International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Real-Time Emotion Recognition System Using Facial Expressions

Mr. Ankit Singh, Ms. Vaishnavi Mahalim, Ms. Divya Kundgar, Ms. Seema Admane

Under The Guidance of Prof. Vikram Popat Deokate

Department of Computer Engineering

Al Ameen Educational & Medical Foundation's College Engineering

Koregoan Bhima, Pune-412216

ABSTRACT

This paper describes an emotion detection system based on real-time detection using image processing with human-friendly machine interaction. Facial detection has been around for decades. Taking a step ahead, Human expressions displayed by face and felt by the brain, captured via video, electric signal, or image form can be approximated. To recognize emotions via images or videos is a difficult task for the human eye and challenging for machines thus detection of emotion by a machine requires many image processing techniques for feature extraction.

This paper proposes a system that has two main processes such as face detection and Facial expression recognition (FER). This research focuses on an experimental study on identifying facial emotions.

The flow for an emotion detection system includes the image acquisition, preprocessing of an image, Face detection, feature extraction, and classification. To identify such emotions, the emotion detection system uses KNN Classifier for image classification, and Haar cascade algorithm an Object Detection Algorithm to identify faces in an image or a real-time video.

This system works by taking live images from the webcam. The objective of this research is to produce an automatic facial emotion detection system to identify different emotions based on these experiments the system could identify several people that are **sad**, **surprised**, and **happy**, in **fear**, are **angry**, **disgust** etc.

CHAPTER 1

INTRODUCTION

Human Emotion Detection is used in various situations when extra security or personal information is crucial. Upon setup, the secondary security layer provides the ability to identify faces with emotions and can also be helpful in confirming if the image is of a specific person or a two-dimensional approximation of the person in front of the camera. Aside from this, business marketing are yet another benefit of utilizing EMS with machine learning. Customer feedback on services or products, such movie streaming services, is a major source of revenue for many large-scale organizations.

The goal is to develop a graphical user interface (GUI) that can recognize a person's facial expression and generate an output based on that calculation. Real-time picture data can be used to calculate the outcome. Currently, for the software to function properly, the camera must be positioned precisely in front of the person listed in the program. If everything proceeds according to plan, we shall receive the result.

Emotions are easy for humans to understand, but not so much for machines. Thus, we are attempting to identify emotions that go beyond just facial expressions.

With a degree of automation that allows for seamless device-human interaction, the ubiquitous computing paradigm is starting to take shape. Paradoxically, a primary obstacle is the complexity of these systems, which makes it hard for users to interact with them. Therefore, it is crucial for the next generation of user interfaces to enable machines to sense user emotions, particularly those of frustration, fear, or dislike.

A facial expression is one or more motions or positions of the muscles beneath the skin of the face. According to one set of controversial theories, these movements convey the emotional state of an individual to observers. Facial expressions are a form of nonverbal communication. Facial expressions are vital to social communication between humans. Facial expression classifiers generalize the learned features to recognize different expressions from unseen faces. With advancements in computer vision and deep learning, it is now possible to detect human emotions

from images in an improved way. A Convolutional Neural Network (CNN) is a Deep Leaning Learning Algorithm which takes an image as the input, assigning importance to various objects in

the image so that it can differentiate it from others. The preprocessing required in a CNN model is lower as compared to other classification models. We are using CNN because it has an ability which automatically detects the important features without any human supervision

1.1 Existing System

The study of nonverbal communication via emotions originated with Darwin's claim that emotion expressions evolved in humans from pre-human nonverbal displays. Furthermore, according to Ekman, there are seven basic emotions which have acquired a special status among the scientific community: Anger, Disgust, Fear, Happiness, Neutral, Sadness, and Surprise. The compactness of emojis reduces the effort of input to express not only emotions, but also serves to adjust message tone, increase message engagement, manage conversations and maintain social relationships. Moreover, emojis do not have language barriers, making it possible for users across countries and cultural backgrounds to communicate. In a study by Barbieri et al., they found that the overall semantics of the subset of the emojis they studied is preserved across US English, UK English, Spanish, and Italian. As validation of the usefulness of mapping emojis to emotions, preliminary investigations reported by Jaeger et al. suggest that emoji may have potential as a method for direct measurement of emotional associations to foods and beverages.

There are several datasets available for research in the field of Facial Expression Recognition, such as the Japanese Female Facial Expressions (JAFFE), Extended Cohn Kanade dataset (CK+), and the FER2013 dataset (Canade, Cohn & Tian, 2000; Lucey et al., 2010; Goodfellow et al., 2013). The type and number of images, the method of labelling the images varies in each dataset. The CK+ dataset uses the FACS system for labelling faces and contains the Action Units (AU's) for each facial image.

