Journal of Industrial Information Technology and Application
Vol.9 No. 3
ISSN(Online): 2586-0852

Facial Expression Analysis for Emotion Detection Using Convolutional Neural Networks

Vikas Jangra, and Sumeet Gill)

Department of Mathematics, M.D. University Rohtak

Abstract. Emotion is a way humans express their feelings, typically through facial expressions, body language, and tone of voice. Among these, facial expressions are the most powerful, natural, and universal signals for conveying emotional states. However, recognizing emotions based on facial expressions can be challenging due to the similarities between certain expressions. For example, fear and surprise often share overlapping patterns, making it difficult to distinguish them with the naked eye. This study focuses on developing a mobile-based application for realtime emotion recognition using facial expressions. It employs a Deep Learning approach, specifically a Convolutional Neural Network (CNN), with the MobileNet algorithm utilized to train the recognition model. The application is designed to identify four types of emotions: happy, sad, surprised, and disgusted. The results demonstrate a high level of accuracy, achieving state-of-the-art performance. Future improvements could involve expanding the range of facial expression categories or exploring alternative deep learning approaches to enhance the model's capabilities.

Keywords; Emotion Recognition, Facial Expressions, Convolutional Neural Networks (CNN), MobileNet Algorithm, Real-Time Application

Cite this paper as : Vikas Jangra, and Sumeet Gill (2025) “Facial Expression Analysis for Emotion Detection Using Convolutional Neural Networks”, Journal of Industrial Information Technology and Application, Vol. 9. No. 3, pp. 1122 - 1129

*Corresponding author: drsumeetgill@mdurohtak.ac.in
Received: Mar. 17. 2025 Accepted: Jun. 24. 2025 Published: Sep. 31. 2025

1. Introduction

Facial expressions serve as a powerful medium to convey human emotions, offering valuable insights into an individual's feelings. The ability to express emotions is an innate characteristic of humans, and people often use emotions as a direct means of communication. Recognizing human emotions is a critical aspect of human-computer interaction, enabling more intuitive and responsive systems. However, emotions are highly influenced by an individual's physical condition and mental state, adding complexity to their interpretation. While facial expressions are a primary method of expressing emotions, the similarity in patterns across different emotions, such as fear and surprise, poses a significant challenge in accurately identifying them. Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have demonstrated exceptional performance in emotion recognition tasks, offering high accuracy and fast processing speeds. CNNs have been extensively utilized in automatic emotion recognition, achieving impressive accuracy rates exceeding 90%. Recognizing the immense potential of CNNs, this study proposes a mobile-based emotion recognition system leveraging CNN technology for real-time performance. In comparison, K-Nearest Neighbors (KNN) has also been widely applied in emotion recognition, achieving accuracy levels above 85%. However, despite its effectiveness, KNN has notable limitations, including high memory requirements and slower performance, which make it less suitable for real-time applications.

2. Methodology

This study starts with inputting the real-time images, followed by the implementation of Convolutional Neural Network (CNN) for recognizing the emotion. Succeeding, the recognized emotion will be displayed.

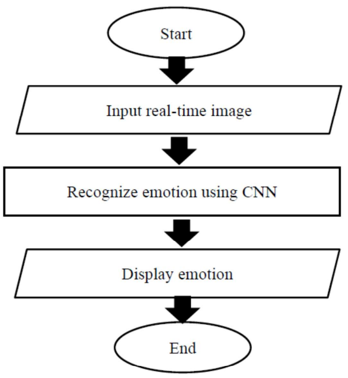

Figure 1. Proposed flowchart of this study

2.1 Facial Expression Real-Time Images

This study covers four types of facial expression which are happy, sad, surprise and disgusting. Figure 2 shows example of these four types of facial expression.

Figure 2. Facial expression

2.2 Recognize Emotion Using Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) are a deep learning-based technology renowned for their ability to achieve high precision in recognition tasks, as highlighted by Liam Schonevel (2021). CNNs consist of multiple layers, each designed to perform specific transformation functions essential for feature extraction and classification.

The convolutional layer is the first layer in a CNN, responsible for extracting features from input images. It retains the relationships between pixels by learning image features through small regions of input data. By applying various filters, the convolutional layer can perform operations such as edge detection, blurring, and sharpening, which are critical for understanding image patterns.

