Facial Expression Recognition using CNN

NAME : Neelansh Pandey (1805210028)

Ritesh Kumar Chaurasia (1805210043)

Eshan Singh (1805210019)

BRANCH : C.S.E

SEMESTER : VIII

Supervisor: Prof YN Singh

Co-Supervisor: Ms. Deepali Awasthi

Table of Content

1) Introduction

2) Need and Purpose for FER

3) Methods of Facial Expression Recognition

4) Requirements for Creating Model

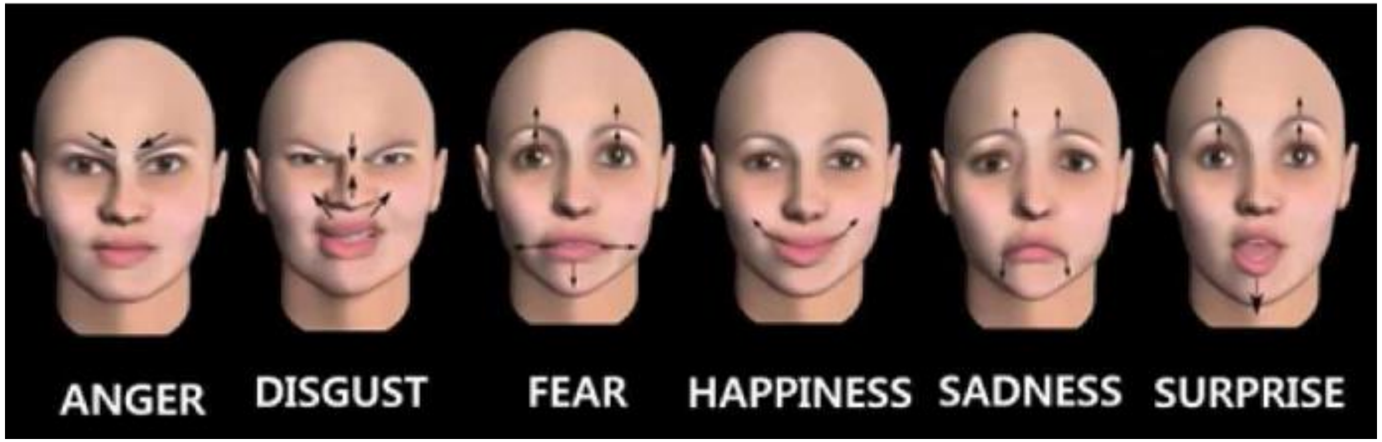
5) Process Model
    1) Selection of Dataset
    2) Image Data Preprocessing
    3) Image Data Augmentation
    4) Features Extraction and Training
        1) Convolution Neural Network
        2) Creating The CNN Model
        3) Components of CNN Model

6) Validation and Output
7) Results
8) Conclusion
9) References

INTRODUCTION

Facial Expression recognition is the process of identifying human emotion using facial expressions. People vary widely in their accuracy at recognizing the emotions of others.

Facial expression recognition software is a technology which uses biometric markers to detect emotions in human faces. More precisely, this technology is a sentiment analysis tool and is able to automatically detect the six basic or universal expressions: happiness, sadness, anger, surprise, fear, and disgust.



Need and Purpose of FER

Facial Expression Recognition can play very vital in modern world technologies. It could prove to be very useful in todays and future technology. Some of them are as follows:

- **Human-Machine interaction:** If we want human and Machines interact with each other it is necessary for a machine to know the current emotion of human from whom it is interacting. In this case FER becomes a necessity.

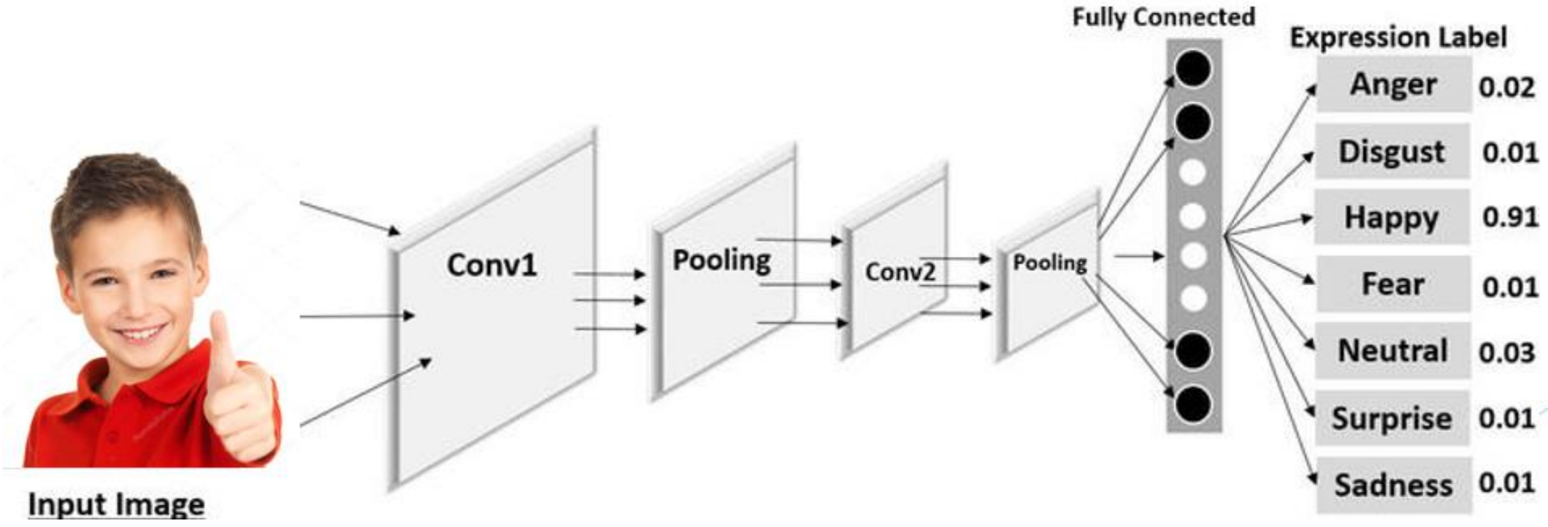
- **Interviews:** FER can be very useful in analyzing the sentiments of the candidate when certain questions are asked and overall analysis of emotions during interview.

