Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

# Semantic Feature Extraction and Deep Convolutional Neural Network-based Face Sentimental Analysis

## Mishmala Sushith

Associate Professor & Head, Department of Information Technology, Adithya Institute of Technology, Coimbatore

**E-mail:** mishmala@gmail.com

### Abstract

Police and government agencies make use of facial recognition technology in order to determine the truth about the criminal. Though this might seem skeptical, it requires the support of the public in order to come into action. In this regard the media plays a major role in molding the public agenda and attracting sentiments and attitudes towards this topic. In this work, various perspectives are taken into consideration in order to determine the impact of social media on the public, police and government with the help of face recognition technology. A total of 443 videos have been analyzed and the outcome showed to be positive for this technology to be incorporated. Close examination of emotional language indicated several levels of anticipation and surprise along with fear and sadness. It is worth noting that trust is in emotion expressed in low levels only. Deep learning based CNN technique is used for categorization. Based on the information obtained and by incorporating new methodologies, conclusions are drawn, strategies are incorporated and recorded.

**Keywords:** Deep learning, convolution neural network, sentiment analysis, emotion recognition, image processing

### 1. Introduction

In everyday human interactions emotions are self explanatory and efficacious. The facial expressions are the primary projection of human emotions [1] and are easily noticeable. However facial expression recognition is a serious process and complex in nature. This fines application in several areas such as human-computer, emotionally driven robots and healthcare [2]. Technological advancement in FER show cases and increase in its performance, but attaining high efficiency remains a big challenge. Surprise, fear, disgust,


*DOI: https://doi.org/10.36548/jiip.2022.3.003*
*Received: 28.07.2022, received in revised form: 19.08.2022, accepted: 05.09.2022, published: 15.09.2022*

sadness, happiness and anger are 6 of the most generic human emotions. Along with this another basic emotion called contempt is added. FER is a critical analysis task and the parameters chosen determine the accuracy of the outcome [3]. Occlusion illustration factors etc are such parameters which include obstruction on the face like sunglasses age and hand. These parameters are taken to consideration by the researchers when they carry out the FER model in order to attain a high level of accuracy. In FER some of the important parameters include:

- **Occlusion:** in the presence of occlusion it is difficult to extract features from the image at the occluded part because of precise feature location and accurate face alignment.
- **Expression Intensity:** This is based on the intensity of the expression. Here when expression is less facial, it is possible to recognise the expression in a more accurate manner.
- **Illuminator Factor:** Classification of the model is affected by the light intensity that is displaced on the object. Due to increase in contrast or decrease in distance between classes false acceptance rate is high with respect to textual values.

FER system is either dynamic or static [4], depending on the image such that the Dynamic Image FER uses continuous frames in temporal information and static FER uses feature representation of one image with face point location information.

## 2. Literature Review

The term FER was initially coined in the mid 1980s [5]. Ever since then several methodologies such as artificial neural networks, random forest classifiers etc., have been incorporated in order to identify the basic 7 emotions [6]. Such as have also been promising in producing effective and good outcomes. In the field of surveillance and security applications human emotion detection is an important part. To accurately identify these emotions research is being carried out in this field [7]. Overfitting of models occlusion in data sets etc are some of the challenges that are faced while incorporating FER. In recent days some FER approaches use AI perception in order to identify emotions in an exhaustive manner. According to the literature surveys and research outcomes analysed by several authors, the artificial intelligence methodology [8] is incorporated in this paper with respect to machine learning and DL approaches [9].

One of the most important aspects in the computer vision area is Facial Sentiment Analysis [10]. The number of research on these topics have been published by several researchers from around the globe. However the issues and challenges in FER are being researched by many researchers resulting in the publication of temples and surveys on sentiment analysis. These techniques have focused primarily on artificial neural networks (ANN), decision tree classifiers (DTC) and support vector machines (SVM). The DL methods have not been predominantly researched by researchers. Hence facial sentiment is analysed and a comparative note is further presented in this work. In [11] the authors have observed several methodologies for classification of FER, facial feature extraction and facial detection. The authors in [12] and [13] have also presented a survey on FER. However they fail to represent the data set used during their survey process to determine the emotions surveyed. Similarly several of the traditional feature extraction methodologies like Locally Linear Embedding (LLE), Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) have been analysed and compared followed by the proposal of an ensemble classifier. However, the DL approach was not compared and is also the technique that is used by FER. Authors in [14] also examine the various features in details for facial recognition technology. In recent years, FER approaches which are DL-based have been explored in [15].

