ISSN: 2278-4632
Vol-13, Issue-12, December 2023

FACIAL EMOTION RECOGNITION USING CONVOLUTION NEUTRAL NETWORKS

[1] J. I. CHAKRAVARTHY, B. Tech, M.Tech. VAMSHI KRISHNA [2], NAVEEN KUMAR [3], PREM KUMAR[4], RAGHAVENDRA [5.]

[1] ASSOCIATE PROFESSOR OF ECE IN MALLA REDDY INSTITUTE OF TECHNOLOGY & SCIENCE, MAISAMMAGUDA, MEDCHAL (M), HYDERABAD-500100, T. S.

[2, 3, 4, 5.] FINALYEAR STUDENTS FROM DEPT OF ECE IN MALLA REDDY INSTITUTE OF TECHNOLOGY & SCIENCE, MAISAMMAGUDA, MEDCHAL (M), HYDERABAD-500100, T. S.

Abstract: Human & computer interaction has been an important field of study for ages. Humans share universal and fundamental set of emotions which are exhibited through consistent facial expressions or emotion. If computer could understand the feelings of humans, it can give the proper services based on the feedback received. An algorithm that performs detection, extraction, and evaluation of these facial expressions will allow for automatic recognition of human emotion in images and videos. Automatic recognition of facial expressions can be an important component of natural human-machine interfaces; it may also be used in behavioral science and in clinical practices. In this model we give the overview of the work done in the past related to Emotion Recognition using Facial expressions along with our approach towards solving the problem. The approaches used for facial expression include classifiers like Support Vector Machine (SVM), Convolution Neural Network (CNN) are used to classify emotions based on certain regions of interest on the face like lips, lower jaw, eyebrows, cheeks and many more. Kaggle facial expression dataset with seven facial expression labels as happy, sad, surprise, fear, anger, disgust, and neutral is used in this project. The system achieved 56.77 % accuracy and 0.57 precision on testing dataset.

Keywords: Facial Expression Recognition, Convolutional Neural Network, Deep Learning.

I. INTRODUCTION

Emotion recognition is a process of identifying the human emotions most likely from facial expressions as well as from speech. The application of emotion recognition system is that it promotes emotion translation between cultures that can be used in multi-cultural communication systems. After extensive research, it is now generally accepted that humans share seven facial expressions that reflect the experiencing of fundamental emotions. These fundamental emotions are anger, neutral, disgust, fear, happiness, sadness, and surprise. Human beings have the capability to recognize emotions easily, but it is difficult for the computers to do the same.

If computers could recognize these emotional inputs, they could give specific and appropriate help to users in ways that are more in tune with the user's needs and preferences. It can be used to find application where efficiency and automation can be useful, including in entertainment, social media, content analysis, criminal justice, and healthcare. For example, content providers can determine the reactions of a consumer and adjust their preferences accordingly. Facial expressions help computers in detecting emotions. This paper deals with helping computers to recognize human emotions in real-time.

II. BACKGROUND

A Facial expression is the visible manifestation of the affective state, cognitive activity, intention, personality and psychopathology of a person and plays a communicative role in interpersonal relations. Human facial expressions can be easily classified into 7 basic emotions: happy, sad, surprise, fear, anger, disgust, and neutral. Our facial emotions are expressed by activation of specific sets of facial muscles. These sometimes subtle, yet complex, signals in an expression often contain an abundant amount of information about our state of mind. Automatic recognition of facial expressions can be an important component of natural human machine interfaces; it may also be used in behavioural science and in clinical practice. It have been studied for a long period of time and obtaining the progress recent decades. Though much progress has been made, recognizing facial expression with a high accuracy remains to be difficult due to the complexity and varieties of facial expressions. On a day to day basics, humans commonly recognize emotions by characteristic features, displayed as a part of a facial expression. For instance happiness is always associated with a smile or an upward movement of the corners of the lips. Similarly other emotions are characterized by other deformations typical to a particular expression.

ISSN: 2278-4632
Vol-13, Issue-12, December 2023

In machine learning, a convolutional neural network (CNN) is a type of feed forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex. Individual cortical neurons respond to stimuli in a restricted region of space known as the receptive field. The receptive fields of different neurons partially overlap such that they tile the visual field. The response of an individual neuron to stimuli within its receptive field can be approximated mathematically by a convolution operation. Convolutional networks were inspired by biological processes and are variations of multilayer perceptron designed to use minimal amounts of pre-processing. They have wide applications in image and video recognition, recommender systems and natural language processing. The convolutional neural network is also known as shift invariant or space invariant artificial neural network (SIANN), which is named based on its shared weights architecture and translation invariance characteristics.

III. AIMS AND OBJECTIVE

The objective of this project is to develop the facial expression recognition system.

Expected achievements in order to fulfill the objectives are:

1) To detect the face segment at real time.
2) To extract the useful features from the face detected.
3) To detect the facial expression from the image.
4) To implement Convolutional Neural Networks for classification of facial expression
5) To classify different emotion such as happy, sad, etc from the image.