Other usages may be Advertisements, can be used by psychologist in diagnosis of Mental disorders, etc.

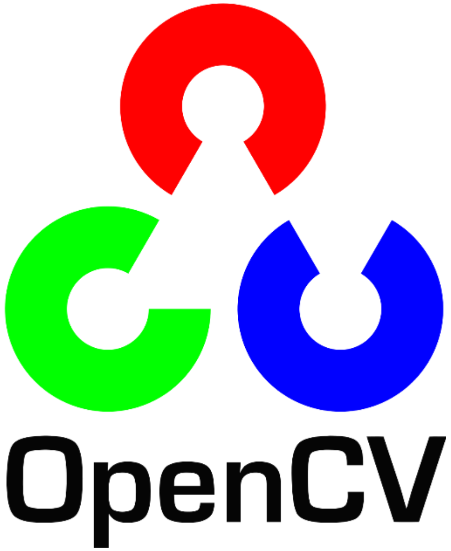
Methods for FER

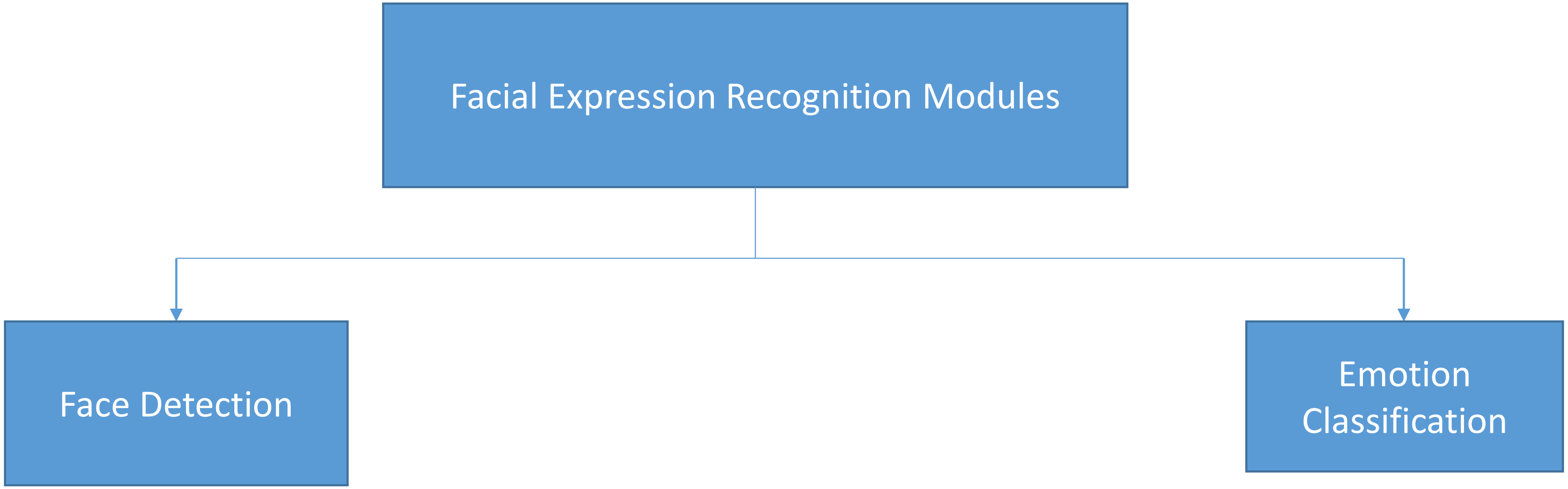
Facial Expression Recognition can be done using Principal Component Analysis, Support Vector Machine and Convolution Neural Network (Artificial Neural Network Approach).