There are several challenges with implementing the FER system. Most datasets consist of images of posed people with a certain expression. This is the first challenge; as real time applications require a model with expressions which are not posed or directed. The second challenge is that the labels in the datasets are broadly classified, which means that in real time there might be some expressions which the system might be able to classify correctly.

1.2 Scope

The scope of this system is to tackle with the problems that can arise in day to day life.

Some of the scopes are:

1. The user's mental state can be identified and monitored by the system.

2. Mini-marts and shopping centers can use the system to view customer
3. feedback and improve their operations.
4. The technology can be installed to detect people's faces and facial expressions in crowded locations like bus, train, and airport terminals. The system may sound an internal alarm if any faces were seen to be suspicious, such as ones that showed anger or fear.
5. The system can also be used for educational purposes, including providing feedback on how students are responding in class.
6. During an interrogation, this system can be used to detect lies from criminal suspects.
7. This system can assist researchers studying emotions in better processing emotion data.
8. Using an individual's emotional knowledge, which this system can identify, clever marketing is possible The primary scope of this project is to establish a model that can classify seven basic emotions: happy, sad, surprise, angry, disgust, neutral, and fear and to achieve the accuracy

better than the baseline. In addition to this, our project also aims to analyze the results of our model in terms of accuracy for each class. In the future, the model is expected to perform wild emotion recognition that has more complex variance of condition than lab condition images.

The primary scope of this project is to establish a model that can classify seven basic emotions: happy, sad, surprise, angry, disgust, neutral, and fear and to achieve the accuracy better than the baseline. In addition to this, our project also aims to analyze the results of our model in terms of accuracy for each class. In the future, the model is expected to perform wild emotion recognition that has more complex variance of condition than lab condition images.

The use of Emoji in marketing activities can enhance the appeal of these activities and bring them closer to the younger generation. It can also have an impact on consumers, including optimizing consumer experience, improving purchase intention, and changing perceptions of brands. Emoji can be used to measure users' emotions and depict the portraits of users.

We have also been motivated observing the benefits of physically handicapped people like deaf and dumb. But if any normal human being or an automated system can understand their needs by observing their facial expression then it becomes a lot easier for them to make the fellow human or automated system understand their needs. Significant debate has risen in the past regarding the emotions portrayed in the world-famous masterpiece of Mona Lisa. British Weekly „New Scientist‟ has stated that she is in fact a blend of many different emotions, 83%happy, 9% disgusted, 6% fearful, 2% angry.

1.3 Operating Environment

1.3.1 USER INTERFACES

This tells about user interfaces how it will work and how it will be display like that. User interface is part of software and is designed such a way that it is expected to provide the user insight of the software. Ul provides fundamental platform for human-computer interaction.

Backend: python 3.10

Front End: python, PyQt 6

1.3.2 SOFTWARE INTERFACES

1. Microsoft Word
2. Dataset Storage
3. Operating System: Windows10

1.3.3 HARDWARE INTERFACES

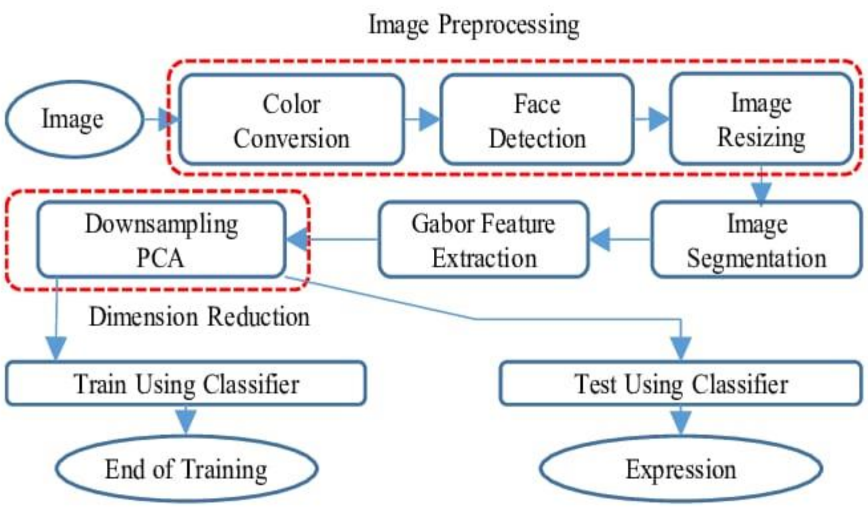
The most common set of requirements defined by any operating system or software application is the physic computer resources, also known as hardware, a hardware requirements list is often accompanied by a hardware capability list (HCL), especially in case of operating systems. An HCL lists tested, compatible, and sometimes incompatible hardware devices for a particular operating system or application. The following subsection discuss the various aspects of hardware requirements.

