Face Emotion Recognition Based on Images Using the Haar-Cascade Front End Approach

G. Gowri Pushpa[1](✉), Jayasri Kotti[2], and Ch. Bindumadhuri[3]

[1] Department of Computer Science and Engineering, Anil Neeru Konda Institute of Technology and Sciences, Visakhapatnam, Andhra Pradesh, India
gowripushpa11@gmail.com

[2] Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India

[3] Department of Information and Technology, University College of Engineering Vizianagaram, JNTUK, Vizianagaram, Andhra Pradesh, India
chbmadhuri.it@jntukucev.ac.in

Abstract. Facial expression recognition (FER) has emerged as a major research topic, with human-computer interactions. Nonverbal messages which are face expressions are crucial in our day-to-day life, which is an example of non-verbal communication. As a human, detecting facial expressions and understanding human emotions is a simple process, but doing it with the assistance of a machine is more challenging. With the remarkable success of deep learning, the different types of architectures of this technique are exploited to achieve a better performance with an accuracy of 87%. In this research, our proposed model shows how to identify and recognize facial emotions from images using neural networks with help of preprocessing techniques and the whole process comprises various stages of classifying the detected features, involving human face detection and classifying them then into any of the seven basic emotion classes using the convolutional neural networks (CNN). Haar-cascade frontal face algorithm was utilized in order to detect human faces from the images. Our model was trained and tested on the FER-2013 dataset.

Keywords: Haar-cascade frontal face algorithm · FER-2013 · Facial expression recognition · Face detection · Convolution neural network (CNN)

1 Introduction

Facial Emotions are mostly communicated through these facial expressions, gestures, and physiological signals. With better deep learning techniques, we can solve these types of complex problems using neural networks. Numerous uses exist in the field of human-computer interface for recognizing these human facial expressions [9]. All the non-verbal communications basically fall under facial emotion recognition. A person's facial gestures can be used to assess both his or her emotional state and cerebral outlook. The basic facial emotions include Angry, Disgust, Sad, Happy, Fear, Neutral, Surprise.

P. Pareek et al. (Eds.): IC4S 2023, LNICST 536, pp. 331–339, 2024.
https://doi.org/10.1007/978-3-031-48888-7_28

Convolutional Neural Networks are widely applied in the field of image recognition [1]. As it can have a huge number of network layers and it also extracts more high-level features [2]. Facial recognition requires several phases which are detection of faces from images, pre-processing, retrieval of those facial features, and identification of emotions in the facial images. For detection of human faces there are widely used other algorithms like Local binary Pattern (LBP) and MTCNN and haar cascade which are advantageous in their own way based on the requirements. The purpose of this study is to characterise an emotional face input image using efficient deep learning techniques.

2 Literature Review

Face emotions play a significant role in our everyday lives, as is universally acknowledged. We need a system that can recognize our facial expressions of emotion and react accordingly. For recognizing face expressions, a compact convolutional neural network is created. The completely linked layer from the conventional convolutional neural network is replaced with a global average pooling layer. The completely connected layer's [17] "black box" characteristics are somewhat eliminated by this technique, which can also be connected to global data to learn more in-depth and complete aspects of facial expressions.

Additionally, the pooling layer is devoid of any parameters that would enable parameter reduction and prevent overfitting. This model is a MTCNN [10] that accomplishes the task of detecting frontal faces which has more complexity of time and space. Many of the models were built using RESNET, ALEXNET, MOBILENETV2, INCEPTION architectures which are pre trained with better weights and resulting in an accuracy of around 70%-75%. In order to avoid overfitting, norms were added to the weight coefficient that not only prevents overfitting but makes the model stable and fast. At last, the final model is made combining a neural network with four convolutions and an MTCNN detection [10] to accomplish the aim of human emotion recognition. Batches and a RELU activation function [16] follow each of these four convolutions, while a soft-max activation function and global average pooling are used in the final layer to classify data. This system has a total of 58423 characteristics, 56951 of which are trainable. As a consequence, when attempting to categorize the facial gestures into one of the seven basic feelings, the prediction accuracy is 67%.

