

Emotion Recognition using ML

Shivansh Mishra¹, Shubham Kushwaha² and Vinita Chauhan³

¹⁻²ABES Engineering College, CSE Department, Ghaziabad, India

Email: shivansh.19b101151@abes.ac.in, shubham.19b101122@abes.ac.in

³Assistant Professor CSE Department, ABES Engineering College, Ghaziabad, India
Email: vinita@abes.ac.in

Abstract—Being a social animal by nature communication is a vital part of human existence. This communication helps people understand each other. Understanding the emotional state of the person standing in front of you is one of the most essential parts in efficient communication. This can be expressed through facial expression, body language, pitch of voice etc. With the increase in advancements in technology, it is expected that machines could also understand the emotional state of users. This study aims to develop an EDR (Emotion Detection Recognition) system that uses CNN (Convolutional Neural Network) and predicts the basic emotional states of the person by reading facial expressions. The term here basic expression refers to a set of expressions that convey universal emotions, usually anger, disgust, fear, happiness, sadness, and surprise.

Index Terms— Emotion Detection Recognition, Convolution Neural Network, Machine Learning, Support Vector Machine, K-nearest neighbour.

I. INTRODUCTION

Emotion recognition is one of the most important part of the communication which we humans perform while interacting with each other. The state of emotion in which a person is can be observed through many ways like body language, facial expression, gestures, the tone in which a person standing in front of you is speaking etc.

This communication is basically divided into two parts: verbal communication and non-verbal communication. By verbal communication, we mean oral communication and non-verbal communication is communication without use of any verbal language. The indicators like body language and facial expressions fall in the category of non-verbal communication. In an ideal situation, this observation and prediction of the emotional state of humans is easily done by any other human and we do this every time we interact with other people.

With the advancements that we are acquiring in technologies like data science and deep learning it is expected that a machine should also be able to understand the human emotional state even more efficiently than any other human could percept. The idea of this study is to come up with a model that can learn the patterns of a human's facial expressions and could predict his current emotional state with as much accuracy as possible.

This FER (Facial Expression Recognition) system could be applied to a wide variety of research areas such as medical diagnosis, HCI (Human Computer Interaction) systems more advanced etc. Recognizing expression under general situations is however more challenging. This is due to the variety of possible poses and lights. There were many algorithms possible like kNN, SVM etc but CNNs have the potential to overcome the previously stated problems also many recent works on FER used CNNs efficiently specifically when it comes to feature extraction and inference.



Fig-1 FER2013 dataset, showing different expressions

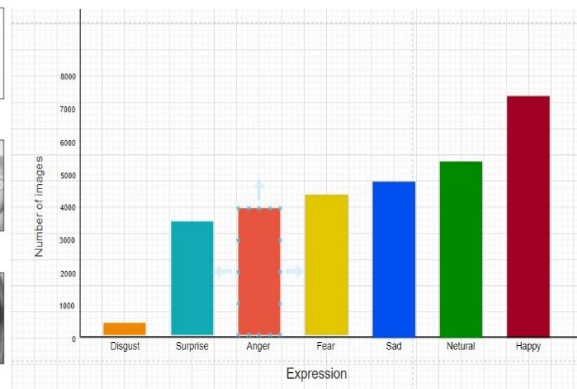


Fig-2 This figure expresses statistically the training data

A. Motivation

Recognizing facial expressions is a key to non-verbal communication between humans. As with time the collaboration/interaction of humans increases with machines, more research are surfacing with the aim to improve human-computer interaction (HCI) and not just it's efficiency but flexibility as well which led to Facial Emotion Recognition technology. The progress in FER would lead to the development of enhanced HCI systems and would result in a better quality of user experience. FER has resulted in the recognition of basic expressions in considered conditions, a possible task. The term basic expression refers to a set of expressions that convey universal emotions, usually anger, disgust, fear, happiness, sadness, and surprise.

B. Scope

Emotion detection is a method that detects the emotional state of humans by recognizing various patterns such as facial expressions, voice intensity, etc. This study aims to develop a project that can recognize basic emotions (contempt, disgust, fear, joy, anger, sadness, and surprise) and make a significant contribution to improving human-computer interaction.

II. LITERATURE REVIEW

During the survey, we found many research papers related to emotion recognition which used different datasets, methodologies, and input formats. There were some models which used EEG rhythms, RNN (Recurrent Neural Network), a combination of RNN and CNN etc. Also, there were some models which did not use any form of neural network. But CNN is perfect for the objective because it's able to cope with the varieties in the images. In 2019, some developers [1] tried to make a real-time emotion detection model using the CNN architecture. Training loss was found to be 0.0887; training accuracy was found to be 76.43%; validation loss was found to be 0.2725; and validation accuracy was found to be 73.81%.

