IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 23, Issue 5, Ser. I (Sep. –Oct. 2021), PP 41-49
www.iosrjournals.org

Facial Emotion Recognition using Machine Learning Techniques

Nisha V[1], Shilpa T[2]

[1](*Dept. of Information Science and Engineering, Bangalore Institute of Technology, VV Puram, Bangalore, India)*

[2](*Dept. of Information Science and Engineering, Bangalore Institute of Technology, VV Puram, Bangalore, India)*

Abstract: *The Human facial expressions transmit a great deal of information visually rather than verbally. In the field of human-machine interaction, facial expression recognition is critical. Many applications exist for automatic facial expression recognition systems, but not limited to human behavior understanding, diagnosis of mental illnesses, and synthetic human emotions. Facial expression identification by computer with a high recognition rate remains a difficult challenge. Geometry and appearance are two common techniques used in the literature for automatic FER systems. Pre-processing, face detection, feature extraction, and expression classification are the four phases of facial expression recognition. Convolutional Neural Networks are used to identify the core seven human emotions of the project: anger, disgust, fear, happiness, sad, surprise, and neutrality. FER-2013 dataset is used to train and test the model which is freely available on Kaggle's FER challenge. The dataset contains 35887 face crops, with the training set containing 28709 face crops and testing set containing 7178 face crops. The model has an accuracy of 66% over validation data.*

Materials and Methods*: Facial Emotion Recognition (FER) is the innovation that analyzes emotions from both static images and video to uncover data on one's passionate state. The dataset used for facial emotion identification is FER-2013, which was produced by compiling the results of each emotion's Google image search. FER-2013 is a big dataset and is freely available on Kaggle's FER challenge. There are 3995, 436, 4097, 7215, 4965, 4830 and 3171 face crops used for training the emotions such as anger, contempt(disgust), fear, happiness, neutrality, sadness, and surprise. Similarly, there are 958, 111, 1024, 1774, 1233, 1247 and 831 face crops used for testing the emotions such as anger, contempt(disgust), fear, happiness, neutrality, sadness, and surprise. The data is in grayscale and has a resolution of 48 × 48 pixels. All of the pictures in the collection are automatically registered and take up the same amount of space. The main objective is to categorize each face into one of seven categories with matching labels depending on the reaction of the facial expression (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The dataset contains 35,887 face crops, with the training set containing 28,709 face crops and the testing set containing 7178 face crops. The testing is separated into testing and validation, with the testing set consisting of 3589 face crops and the remaining 3589 face crops being utilized for validation.*

Results*: Emotion Recognition has been an important aspect for human-to-human interaction. For human-to-machine interaction the simple and effective model is built where those models are trained and tested for the implementation process of emotion recognition. The face emotions are recognized using the models which are trained and tested.*

Conclusion: *A Facial Sentiment Acknowledgement framework to distinguish and characterize facial feelings is developed. The classifier model comprised of a pre- trained neural network for feature extraction and SVM for the classification of emotions. This model has an accuracy of 66% over validation data. The real-world application of the identified emotions in affective computations are tried, utilizing an arrangement to control different applications that help in developing driver support frameworks, such as, speed controller.*

Key Word*: FER, CNN, Recognition, Emotion.*

Date of Submission: 28-09-2021 Date of Acceptance: 12-10-2021

I. Introduction

The method of recognizing human sentiments from facial expressions is known as facial emotion recognition. This technology is getting better by showing the accurate results and by recognizing the emotions of an individual. In a framework including human-machine interaction, the emotion recognition has consistently been a wide space of examination since the machines can never break down the emotion of a person. In spite of the progression made in recognition of face emotion, it is yet far from having a markable relationship among human and machine, that the machine doesn't understand the state of a person. Recognizing emotions is naturally performed by people. It is the most essential process for human-to-human interaction, and subsequently, to accomplish better human-machine interaction, emotions should be considered. Human face expressions may be readily categorized into seven main emotions along with emojis as shown in the Figure 1. The emotions are: anger, surprise, neutral, fear, happy and sad.