3 Dataset

FER-2013 has been utilized for the purpose of training and testing this model. This FER-2013 data set is available in both images and pixel values. We have implemented the model using the pixel values which are in csv format. These pixel values comprise almost every sort of the 7 recognized facial emotions (Fig. 1).

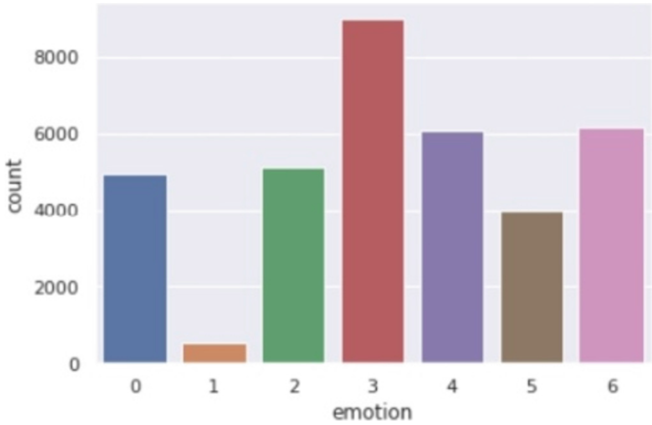

Fig. 1. Training data set count

4 Proposed Methodology

We proposed an approach to recognize the human from the images using traditional CNN [2] which is capable of classifying seven different types of emotions. The proposed traditional CNN is composed of 5 convolutional layers which have conv2d, batch normalization [4], maxpooling2d and dropout. A dense layer of 200 neurons attached to the completely connected layer and a SoftMax activation function are used to categorize human emotions (Fig. 2).

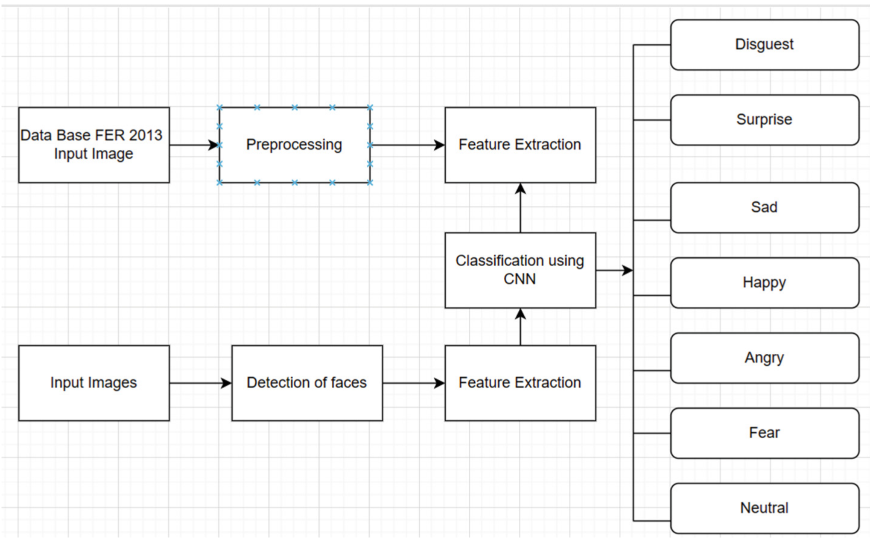

Fig. 2. Flow chart of general facial expression recognition

4.1 Data Pre-Processing

The proposed model has certain pre-processing steps such as normalizing the images, adjusting the dimensions of the image, random sampling techniques all of which are considered pre-processing strategies for image classification [12]. After normalizing the picture, it has a pixel value range from 0 to 1, as a picture can have a pixel value of max 255 and by splitting them by 255, they fall into the range of 0 and 1, making math computations much simpler and the model's intricacy lower. Then the reshaping of the dimensions of data is performed by adding extra dimensions for pixel values and we generate a shape of (-1,48,48,1).

Then the random over sampler is performed so that the majority of the class will be shuffled randomly so that it avoids the model to memorize and helps in generalizing. This sampler can also be used for minority classes based on the requirement of the model. Once these pre-processing methods have been successfully completed, we will divide the data set into both testing and training with the aid of the test size parameter and random state parameter, which shuffles the training data. We had introduced a shuffle size of 45 i.e., for every iteration a 45 new image pixel values will be generated replacing the old pixel values.