IV. SURVEY

In a paper [1], a hybrid approach in which multi modal information for facial emotion recognition is used. In the experiment conducted by authors, they chose two different speakers using two different languages. The evaluation is carried out with three different media clips, (1) audio information of the emotions only, (2) video information of the emotions only, (3) both audio and video information (original video clip).

Video and audio dominance of each type of emotion is recorded and compared. The results of audio and facial recognition are provided as input to the weighing matrix. Inside the weighing matrix computations are made and the expression whose computed value is maximum is the result.

According to a paper [2], the problem that was solved is about Emotion recognition using facial expression. Microsoft Kinect was used for 3D modelling of the face. Microsoft Kinect has 2 cameras. One works with visible light and the other one works with infrared light. It gives three-dimensional co-ordinates of specific face muscles. Facial Action Coding System (FACS) was used to return special coefficients called Action Units (AU).

There are 6 Action Units. These Action Units (AU) represent different region of face. Six men of the age group 26-50 participated and tried to mimic the emotions specified to them. Each person had 2 sessions and each session had 3 trials. 3-NN had an accuracy of 96%.MLP had an accuracy of 90%.

According to the paper [3], CERT can detect 19 different facial actions, 6 different prototypical emotions and 3D head orientation using Facial Action Unit Coding System (FACS) and three emotion modules. It follows 6 stages: (1) Face Detection using Gentle Boost as boosting algorithm, (2) Facial Feature Detection – Specific location estimates are estimated by combining log likelihood ratio and feature specific prior at that location, and these location estimates are refined using Linear regressor, (3) Face Registration – affine wrap is made and L2 Norm is minimized between wrapped facial feature position and canonical position from GENKI dataset, (4) Feature Extraction – feature vector is obtained using Gabor filter on face patch from previous patch, (5) Action Unit Recognition – feature vector is fed to Support Vector machine to obtain Action Unit Intensities, (6) Expression Intensity and Dynamics – Empirically CERT outputs significantly correlates with facial actions.

In a paper [4], Psychological theories state that all human emotions can be classified into six basic emotions: sadness, happiness, fear, anger, neutral and surprise.

Three systems were built- one with audio, another with face recognition and one more with both. The performances of all the systems were compared. Features used for speech-global prosodic features, for facedata from 102 markers on face. Both feature level and decision level integration were implemented. The result proved that performance of both the systems was similar. However, recognition rate for specific emotions presented significant errors. The type of integration to be used is dependent on the nature of the application.

ISSN: 2278-4632
Vol-13, Issue-12, December 2023

V. ADVANTAGES & DISADVANTAGES OF DIFFERENT ALGORITHMS

Face Detection Algorithm	Advantages	Disadvantages
Viola Jones Algorithm	1. High detection speed 2. High Accuracy	1. Long training time. 2. Limited Head Pose. 3. Not able to detect dark faces.
Local Binary Pattern Histogram	1. Simple Computation 2. High tolerance against the monotonic	1. Only used for binary grey images 2. Overall performance is inaccurate compared to Viola-Jones Algorithm.
Ada Boost Algorithm	Need not to have any prior knowledge about face structure.	The result highly depends the training data and affected by weak classified.
SMQT features and SNOW classifier method	1. Capable to deal with lighting problem in object detection. 2. Efficient in computation	The region contain very similar to grey value regions will be misidentified as face.
Neural-Network	High accuracy only if large size of image were trained.	1.Detection process is slow and computation is complex. 2. Overall performance is weaker than Viola-Jones algorithm.

VI. METHODOLOGY

The facial expression recognition system is implemented using convolutional neural network. The block diagram of the system is shown in following figures:

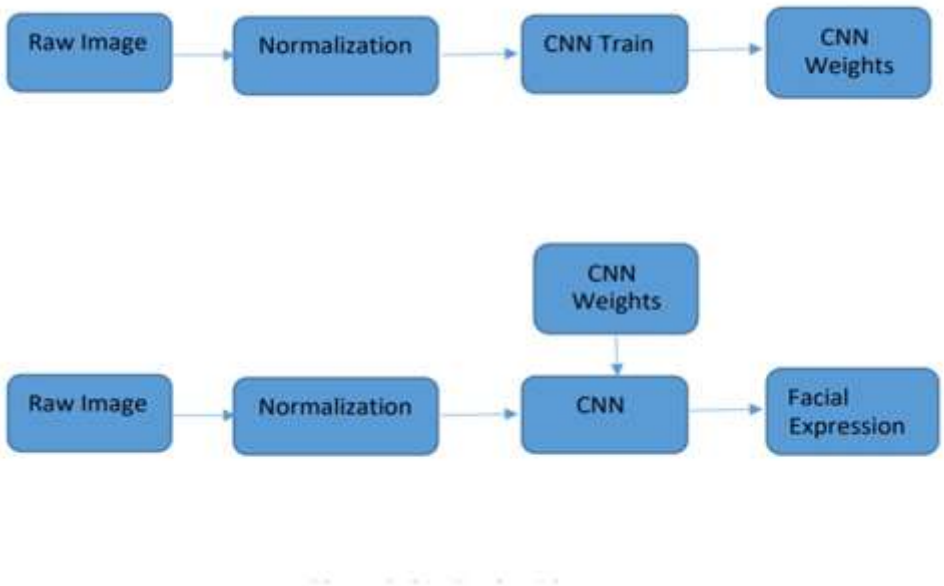
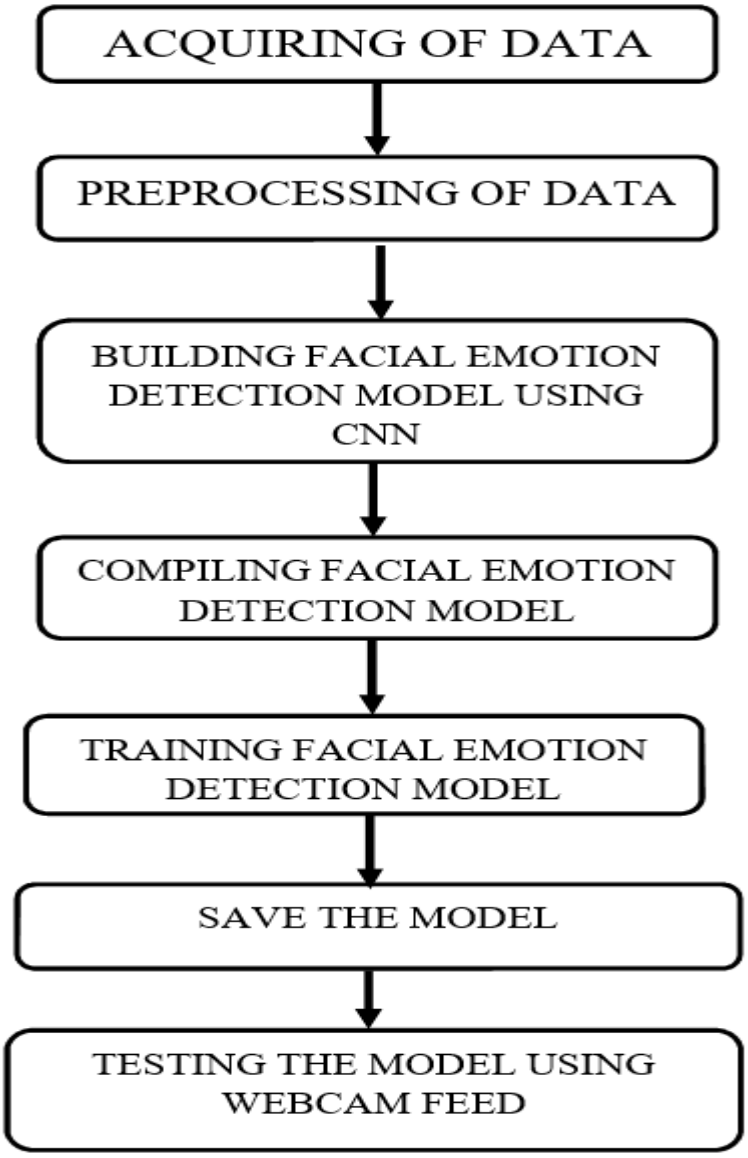


Fig: facial expression recognition system.

ISSN: 2278-4632
Vol-13, Issue-12, December 2023

During training, the system received a training data comprising grayscale images of faces with their respective expression label and learns a set of weights for the network. The training step took as input an image with a face. Thereafter, an intensity normalization is applied to the image. The normalized images are used to train the Convolutional Network. To ensure that the training performance is not affected by the order of presentation of the examples, validation dataset is used to choose the final best set of weights out of a set of trainings performed with samples presented in different orders. The output of the training step is a set of weights that achieve the best result with the training data. During test, the system received a grayscale image of a face from test dataset, and output the predicted expression by using the final network weights learned during training. Its output is a single number that represents one of the seven basic expressions.

VII. FLOW CHART



ISSN: 2278-4632
Vol-13, Issue-12, December 2023

VIII. STAGES OF COMPUTATION

A. *Dataset*

The dataset from a Kaggle Facial Expression is used for the training and testing. It is a csv file consisting of image in pixels (where each pixel varies from 0 - 255) along with a label indicating the emotion of the person in image. It comprises pre-cropped, 48-by-48-pixel grayscale images of faces each labelled with one of the 7 emotion classes: anger, disgust, fear, happiness, sadness, surprise, and neutral. Dataset has training set of 35,887 facial images with facial expression labels. The dataset has class imbalance issue, since some classes have large number of examples while some has few. The dataset is balanced using oversampling, by increasing numbers in minority classes. The balanced dataset contains 40263 images, from which 29263 images are used for training, 6000 images are used for testing, and 5000 images are used for validation.