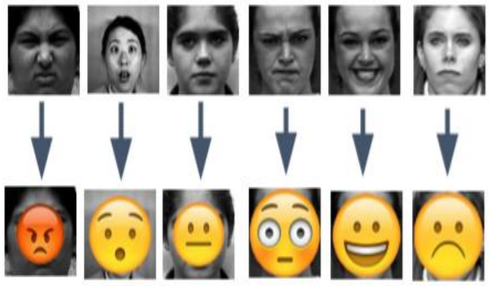


Figure 1: Emojis imposed on input images through webcam

Convolutional Neural Network (CNN) was initially suggested in 1960s by Hubel and Wiesel, who revealed its exclusive network topology, which may significantly minimize network complexity. CNN can directly input the face using a web cam, making image preprocessing straightforward. To minimize the training requirements, CNN used the methods of local detection field, mass distribution, and merging. CNN has picture translation, rotation, and coloring invariance all at the same time. CNN is specifically used to computer vision, including image classification and object identification, in the investigation of deep learning. Prior to the widespread usage of CNN, most pattern recognitions were performed using a physical feature removal and classifier [3]. However, the emergence of CNN completely altered pattern recognition. It employs the raw pixel strength of the input picture as a level vector rather than a conventional manual feature extraction approach. Convolutional Neural Networks have an incredible ability for processing two-dimensional data as a whole. The benefits of CNN are as follows: the features of the face may be recorded regardless of their position, and users do not need to arrange the filters to extract the characteristics.

II. Problem Definition

In Facial Emotion Recognition (FER), it includes two parts, one is to identify emotion in images. The second part is to apply them on real time video section. The facial emotion recognition in images is trying to identify one of the 7 facial emotions: happy, angry, neutral, surprised, sad, fearful and disgust [9]. A lot of facial emotion images with different test data is used in order to train a neural network and effectively identify the features representing the facial emotions. Then it is applied to a series of unseen test images. It is then applied to recognize the emotions of faces in real time.

III. Objective

The primary goal of Facial Emotion Recognition is to create a system that can recognise diverse human emotional states by studying facial expressions, which may then be utilised to improve Human-Computer Interaction (HCI)[2]. It also evaluates the outcomes for each class in terms of accuracy. As illustrated in Figure 2, it is capable of automatically detecting the seven fundamental or universal expressions: joyful, sad, angry, surprise, fear, disgust, and neutral.

Figure 2: Different human emotions

IV. Material and Methods

Procedure Methodology

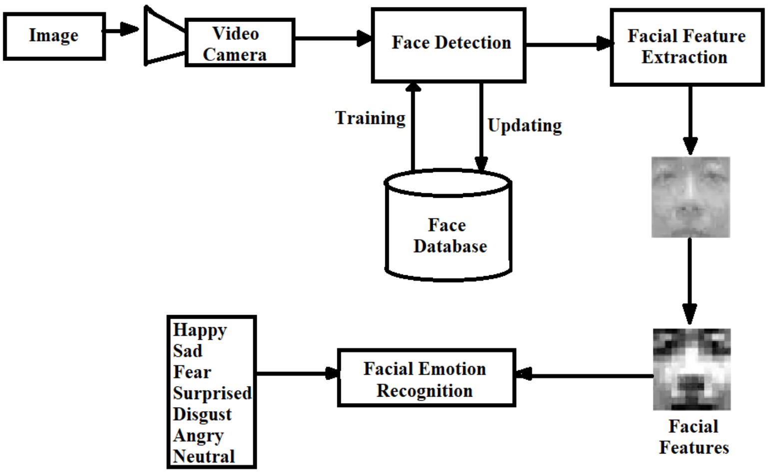


Figure 3: System Architecture of Facial Emotion Recognition

The Figure 3 shows the architecture of Facial Emotion Recognition. Live streaming concept starts here, that is, ultrasonic wave frequency is initiated in each and every second, to detect the presence of human. As soon as, the video is on, video streaming reads and stores the loading of face emotions dataset [3]. The stored frames resize the image captured by it, to make a proper frame based on streaming. Later it detects the faces captured in video and it extracts the face region. After extracting, classification is done based on one of the basic 7 emotion shown by the person on the webcam [7].