4.2 Model Training

This Traditional CNN [14] design consists of 5 convolutional layers and a flatten and finally three fully connected layers. Each has a Conv2D, Activation, Batch normalization, dropout and Maxpooling2D. The first layer is the input layer which takes input in with padding value 1, stride value 1 and kernel size of 3X3. The Conv2D extracts all possible feature maps and identifies the hidden internal representations. This was managed by the filters of a kernel of conv2d. This is also known as a feature extraction layer (Fig. 3).

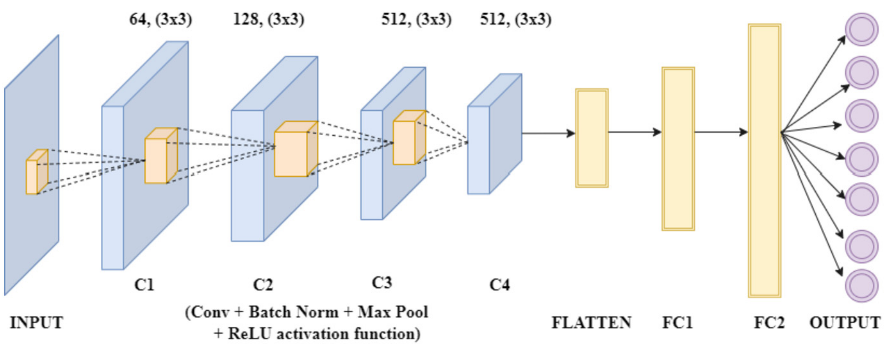

Fig. 3. Proposed Architecture of CNN model

Batch normalization [5] functions on extracted data by conv2d which takes as input batches. This makes the model learn much faster. We had introduced a batch size of 128 for our particular model. Maxpooling2D helps in finding all prominent places of the feature maps and thereby reducing the dimensions by extracting these prominent features. The RELU Activation function [7] brings the nonlinearity for the feature maps generated by convolutional layers. A dropout layer [3] is used to avoid the over-fitting of the model. We had dropped 20% of neurons for every new iteration. The flattened layer then converts into the 1d array and then the fully connected layer and makes its predictions according to the input.

4.3 Model Testing

During the testing phase, validation accuracy and loss of the proposed model is saved in their best state having higher accuracy. We have implemented certain call backs like early stopping, lr scheduler and model checkpoint in order to stop the model from over-fitting and save it in the best possible condition with a higher learning rate. We ran the model for 80 epochs using the Adam algorithm and category cross entropy [6] as the loss function. For prediction purposes, test data will be fed to the trained model to find final accuracy. This model implemented a haar-cascade frontal face feature [12] which detects the frontal faces present in the images. Once detecting the faces by this frontal face algorithm [8] they are reshaped and pixel values are divided by 255 and sent to the model to make predictions by recognizing facial emotions. The prediction is in the following way (Fig. 4).

Fig. 4. Predictions of model for test data set

5 Experimental Results and Analysis

The model was evaluated, and our research's findings indicate that it has a training accuracy of 91%, a testing accuracy of 87%, a training loss of 0.2469, and a testing loss of 0.5012. By The values we can state that this model has predicted effectively with higher accuracy and loss values. With the help of loss curve, it is possible to detect if the model has got overfit or underfit or optimal fit [11].

From the below loss curves we can say that the proposed model is perfectly designed with no overfitting or underfitting of the data with the better loss values and accuracy (Figs. 5 and 6).