Accuracy of Convolution neural network is higher than any other methods and its easy to implement. So in this project we will be using Convolution neural network in order to achieve Facial Expression Recognition.



Requirements for Creating Model





Facial Expression Recognition Modules

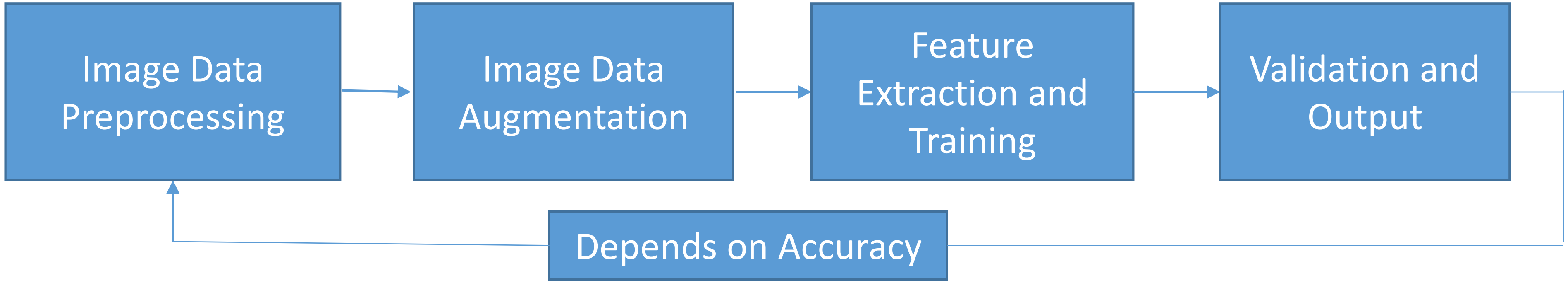
Face Detection

Emotion Classification

Face Detection: Face Detection users Haar Cascade features with help of Haar cascade frontal face default file which is open source.

Emotion Classification: Emotion Classification will be done by training and testing a convolution neural network using dataset containing various emotions.

Process Model

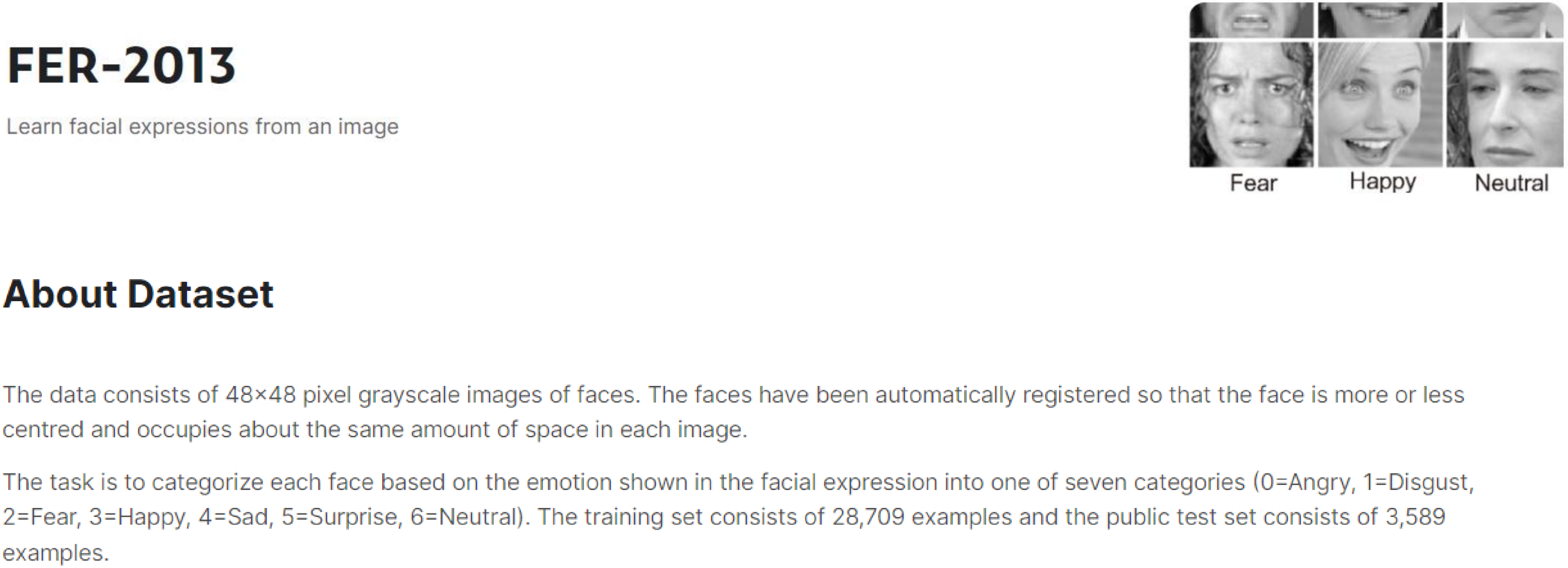


After selection of dataset steps in Facial Expression Recognition Using CNN

1) Image Data Preprocessing
2) Image Data Augmentation
3) Feature Extraction and Training
4) Validation and Output

If not satisfied with accuracy do hyper parameter tuning by adding hidden layer, changing learning rate, number of epochs, batch size, etc.

