Facial Emotion Recognition using Modified VGG16 Model through Transfer Learning

Rakshith M D

Computer Science and Engineering

Canara Engineering College

Bantwal, India

rakshithmd@canaraengineering.in

Harish H Kenchannavar

Information Science and Engineering

Gogte Institute of Technology

Belagavi, India

harishhk@git.edu

ABSTRACT

Emotion is a modality through which the intention and state of the person can be inferred. It plays a significant part in maintaining interpersonal relations. The meaningful aspects of human behavior in society can be exhibited through emotions. Recognizing the emotions of a person through his facial expressions has become a challenging task in different applications like video surveillance, entertainment, education, and healthcare. In the proposed research, following datasets: Extended Cohn -Kanade (CK+) and FER2013 were considered. Transfer learning based VGG16-M model, a variant of Deep Convolutional Neural Network (DCNN) is applied for performing the emotion recognition task. The presented model provides the test accuracy of 97.15% and 58.83% on CK+ and FER2013 datasets with 50 epochs. The limitation of the present work is that it only focuses on recognizing expressions from static images rather than dynamic ones.

Keywords— Emotion, Facial expression, Transfer learning, Convolution.

# INTRODUCTION

Emotions are internal conscious states that we infer in others and ourselves. The observable behaviour associated with emotion is inferred to gain private experiences. Emotion is a significant life event that involves four components: feelings, social-expressive, bodily arousal, and sense of purpose. Feelings vary in intensity and quality. During the occurrence of emotion, the adaptive coping behaviour is activated by bodily arousal. The purposive component gives emotion its goal directed force. Social-expressive components like gestures, vocalizations and facial expressions make our emotions public. Mainly there are two categories of emotions: 1) Anticipatory emotions involve desire and fear. 2) Outcome emotions deal with happiness, sadness, regret, relief, and anxiety. Factors affecting emotions are personality, culture, weather, stress, age, gender, environment, organization, and marital relation. (Friesen et al., 1978) developed FACS for emotion recognition. (Aneja et al., 2016), (Edwards, Jackson, & Pattison, 2002) proved that the impression of a person’s emotion can be learnt through his face in different applications like animation and health. (Chu et al., 2017). illustrated about the disorders diagnostics in children. Because of the low-resolution images and varying backgrounds, emotion recognition using facial images has become a challenging task. (Mollahosseini, Chan, & Mahoor, 2016) achieved the task of emotion recognition through facial expressions. Convolutional neural networks (CNNs) are drastically used for extracting the features from facial expressions thereby contributes in developing efficient facial expression recognition system. (Gao, Luo, & Ma, 2021) presented an OpenCV based method to detect emotions from a video stream by extracting eye and mouth features. (Stewart et al., 2005) designed a hybrid approach to classify facial expressions from multiple pose images. (Wang et al., 2020) performed the task of fingerprint minutiae matching by applying deep convolutional neural network. (Ali et al., 2020) presented a method to classify facial expressions from multiple pose images. In the presented work, by taking into account the above observations a deep learning based model with transfer learning is incorporated for recognizing facial expressions.

We make an attempt to illustrate that by using transfer learning even with less number of epochs a network will be able to obtain high accuracy. The contributions of the presented research are as follows:

1) A transfer learning based deep CNN model is proposed to recognize emotions from facial expressions and also to obtain higher accuracy than the existing techniques.

2) A compact CNN model for recognizing seven facial expressions is proposed. Longer training sets are considered for illustrating consistency.

3) The proposed model provides an accuracy of about 95% with respect to testing in the minimum number of epochs.

# ARCHITECTURE OF FACIAL EXPRESSION RECOGNITION (FER) SYSTEM

Artificial intelligence (AI) is a broad domain which involves transfer learning. It enables the transfer of learning obtained from pre-trained models to the data of interest thereby reusing the knowledge. It is the most widely used technique for building the models with reasonably less time and learning the patterns from the already obtained information.