The confusion matrix, which displays both the number of accurate forecasts and the number of inaccurate guesses produced by the proposed CNN algorithm, is shown

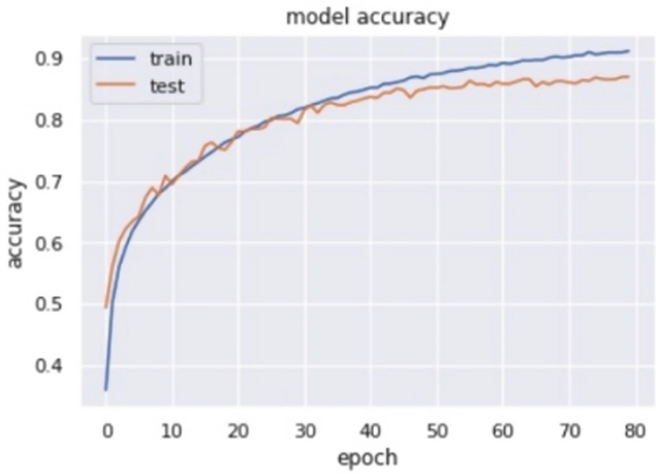

Fig. 5. Loss curve model.

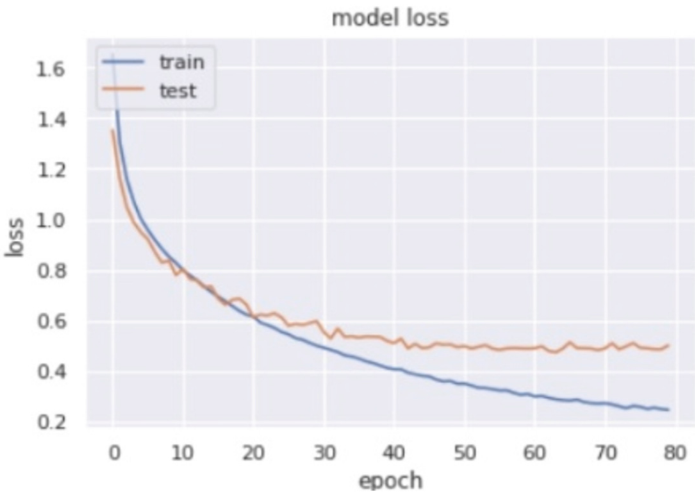

Fig. 6. Accuracy curve model

below, can be used to analyze how well this model predicted nearly every type of human emotion. Each model has a total of 4,478,278 params where 4,475,759 are trainable params and 3,968 are non-trainable (Fig. 7).

Model is shown in the confusion matrix above, along with the model's total number of incorrect predictions. The model is evaluated using the accuracy, recall, precision and f-1 score which are shown in the Fig. 8 i.e., classification report [13].

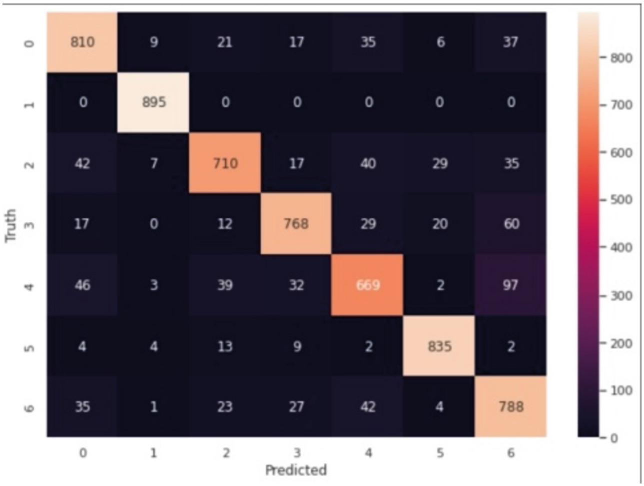

Fig. 7. Confusion matrix

	precision	recall	f1-score	support
0	0.86	0.87	0.87	935
1	0.99	1.00	0.99	895
2	0.83	0.82	0.82	880
3	0.87	0.86	0.87	906
4	0.77	0.76	0.76	888
5	0.92	0.96	0.94	869
6	0.81	0.80	0.80	920
accuracy			0.87	6293
macro avg	0.87	0.87	0.87	6293
weighted avg	0.86	0.87	0.87	6293

Fig. 8. Classification report of Proposed model

6 Conclusion and Feature Work

This article discussed current FER study and informed us of the most recent advancements in this field. The FER system in this article offers a method for identifying emotions from images that is more effective than all other conventional methodologies, with an accuracy of 87% and a prediction rate that is higher. We also have a study that showcases the high rate at which researchers were able to figure out that machines will become better at interpreting feelings in the future, suggesting that interactions between humans and machines will become more natural over time. This method's design does not include any pre-trained models. The suggested work can be improved as a video summarization and broadened to recognize human face emotions based on speech and video.