CHAPTER 2

SYSTEM

2.1 Proposed system

The block diagram of emotion recognition system is shown in below fig. and the working of each block is explained in detail in following subsections.



2.1.1 IMAGE PREPROCESSING:

Pre-processing is a common name for operations with images at the lowest level of abstraction — both input and output are intensity images. The aim of pre-processing is an improvement of the image data that suppresses unwilling distortions or enhances some image features important for further processing, although geometric transformations of images (e.g. rotation, scaling, translation) are classified among pre-processing methods here since similar techniques are used.

i) Color conversion:

The process of converting colors from one color space to another is called color conversion. All colors in a color space are fixed relative to the color space's white point. The color conversion can be done by following steps :

- Take a Red-Green-Blue (RGB) image as input into Python and separate individual color channels
- Recombine RGB color channels to reconstruct the initial image
- Convert an RGB image and convert it to a gray scale image in Python.

ii) Face detection:

Face detection is a computer technology being used in a variety of applications that identifies human faces in digital image Face-detection algorithms focus on the detection of frontal human faces. It is analogous to image detection in which the image of a person is matched bit by bit. Image matches with the image stores in database. Any facial feature changes in the database will invalidate the matching process.

Firstly, the possible human eye regions are detected by testing all the valley regions in the gray-level image. Then the genetic algorithm is used to generate all the possible face regions which include the eyebrows, the iris, the nostril and the mouth corners.

iii) Image Resizing:

Image resizing refers to the scaling of images. Scaling comes in handy in many image processing as well as machine learning applications. It helps in reducing the number of pixels from an image and that has several advantages. It also helps in zooming in on images. Many times, we need to resize the image i.e. either shrink it or scale it up to meet the size requirements.

OpenCV provides several interpolation methods for resizing an image as follows:

- cv2.INTER_AREA: This is used when we need to shrink an image.
- cv2.INTER_CUBIC: This is slow but more efficient.
- cv2.INTER_LINEAR: This is primarily used when zooming is required. This is the default interpolation technique in OpenCV.

2.1.2 IMAGE SEGMENTATION:

Image segmentation involves converting an image into a collection of regions of pixels that are represented by a mask or a labeled image. By dividing an image into segments, you can process only the important segments of the image instead of processing the entire image. A common technique is to look for abrupt discontinuities in pixel values, which typically indicate edges that define a region.

2.1.3 GABOR FEATURE EXTRACTION:

A Gabor filter is a linear filter used in image processing for edge detection, texture classification, feature extraction and disparity estimation. It is a bandpass filter, i.e. it passes frequencies in a certain band and attenuates the other frequencies outside such band.

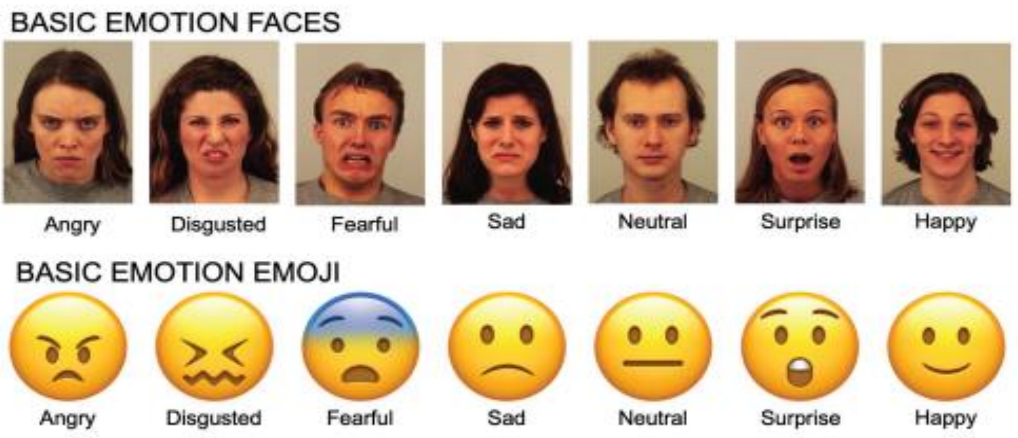
2.1.4 DOWN SAMPLING PCA :

Down sampling is the reduction in spatial resolution while keeping the same two-dimensional (2D) representation.

- It makes the data of a more manageable size.
- Reduces the dimensionality of the data thus enabling in faster processing of the data (image).
- Reducing the storage size of the data.
- There are also some other uses of this technique depending on the usage.