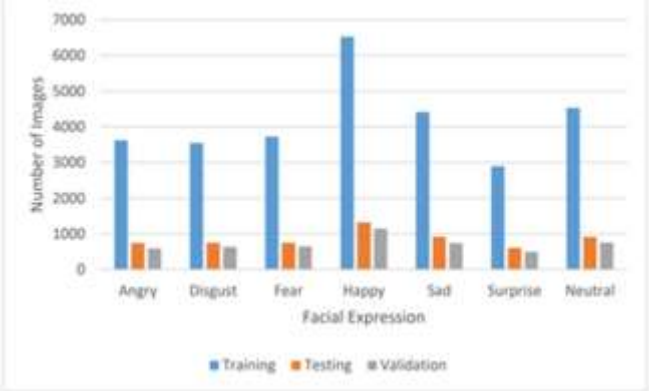


Figure 2

B. *Pre-processing*

Firstly, we collected the dataset from the public domain containing images of facial expression of different person. These images collected from the public domain are raw images. These raw images are to be processed before they are ready to use. Each image is different from other image based on size. More uniformity in data means more accuracy of the model. Therefore, we pre-process the data by rescaling or normalizing the data

C. *Training and Testing of Data*

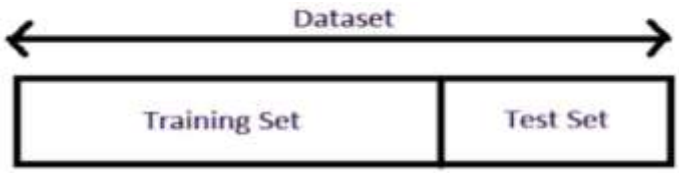


Figure 3

Training data and test data are two important concepts in machine learning.

1) *Training Data:* The observations in the training set form the experience that the algorithm uses to learn. In supervised learning problems, each observation consists of an observed output variable and one or more observed input variables.
2) *Test Data:* The test set is a set of observations used to evaluate the performance of the model using some performance metric. It is important that no observations from the training set are included in the test set. If the test set does contain examples from the training set, it will be difficult to assess whether the algorithm has learned to generalize from the training set or has simply memorized it A program that generalizes well will be able to effectively perform a task with new data. In contrast, a program that memorizes the training data by learning an overly complex model could predict the values of the response variable for the training set accurately, but will fail to predict the value of the response variable for new examples. Memorizing the training set is called over-fitting. A program that memorizes its observations may not perform its task well, as it could memorize relations and structures that are noise or coincidence. Balancing memorization and generalization, or over-fitting and under-fitting, is a problem common to many machine learning algorithms. Regularization may be applied to many models to reduce over-fitting.

ISSN: 2278-4632
Vol-13, Issue-12, December 2023

D. Classification

Previously built Convolution Neural Network (CNN) model is used. The trained data is used to predict emotion. A list with probabilities of all 7 emotions is obtained as an output. The required output is the maximum of these values and the corresponding emotion is predicted as the final output.

IX. ARCHITECTURE OF CNN

Convolutional Neural Networks (ConvNets or CNNs) are a category of Neural Networks that have proven very effective in areas such as image recognition and classification. ConvNets have been successful in identifying faces, objects and traffic signs apart from powering vision in robots and self-driving cars.

Convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analysing visual imagery. CNNs are regularized versions of multilayer perceptrons. Multilayer perceptron usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer.

Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field.

A typical architecture of a convolutional neural network contains an input layer, some convolutional layers, some fully-connected layers, and an output layer. CNN is designed with some modification on LeNet Architecture. It has 6 layers without considering input and output. The architecture of the Convolution Neural Network used in the project is shown in the following figure 4.

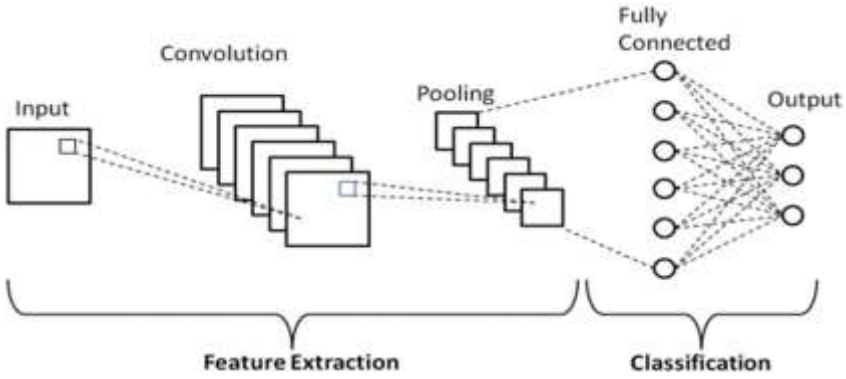


Fig: Convolutional Neural Networks

A. Input Layer

The input layer has pre-determined, fixed dimensions, so the image must be pre-processed before it can be fed into the layer. Normalized gray scale images of size 48 X 48 pixels from Kaggle dataset are used for training, validation and testing. For testing propose laptop webcam images are also used, in which face is detected and cropped using OpenCV Haar Cascade Classifier and normalized.

