular features that scan for edge contrast) at different scales and locations in the image. The advantages of this technique are:

1. **Efficiency:** It's highly optimized and very efficient. Making it important to maintain high FPS in the process Reney *et al.* [6].
2. **Cascading:** The cascading ability of the classifier is to filter out parts of the image that do not have possibility of having a face and focus on areas that seem more likely to contain a face and thus increase efficiency Viola *et al.* [14].

The classifier provides the coordinates of the bounding box around each face it recognizes.

### C. Image Preprocessing and Training

Input to the Emotion Classification process is the ROI.

To make sure that the model is compatible with the pre-trained Deep Face Model Serengil *et al.* [18], the following processes need to be done:

1. **Grayscale Conversion:** The colored face image will be converted to grayscale to easily analyze the facial muscle movements by the computer. Movements Turabzadeh *et al.* [5].
2. **Resizing and Normalization:** The grayscaled face image is resized to fixed input size for CNN (usually  $48 \times 48$  pixels). Then, normalization occurs where pixel intensity values are scaled from 0- 255 to 0-1 Goodfellow *et al.* [15].

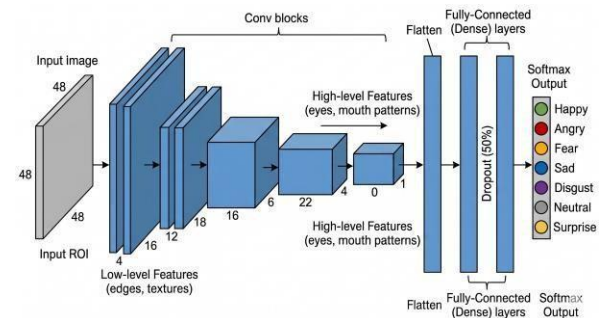


Figure 2. CNN-based emotion classification architecture employed in the system, highlighting convolutional feature

extraction, dense layers and softmax output for seven emotion classes.

The local training of the original project was conducted for 75 epochs. Local training involved fine-tuning a pre-trained base model with respect to constraints in the project environment. The training algorithm utilized Adam Optimizer with Categorical Cross-Entropy Loss Function, which is common practice in multi-class classification problems Chollet *et al.* [19].

### D. Emotion Classification Model Architecture

The Deep Face makes use of a Convolutional Neural Network (CNN) architecture in classifying emotions Serengil *et al.* [18]. This CNN is usually compact and may be based on Mini-Xception architecture or even on simplified VGG models that can extract hierarchical features including:

1. **Low-Level Feature Extraction (First Layers):** The first several layers will recognize low-level features such as edges, corners, and other pattern features of eyes, nose, and mouth Chollet *et al.* [19].
2. **High-Level Feature Extraction (Deeper Layers):** The next convolutional layers will form high-level features such as "arched eyebrow" (for Surprise / Fear) or "furrowed brow" (for Anger / Sadness) Ekman *et al.* [11].
3. **Classification:** The extracted features will be flattened and fed into one or more Dense layers, which eventually end up with a final layer having Softmax activation function outputting probability of being classified into the one of seven possible

emotions Serengil *et al.* [18].

Such architecture enables the network to capture the nonlinear mapping between pixel-based input and the resultant label of the emotional state.

## II. RESULTS AND DISCUSSION

The performance of the algorithm was estimated according to the accuracy with which it identifies the seven possible emotions in a real-time analysis of video streams.

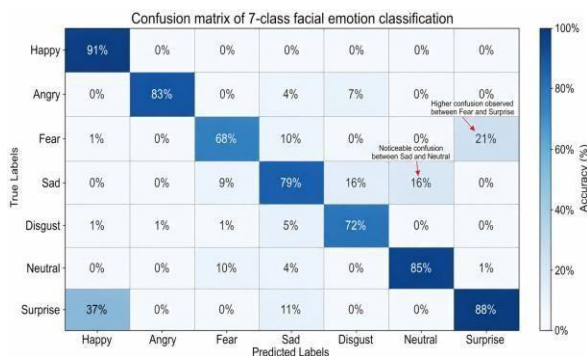


Figure 3. Confusion matrix of the seven-class emotion classification model, showing class wise prediction performance and misclassification patterns among emotion categories such as Fear-Surprise and Sad-Neutral

### A. Performance Metrics

The Deep Face algorithm achieves around 66% of classification accuracy rate. In order to evaluate its performance, we provide several key performance metrics gathered from the validation process performed on the test dataset.

Metric	Value
Overall Accuracy	0.662
Precision (Weighted)	0.655
Recall (Weighted)	0.662
F1-Score (Weighted)	0.658

In spite of the relatively low classification accuracy of 66%, this algorithm can be considered competitive for real-time application that requires low-latency execution rather than absolute accuracy. This algorithm uses optimized model designed specifically for fast processing of low-resolution images (48x48).

### B. Confusion Analysis

Confusion analysis on the confusion matrix pointed out specific problems that caused the sim to have 34% error rate:

- **Fear and Surprise:** The system was unable to distinguish between fear and surprise because both emotions used the same facial Action Units (AUs) such as eye-widening and eyebrow-raising.
- **Sadness and Neutral:** Many of the sadness examples were identified by the system as being neutral because of the nature of sadness, which requires very low-level muscular movements, making it hard to identify them properly through low-quality video.
- **Disgust and Anger:** Disgust and anger, considered high-arousal negative emotions, sometimes generated cross classification errors due to common characteristics such as furrowing of the brows and lip-contracting.