## 3. Proposed Methodology

Facial recognition, law enforcement, entertainment, politics, news, education and cooking are the keywords used for choosing 443 videos from YouTube during data selection. The data is collected for a duration of one year from January 2021 to December 2021. Videos that did not suit the requirements of the research were removed from the database through a manual check. In late 2017, this technology was used by the police for certain investigations and had been trialed and adopted in several countries. 500 YouTube videos were identified in the data selection process of which, 53 were eliminated as they did not contain the essential features for analysis. The remaining 447 videos, underwent an analysis process during which two of them were removed by the host. Further, two more videos were removed as the content was different despite the title. 443 videos were retained in the final data set. The length of the videos ranged from 20 seconds to 59 minutes. The overall videos ran for a duration of 15 hours, 18 minutes and 52 seconds while the average video length was about 4 minutes and 34 seconds. The videos were mostly focused on the Indian regional languages while the majority were from the southern regions of India. Seven major keywords are used for identifying the videos. The findings and inaccuracies are recorded for further analysis and application.

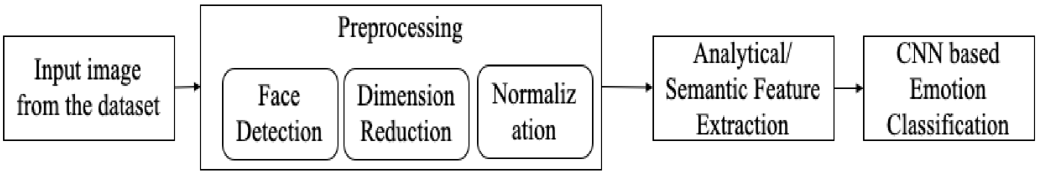


**Figure 1.** Block diagram of the emotion classification module

Figure 1 provides the block diagram of the proposed emotion classification module. The input image from the dataset is provided to the module where pre-processing is performed with face detection, dimension reduction and normalization. Further, the pre-processed data undergoes semantic and analytical feature extraction. The extracted features are fed to the deep learning-based CNN emotion classification module that categorizes the emotions as joy, surprise, trust, anticipation, sadness, anger, fear, disgust and neutrality. Jupyter notebook and python script was used in the Google developer pannel to access the closed caption data feature for the YouTube videos that were accessed. The data was converted into csv files using unique identification (UIDs) that were assigned to the list of short URLs for each video.

Each video is considered as a unit of analysis. The NRC lexicon and syuzhet package are used for conducting the sentiment analysis. In the VOSOSN Dash application, for Reddit, YouTube, and Twitter data analysis, the NCR lexicon and syuzhet packages are used successfully as seen in the literature. Commercial entities, public institution, political figures and other such target entities and the sentiment towards them can be tracked in a codified way using the NRC lexicon model. With this, it is also possible to identify the kind of emotion that a media article or newspaper headline evokes in the audience. Each word is assigned with a sentiment or emotion by participants through a crowd sourcing platform while assembling the lexicon. Surprise, disgust, anger, trust, fear, anticipation, joy and sadness are the eight major emotions contained in the lexicon. The deep learning-based CNN emotion classification module categorizes images based on the observed emotions.

## 4. Experiments and Results

Mostly positive emotions and sentiments were observed all the videos in the overall language as shown by the sentiment analysis. On average, the overall positive videos were found to be a higher value of about 56% from the sentiment analysis performed. The negative sentiment expressed in the videos was around 39% on average, which is more than one-third of the total videos. Around 5% of the videos expressed neutral emotions. Figure 2 expresses the comparison of the analyzed emotions namely joy, surprise, trust, anticipation, sadness, anger, fear, disgust and neutrality. The highest average achieved with joy is about 31%, while sadness is 14%. Surprise, trust, and anticipation were at 6%, 7% and 12% respectively. Further, anger, fear and disgust were recorded at 10%, 11% and 4% respectively.