2.2 Objectives of system

In this deep learning project, we have built a convolution neural network to recognize facial emotions. We have trained our model on the FER2013 dataset. Then we are mapping those emotions with the corresponding emojis or avatars. We are using OpenCV's HAAR cascade xml we are getting the bounding box of the faces in the webcam. Then we feed these boxes to the trained model for classification.



The emoji are correctly imposed on their corresponding faces. The input will be raw image of the expression, and output will be shown as above.

2.3 System Functionality

Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions. [1] Recently, generative artificial neural networks have been able to surpass many previous approaches in performance.

Facial expression recognition is a process performed by humans or computers, which consists of:

1. Locating faces in the scene (e.g., in an image; this step is also referred to as face detection)
2. Extracting facial features from the detected face region (e.g., detecting the shape of facial components or describing the texture of the skin in a facial area; this step is referred to as facial feature extraction).
3. Analyzing the motion of facial features and/or the changes in the appearance of facial features and classifying this information into some facial-expression interpretation categories such as facial muscle activations like smile or frown, emotion (affect)categories like happiness or anger, attitude categories like (dis)liking or ambivalence, etc. (this step is also referred to as facial expression interpretation).
4. Converting the detected facial expression into emojis by mapping the predicted results.

Several Projects have already been done in this field and our goal will not only be to develop an Automatic Facial Expression Emoji Generation System but also improving the accuracy of this system compared to the other available system.

2.4 Library and Packages

OpenCV: OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in commercial products. Being a BSD-licensed product, OpenCV makes it easy for businesses to utilize and modify the code. The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms

- OpenCV's application areas include:
  - 2D and 3D feature toolkits
  - Facial recognition system
  - Gesture recognition
  - Human–computer interaction (HCI)
  - Object identification
  - Motion tracking

To support some of the above areas, OpenCV includes a statistical machine learning library that contains

- Decision tree learning
- k-nearest neighbor algorithm
- Artificial neural networks
- Random forest - Support vector machine (SVM)
- Deep neural networks (DNN)

- **NumPy:** NumPy is an acronym for "Numeric Python" or "Numerical Python". It is an open source extension module for Python, which provides fast precompiled functions for mathematical and numerical routines. Furthermore, NumPy enriches the programming language Python with powerful data structures for efficient computation of multi-dimensional arrays and matrices. The implementation is even aiming at huge matrices and arrays. Besides that, the module supplies a large library of high-level mathematical functions to operate on these matrices and arrays.

  It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

  - A powerful N-dimensional array object
  - Sophisticated (broadcasting) functions
  - Tools for integrating C/C++ and Fortran code
  - Useful linear algebra, Fourier Transform and random number capabilities

- **SciPy:** SciPy (Scientific Python) is often mentioned in the same breath with

  NumPy. SciPy extends the capabilities of NumPy with further useful functions for minimization, regression, Fourier-transformation and many others. NumPy is based on two earlier Python modules dealing with arrays. One of these is Numeric. Numeric is like NumPy, a Python module for high-performance, numeric computing, but it is obsolete nowadays. Another predecessor of NumPy is Numarray, which is a complete rewrite of Numeric but is deprecated as well. NumPy is a merger of those two, i.e. it is built on the code of Numeric and the features of Numarray.

- **Haar Cascade Classifier in OpenCV:** Haar feature-based cascade classifiers is an effectual machine learning based approach, in which a cascade function is trained using a sample that contains a lot of positive and negative images. The outcome of AdaBoost classifiers is that the strong classifiers are divided into stages to form cascade classifiers. The term “cascade” means that the classifier thus produced consists of a set of simpler

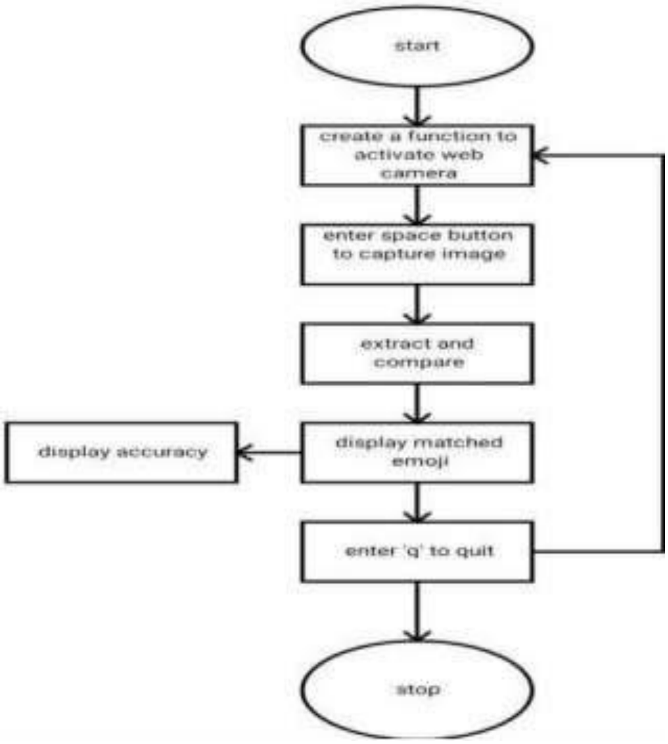
  classifiers which are applied to the region of interest until the selected object is discarded or passed. The cascade classifier splits the classification work into two stages: training and detection. The training stage does the work of gathering the samples which can be classified as positive and negative. The cascade classifier employs some supporting functions to generate a training dataset and to evaluate the prominence of classifiers. In order to train the cascade classifier, we need a set of positive and negative samples.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