Block Diagram

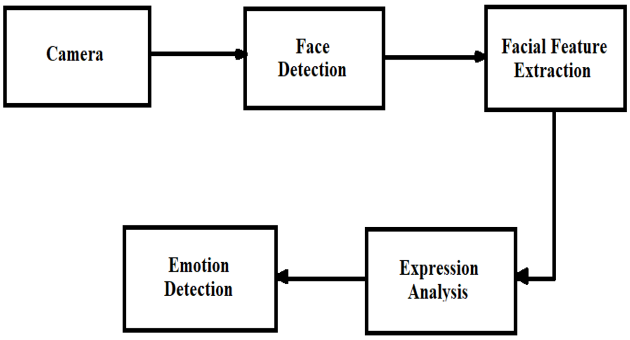


Figure 4: Block diagram of Facial Emotion Recognition

Facial emotion recognition steps are shown in Figure 4 which identifies the emotion of a person's face displayed on the webcam. It reads ultrasonic sensor for the detection of human presence, which is setup, to check each and every second. After the presence of human is detected, then the camera will focus on captured face image alone. Using image processing concept (Neural Network) an identification is done to know what kind of expression is the person showing along with accuracy of that particular emotion class. If the person does not show any expressions, then the emotion named neutral is displayed with the accuracy and emoji, which is one of the emotions for which the model is trained [3,6].

V. Algorithms

Algorithm for Pre- processing and training on dataset

Data preprocessing is the process of preparing raw data for use with a machine learning model. It is the first and most important stage in developing a machine learning model. When developing a machine learning project, it does not always come across clean and structured data. And, before doing any action on data, it must be cleaned and formatted. As a result, the data preparation job is employed for this. The first thing needed to construct a machine learning model was a dataset, because a machine learning model is entirely dependent on data. The dataset is a collection of data in a certain format for a specific task.

Inputs: Images along with their pixel values.
Output: Training model.
Step 1: Load image and their pixel values.
Step 2: Process the image i.e., resizing, normalization and conversion to 1D array.
Step 3: Load the filenames and their respective labels.
Step 4: Perform data augmentation and then split data into training and testing batches.
Step 5: Load the model from Keras, train it on training batches and complete it using optimizers.
Save the model for future use.

Algorithm for Optimal Face Detector

Object detection is a computer technology related to image processing and computer vision. It is focused with detecting occurrences of an item such as human faces, buildings, trees, automobiles, and so on. The basic goal of face detection algorithms is to assess whether or not a face exists in a picture. Given a picture (this technique only works with grayscale images), the programmer examines several smaller subregions and attempts to locate a face by looking for certain traits in each subregion. Because a picture might have multiple faces of varying sizes, it must verify many possible locations and scales. In this method, Haar-like characteristics is employed to recognize faces.

i. Input
   Image ← input image
   D_{fast} ← single stage detector
   D_{slow} ← two stage detectors
   CP ← image complexity predictor
   τ← hardness threshold
ii. Computation
   If (CP (Image) > τ)
      R ← D_{slow}(Image)
   else
      R ← D_{fast}(Image)
iii. Output
   R ← set of region proposals

Algorithm for Feature Extraction

The face recognition technique is used to recognize a person by using some of that person's characteristics and matching those traits with a digital image. Eyes, nose, skin, iris, finger, and other characteristics are retrieved for face identification. According to some authors, “Face recognition is considered to be an important part of the biometrics technique or software application by which it can be analyzed, identified, or verify digital image of the person by using the feature of the person's face that are unique characteristics of each person” [1].