# 

**Figure 1: Architecture of Transfer learning based FER System**

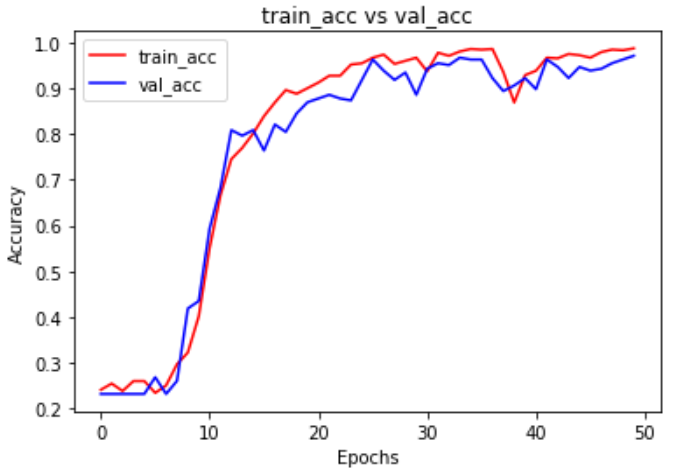
VGG16 involves 16 layers which are combination of convolution and fully connected layers. The VGGNet-16 is exploited as features extractor for representing the images as dimensional feature vector. The VGG-16 model is designed with 13 and 3 layers which are convolutional and fully connected. The output feature vector gained from the final fully connected layer represents final features. The first, second convolutional layers composed of 64 feature kernel filters with filter size 3×3.Transforming input image into first, second convolutional layers, results in image with dimensions 224×224×64. Then the obtained outcome is transformed with a stride 2 in the max pooling layer. 124 feature kernel filters with the filter size 3×3 are indicated in the third and fourth convolutional layers which are succeeded by layers of max pooling holding stride 2. The obtained outcome is minimized into 56×56×128. The kernel size of 3×3 is possessed in the fifth, sixth and seventh convolutional layers. All the three layers utilize 256 feature maps and they are pursued by the max pooling layer holding stride 2.The two convolutional layer set are possessed from eighth to thirteen and they have 512 kernel filters whereas these layers are pursued by layers of maxpooling with a stride 1. Fourteen to sixteenth layers extracts the valuable features. These are fully connected in nature and responsible for enhancing the classification accuracy. The proposed FER system is illustrated in Figure 1. The VGG16 model is applied on the ImageNet dataset comprising of 1000 classes. For training 1.28 million images were used. Testing was incorporated with 50,000 images. This model is refined to obtain VGG16-M model by adding 2 dense layers with 128 and 64 hidden units. Activation function used is ‘relu’. The proposed model is applied on the corresponding data and its recognition accuracy is observed. The table 1 illustrates the base VGG16 model and the modifications that have been incorporated to derive VGG16-M model.

**Table 1: Summary of the Proposed Model**

|  |  |  |
| --- | --- | --- |
| **Layer (Type)** | **Output Shape** | **Parameters** |
| VGG16 (the base model) | (None, 7, 7, 512) | 14,789,063 |
| Dense (ReLU) | 128 |
| Dense (ReLU) | 64 |
| Dense(Softmax) | 7 |

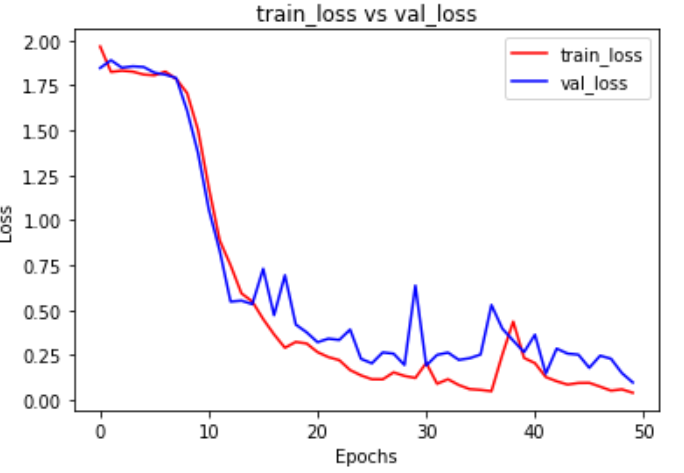
# RESULTS

The performance of the presented FER system using TL on CK+ (Lucey et al., 2010) and the Facial Expression Recognition 2013 (FER2013) (Courville et al., 2013) datasets is analyzed in this section. Primarily, the datasets have been described. Secondarily, the presented model’s performance is compared with the existing methods to understand its efficiency. CK+ dataset includes spontaneous expressions of 123 subjects with 593 sequences. FER2013 dataset incorporates 35,887 images with size 48 × 48. In CK+ dataset, training involved 735 images and testing is incorporated with 246 images. In the case of FER2013 dataset, training phase involved 28,709 images and testing phase incorporates 7178 images. The split ratio considered for training and testing was 80:20. Adam optimizer was used for training the model with 50 epochs.