CHAPTER 3

SOFTWARE DEVELOPMENT METHODOLOGY

3.1 Description of Diagram



In the Fig 3.1, Facial expressions can be described as the arrangement of facial muscles to convey a certain emotional state to the observer in simple words. Emotions can be divided into six broad categories—Anger, Disgust, Fear, Happy, Sad, Surprise, and Neutral. In this, train a model to differentiate between these, train a convolutional neural network using the FER2013 dataset and will use various hyper-parameters to fine-tune the model.

The design starts with the initializing CNN model by taking an input image (static or dynamic) by adding a convolution layer, pooling layer, flatten layers, and dense layers. Convolution layers will be added for better accuracy for large datasets. The dataset is collected from a CSV file (in pixel format) and it's converted into images and then classify emotions with respective expressions.

3.2 System Design

System design shows the overall design of the system. In this section we discuss in detail the design aspects of the system:

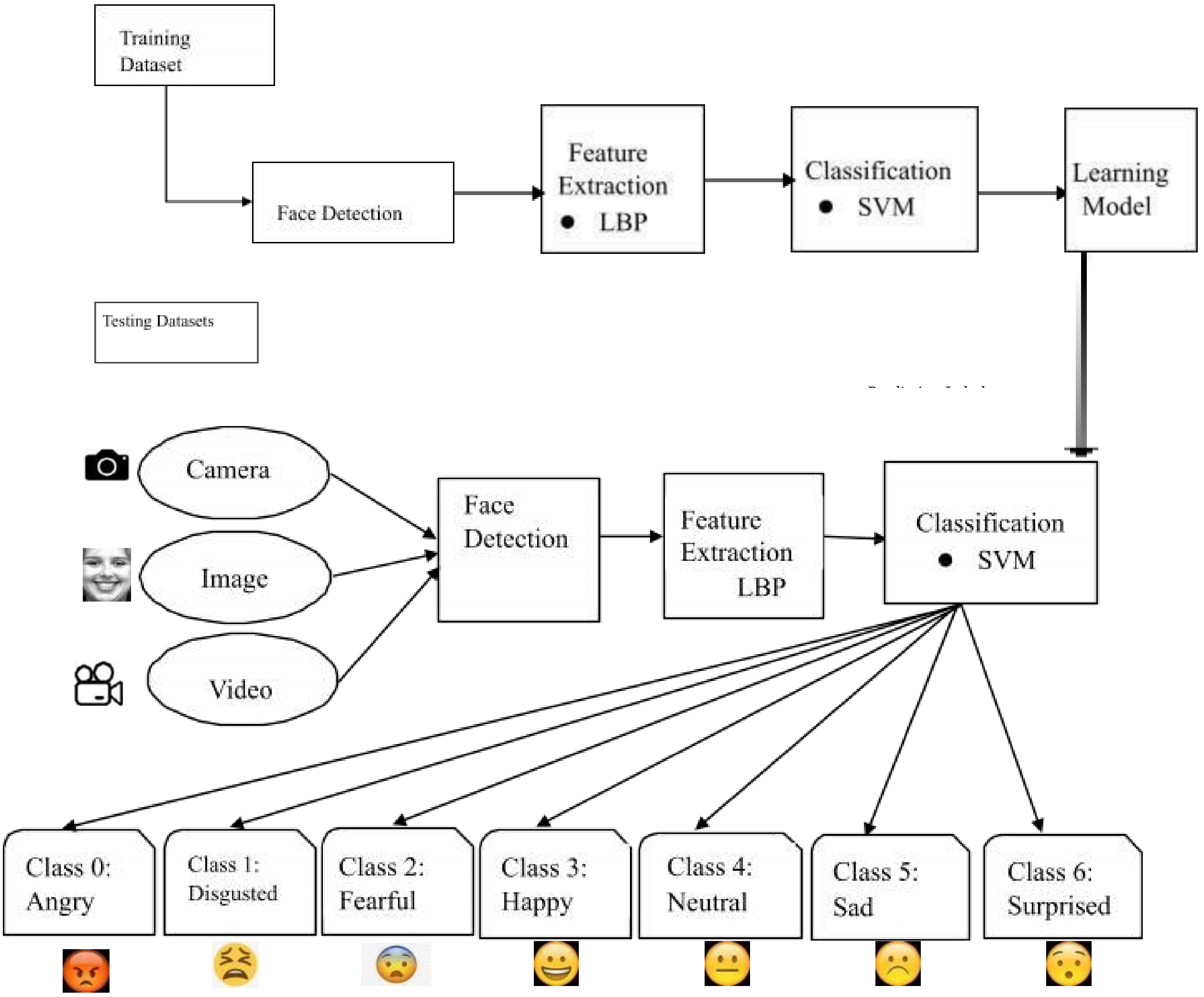


FIG 4.2 SYSTEM ARCHITECTURE

Here emotions are classified as happy, sad, angry, surprise, neutral, disgust, and fear with 34,488 images for the training dataset and 1,250 for testing. Each emotion is expressed with different facial features like eyebrows, opening the mouth, raised cheeks, wrinkles around the nose, wide-open eyelids and many others. Trained the large dataset for better accuracy and

result that is the object class for an input image. Based on those features it performs convolution layers and max pooling. These are the seven different universal emotions with the following expressions above Fig 4.2.

