

# Using Deep Learning to Identify Emotions in Real Time

Dr. Anil V Turukmane<sup>1</sup>

<sup>1</sup>Professor, School of Computer Science and Engineering  
VIT-AP University, Vijaywada, AP, India  
Email: anilturukmane@gmail.com

**Abstract—** One of the most important parts of communication is nonverbal conduct. Body language and other nonverbal cues are often more effective in conveying meaning than the words we use. One example of nonverbal body language that might influence workplace communication is facial expression. It is possible to read a person's emotional state by observing their facial expressions. Computers can be trained to recognize human emotions based on facial expressions. The model is trained using OpenCV from Computer Vision, deepface from Deep Learning's deep CNN, and the human emotions and expressions they reveal can then be captured.

**Index Terms—** Facial expressions, OpenCV, Deep Face, Convolutional Neural Network, ComputerVision.

## I. INTRODUCTION

Daily life brings about a wide range of human emotions. Relationship development and clear communication are both aided by people's emotional states. According to research conducted by Mehrabian [1], facial expressions account for 55% of the emotional message, while written information accounts for only 7% and vocal intonation accounts for 38%. Because of this, facial expressions are widely regarded as a crucial channel in human-computer connection. Human-computer interaction has many potential applications in the realms of entertainment, medicine, and security as a result of technological advancements. Convolutional neural networks are essential for automated facial recognition (CNN). There are several levels in these CNNs. Many filters are used in each layer to accomplish certain goals, such as pattern recognition or data reduction. Convolutional neural networks are useful for analyzing images in many fields, including medicine, social media, and automotive software.

When it comes to facial expressions, the mouth, eyes, eyebrows, etc. are where the bulk of the information is conveyed. Automatic emotion analysis from facial expressions has numerous practical applications, including in intelligent educational systems.

empathetic robots, monitoring for driver fatigue, interactive gaming, and retrieval, categorization, and management of data based on emotions. There are seven emotions that are generally recognized by everyone: anger, disgust, fear, happiness, sadness, neutral, and surprise. There are primarily three phases to an automatic Face detection, facial feature extraction, and emotion recognition are the three main functions of a facial emotion detection system. While Open CV is used for real-time face identification, Deep Learning is utilized for feature extraction and emotion recognition (CNN).

## II. RELATED WORK

Smarter living is possible now that technology can read human emotions and adapt their responses to specific

situations. Humans rely primarily on their facial expressions to convey their feelings. The authors propose that by 2020, [2] Mayur Surve et al. will have perfected a model that can capture real-time images from a camera. After that, several face detection and identification algorithms are put to use. In addition, they developed a single-click user interface (GUI) to collect photographs, shape the dataset, and embed the dataset. The image's face was identified using the Haar cascade algorithm.

Two methods were proposed and compared by Schuller et al. [3]. With the use of generated features from the initial speech signal's pitch and energy curve, the first method classifies the sample's global statistical composition using the Gaussian mixture model. Then, temporal complexity is increased by permanently concealing information in a series of steps. Markov models that take into account a large number of states by relying on low-level characteristics of samples at that time rather than global statistics. This paper shows the framework of modern recognition engines and the outcomes of the suggested approaches. emphasizes the use of audio, visual, and textual information for multi-modal impact analysis, as these three modes appear in about 90% of the relevant literature. Multimodal affect analysis can only be performed if there are multiple channels or modes, such as visual, auditory, textual, gestural, etc.

In a paper by George Karaidakis et al. [5], the authors provided a multimodal method for recognizing eight emotions based on the interplay of facial expressions, body language, and vocalizations.

The model was developed using data from 8 different moods and 10 different people.

Public challenges like FER2013's were created to help get the word out about developments in the field of image-based facial expression recognition, which Yang Y et al. [6] described. The winning entry in the FER2013 competition was one of the earliest convolutional neural network models trained to recognize emotions in human faces. The ensemble CNN model was trained with the square hinge loss as the objective function. Data augmentation throughout both training and testing helped optimize the model.

Balahur Alexandra et al. [7] created a number of models that could recognize seven common emotions from facial expressions, including happiness, sadness, disgust, fear, neutral, anger, and surprise. Their data source was the K+ dataset. An emotion recognition system that relied on a model trained with image data was successful. It would accept an input video stream and continuously recognize faces, classifying the primary emotions of those faces. Although the real-time system can differentiate between some of the seven emotions, it still needs more work to become truly effective outside of the lab.

In this article by Medhat Wlaaa et al. [8], the authors compare and analyze the performance of two established methods for emotion detection (supervised and lexical knowledge based) against a proposed method, grounded in commonsense knowledge stored in the EmotiNet knowledge base. Current methods, which tend to focus on word-level analysis of texts, are able to unearth relatively few emotional overtones.