Selection of Dataset



FER-2013

Learn facial expressions from an image

About Dataset

The data consists of 48×48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centred 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.

Image Data Preprocessing

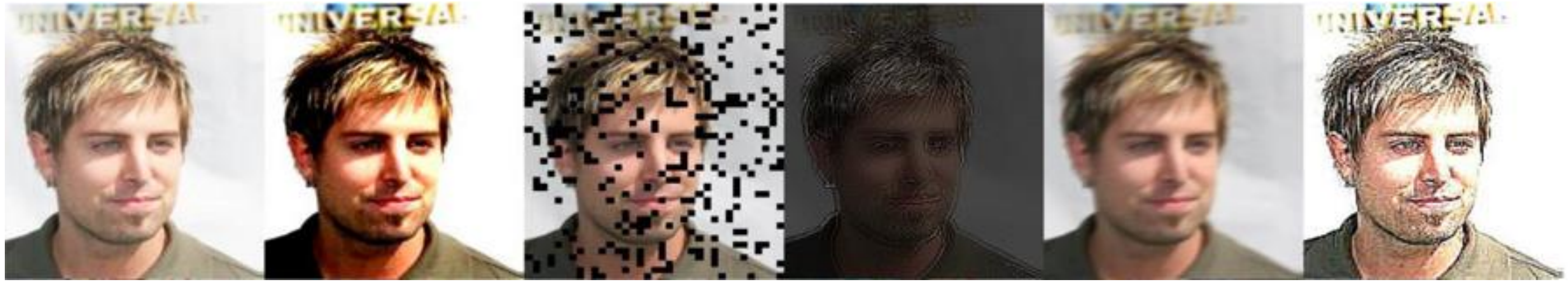
Image preprocessing are the steps taken to format images before they are used by model training and inference. This includes, but is not limited to, resizing, orienting, and color corrections. Preprocessing is required to clean image data for model input. Image preprocessing may also decrease model training time and increase model inference speed. Image Preprocessing can be done using OpenCV.

E.g. - Fully connected layers in convolutional neural networks required that all images are the same sized arrays which can be done by Image Preprocessing.

Image Data Augmentation

Image augmentation are manipulations applied to images to create different versions of similar content in order to expose the model to a wider array of training examples. For example, randomly altering rotation, brightness, or scale of input image requires that a model consider what an image subject looks like in a variety of situations.

Image augmentation creates new training examples out of existing training data. This is particularly important when collected datasets may be small. It is done using Image Data Generator.



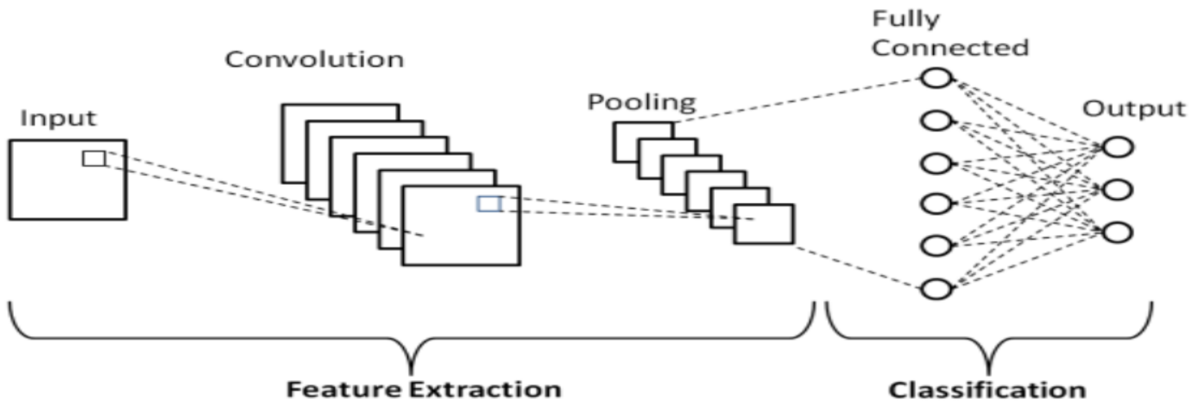
Feature Extraction and Training

Feature Extraction and Training will be done using Convolution Neural Network.