Input: Extracted face boundary from the previous layer
Output: 1x30 matrix (x and y coordinates for 15 feature points)
Step 1: Import Kaggle dataset
Step 2: Pre-process dataset
Step 3: Split dataset into train and test
Step 4: Crop face from the boundary (a subset of image matrix)
Step 5: Convert image to grayscale
Step 6: Use OpenCV library function to transform an image to the parameterized location
      Face ← cv2.resize(cropped face,(96,96))
Step 7: Load a pre- trained Keras model
      Model ← loadModel (“file_location”)

- Add BatchNormalization(input)layer
- Add Convolutional2D layers
- Apply ReLu activation function
- Add MaxPooling2D layers
- Aggregate using GlobalAveragePooling2D layer
- Add Dense layers
- Compile model with parameters: optimizer (adam), loss(mse), metrics(accuracy)
- Add modelCheckpoint for the best trained model
- Save the model

Step 8: Prediction ← Model.predict(Face,batch_size=1)

Algorithm for Support Vector Machine

Support vector machines are supervised machine learning algorithm which will be used for multiple selected categorization or regression challenges. SVM is a machine learning model that is able to discover between two various classes if the set of specific data is provided in the training set of algorithms.

Input: k, m, q, C, γ, and termination criterion

Output: Optimal value for SVM parameters and classification accuracy

```
Begin
    Initialize k solutions
    sample selected S
    store newly generated solutions
    call SVM algorithm to evaluate k solutions
        T= Sort (S1,…….., SK)
    while classification accuracy ≠ 100% or number of iterations ≠ 10 do
        for i = 1 to m do
            select S according to its weight
            call SVM algorithm to evaluate newly generated solutions
        end
        T = Best (Sort S1, ………., Sk + m), K)
    end
End
```

VI. Implementation

This system for recognizing the emotions from face was inspired by certain works around Convolutional Neural Network approach. These works encouraged us to come up with our own ideologies. The related work for the present model has been summarized below.

Face Detection System

Edwards et al. explains the recent approach of him, in the form of 2 Dimensional or 3 Dimensional recognitions of face in progress, focusing mostly approach based on local, holistic (subspace), and hybrid quality. A comparative work between these approaches in terms of refining time, complexity, discrimination, and fitness will be carried out [6]. They had concluded that local feature methods, which were the best option based on discrimination, rotation, relocation, convolution, and accuracy. They had a hope that this might further encourage researchers in this field to regulate and pay more attention to the usage of local procedures for face recognition systems.

Face Recognition model

Facial Emotion Recognition (FER) is the innovation that analyzes emotions from both static images and video to uncover data on one's passionate state. The complexity of emotions, the expected utilization of the innovation in any specific situation, and the contribution of new advances, for example, man-made reasoning raises critical protection risks. It has a place with the group of innovations which is referred to 'affective computing', a multidisciplinary field of examination on computer's abilities to perceive and decipher human feelings and emotional states and it frequently expands on Artificial Intelligence innovations.

Convolutional Neural Network (CNN)

Convolutional Neural Network is an algorithm in the research of Deep Learning, which is outlined for preparing organized varieties of information. They are extremely acceptable at getting on design in the input image such as lines, circles or even eyes and faces. This characteristic makes CNN powerful for computer vision. It can run undeviatingly on an under gone image without worrying about any pre- processing. It a feed-forward neural network. The strength of CNN comes from a specific sort of layer called the convolutional layer. It contains numerous convolutional layers gathered on top of one another, each one equipped of perceiving more refined shapes. With three or four convolutional layers it is possible to perceive manually written digits and with 25 layers it is feasible to differentiate human faces and the voice of a speaker to recognize the emotions. The main objective is to train and activate machines to see the world as people do. The Figure 5 shows the architecture of CNN, each layer in the neural network performs different operations which are already predefined.

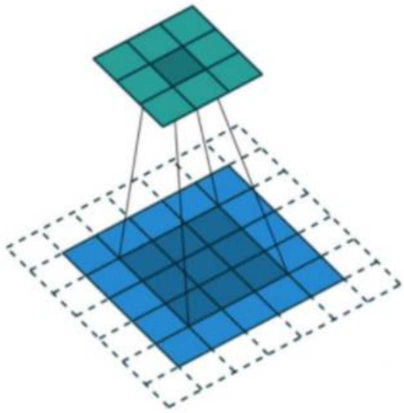


Figure 5: Convolutional Neural Network Architecture

a. Feature Extraction

In the convolutional level, the channel starts over the upper left most space of the information picture & finishes grid augmentation activity among the qualities of channel in addition to the pixel esteems in space of picture that is floating over. The outcome starting with the estimation is to put over the suitable space on the output "include map" (the pink 3 x 3 lattice). The frequency then, at that point that prints one pixel to one side then changes the interaction. The process keeps on working up until it can presently don't move right, whereby it rather moves to outermost left space of image & downcast one pixel [5]. This channel resolves to finish an aggregate of 9 grid duplications which will fill every passage in the 3 x 3 feature plan as displayed in Figure 6.