### III. PROPOSED METHODOLOGY

#### A. Face detection

The field of face detection has recently gained a lot of attention due to the wide variety of uses it has.

Cascaded Haar wavelets for object detection: Paula Viola and Michael Jones created a method for detecting objects called the Haar Cascade classifier. The Haar cascade is a classifier built from a series of Haar-like features. The feature is calculated by subtracting the value of each pixel in the written area from each pixel in the unwritten area. To begin, the face detector uses a 24x24 grid as its foundation. There are around 160k unique Haar-Like Features that can be derived from this initial face detector. Yet, not every one of these options is actively employed.

An example of deriving the Haar value from a rectangular slice of an image has been provided. Pixels with a value of 1 are the darker parts of the haar feature, while those with a value of 0 are the brighter parts. Each of these looks for a different feature in the picture. An example of an abrupt intensity change in an image might be an edge, line, or other image structure.

OpenCV (Open Source Computer Vision Library) is a free, publicly available, and widely used computer vision and machine learning software library. OpenCV's `detectMultiScale()` method, which belongs to the `CascadeClassifier` class, is used to locate human faces in a picture. You can pass in an image as input to this method, and the recognized faces can be stored in a `MatOfRect` object.

In this case, OpenCV is used for the actual implementation. The `object VideoCapture()` is used to record the video feed from the camera as it is being streamed live. The images are recorded and then transformed to grayscale. After that, the grayscale frames were transformed into BGR. Emotion in an image can be identified by capturing a sequence of frames and then applying haar cascade.

#### B. Facial recognition

Human faces are the starting point for facial recognition because they may be used to learn about other people by observing and learning from their unique facial traits and patterns. In order to recognize people's faces in digital photos, DeepFace employs a deep learning technique. It is possible to train facial recognition software to recognize human faces. To do this, deep neural network and machine learning techniques are applied to a database of pictures of people's faces taken from a variety of various perspectives.

Human eye detection is the first step of a procedure that also includes identifying other facial features. The generated model looks to the facial region for clues by calculating the width of the nose, the distance between the eyes, and the shape and size of the lips. Once a face has been identified, computer vision techniques are used to hone the model's ability to recognize distinctive facial features, such as the corners of the eyes, the distance between the eyes, the bridge of the nose, the corners of the mouth, etc. To be exact, we treat each attribute as a nodal point, and there are roughly 80 of them on each face.

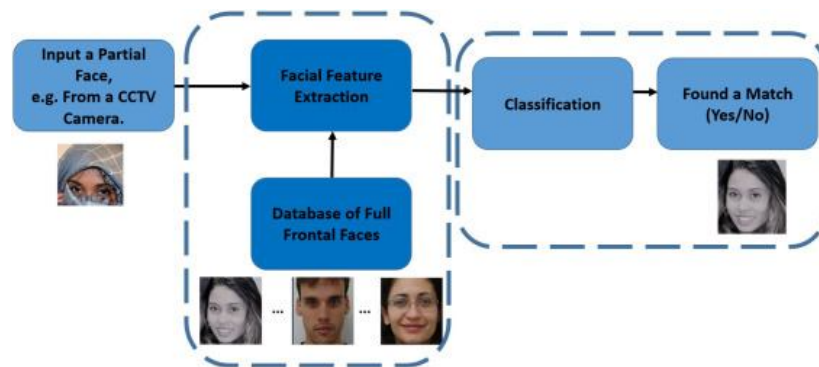


Fig. 3.1 face detection video dataset using deep face

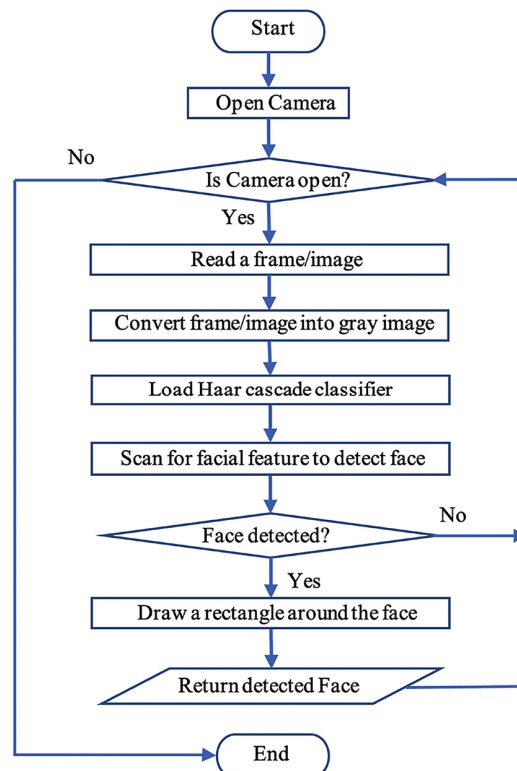


Fig 3.2 Generic Process of Emotion Detection

Here, we utilize Deepface, a Python framework for lightweight face recognition and facial attribute analysis (age, gender, emotion, and race).