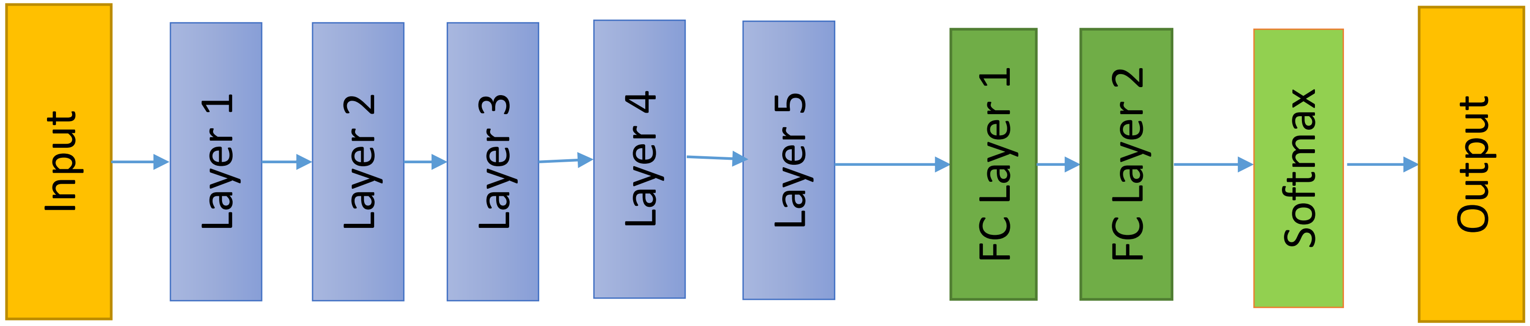
Convolution Neural Network (CNN)

A convolutional neural network (CNN) is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data.

The convolution layers are used to help the computer determine features that could be missed in simply flattening an image into its pixel values.



Creating the CNN model:



Creating the Convolutional Neural Network (CNN)

Input: The dataset we used in training of our Model.

Layer: Each layer contains Convolution layer, Batch Normalization layer, ReLU as Activation function, Max pooling and dropout layer. Five such layers are used in the model.

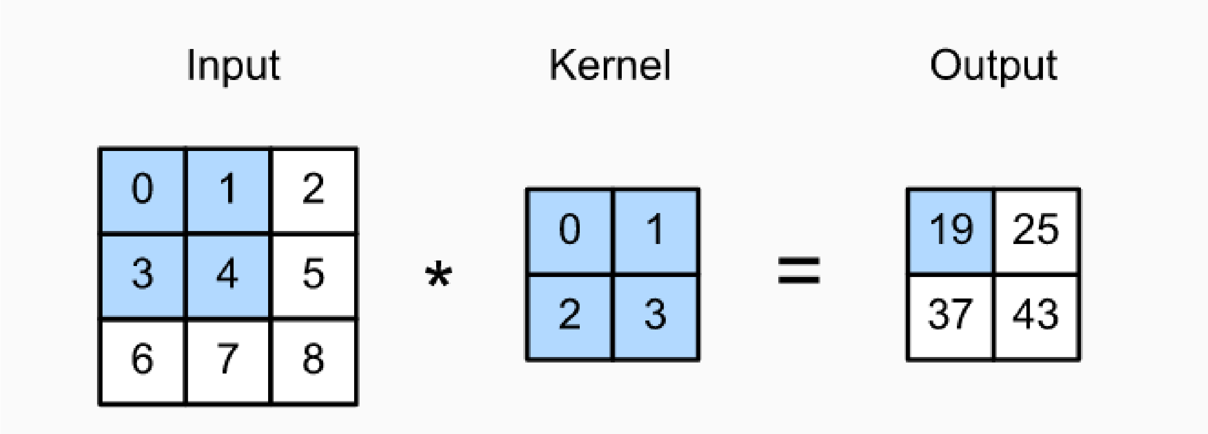
FC layer: FC layer means Fully connected layer. Each of this layer uses a Fully connected neural network, then a layer of Batch Normalization, ReLU as Activation function and dropout layer.

Softmax layer and Output: It contains Softmax as a Activation Function and for output seven neuron representing each classes.

Components of the Model

Convolution 2D

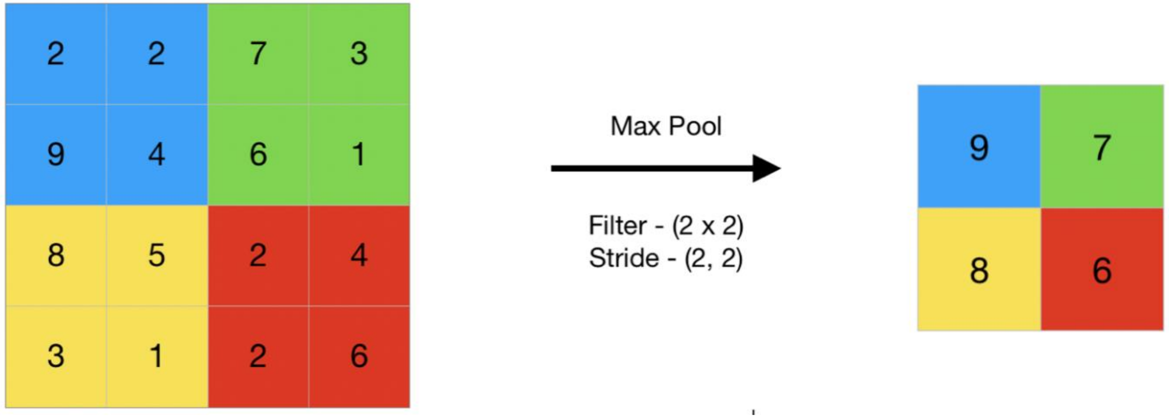
Convolution is using a 'kernel' to extract certain 'features' from an input image. A kernel is a matrix, which is slid across the image and multiplied with the input such that the output is enhanced in a certain desirable manner.



$$\begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix} * \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 19 & 25 \\ 37 & 43 \end{bmatrix}$$

Max Pooling

Max Pooling is a pooling operation that calculates the maximum value for patches of a feature map, and uses it to create a down sampled (pooled) feature map. It is usually used after a convolutional layer.