B. Convolution

The primary purpose of Convolution in case of a CNN is to extract features from the input image. Convolution preserves the spatial relationship between pixels by learning image features using small squares of input data. The convolution layer's parameters consist of a set of learnable filters. Every filter is small spatially (along width and height), but extends through the full depth of the input volume. For example, a typical filter on a first layer of a CNN might have size 3x5x5 (i.e. images have depth 3 i.e. the color channels, 5 pixels width and height). During the forward pass, each filter is convolved across the width and height of the input volume and compute dot products between the entries of the filter and the input at any position. As the filter convolve over the width and height of the input volume it produces a 2-dimensional activation map that gives the responses of that filter at every spatial position. Intuitively, the network will learn filters that activate when they see some type of visual feature such as an edge of some orientation or a blotch of some color on the first layer, or eventually entire honeycomb or wheel-like patterns on higher layers of the network. Now, there will be an entire set of filters in each convolution layer (e.g. 20 filters), and each of them will produce a separate 2-dimensional activation map.

ISSN: 2278-4632
Vol-13, Issue-12, December 2023

The 2-dimensional convolution between image A and Filter B can be given as:

$$C(i,j) = \sum_{m=0}^{Ma-1} \sum_{n=0}^{Na-1} A(m, n) * B(i - m, j - n)$$

where size of A is (M_a x N_a),
size of B is (M_b x N_b), $0 \leq i < M_a + M_b - 1 \wedge 0 \leq j < N_a + N_b - 1$

A filter convolves with the input image to produce a feature map. The convolution of another filter over the same image gives a different feature map. Convolution operation captures the local dependencies in the original image. A CNN learns the values of these filters on its own during the training process (although parameters such as number of filters, filter size, architecture of the network etc. still needed to specify before the training process). The more number of filters, the more image features get extracted and the better network becomes at recognizing patterns in unseen images.
The size of the Feature Map (Convolved Feature) is controlled by three parameters

1) *Depth:* Depth corresponds to the number of filters we use for the convolution operation.
2) *Stride:* Stride is the size of the filter, if the size of the filter is 5x5 then stride is 5.
3) *Zero-padding:* Sometimes, it is convenient to pad the input matrix with zeros around the border, so that filter can be applied to bordering elements of input image matrix. Using zero padding size of the feature map can be controlled.

C. Rectified Linear Unit

An additional operation called ReLU has been used after every Convolution operation. A Rectified Linear Unit (ReLU) is a cell of a neural network which uses the following activation function to calculate its output given x:

$$R(x) = Max(0,x)$$

Using these cells is more efficient than sigmoid and still forwards more information compared to binary units. When initializing the weights uniformly, half of the weights are negative. This 4 helps creating a sparse feature representation. Another positive aspect is the relatively cheap computation. No exponential function has to be calculated. This function also prevents the vanishing gradient error, since the gradients are linear functions or zero but in no case non-linear functions.

D. Pooling (sub-sampling)

Spatial Pooling (also called subsampling or down-sampling) reduces the dimensionality of each feature map but retains the most important information. Spatial Pooling can be of different types: Max, Average, Sum etc. In case of Max Pooling, a spatial neighbourhood (for example, a 2×2 window) is defined and the largest element is taken from the rectified feature map within that window. In case of average pooling the average or sum of all elements in that window is taken. In practice, Max Pooling has been shown to work better.
Max Pooling reduces the input by applying the maximum function over the input x_i. Let m be the size of the filter, then the output calculates as follows:

$$M(x_i) = \max \{x_i+k,+l \, |k| \leq m/2 \, , \, |l| \leq m/2 \, k, \, lc \, \mathbb{N}\}$$