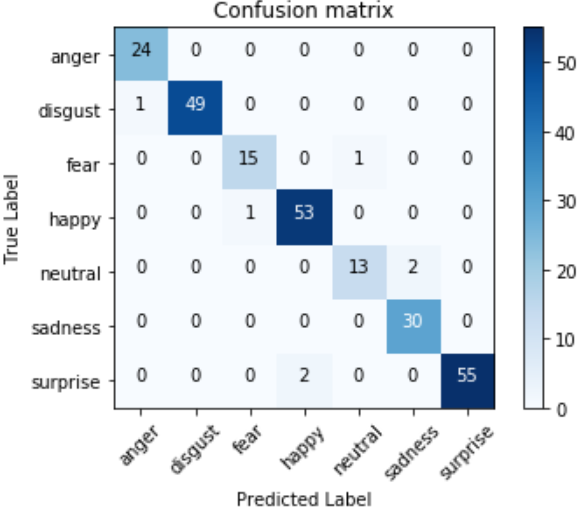
**Figure 2: Accuracy on CK+**

Figure 2 depicts the accuracy obtained by the model on CK+. Similarly, Figure 3 illustrates the loss obtained. Sum of errors obtained for each example in the test set is indicated by loss. Accuracy and the loss are the two most important parameters required for testing the performance of a deep learning model.



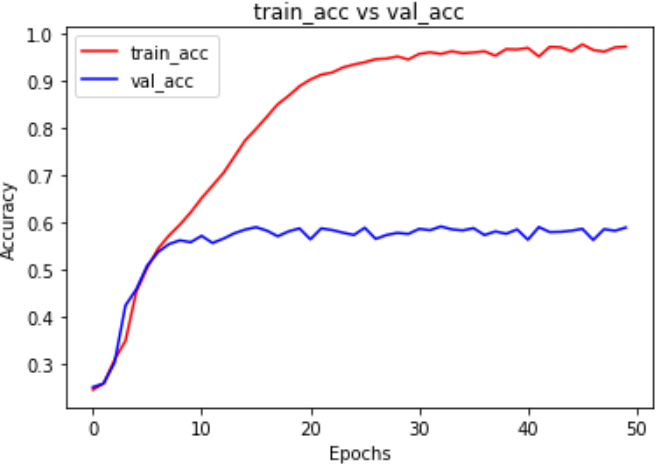
**Figure 3: Loss on CK+**

Figure 4 illustrates the confusion matrix obtained by the proposed model on the test set. For each class of emotion the test set involves the following number of images: anger: 24, disgust: 50, fear: 16, happy: 54, neutral: 15, sadness: 30, and surprise: 57.

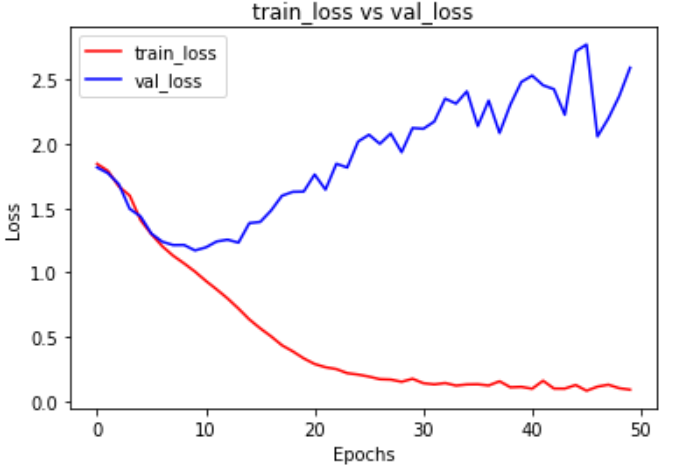


**Figure 4: Confusion matrix for CK+**

Figure 5 illustrates the training and validation accuracy obtained by the proposed model on FER2013 dataset. Similarly, Figure 6 depicts the loss obtained.