References

1. Sariyanidi, E., Gunes, H., Cavallaro, A.: Automatic analysis of facial affect: a survey of registration, representation, and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **37**(6), 1113–1133 (2015)
2. O'Shea, K., Nash, R.: An Introduction to Convolutional Neural Networks. arXiv:1511.08458v1 [cs.NE], 26 Nov 2015
3. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(56), 1929−1958 (2014)
4. Daneshmand, H., Kohler, J., Bach, F., Hofmann, T., Lucchi, A.: Batch normalization provably avoids rank collapse for randomly initialised deep networks. arXiv:2003.01652v3 [stat.ML11], June 2020
5. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167v3 [cs.LG], 2 March 2015
6. Valeria, A., Nadiia, S.: Generalization of cross-entropy loss function for image classification. Mohyla Math. J. **3**, 3–10 (2021)
7. Agarap, A.F.M.: Deep Learning using Rectified Linear Units (ReLU). arXiv:1803.08375v2 [cs.NE], 7 February 2019
8. Padilla, R., Costa Filho, C.F.F., Costa, M.G.F.: Costa evaluation of Haar cascade classifiers designed for face detection international science index. Comput. Inf. Eng. **6** (2012)
9. Bansal, H., Khan, R.: A review paper on human computer interaction. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **8**, 53–56 (2018)
10. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Sig. Process. Lett.**23**(10), 1499–1503 (2016)
11. Ekundayo, O., Viriri, S.: Facial Expression recognition: a review of methods, performances and limitations**.** In: Conference on Information Communications Technology and Society (ICTAS). IEEE, Durban, South Africa 2019
12. Pitaloka, D.A., Wulandari, A., Basaruddin, T., Liliana, D.Y.: Enhancing CNN with preprocessing stage in automatic emotion recognition. In: 2nd International Conference on Computer Science and Computational Intelligence (ICCSCI), pp. 523–529, Bali, Indonesia (2017)
13. Jumani, S.Z., Ali, F., Guriro, S., Kandhro, I.A., Khan, A., Zaidi, A.: Facial expression recognition with histogram of oriented gradients using CNN. Ind. J. Sci. Technol. **12**, 1–9 (2019)
14. Mehendale, N.: Facial emotion recognition using convolutional neural networks (FERC). SN Appl. Sci. **2**, 446 (2020)
15. Lin, G., Shen, W.: Research on convolutional neural network based on improved Relu piecewise activation function. In: 8th International Congress of Information and Communication Technology (ICICT), pp. 977–984 (2018)
16. Wang, Y., Wu, J., Hoashi, K.: Lightweight deep convolutional neural networks for facial expression recognition. IEEE, Kuala Lumpur, Malaysia (2019)
17. Chang, T., Wen, G., Hu, Y., Ma, J.: Facial expression recognition based on complexity perception classification algorithm, arXiv:1803.00185 [cs.CV] (2018)
18. Abdulsalam, W.H., Alhamdani, R.S., Abdullah, M.N.: Facial emotion recognition from videos using deep convolutional neural networks. Int. J. Mach. Learn. Comput. **9**, 14–19 (2019)
19. Pantic, M., Rothkrantz. L.J.M.: Towards an affect – sensitive multimodal human-computer interaction. In: Proceedings of the IEEE, vol. 91, pp. 1370–1390, September 2003
20. Debnath, T., Reza, M.M., Rahman, A., Beheshti, A., Band, S.S., Alinejad-Rokny, H.: Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity. Sci. Rep. (2022)

21. Mellouka, W., Handouzia, W.: Facial emotion recognition using deep learning: review and insights. In: The 2nd International Workshop on the Future of Internet of Everything (FIoE), Leuven, Belgium, pp. 9–12, August 2020
22. Bie, M., Xu, H., Gao, Y., Che, X.: Facial expression recognition from a single face image based on deep learning and broad learning. Wirel. Commun. Mob. Comput. 1–10 (2022)