To enhance the model's learning capabilities, the Rectified Linear Unit (ReLU) activation function is employed. ReLU introduces non-linearity into the network, enabling it to model complex relationships within the data. Additionally, it ensures that the ConvNet focuses on non-negative linear values, aligning with the characteristics of real-world data. This combination of layers and transformations makes CNNs highly effective for emotion recognition tasks.

The pooling layer plays a crucial role in reducing the number of parameters when processing large images. This layer helps simplify computations while retaining the essential features of the image.

Spatial pooling, also known as subsampling or downsampling, reduces the dimensionality of feature maps while preserving critical information. It can be performed using different methods, including max pooling, average pooling, or sum pooling, depending on the specific requirements of the task.

The fully connected layer, on the other hand, takes the output from the previous layers and flattens the feature maps into a one-dimensional vector. This vector is then fed into a fully connected neural network for further processing and classification. The figure illustrates the overall architecture of a CNN, highlighting these interconnected layers.

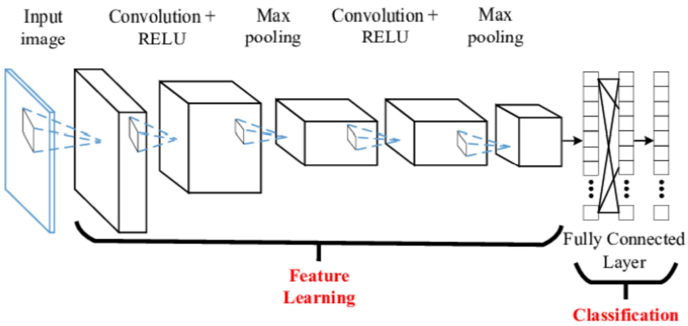

Figure 3. CNN Architecture

a) Training the CNN Model

Before initiating the training process, all images undergo preprocessing on the Roboflow platform using an augmentation process. This process involves applying transformations such as horizontal and vertical flipping, rotation, saturation adjustments, and blurring. These augmentations are designed to enhance the dataset and improve the model's ability to generalize.

Once the augmentation is complete, the training process begins. The model achieved its highest accuracy of 80% during training, which was conducted over 100 epochs. The number of epochs significantly impacts the model's accuracy. If the number of epochs is less than 100, the accuracy fails to reach 80%, potentially leading to suboptimal recognition performance.

This analysis aims to determine the most suitable and effective ratio for splitting images into separate sets for training, testing, and validation. The findings reveal that the optimal ratio is 90% for training images and 5% each for testing and validation images. This configuration achieved the lowest error rate of 19.56%, making it the most effective split for this study.

Table 1. Analysis of the Accuracy of Training

	Splitting images (%)	Accuracy (%)	Loss (%)	Error Rate	Error Rate (%)
Train	**90**	**94**	**53**		
Test	**5**	**80**	**94**	**9/46**	**19.56**
Validation	**5**	**84**	**66**		
Train	80	94	54		
Test	10	78	79	19/91	20.87
Validation	10	89	66		
Train	70	93	53		
Test	15	78	79	28/100	28
Validation	15	79	79		

3. Result and Analysis

The accuracy of the application is evaluated based on emotions detected from facial expression images captured through a mobile phone camera. For this evaluation, a total of 80 images were used during the testing phase.

As shown in Figure 2, some instances resulted in incorrect predictions, where the application failed to accurately identify the emotion. These errors occur because the application misinterprets the facial expression in the image. For example, while the expected emotion might be "sad," the application could incorrectly recognize it as "happy." This misclassification is primarily due to the similarities between certain facial expressions, which make accurate recognition challenging.

To conclude overall accuracy performance, the average accuracy is calculated. Equation 1 shows the formula for accuracy calculation. The overall accuracy result for this application is 85%.

$$Accuracy = \frac{Number\ of\ correct\ prediction}{Total\ number\ of\ all\ cases} * 100$$

3.1 Confusion Matrix

The accuracy testing of the application is conducted using 80 testing images which composed of 20 images from each emotion category. Table 2 tabulates the confusion matrix result from the testing conducted.

Table 2. Summarization of confusion matrix result

		Actual				
		Happy	Sad	Surprise	Disgusting	Total
Predicted	Happy	**18**	0	1	1	20
	Sad	1	**16**	0	3	20
	Surprise	1	0	**17**	2	20
	Disgusting	2	0	1	*17*	20
	Total	22	16	19	23	**80**

Table 3. Results of True False Prediction for emotion class

	Happy	Sad	Surprise	Disgusting
True Positive (TP)	18	16	17	17
True Negative (TN)	56	60	58	54
False Positive (FP)	4	0	2	6
False Negative (FN)	2	4	3	3

3.2 Sensitivity and Specificity Calculation

Sensitivity (also known as Recall or True Positive Rate) measures the proportion of actual positive samples that are correctly identified by the model.