TABLE I. NUMERICAL VALUE OF NO OF IMAGES IN EMOTIONS

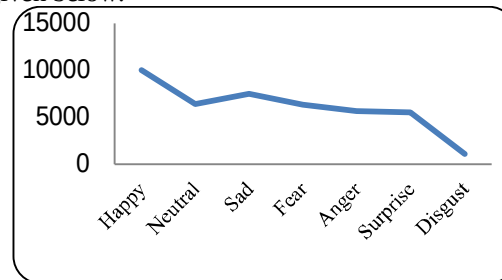
|          |      |
|----------|------|
| Happy    | 9989 |
| Neutral  | 6348 |
| Sad      | 7437 |
| Fear     | 6325 |
| Anger    | 5645 |
| Surprise | 5463 |
| Disgust  | 1093 |

#### IV. IMPLEMENTATION

Face Emotion Recognition 2013 is the dataset utilized for these studies (FER2013). There were a grand total of 27473 possible values used in the development of these sets. Face photos in 48x48 grayscale are the only type of data currently available. Contrarily, there are only 70 instances in the test data set.

OpenCV was used to detect faces. The face detector looks at every part of an image given a live feed in real time. Faces are identified in a succession of camera-extracted shots. The classifier looks to an XML file for guidance on how to label each pixel in a picture. Specifically, Open CV's Haar Cascade classifier file is used to spot XML data. A window is used to display the live feed, and a bounding box is drawn around the largest face, with the associated emotion label shown above the box in real time. Emotions like joy, sorrow, surprise, anger, apathy, and fear are all in the model's database, and it can pick out the one that is most prevalent. These are examples of the system-recognized range of human emotions.

The plot for dominant emotion is given below.



#### VI. CONCLUSION AND FUTURE SCOPE

The goal of this study was to utilize deep learning to develop face emotion identification that can identify the prevalent emotion. We've shown how to put into action a deep learning architecture that can read and predict human expressions and emotions. A deep face model is trained and evaluated to carry out this task. Improved accuracy is anticipated to result from utilizing a wider variety of data augmentation techniques, a deeper network, and the use of regularization.

Occlusion sensitivity, or the ability to deduce a person's emotional state from a photograph of their face even if part of their face is obscured, is a potential future direction for this study. This work will pave the way for an improved version of the model that can read and interpret a wide range of human facial expressions, regardless of lighting or background.

#### REFERENCES

- [1] Mehrabian. Communication without words, Psychological today, vol. 2, p. 53–55, 1968. [1] Duncan, Dan, Gautam Shine, and Chris English. "Facial emotion recognition in real time." Stanford University (2016)
- [2] J. D. Bodapati and N. Veeranjanyulu, "Facial Emotion Recognition Using Deep Cnn Based Features", 2019 International Journal of
- [3] Schuller, G. Rigoll M. Lang, "Hidden Markov model-based speech emotion recognition", 2003, Proceedings. International Conference on Multimedia and Expo, volume 2, pp. 401-404
- [4] Poria S, Cambria E, Bajpai R & Hussain A (2017) A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion, 37, pp. 98-125.

- [5] Turukmane, Anil V. "Forecasting the IoT-based cyber threats using the hybrid forage dependent ensemble classifier." *Concurrency and Computation: Practice and Experience* (2023): e7460.
- [6] Caridakis George, Castellano Ginevra, Kessous Loic, Raouzaïou Amaryllis, Malatesta Lori, steriadis Stelios, Karpouzis Kostas "Multimodal emotion recognition from expressive faces, body gestures and speech", 19 September 2007, IFIP *Advances in Information and Communication Technology (AICT)*, Volume 247, Issue 247, pages 375-388.
- [7] Sudhakar, P., Kamble, N. K., Geetha, K., Turukmane, A. V., Perli, S. B., & Jayaraman, P. (2023). Faulty diagnostics model for wind power plant application using AI. *Measurement: Sensors*, 25, 100621.
- [8] Balahur Alexandra, Hermida Jesus M, Montoyo A. "Detecting implicit expressions of emotion in text: A comparative analysis", November 2012 *Decision Support Systems*, 53 (4), pp. 742-753
- [9] Walaa Medhat, Ahmed Hassan, Hoda Korashy, *Sentiment analysis algorithms and applications: A survey*, Ain Shams Engineering Journal, Volume 5, Issue 4, 2014, Pages 1093-1113.