3.3 System Flowchart

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

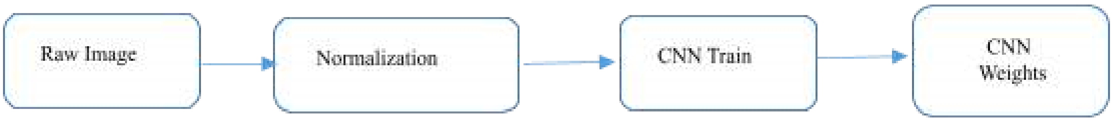


FIG 4.3(A) FLOWCHART OF TRAINING PHASE

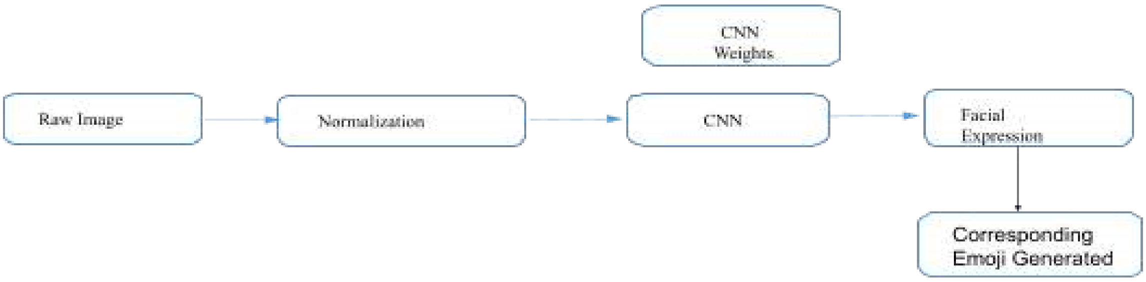


FIG 4.3(B) FLOWCHART OF TESTING PHASE

During training, the system receives 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 training exercises 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 the test, the system received a grayscale image of a face from the 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.

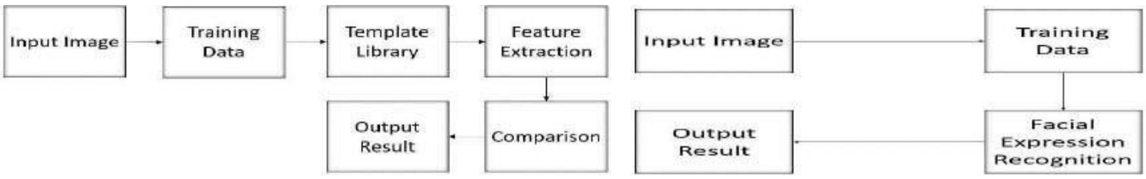


Fig 4.3(c) Schema of facial expression recognition

The original structure contains six steps which are input image, training data, template library, feature extraction, comparison and output result, as shown in Figure 4.3(c). However, a simplified structure that is used in this paper only has four steps after we combine the step of template library, feature extraction and comparison to facial expression recognition, as shown in Figure above. It will greatly increase the efficiency and reduce the running time.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

CHAPTER 4

FACIAL EMOTION RECOGNITION USING CNN

4.1 Dataset

The dataset from a Kaggle Facial Expression Recognition Challenge (FER2013) is used for the training and testing. It comprises pre-cropped, 48-by-48-pixel grayscale images of faces each labeled with one of the 7 emotion classes: anger, disgust, fear, happiness, sadness, surprise, and neutral. Dataset has a training set of 35887 facial images with facial expression labels. The dataset has class imbalance issues, since some classes have a large number of examples while some have 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.