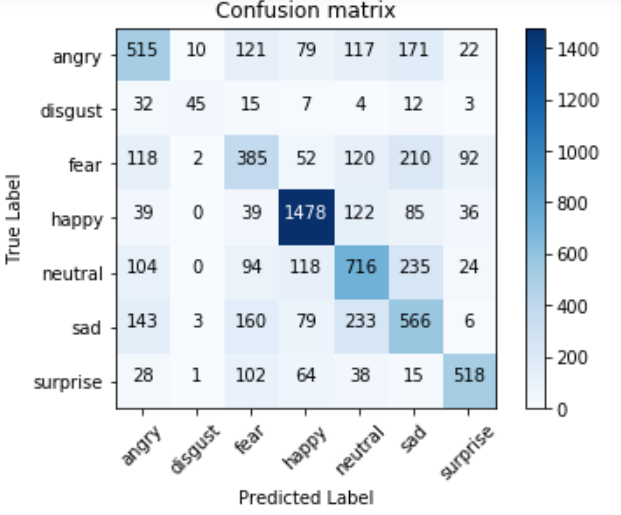


**Figure 5: Accuracy on CK+**



**Figure 6: Loss on FER2013**

Figure 7 shows the confusion matrix obtained by the proposed model on the test set. For each class of emotion the test set involves the following number of images: angry: 1035, disgust: 118, fear: 979, happy: 1799, neutral: 1291, sad: 1190, and surprise: 766.



**Figure 7: Confusion matrix for FER2013**

The proposed model’s outcomes with respect to performance by considering accuracy metric is represented in the Table 2.

**Table 2: Comparative Analysis for CK+**

|  |  |
| --- | --- |
| **Technique** | **Accuracy** |
| Deep features + HOG  (Hao et al., 2020) | 90.58% |
| Single Deep CNNs  (Jain et al., 2019) | 93.24% |
| CNN+LSTM  (Li et al., 2019) | 90.51% |
| Proposed model(VGG16 -M) | 97.15% |

Table 3 shows the performance analysis of the proposed model on FER2013 dataset.

**Table 3: Comparative Analysis for FER2013**

|  |  |
| --- | --- |
| **Technique** | **Accuracy** |
| VGG+SVM (MarianaIuliana, 2018) | 66.31% |
| Bag of Words (Tudor et al., 2013) | 67.4% |
| GoogleNet (Panagiotis et al., 2018) | 65.2% |
| Proposed model(VGG16-M) | 58.83% |

# CONCLUSION

A DCNN model using TL for recognizing the emotions from facial expressions is presented in this paper. Experimental results illustrate that the model VGG16-M provides recognition accuracy of 97.15% on CK+ and 58.83% on FER2013 datasets for 50 epochs. However, the task of recognizing emotions can also be performed with the help of modalities such as speech, gesture, and physiological signals in real-time applications. From the experimental results it is inferred that the proposed model obtain higher accuracy on CK+ when compared with FER2013.

**REFERENCES**

[1] A.M. Ali, H. Zhuang and A.K. Ibrahim, “Multi-pose facial expression recognition using rectangular HOG feature extractor and label-consistent KSVD classifier,” International Journal of Biometrics, Vol.12 No.2, pp.147– 162, 2020.

[2] D. Aneja D, A. Colburn, G. Faigin, L. Shapiro, B. Mones, “Modeling stylized character expressions via deep learning,” in Asian Conference on Computer Vision. Springer: Cham, Switzerland, pp. 136-153, 2016.

[3] H.C. Chu, W.W. Tsai, M.J. Liao, Y.M. Chen Y.M, “Facial emotion recognition with transition detection for students with high functioning autism in adaptive e-learning,” Soft Comput 22: 2973–2999, 2017.

[4] P.L.C. Courville, A. Goodfellow, I.J.M. Mirza, Y. Bengio, “FER-2013 Face Database,” Universit de Montreal: Montréal, QC, Canada, 2013.

[5] E. Edwards, H.J. Jackson, P.E. Pattison, “Emotion recognition via facial expression and affective prosody in schizophrenia: A methodological review,” Clin. Psychol. Rev 22: 789–832, 2002. [CrossRef].

[6] E. Friesen, P. Ekman, W. Friesen, J. Hager, “Facial Action Coding System: A Technique for the Measurement of Facial Movement,” Psychologists Press: Hove, UK, 1978.