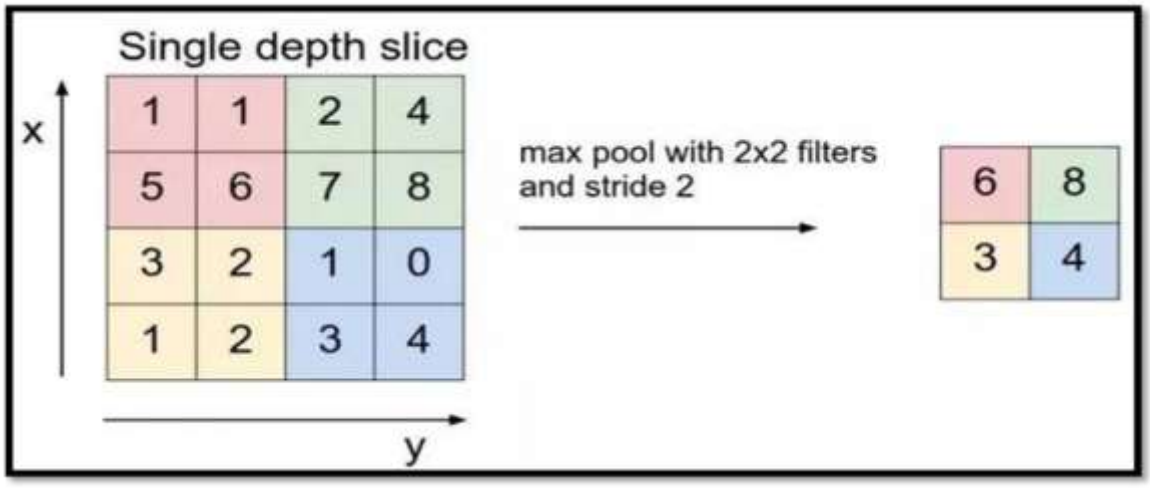


Fig: *Pooling (sub-sampling)*

ISSN: 2278-4632
Vol-13, Issue-12, December 2023

The function of Pooling is to progressively reduce the spatial size of the input representation. In particular, pooling

1) Makes the input representations (feature dimension) smaller and more manageable 5
2) Reduces the number of parameters and computations in the network, therefore, controlling over-fitting
3) Makes the network invariant to small transformations, distortions and translations in the input image (a small distortion in input will not change the output of Pooling.
4) Helps us arrive at an almost scale invariant representation. This is very powerful since objects can be detected in an image no matter where they are located.

E. Classification (Multilayer Perceptron)

The Fully Connected layer is a traditional Multi-Layer Perceptron that uses a soft-max activation function in the output layer. The term "Fully Connected" implies that every neuron in the previous layer is connected to every neuron on the next layer. The output from the convolutional and pooling layers represent high-level features of the input image. The purpose of the Fully Connected layer is to use these features for classifying the input image into various classes based on the training dataset.

Soft-max is used for activation function. It treats the outputs as scores for each class. In the Soft-max, the function mapping stayed unchanged and these scores are interpreted as the unnormalized log probabilities for each class. Soft-max is calculated as:

$$f(z)_j = \frac{exp(zj)}{\sum_{k=1}^{K} exp(zk)}$$

where j is index for image and k is number of total facial expression class.

Apart from classification, adding a fully-connected layer is also a (usually) cheap way of learning non-linear combinations of these features. Most of the features from convolutional and pooling layers may be good for the classification task, but combinations of those features might be even better. The sum of output probabilities from the Fully Connected Layer is 1. This is ensured by using the as the activation function in the output layer of the Fully Connected Layer. The Soft-max function takes a vector of arbitrary real-valued scores and squashes it to a vector of values between zero and one that sum to one.

F. Fully Connected Layer

This layer is inspired by the way neurons transmit signals through the brain. It takes a large number of input features and transform features through layers connected with trainable weights. Two hidden layers of size 500 and 300 unit are used in fully-connected layer. The weights of these layers are trained by forward propagation of training data then backward propagation of its errors. Back propagation starts from evaluating the difference between prediction and true value, and back calculates the weight adjustment needed to every layer before. We can control the training speed and the complexity of the architecture by tuning the hyper-parameters, such as learning rate and network density. Hyper-parameters for this layer include learning rate, momentum, regularization parameter, and decay.

The output from the second pooling layer is of size Nx20x9x9 and input of first hidden layer of fully-connected layer is of size Nx500. So, output of pooling layer is flattened to Nx9216 size and fed to first hidden layer. Output from first hidden layer is fed to second hidden layer. Second hidden layer is of size Nx128 and its output is fed to output layer of size equal to number of facial expression classes.

G. Output Layer

Output from the second hidden layer is connected to output layer having seven distinct classes. Using Soft-max activation function, output is obtained using the probabilities for each of the seven class. The class with the highest probability is the predicted class.

X. RESULT

CNN architecture for facial expression recognition as mentioned above was implemented in Python. Along with Python programming language, Numpy , pandas, keras ,tensorflow libraries were used.

Training image batch size was taken as 64, while filter map is of size 64x5x5 for both convolution layer. Validation set was used to validate the training process. Input parameters for training are image set and corresponding output labels. The training process updated the weights of feature maps and hidden layers based on hyper-parameters such as learning rate, momentum, regularization and decay. We got the validation accuracy of 61.6% from the classifier.

ISSN: 2278-4632
Vol-13, Issue-12, December 2023

CNN Classifier is then used to classify image taken from webcam in Laptop. Face is detected in webcam frames using Haar cascade classifier from OpenCV. Then detected face is cropped and normalized and fed to CNN Classifier
These are the following 7 emotions that we have got as output from the model:

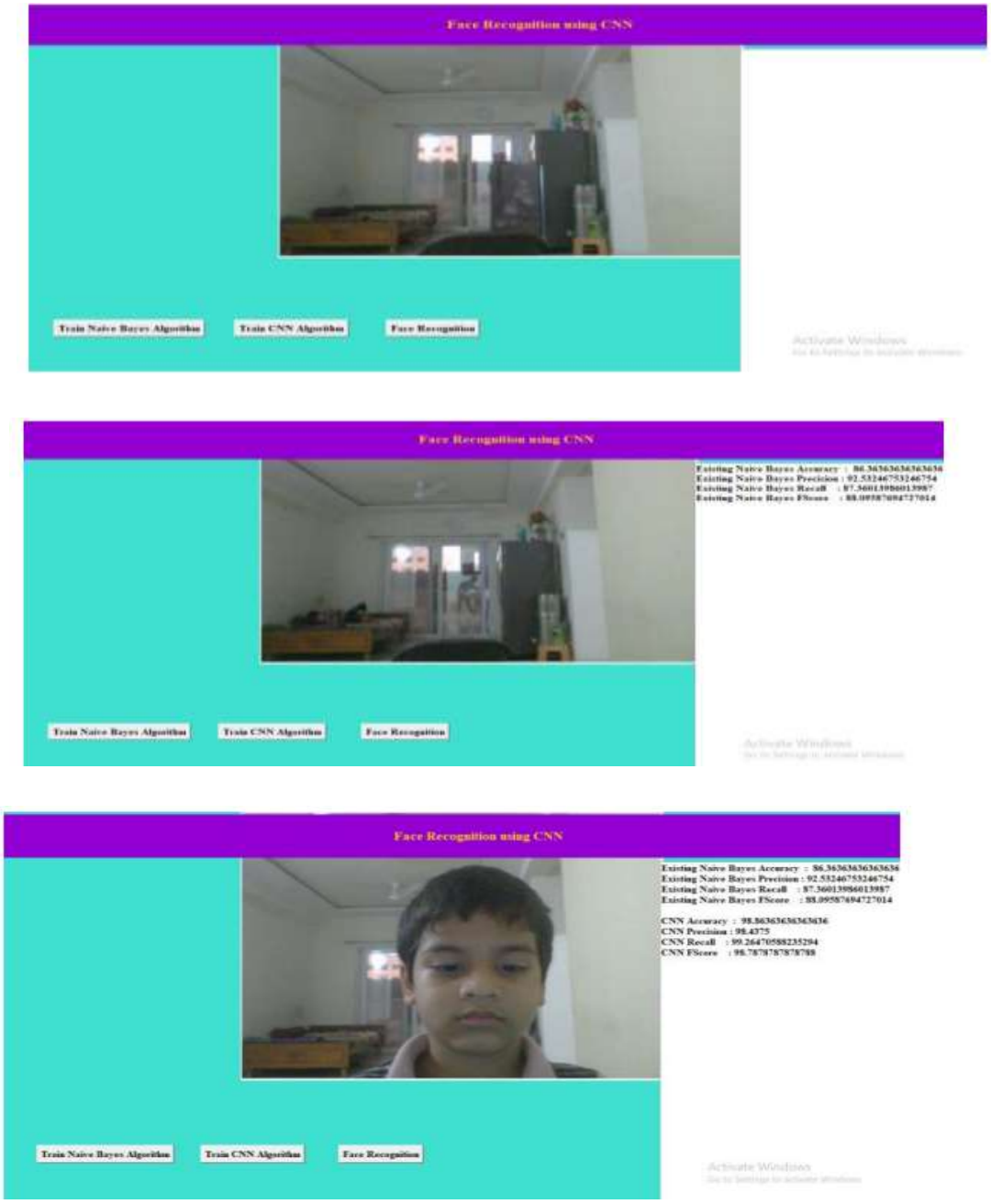
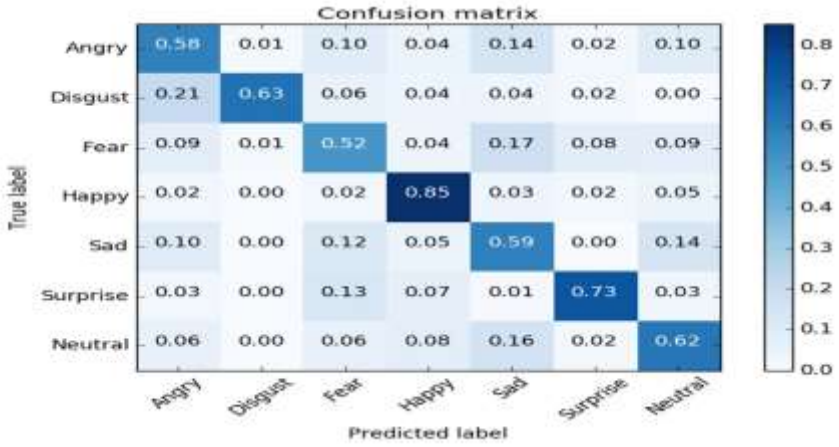


Fig: Train CNN algorithm and Face recognition output

ISSN: 2278-4632
Vol-13, Issue-12, December 2023

Fig: Prediction Probability

In above screen in blue colour text we can see person recognized as 'Person 6' and in text area we can see prediction probability % as 0.93. Similarly you can recognize all persons given in dataset and application may recognized output side persons also and for that we need to set some thresholds. While testing you note down prediction probability of dataset persons and unknown persons and based on those values will put some constant threshold. For example if prediction probability > 0.95% then only recognized else display unknown person.