2	2	7	3
9	4	6	1
8	5	2	4
3	1	2	6

Max Pool

Filter - (2 x 2)
Stride - (2, 2)

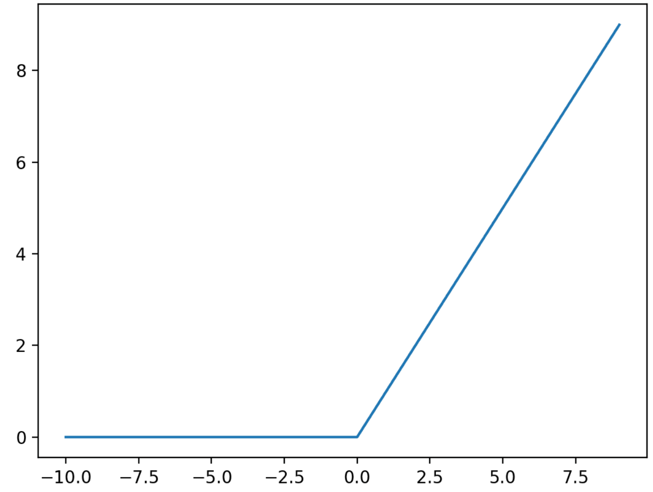
9	7
8	6

Dropout

Dropout is a technique used to prevent a model from overfitting. Dropout works by randomly setting the outgoing edges of hidden units (neurons that make up hidden layers) to 0 at each update of the training phase.

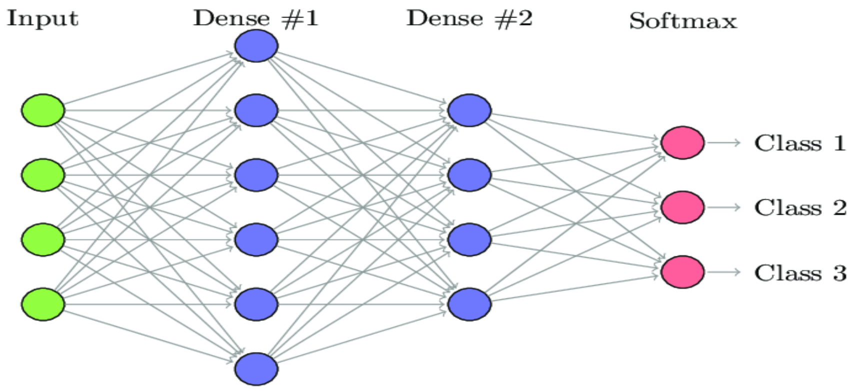
Activation Function (ReLU)

The rectified linear activation function or ReLU for short is a piecewise linear function that will output the input directly if it is positive, otherwise, it will output zero. It achieves better performance than most Activation Functions.



Fully Connected Neural Network

Array made from features of Images is flattened and passed through FCNN and in this way neural network is trained further to produce output.