[7] F. Gao, D. Luo, X. Ma, “Research on facial expression recognition of video stream based on OpenCV,” International Journal of Biometrics, Vol.13 No.1, pp.114 – 129,2021. DOI:10.1504/IJBM.2021.112219.

[8] W. Hao, S. Wei, B. Fang, “Facial expression recognition using iterative fusion of MO-HOG and deep features,” J. Supercomput. 2020, 76, 3211–3221.

[9] N. Jain, S. Kumar, P. Shamsolmoali and M. Zareapoor, “Hybrid deep neural networks for facial emotion recognition,” in: Pattern Recognition Letters, 115, pp.101-106, 2019.

[10] T.S. Li, P. Kuo, T. Tsai, and P. Luan, “CNN and LSTM Based Facial Expression Analysis Model for a Humanoid Robot,” IEEE Access, 7, 93998 -94011, 2019.

[11] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, “The extended cohn -kanade dataset (ck+): A complete dataset for action unit and emotion -specified expression,” In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), San Francisco, CA, USA, 13–18 June 2010. [CrossRef].

[12] G. Mariana–Iuliana, R.T. Lonescu, M. Popescu, “Local Learning with Deep and Handcrafted Features for Facial Expression Recognition,” 2018. arXiv:1804.10892.

[13] A. Mollahosseini, D. Chan, M.H. Mahoor, “Going deeper in facial expression recognition using deep neural networks,” In Proceedings of the IEEE 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7 –10, 2016.

[14] G. Panagiotis, I. Perikos, I. Hatzilygeroudis, “Deep Learning Approaches for Facial Emotion Recognition: A Case Study on FER -2013,” In Advances in Hybridization of Intelligent Methods; Springer: Cham, Switzerland, 2018; pp. 1 –16.

[15] M.B. Stewart, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, J. Movellan, “Recognizing facial expression: Machine learning and application to spontaneous behavior,” In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR, San Diego, CA, USA, 20 –25 June 2005; Volume 2, pp. 568 –573.

[16] R.I. Tudor, M. Popescu, C. Grozea, “Local learning to improve bag of visual words model for facial expression recognition,” In Proceedings of the ICML Workshop on Challenges in Representation Learning, 2013.

[17] S.H. Wang, M.J. Cheng, Z.Q. Ma, X.X. Sun, “Research on the application of convolutional-deep neural networks in parallel fingerprint minutiae matching. International Journal of Biometrics,” Vol.13 No.1, pp.96–113, DOI: 10.1504/IJBM.2021.112220.

1. Yahia Said, and Mohammad Barr, “Human emotion recognition based on facial expressions via deep learning on high-resolution images,” Multimedia Tools and Applications, vol. 80, no. 16, pp. 25241-25253, 2021.

[19] Jun Liu, Yanjun Feng, and Hongxia Wang, “Facial expression recognition using pose-guided face alignment and discriminative features based on deep learning,” *IEEE Access*, vol. 9, pp. 69267-69277, 2021.

[20] N. Kumari and R. Bhatia, “Efficient facial emotion recognition model using deep convolutional neural network and modified joint trilateral filter,” Soft Computing, pp.1-14, 2022.

[21] Ye Ming, Hu Qian, Liu Guangyuan, “CNN-LSTM Facial Expression Recognition Method Fused with Two-Layer Attention Mechanism”, *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 7450637, 9 pages, 2022. <https://doi.org/10.1155/2022/7450637>.

[22] B. Li, D. Lima, “Facial expression recognition via ResNet-50,” International Journal of Cognitive Computing in Engineering 2, pp. 57–64, 2021.

[23] N. Kumar H.N, S.A. Kumar, G. Prasad M.S, M.A. Shah, “Automatic facial expression recognition combining texture and shape features from prominent facial regions,” IET Image Process. pp. 1–15, 2022. <https://doi.org/10.1049/ipr2.12700>.

[24] A. Agrawal, N. Mittal, “Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy”, *Visual Computer*, vol. 36, issue 2, pp.405-412, 2020.

# [25] C. Jin, “Cross-database facial expression recognition based on hybrid improved unsupervised domain adaptation,” *Multimed Tools Appl* 82, 1105–1129 (2023). https://doi.org/10.1007/s11042-022-13311-2.