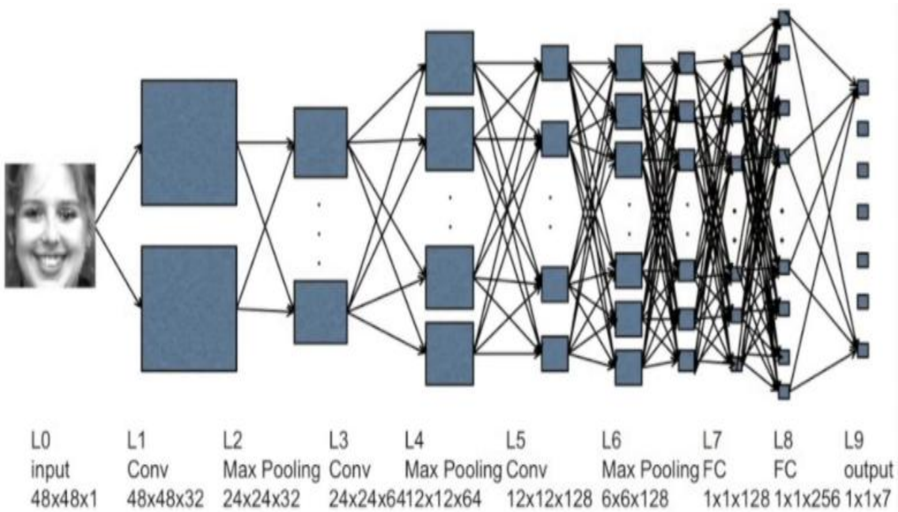


Figure 6: Feature Map

b. Feature Classification

Haar is a feature similar to Kernels that is often used to identify edges. Some features are shared by all human faces, such as the fact that the eye region is darker than the upper check region and the nose region is brighter than the eye region. Their size and position, along with this matchable trait, will let us detect a face. Here are several Haar features that may be utilised to assess whether or not a face exists. According to the Haar feature, the black region is represented by +1 and the white region by -1. In a 24x24 window, it shows an image. Each feature is a single value calculated by subtracting the sum of the pixels under the white rectangle from the sum of the pixels under the black rectangle.

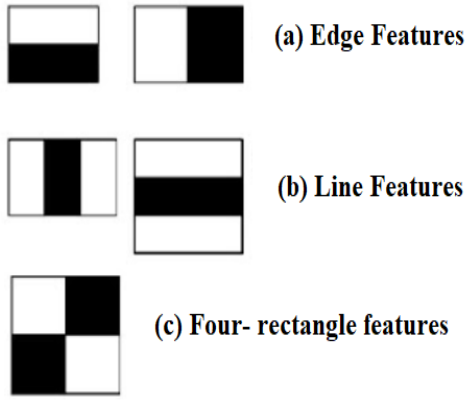


Figure 7: Haar Features

The Figure 7 illustrates the Haar Cascade Classifier's features. The possible sizes and locations of each kernel are now used to define a variety of properties. The number of pixels under white and black rectangles must be determined for each feature computation. A 24x24 window will have 160000+ Haar features, which is a huge number [9]. To solve this difficulty, they utilised integral images. It simplifies the computation of the sum of pixels to a four-pixel operation, regardless of the number of pixels.