Specificity (also known as True Negative Rate) measures the proportion of actual negative samples that are correctly identified.

These metrics allow for a comprehensive evaluation of the model's performance on each emotion class, providing insights into how well the model identifies both positive and negative instances for each emotion, which is especially critical in emotion recognition tasks. To calculate the sensitivity, and specificity for each emotion class, the following formulas are used:

$$Sensitivity = \frac{TP}{TP\ \ FN}$$

$$Specificity = \frac{TP}{TP\ \ FN}$$

Table 4. Results for Accuracy, Sensitivity and Specificity Calculation for each Class of Emotion

Type of Emotion	Accuracy(%)	Sensitivity(%)	Specificity(%)
Happy	92.50	90	93.33
Sad	95.00	80	100.00
Surprise	93.75	85	96.66
Disgusting	88.75	85	90.00
Average	**92.50**	**85.00**	**95.00**

4. Conclusion

An emotion recognition application utilizing Convolutional Neural Network (CNN) was successfully developed. This application is capable of identifying four distinct emotions: happy, sad, surprised, and disgusted. The CNN model employs the MobileNet algorithm trained on a custom dataset and was evaluated using confusion metrics. The application demonstrated impressive performance, achieving an average accuracy of 92.50%. Additionally, it attained a sensitivity of 85.00% and a specificity of 95.00%, highlighting its effectiveness in accurately recognizing emotions.

References

[1] Liam Schoneveld, Alice Othmani, Hazem Abdelkawy. (2021). Laveraging recent advances in deep learning for audio-Visual emotion recognition. Pattern Recognition Letters 146, 1-7.

[2] Yishu Liu, Guifang Fu. (2021). Emotion recognition by deeply learned multi-channel textual and EEG features. Futures Generation Computer Systems 119, 1-6.

[3] Neha Jain, Shishir Kumar, Amit Kumar, Pourya Shamsolmoali, Masoumeh Zareapoor. (2018). Hybrid deep neural networks for face emotion recognition. Pattern Recognition Letters 115, 101-106.

[4] Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, Remigiusz J. Rak. Emotion recognition using facial expressions. International Conference on Computational Science (2017).

[5] Shervin Minaee, Amirali Abdolrashidi. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network (2019).

[6] Nithya Roopa. S. Emotion Recognition from Facial Expression using Deep Learning. International Journal of Engineering and Advanced Technology (IJEAT) (2019); 8(6S):.

[7] Arkaprabha Bhandari, Nikhil R Pal. (2021). Can edges help convolutional neural networks in emotion recognition?. Neurocomputing 433, 162-168.

[8] Poorna S S, Anjana S, Parvathy Varma, Anjana Sajeev, Arya K Ca, Renjith Sb , G J Nair a. Facial Emotion Recognition using DWT based Similarity and Difference features. Proceedings of the Second International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (2018).

[9] Firoz Mahmud, Bayezid Islam, Arfat Hossain, Pushpen Bikash Goala. Facial Region Segmentation Based Emotion Recognition Using K-Nearest Neighbors . Proceeding of the International Conference on Innovation in Engineering and Technology (ICIET) (2018).

[10] Marina Chatterjee. A Quick Introduction to KNN Algorithm. https://www.mygreatlearning.com/blog/knn-algorithm-introduction/ (accessed 10 May 2020).

[11] Wafa Mellouk, Wahida Handouzi. Facial emotion recognition using deep learning: review and insights. Procedia Computer Science (2020); 175.

[12] Ratnalata Gupta, Vishwamitra LK. (2020). Facial expression recognition from videos using CNN and feature aggregation. Materials Today: Proceedings.

[13] Jagli, P. M., & Shetty, M. P. (2019). Human Emotion Recognition using Machine Learning. International Journal of Trend in Scientific Research and Development, 101-103.

[14] Rassadin, & Savchenko. (2017). Compressing deep convolutional neural networks in visual emotion recognition. National Research University Higher School of Economics, Laboratory of Algorithms and Technologies for Network Analysis, 207-213.

[15] Stephanie Glen. Sensitivity vs Specificity and Predictive Value. https://www.statisticshowto.com/probability-and-statistics/statisticsdefinitions/sensitivity-vs-specificity-statistics/(2020).