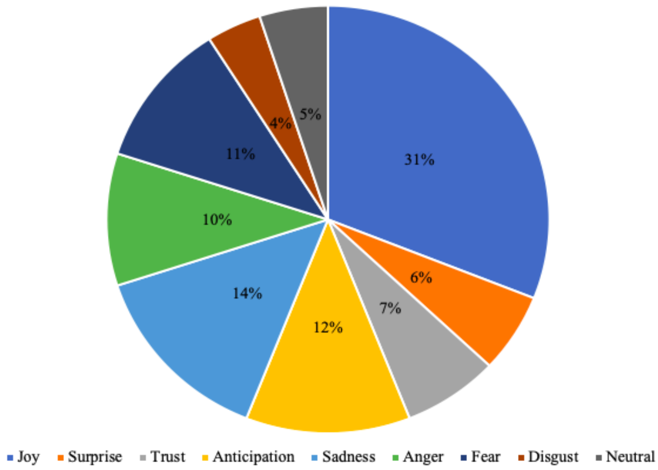


**Figure 2.** Comparison of the percentage of overall emotions

### 4.1 Classification of channels

Most of the channels were classified as individual posts which contributed to 45% and media outlet depiction contributed to about 27%. The channel classification showed an overall positive sentiment according to the result of the analysis. Further, categories such as entertainment, politics, news, education, cooking, and so on were analyzed on an individual basis to estimate the emotions in those videos. It is observed that during entertainment, education and cooking, the overall positive emotions are higher, while it is compromised with politics and news.

Over a period of one year (2021), the data is captured across 443 videos and the sentiment expressed was more positive. This data is tabulated in table 1. The comparison of frequency in which the videos were posted is shown in figure 3. The videos are taken from Tamil Nadu, Kerala, Karnataka, Andhra Pradesh, Maharashtra and Delhi. It is also observed that the number of videos posted during April is high owing to the summer vacation time in these regions. The image pixel values are provided as input to the CNN module. The convolution, polling, and fully connected layers analyse and categorize emotions. On average, the overall videos are higher between April to December when compared to January to March.

**Table 1:** Sentiment analysis for youtube thumbnail images

| Category | Joy | Suprise | Trust | Anticipation | Sadness | Anger | Fear | Disgust | Neutral | Positive Sentiment % | Negative Sentiment % | Neutral sentiment % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Tamil Nadu** | 40.92 | 17.98 | 13.02 | 20.77 | 22.01 | 22.63 | 40.92 | 17.98 | 13.02 | 56 | 39 | 5 |
| **Kerala** | 7.92 | 3.48 | 2.52 | 4.02 | 4.26 | 4.38 | 7.92 | 3.48 | 2.52 | 54 | 41 | 5 |
| **Karnataka** | 9.24 | 4.06 | 2.94 | 4.69 | 4.97 | 5.11 | 9.24 | 4.06 | 2.94 | 49 | 45 | 6 |
| **Andhra Pradesh** | 15.84 | 6.96 | 5.04 | 8.04 | 8.52 | 8.76 | 15.84 | 6.96 | 5.04 | 61 | 34 | 5 |
| **Maharashtra** | 18.48 | 8.12 | 5.88 | 9.38 | 9.94 | 10.22 | 18.48 | 8.12 | 5.88 | 59 | 33 | 8 |
| **Delhi** | 13.2 | 5.8 | 4.2 | 6.7 | 7.1 | 7.3 | 13.2 | 5.8 | 4.2 | 52 | 42 | 6 |

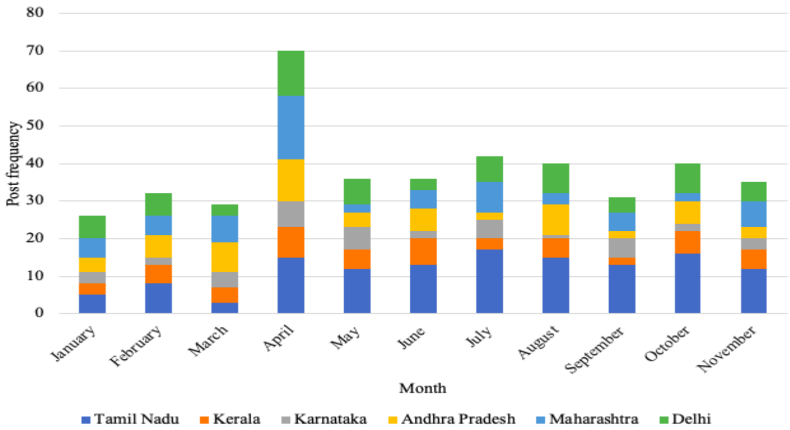


**Figure 3.** Post frequency for one selected topic among the pool of videos

Illumination factor, expression intensity and occlusion contribute to the quality of the image, which is also a crucial factor in the appropriate identification of emotion. The expressions also may vary from region to region which is another significant challenge in sentiment analysis. This can be overcome with a deep learning technique where analysis is done region-wise by first identifying the region to which the person may belong based on their skin tone and facial structure rather than performing a general overall analysis. A video outburst is also observed when there is sensational news in the media that affects the entire nation. Some of the issues faced while performing emotion recognition include occluded data and occlusion, the difference in modalities, infrared data, visualization mismatch, and fake emotion recognition.