The above table, showed that this model gives highest accuracy for happy emotion with 85 %, followed by surprise with 73 %, disgust with 63 %, neutral with 62 %, sad with 59 %, angry with 58% and lowest accuracy for fear emotion as 52 %.

ISSN: 2278-4632
Vol-13, Issue-12, December 2023

XI. IMPROVEMENTS &USES

For future work, our system's precision and accuracy can be improved by collecting more data from different sources & training our model with it. Additionally, in the future we can use different techniques to extract more features from the images or dataset. So that, our model can very easily detect the proper emotions for all the faces. Nowadays, emotion recognition is used for various purposes that some people do not even notice on a daily basis. Here are few areas where emotion recognition is used:

A. Security Measures

Emotion recognition is can used by schools and other institutions since it can help prevent violence and improves the overall security of a place.

B. HR Assistance

There are companies that use AI with emotion recognition system as HR assistants. The system is helpful in determining whether the candidate is honest and truly interested in the position by evaluating intonations, facial expressions, keywords, and creating a report for the human recruiters for final assessment.

C. Customer Service

There are companies that use AI with emotion recognition system as HR assistants. The system is helpful in determining whether the candidate is honest and truly interested in the position by evaluating intonations, facial expressions, keywords, and creating a report for the human recruiters for final assessment.

D. Differently Abled Children

There is a project using a system in Google smart glasses that aims to help autistic children interpret the feelings of people around them. When a child interacts with other people, clues about the other person's emotions are provided using graphics and sound.

E. Audience Engagement

Companies are also using emotion recognition to determine their business outcomes in terms of the audience's emotional responses. Apple also released a new feature in their iPhones where an emoji is designed to mimic a person's facial expressions, called Animoji.

F. Video Game Testing

Video games are tested to gain feedback from the user to determine if the companies have succeeded in their goals. Using emotion recognition during these testing phases, the emotions a user is experiencing in real-time can be understood, and their feedback can be incorporated in making the final product.

G. Healthcare

The healthcare industry sure is taking advantage of facial emotion recognition nowadays. They use it to know if a patient needs medicine or for physicians to know whom to prioritize in seeing first.

XII. CONCLUSION

In this paper, an image processing and classification method has been implemented in which images of the faces are used to train a classifier predictor that predicts the seven basic human emotions for the given test images. The procedure to predict the emotions of a person by processing the image which was taken by web cam through various stages, such as pre-processing, face detection, and classifier using CNN is showcased. The overall validation accuracy of the model is 61.6%.

FUTURE SCOPE: • In this project, six emotions (Happy, Sad, Neutral, Anger, Fear and Surprise) were identified. Various emotions can be added for further classification. • Facial emotion recognition for streaming videos can be developed. A mobile application can be developed for facial emotion detection.

ISSN: 2278-4632
Vol-13, Issue-12, December 2023

REFERENCES

[1] Anil, J., and L. Padma Suresh. "Literature survey on face and face expression recognition." Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference on. IEEE, 2016.

[2] Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski and Remigiusz J. Rak .“Emotion recognition using facial expressions”. International Conference on Computational Science(ICCS), 12- 14 June, 2017.

[3] Gwen Littlewort, Jacob Whitehill, Tingfan Wu, Ian Fasel, Mark Frank, Javier Movellan, and Marian Bartlett. “The Computer Expression Recognition Toolbox (CERT)”, Face and Gesture 2011, 21-25 March 2011.

[4] Björn Schuller, Stephan Reiter, Ronald Müller, Marc Al-Hames, Manfred Lang, Gerhard Rigoll. “Speaker Independent Speech Emotion Recognition by Ensemble Classification”, 2005 IEEE International Conference on Multimedia and Expo, 6-6 July 2005.

[5] Shan, C., Gong, S., & McOwan, P. W. (2005, September). Robust facial expression recognition using local binary patterns. In Image Processing, 2005. ICIP 2005. IEEE International Conference on (Vol. 2, pp. II-370). IEEE.

[6] Michael J. Lyons, Shigeru Akemastu, Miyuki Kamachi, Jiro Gyoba. Coding Facial Expressions with Gabor Wavelets, 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205 (1998)

[7] Shan C, Gong S, McOwan PW. Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis Comput. 27(6):803–816 (2009)

[8] Carcagnì P, Del Coco M, Leo M, Distante C. Facial expression recognition and histograms of oriented gradients: a comprehensive study. SpringerPlus. 4:645. (2015)

[9] Raghuvanshi, Arushi, and Vivek Choksi. "Facial Expression Recognition with Convolutional Neural Networks." Stanford University, 2016

[10] https://ieeexplore.ieee.org/abstract/document/9145558

[11] https://www.academia.edu/3713870/A_Review_Paper_on_FACIAL_RECOGNITION

[12] https://ieeexplore.ieee.org/document/9640896