In 2021 Shrey Modi, and Mohammed Husain Bohara tried to make the FER model using CNN with a CNN method accuracy of 82.51(35 epochs) and Transfer Learning accuracy of 73.58(35 epochs)

III. METHODOLOGY

This cycle of prediction and learning is done using machine learning and this machine learning is basically classified into three types:

- Supervised Learning - Uses labelled data both input and output.
- Unsupervised Learning - Uses unlabelled data.
- Reinforcement Learning – it uses the science of decision-making.

In this model, we have used the CNN algorithm to study the images, by labelled data and predicting the required outcome. The tests were made on the basis of labelled output data. Thus, we have used supervised learning but more specifically CNN falls under the category of deep learning.

A. Hardware and Software Requirements

Python libraries (Packages Needed):

- **Numpy:** Every image can be defined as a matrix of pixels and here to work on these images i.e. matrix Numpy is used which provides a highly efficient multidimensional array of objects and tools.
- **OpenCV:** It is a vast open-source library used in the fields of computer vision, machine learning, etc. It is used for analyzing images.
- **Keras:** It is an API that reduces cognitive load also it acts as an interface for TensorFlow. Used as a Deep-learning framework.
- **TensorFlow:** It is an open-source software library used for machine learning and artificial intelligence
- **Pillow:** It is a python image library used for processing images.
- **PyCharm:** It is a python IDE providing a wide range of vital tools for development.

Hardware Used in Making the model:

- I5 processor
- 8Gb RAM
- Windows OS
- 180 Mb size

B. Complete Process



Fig-3 This figure depicts the flowchart of the entire process through which model is made

Different python libraries were mentioned before being imported. Dataset is selected from the Kaggle FER2013 dataset.

The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image.

The task is to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The training set consists of 28,709 examples and the public test set consists of 3,589 examples.

The image is processed using python libraries and then the CNN layer is made. The CNN model is trained using the training dataset and tested out using testing dataset.

4. Convolutional Neural Network (CNN)

CNN falls under the category of Neural Network. In a regular neural network there are three types of layers:

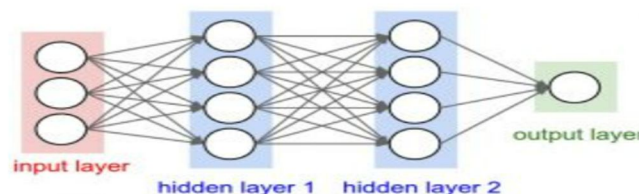


Fig-4 This figure depicts the skeletal structure of neural network

- **Input Layers:** Regular Neural Network consist of an input layer through which give input to our model.
- **Hidden Layer:** The output of hidden layer becomes the input of hidden layer. There can be many hidden layer depending on the model and data size. It is the processing layer of neural network. The addition of learning biases and activation function is done in this layer.

Output Layer: The output from hidden layer is fed to a logistic function be it sigmoid or softmax which coverts the output in a desired format.