Validation and Output

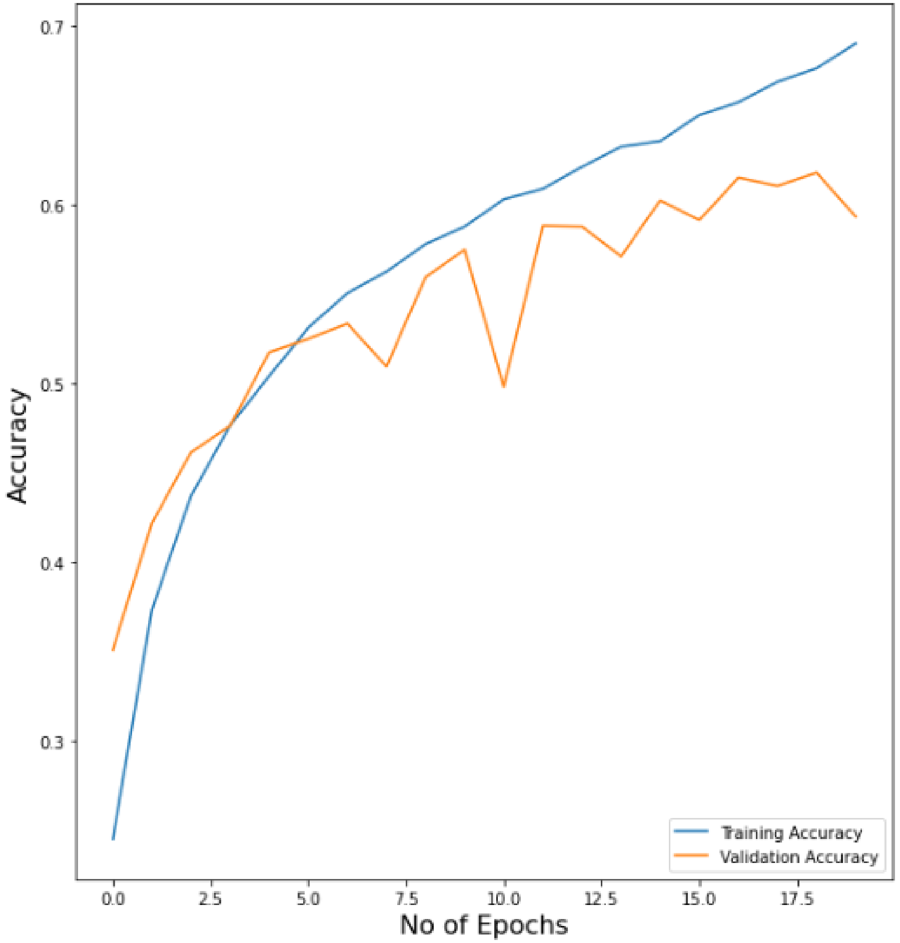
In validation phase we check whether our model is providing accurate results.
In the case of less accurate result we again do feature extraction and training with some changes that can increase accuracy and even if accuracy is very low we will do changes in our dataset.

Validation and output phase can be executed with function in Keras and OpenCV.
In this phase we will also produce output. It will be done using webcam.
It will require Face detection and Facial Expression detection in order to produce correct output.

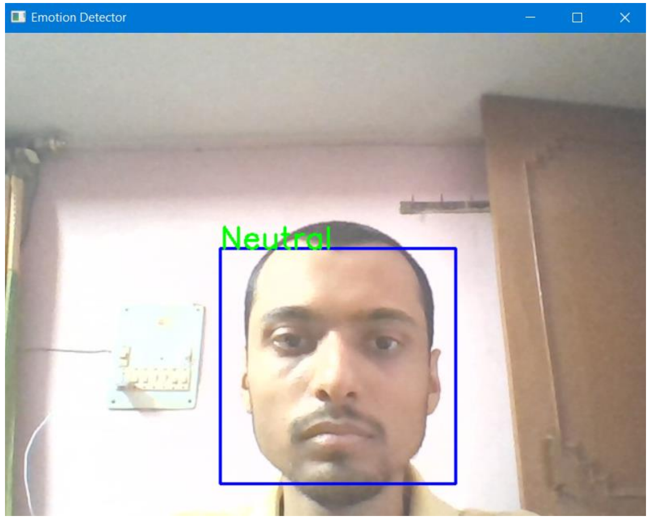
Output will have following labels:

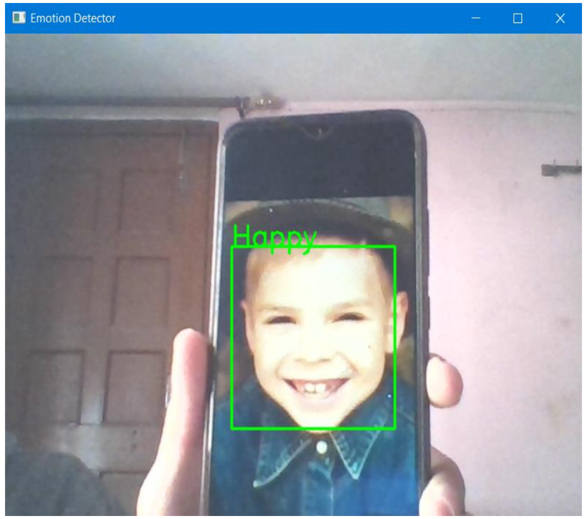
Neutral, Happiness, Sadness, Anger, Surprise, Fear, and Disgust.