Activation Function

ReLu is a non-linear activation function used in multi-layer neural networks and deep neural networks. Figure 8 depicts the graph of the ReLu activation function. This function can be referred to as:

$$f(x) = \max(0,x)$$

According to the previous equation, the output of ReLu is the highest value between 0 and the input value. When the input is negative, the output is equal to zero; when the input is positive, the output is equal to the input value. The preceding equation may be rewritten as:

$$f(x) = \begin{cases} 0, \text{ if } x < 0 \\ 1, \text{ if } x \geq 0 \end{cases}$$

where x is an input value

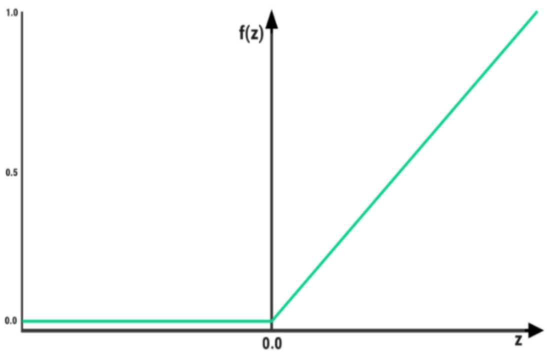


Figure 8: ReLu Activation Function

VII. Experimental Results

Emotion Recognition has been an important aspect for human-to-human interaction. For human-to-machine interaction the simple and effective model is built where those models are trained and tested for the implementation process of emotion recognition. The Figure 6 shows the clear idea of implementation of the project. The face emotions are recognized using the models which are trained and tested.

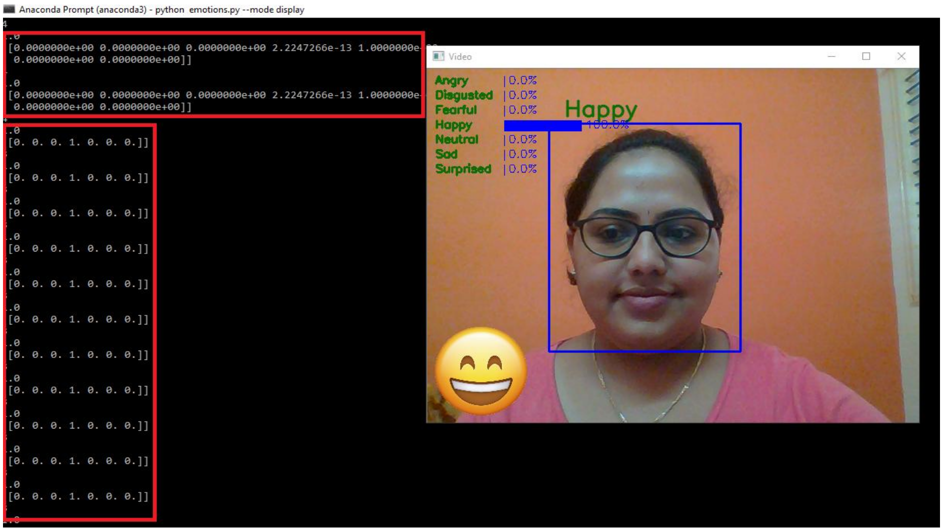
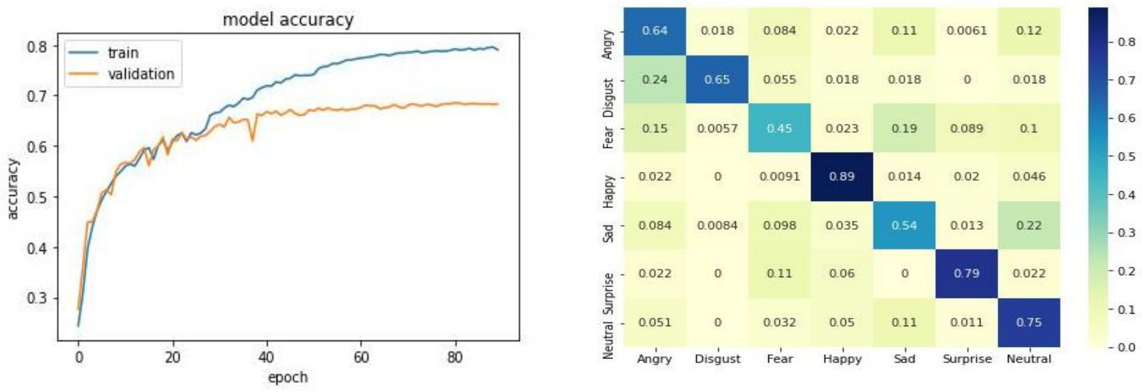


Figure 9: Displaying happy emotion among the 7 emotions

Similar to the Figure 9 all the remaining emotions like neutral, fear, sad, disgust, surprise and angry are displayed along with the emojis, confusion matrix and the accuracy. In order to quit the video "Q or q" as to be pressed on the keyboard.