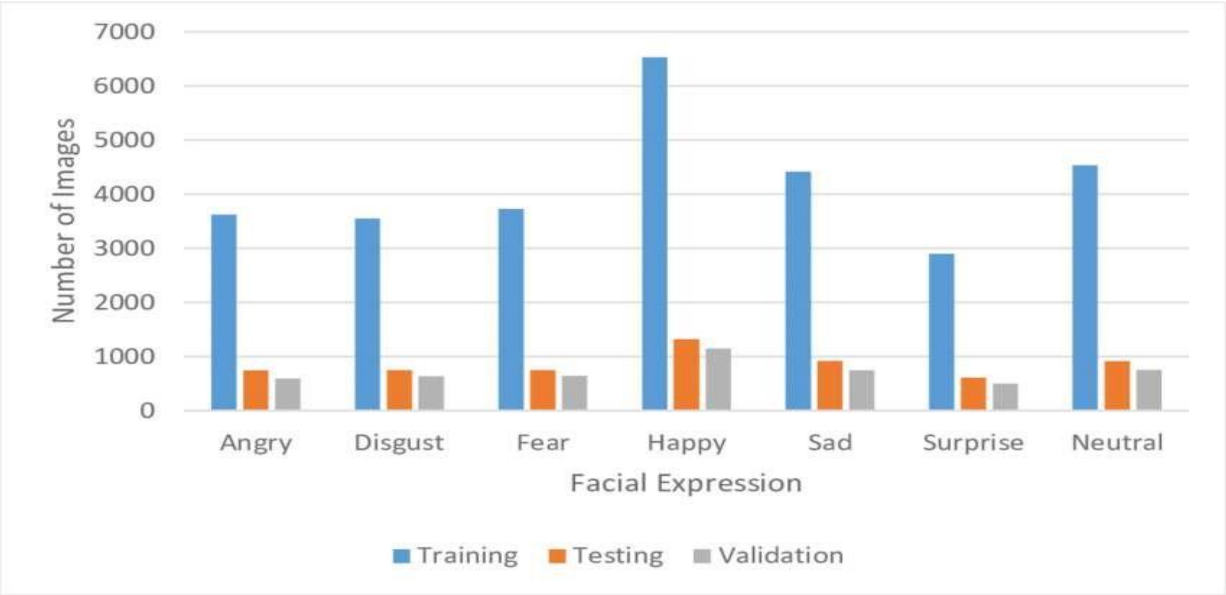


Fig 4.4(a): Training, Testing and Validation Data distribution

Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

The images in the FER2013 dataset have size 48x48 and are black and white images. The FER2013 dataset contains images that vary in viewpoint, lighting, and scale. Fig.4.4(b) shows some sample images from the FER2013 dataset, and Table 4.4(c) illustrates the description of the dataset.

Table 4.1 Description of the FER2013 dataset

Label	Number of images	Emotion
0	4593	Angry
1	547	Disgust
2	5121	Fear
3	8989	Happy
4	6077	Sad
5	4002	Surprise
6	6198	Neutral

4.2 Process of Facial Expression Recognition

The process of FER has three stages. The preprocessing stage consists of preparing the dataset into a form which will work on a generalized algorithm and generate efficient results. In the face detection stage, the face is detected from the images that are captured real time. The emotion classification step consists of implementing the CNN algorithm to classify input images into one of seven classes.

Fig.4.5 Process flow

4.3 Pre-processing

The input image to the FER may contain noise and have variation in illumination, size, and color. To get accurate and faster results on the algorithm, some preprocessing operations were done on the image. The preprocessing strategies used are conversion of image to grayscale, normalization, and resizing of image.

1. Normalization - Normalization of an image is done to remove illumination variations and obtain improved face image

2. Gray scaling - Gray scaling is the process of converting a colored image input into an image whose pixel value depends on the intensity of light on the image.

   Gray scaling is done as colored images are difficult to process by an algorithm.

3. Resizing - The image is resized to remove the unnecessary parts of the image.

   This reduces the memory required and increases computation speed.