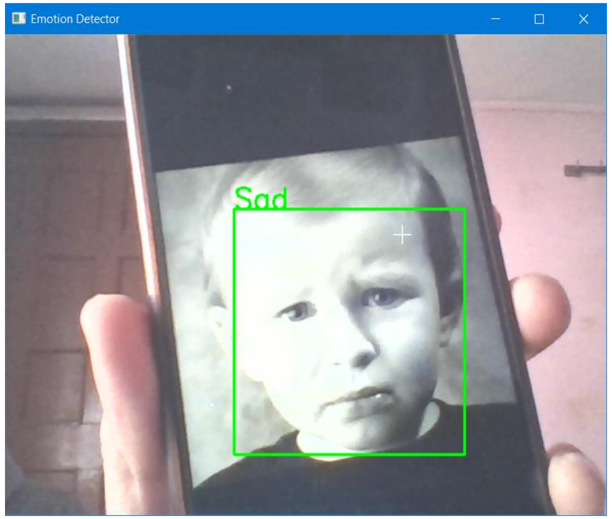
Results

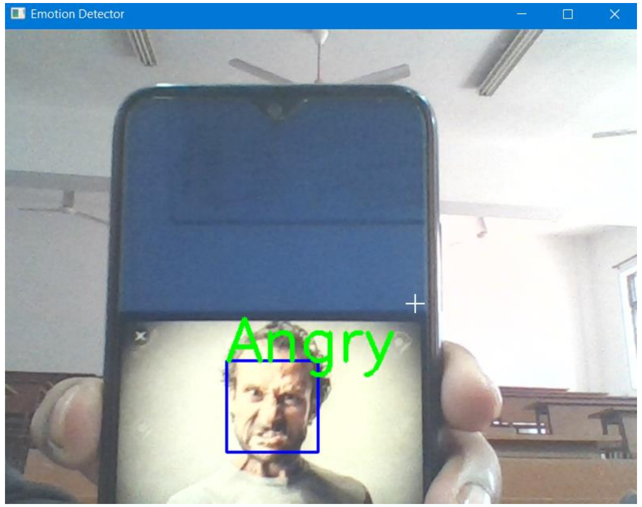


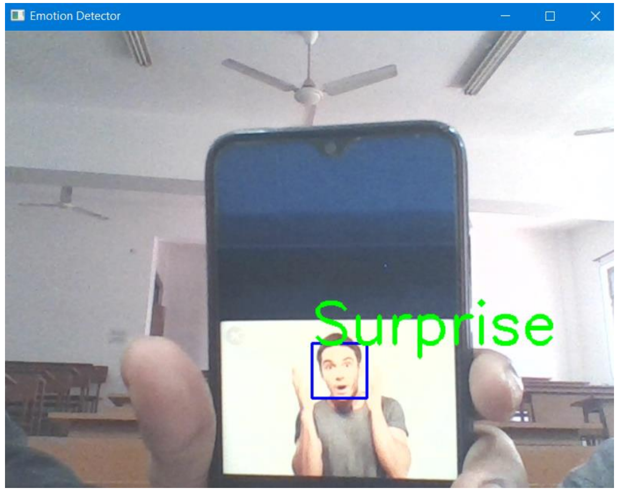
Relationship between Accuracy and number of epochs

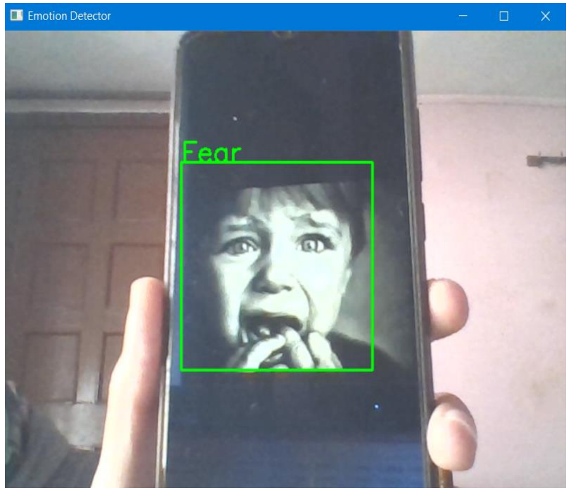

Emotion Detector
Neutral


Emotion Detector
Happy


Emotion Detector
Sad


Emotion Detector
Angry


Emotion Detector
Surprise


Emotion Detector
Fear

Conclusion

1) Objective of our project is Emotion Detection using Facial Expressions.

2) Method used for implementing is Convolution Neural Network (CNN)

3) Steps in Facial Expression Recognition Using CNN

   1) Image Data Preprocessing
   2) Image Data Augmentation
   3) Feature Extraction and Training
   4) Validation and Output

4) Image Data Preprocessing and augmentation will be done using functions present in OpenCV and Tensorflow.

5) Feature Extraction and Training will be done using CNN with the help of Keras and Tensorflow.

6) Validation and Output phase can be done using OpenCV and Keras

7) Training accuracy is 72% and Validation accuracy is 62%.

References

[1] Syed Inthiyaz et al 2021 J. Phys.: Conf. Ser. 1804 012202, “Facial Expression Recognition Using KERAS”, accessed 9 April 2021 <https://iopscience.iop.org/article/10.1088/1742-6596/1804/1/012202>

[2] William Pao, a&s International, “The many application of emotion recognition”, accessed 12 Oct 2017 <https://www.asmag.com/showpost/23883.aspx>

[3] Sumit Saha, “A Comprehensive Guide to Convolutional Neural Networks”, accessed 15 December 2018 <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>

[4] Joseph Nelson, “Why Image Preprocessing and Augmentation Matter”, accessed 26 Jan 2020. <https://blog.roboflow.com/why-preprocess-augment/>

[5] K. Shan, J. Guo, W. You, D. Lu and R. Bie, "Automatic facial expression recognition based on a deep convolutional-neural-network structure," 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), 2017, pp. 123-128, doi: 10.1109/SERA.2017.7965717.

Thank You!!