ISSN: 2582-4252

## 5. Conclusion

When there is a relatively small amount of information on a specific topic, the proposed emotion recognition technique provides additional data in investigating the emotions of people which can be used for several applications. A unique perspective and perception are obtained on the topic using this technique. The police can make use of this technique while investigating criminals, the government can make use of it to understand how the citizens respond to new changes, and hospitals can make use of it in studies related to psychology and psychiatry. Despite an overall discourse provided by sentiment analysis, the technology has certain limitations. In this work, YouTube video thumbnails are used for analyzing the emotions from a total of 443 videos released over a period of one year. Deep learning based CNN technique is used for this purpose. In a reasonable timeframe, larger sample sizes can be analyzed using the automation technique. Enhanced replaceability with minimal need for additional code, interoperability, and reusability are the major benefits of this technique. This work has demonstrated that in spite of data limitations, the data sources can be utilized in a beneficial manner. A comprehensive and complete understanding of various data sources can be analyzed in future research. This can also be extended to analyzing the public comments on the videos to further understand the emotions of the public, enabling the authorities such as police and the government to address the concerns of the citizens.

## References

[1] Andrejevic, M., & Selwyn, N. (2020). Facial recognition technology in schools: critical questions and concerns. Learning, Media and Technology, 45(2), 115-128.

[2] Smeets, D., Claes, P., Vandermeulen, D., & Clement, J. G. (2010). Objective 3D face recognition: Evolution, approaches and challenges. Forensic science international, 201(1-3), 125-132.

[3] Dabbaghchian, S., Ghaemmaghami, M. P., & Aghagolzadeh, A. (2010). Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology. Pattern recognition, 43(4), 1431-1440.

[4] Ciftci, O., Choi, E. K. C., & Berezina, K. (2021). Let’s face it: are customers ready for facial recognition technology at quick-service restaurants?. International Journal of Hospitality Management, 95, 102941.

[5] Shao, X. F., Li, Y., Suseno, Y., Li, R. Y. M., Gouliamos, K., Yue, X. G., & Luo, Y. (2021). How does facial recognition as an urban safety technology affect firm performance? The moderating role of the home country's government subsidies. Safety science, 143, 105434.

[6] Hansen, M. F., Smith, M. L., Smith, L. N., Salter, M. G., Baxter, E. M., Farish, M., & Grieve, B. (2018). Towards on-farm pig face recognition using convolutional neural networks. Computers in Industry, 98, 145-152.

[7] Jiang, D., Hu, Y., Yan, S., Zhang, L., Zhang, H., & Gao, W. (2005). Efficient 3D reconstruction for face recognition. Pattern Recognition, 38(6), 787-798.

[8] Jadhav, D. V., & Holambe, R. S. (2009). Feature extraction using Radon and wavelet transforms with application to face recognition. Neurocomputing, 72(7-9), 1951-1959.

[9] Phillips, P. J., & O'toole, A. J. (2014). Comparison of human and computer performance across face recognition experiments. Image and Vision Computing, 32(1), 74-85.

[10] Xu, Y., Zuo, W., & Fan, Z. (2012). Supervised sparse representation method with a heuristic strategy and face recognition experiments. Neurocomputing, 79, 125-131.

[11] Li, Y., Zheng, W., Cui, Z., & Zhang, T. (2018). Face recognition based on recurrent regression neural network. Neurocomputing, 297, 50-58.

[12] Bah, S. M., & Ming, F. (2020). An improved face recognition algorithm and its application in attendance management system. Array, 5, 100014.

[13] Xu, Y., Zhong, A., Yang, J., & Zhang, D. (2010). LPP solution schemes for use with face recognition. Pattern Recognition, 43(12), 4165-4176.

[14] Garaus, M., Wagner, U., & Rainer, R. C. (2021). Emotional targeting using digital signage systems and facial recognition at the point-of-sale. Journal of Business Research, 131, 747-762.

[15] Shi, L., Wang, X., & Shen, Y. (2020). Research on 3D face recognition method based on LBP and SVM. Optik, 220, 165157.

## Author's biography

**Mishmala Sushith** is currently working as an associate professor and head in the Department of Information Technology, Adithya Institute  of Technology, Coimbatore India. Her area of research includes image processing and networks.

 