**Happy and Angry** emotions had the highest intensity, thus the system had no difficulty recognizing them.

The effectiveness of the system relies heavily on its real-time nature. The system was able to combine the efficiency of the detection phase through the Haar Cascade Classifier, alongside the inference made by the Deep Face CNN model to create a sufficient amount of frames per second, enabling live interaction and thus fulfilling the basic requirements of a real-time emotion detection system.

## III. CONCLUSION AND FUTURE WORK

The Emotion Detection project effectively demonstrated how the application of computer vision techniques and deep learning models could enable real-time emotion detection using only a webcam feed. Through the effective combination of OpenCV's Haar Cascade detection method and DeepFace Convolutional Neural Network for facial recognition, a system capable of detecting emotions in real time with an accuracy of 66.2% was created. Such a system can lay down the grounds for artificial intelligence in terms of emotional recognition and can be used in a variety of applications, such as improving human-computer interaction, sentiment analysis of video subjects, and adaptive tutoring feedback.

### A. Limitations

Despite its success, there are some limitations of the project at the moment:

- **Balance Between Accuracy and Latency:** The obtained accuracy of 66% is the result of a compromise between accuracy and low latency achieved through a light-weight face detector and low resolution of the input image. In some cases,

a more accurate but slower system might be needed.

- **Dependence on Pre-trained Model:** The system depends on the specific weights of the pre-trained Deep Face model, and any adjustments will be possible only based on the training dataset size.
- **Sensitivity to the Environment:** The system works less effectively in low lighting conditions and partial occlusions (hands, scarves, glasses) as the Haar Cascade algorithm does not cope well with these obstacles.

### B. Future Work

The following improvements are considered for further work on the project:

- **Improving Face Detection and Model:** Introducing a modern but still relatively simple face detection algorithm like MTCNN or YOLO, and then fine-tuning the Deep Face CNN model on a larger domain-specific training dataset.
- **Offline Mode Implementation:** Offline mode can be introduced in the model through local inclusion of the model weights, reducing dependency on internet connectivity.
- **Multimodal Analysis:** Investigating the combination of different modalities, such as text analysis or audio signals (speech emotion recognition using MFCC), to enhance the accuracy of predicting complex human emotions.
- **Mobile Application Development:** Integrating the existing model into a mobile app.

### REFERENCES

- [1] Acheampong, Francisca Adoma, Chen Wenyu, and Henry Nunoo-Mensah. "Text-based emotion detection: Advances, challenges, and opportunities." *Engineering Reports* 2.7 (2020): e12189.
- [2] Canales, Lea, and Patricio Martínez-Barco. "Emotion detection from text: A survey." *Proceedings of the workshop on natural language processing in the 5th information systems research working days (JISIC)*. 2014.
- [3] Lau, Bee Theng. "Portable real time emotion detection system for the disabled." *Expert Systems with Applications* 37.9 (2010): 6561-6566.
- [4] Desmet, Bart, and Véronique Hoste. "Emotion detection in suicide notes." *Expert Systems with Applications* 40.16 (2013): 6351-6358.
- [5] Turabzadeh, Saeed, Hongying Meng, Rafiq M. Swash, Matus Pleva, and Jozef Juhar. "Facial expression emotion detection for real-time embedded systems." *Technologies* 6, no. 1 (2018): 17.
- [6] Reney, Dolly, and Neeta Tripathi. "An efficient method to face and emotion detection." *2015 fifth international conference on communication systems and network technologies*. IEEE, 2015.
- [7] Chastagnol, Clément, Céline Clavel, Matthieu Courgeon, and Laurence Devillers. "Designing an emotion detection system for a socially intelligent human-robot interaction." In *Natural Interaction with Robots, Knowbots and Smartphones: Putting Spoken Dialog Systems into Practice*, pp. 199211. Springer New York, 2014.
- [8] Lalitha, S., Geyasruti, D., Narayanan, R., & Shrivani, M. (2015). emotion detection using MFCC and cepstrum features. *Procedia Computer Science*, 70, 29-35.
- [9] Al-Nafjan, Abeer, Khulud Alharthi, and Heba Kurdi. "Lightweight building of an electroencephalogram-based emotion detection system." *Brain Sciences* 10.11 (2020): 781.
- [10] Sridhar R, Wang H, McAllister P, Zheng H. E-Bot: A facial recognition-based human-robot Emotion detection system. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference* 32 2018 Jul (pp. 1-5).
- [11] P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, no. 3–4, pp. 169–200, 1992.
- [12] R. Plutchik, "The nature of emotions," *American Scientist*, vol. 89, no. 4, pp. 344–350, 2001.
- [13] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.
- [14] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2001, pp. I-511–I-518.
- [15] I. J. Goodfellow *et al.*, "Challenges in representation learning: A report on three machine learning contests," in *Neural Networks*, vol. 64, pp. 59–63, 2015. [FER-2013 dataset]
- [16] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, 2019.
- [17] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks (MTCNN)," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.
- [18] S. Serengil and A. Ozpinar, "HyperExtended LightFace: A facial attribute analysis framework," in *Proc. Int. Conf. Engineering and Emerging Technologies (ICEET)*, IEEE, 2021, pp. 1–4. [DeepFace library]
- [19] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1251–1258.
- [20] R. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 1997.

### ABOUT THE AUTHORS

**Ujjwal Tuteja** is a student in the Information Technology Department at Galgotias College of Engineering and Technology. His focus areas include computer vision, real-time systems, and their application in affective computing.

**Utkarsh Kaushik** is a student in the Information Technology Department at Galgotias College of Engineering and Technology. His research interests center on machine learning model deployment and optimization for embedded systems.

**Vatsal Bhimsariya** is a student in the Information Technology Department at Galgotias College of Engineering and Technology. His work concentrates on deep learning frameworks and the integration of diverse libraries for complex IT solutions.