The accuracy shown in the Figure 9 can be graphically represented as shown in the Figure 10(a). Similarly, the confusion matrix displayed in a single line in the Figure 9 can be represented in the form of an actual matrix as shown in the Figure 10(b).



(a) (b)

Figure 10: Displaying the (a) accuracy graph and (b) confusion matrix

VIII. Conclusion and Future Work

A Facial Sentiment Acknowledgement framework to distinguish and characterize facial feelings is developed. The classifier model comprised of a pre- trained neural network for feature extraction and SVM for the classification of emotions. The classifier model is prepared on 35,887 face crops, where the training set includes 28,709 face crops, the testing set includes 7178 face crops. The testing is split into testing and validation where the number of face crops are divided, so the testing set consists of 3589 face crops and the remaining 3589 face crops are used for validation of FER2013 database. The face is identified utilizing face detector procedure and the emotions are arranged utilizing the trained classifier model. This model has an accuracy of 66% over validation data. The real-world application of the identified emotions in affective computations are tried, utilizing an arrangement to control different applications that help in developing driver support frameworks, such as, speed controller.

The future characteristics doesn't function well in low light situations; consequently, this prototype can be enhanced to work appropriately in such conditions. The projected framework shows low precision rate for the recognition of fear and sad appearance. This can be enhanced by preparing the pictures on a bigger record. The accurateness can be expanded by making a personalized file to train the model.

REFERENCES

[1]. Min Shi, Lijun Xu, Xiang Chen. "A Novel Facial Expression Intelligent Recognition Method Using Improved Conventional Neural Network." Digital Object Identifier 10.1109/ACCESS.2020.2982286.

[2]. Xiangjian Chen, Di Li, Pingxin Wang AND Xibei Yang. "A Deep Convolutional Neural Network with Fuzzy Rough Sets for FER." Digital Object Identifier 10.1109/ACCESS.2019.2960769.

[3]. Mariana-Iuliana Georgescu, Radu Tudor Ionescu, (Member, IEEE), and Marius POPESCU. "Local Learning with Deep and Handcrafted Features for Facial Expression Recognition." Digital Object Identifier 10.1109/ACCESS.2019.2917266.

[4]. Ji-Hae Kim, Byung-Gyu Kim, Partha Pratim Roy, Da-Mi Jeong, "Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure." Digital Object Identifier 10.1109/ACCESS.2019.2907327.

[5]. Grudin, Maxim A. "Face recognition systems." Pattern recognition 33.7 (2019): 1161-1177.

[6]. Ahonen, Timo, Abdenour Hadid, and Matti Pietikainen. "Face description with local binary patterns: Application to face recognition." IEEE transactions on pattern analysis and machine intelligence 28.12 (2020): 2037-2041.

[7]. Albawi, Saad, Tareq Abed Mohammed, and Saad Al-Zawi. "Understanding of a convolutional neural network." 2017 International Conference on Engineering and Technology (ICET). IEEE, 2017.

[8]. Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep face recognition." (2018).

[9]. Michael Revina, W.R. Sam Emmanuel, "A Survey on Human Face Expression Recognition Techniques" Journal of King Saud University Computer and Information Sciences, 2018.

[10]. Mrs. Jyothi S Nayak, Preeti G, Manisha Vatsa, Manisha Reddy Kadiri, Samiksha S "Facial Expression Recognition: A Literature Survey International Journal of Computer Trends and Technology (IJCTT)" Volume- 48 Number-1 2017.

Nisha V. "Facial Emotion Recognition using Machine Learning Techniques." *IOSR Journal of Computer Engineering (IOSR-JCE),* 23(5), 2021, pp. 41-49.