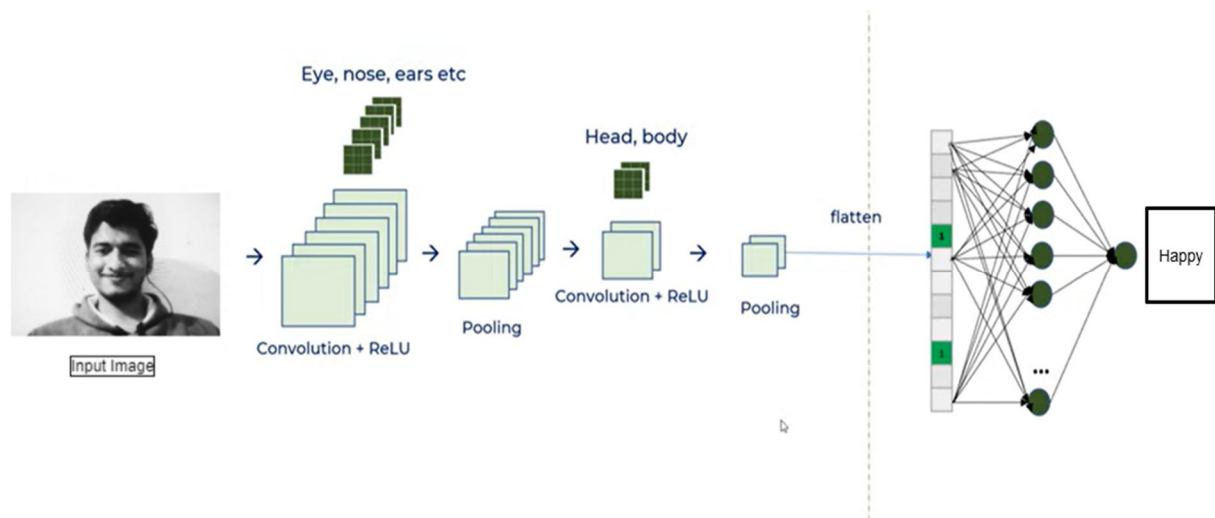


Fig-5 CNN Architecture

In this model we have created CNN with 2 convolutional layers and ReLU[15] as an activation function which makes the model non-linear and 64 and 32 kernels with kernel size equal to 3 and also we flatten the image to a 1D array and these convolutional layers are directly connected to output layers. We have used max pooling[4] with pool size of (3,3).

The pool layers are frequently used to reduce the size of volume resulting in fast computation, reduce memory and also prevent overfitting (a condition in which model works too well in training data but not in testing data).

IV. CONCLUSION

The system successfully predicts the desired outcome not only with the images but also in real-time videos as an input with an accuracy of 86.73 percent.

The model is not only able to predict the emotions but also anger, disgust, fear, happiness, sadness, or surprise but also in the desired speed.

The model would be able to contribute well in the human psychology[7] reaction and medical diagnosis.

REFERENCES

- [1] S. Kak, F. Mustafa, Pedro Valente, "A Review of Person Recognition Based on Face Model", Computer Science, Published in 2018.
- [2] Tayyaba Zamindar, Shradhatai Gangawane, Shital Kalane, Yogita Kalane, "Real-Time Face Recognition and Detection Using Python", Paper Id: IJRASET42939, ISSN : 2321-9653, Publisher Name: IJRASET, Published in 2022.
- [3] Agarwal AK, Singh YN, "Evaluation of face recognition methods in unconstrained environments" Computer Science, Published in 2015.
- [4] Jafri R, Arabnia HR, "A survey of face recognition techniques" J Inf Process Syst, 2009, 5: 41 - 68.
- [5] Introna LD, Nissenbaum H, "Facial Recognition Technology" Published in 2009.
- [6] Melike GÜNAY, Tolga ENSARI, "Comparison of Face Recognition Algorithms", ISSN: 978-1-5090-649, Published in 2017.
- [7] C. Han, "Modular PCA Face Recognition Based on Weighted Average". Modern Applied Science, Vol. 3, No. 11, November 2009.
- [8] D. Chen, and H. Jiu-qiang, "An FPGA- based face recognition using combined 5/3 DWT with PCA methods". Journal of Communication and Computer, ISSN 848- 7709, USA, Volume 6, No. 10 (Serial No.59), Oct. 2009.
- [9] H. Moon, and P. Phillips, "Computational and performance aspects of PCA-based face-recognition algorithms", Perception, 2001, volume 30, pp303-321.
- [10] M. Sharkas, "Application of DCT Blocks with Principal Component Analysis for Face recognition", Proceedings of the 5th WSEAS int. Conf. on Signal, Speech and Image Processing, Corfu, Greece, (pp107- 111), 2005
- [11] Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. 2015. Learning spatiotemporal features with 3d convolutional networks. In 2015 IEEE International Conference on Computer Vision (ICCV) .4489-4497. IEEE.
- [12] Eyben, F., Wöllmer, M., & Schuller, B. (2010, October). Opensmile: the munich versatile and fast open-source audio feature extractor. In Proceedings of the 18th ACM international conference on Multimedia. 1459-1462. ACM.

- [13] Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., & Darrell, T. 2015. Longterm recurrent convolutional networks for visual recognition and description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2625-2634.
- [14] Kahou, S. E., Pal, C., Bouthillier, X., Froumenty, P., Gülçehre, Ç, Memisevic, R. and Mirza, M. 2013. Combinin modality specific deep neural networks for emotion recognition in video. In Proceedings of the 15th ACM on International conference on multimodal interaction. 543-550. ACM.
- [15] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. 2014. Caffe: Convolutional architecture for fast feature embedding. In ACM MM.
- [16] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D. and Rabinovich, A. 2015. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1-9.