4.4 Emotion Classification

In this step, the system classifies the image into one of the seven universal expressions - Happiness, Sadness, Anger, Surprise, Disgust, Fear, and Neutral as labelled in the FER2013 dataset. The training was done using CNN, which is a category of neural networks proved to be productive in image processing. The dataset was first split into training and test datasets, and then it was trained on the training set. Feature extraction process was not done on the data before feeding it into CNN.

The approach followed was to experiment with different architectures on the CNN, to achieve better accuracy with the validation set, with minimum overfitting. The emotion classification step consists of the following phases:

1) Splitting of Data:

   The dataset was split into 3 categories according to the "Usage" label in the

   FER2013 dataset: Training, Public Test, and Private Test. The Training and Public Test set were used for generation of a model, and the Private Test set was used for evaluating the model.

2) Training and Generation of model:

   The neural network architecture consists of the following layers:

   i. Convolution Layer:

      In the convolution layer, a randomly instantiated learnable filter is slid, or convolved over the input. The operation performs the dot product between the filter and each local region of the input. The output is a 3D volume of multiple filters, also called the feature map.

   ii. Max Pooling:

      The pooling layer is used to reduce the spatial size of the input layer to lower the size of input and the computation cost.

   iii. Fully connected layer:

      In the fully connected layer, each neuron from the previous layer is connected to the output neurons. The size of the final output layer is equal to the number of classes in which the input image is to be classified.

   iv. Activation function:

Activation functions are used to reduce the overfitting. In the CNN architecture, the RELU activation function has been used. The advantage of the RELU activation function is that its gradient is always equal to 1, which means that most of the error is passed back during back-propagation.

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

Equation 1: Equation of RELU Activation Function

v. SoftMax:

The SoftMax function takes a vector of N real numbers and normalizes that vector into a range of values between (0, 1).

vi. Batch Normalization:

The batch normalizer speeds up the training process and applies a transformation that maintains the mean activation close to 0 and the activation standard deviation close to 1.

3) Evaluation of model:

The model generated during the training phase was then evaluated on the validation set, which consisted of 3589 images.

4) Using model to classify real time images:

The concept of transfer learning can be used to detect emotion in images captured in real time. The model generated during the training process consists of pretrained weights and values, which can be used for implementation of a new facial expression detection problem. As the model generated already contains weights, FER becomes faster for real time images.

CHAPTER 5

APPLICATIONS AND FUTURE SCOPE

Potential uses of FER cover a wide range of applications, examples of which are listed here be- low in groups by their application field.

Provision of personalized services:

- analyze emotions to display personalized messages in smart environments
- provide personalized recommendations e.g. on music selection or cultural material

- analyze facial expressions to predict individual reaction to movies

Customer behavior analysis and advertising

- analyze customers' emotions while shopping focused on either goods or their arrangement within the shop
- advertising signage at a railway station using a system of recognition and facial tracking for marketing purposes

Healthcare

- detect autism or neurodegenerative diseases
- predict psychotic disorders or depression to
- identify users in need of assistance
- suicide prevention
- detect depression in elderly people
- observe patients' conditions during treatment

Employment

- Help decision-making of recruiters
- identify uninterested candidates in a job interview
- monitor mood and attention of employees

Education

- monitor students' attention
- detect emotional reaction of users to an educative program and adapt the learning path
- design affective tutoring system
- detect engagement in online learning

Public safety

- lie detectors and smart border control
- predictive screening of public spaces to identify emotions triggering potential terrorism threat
- Analyzing footage from crime scenes to indicate potential motives in a crime

Crime detection

- detect and reduce fraudulent insurance claims
- deploy fraud prevention strategies
- spot shoplifters

Other

- driver fatigue detection
- detection of political attitudes

.

Future Work

Face expression recognition systems have improved a lot over the past decade. The focus has definitely shifted from posed expression recognition to spontaneous expression recognition. Promising results can be obtained under face registration errors, fast processing time and significant performance improvements can be obtained in our system. System is fully automatic and has the capability to work with images feed. It is able to recognize spontaneous expressions. The system can be used in Digital Cameras wherein the image can be captured only when the person smiles. In security systems which can identify a person, in any form of expression he presents himself. Doctors can use the system to understand the intensity of pain or illness of a deaf patient. Our system can be used to detect and track a user's state of mind, and in mini-marts, shopping centers to view the feedback of the customers to enhance the business etc.

In all the biometric modalities — fingerprint, expression, gait, behavioural, DNA, and others — face is gaining adoption faster. Because, it is not only convenient for almost all to use, but a face provides a sensor (here, camera) and device with too much "signal" or data as any other tool. For example, fingerprints cannot suit the many differences in the forms, sizes, distinctive marks, and other distinguishing features of a qualified AI program, for all their simplicity and apparent sophistication.

More commonly, the face is often simpler to use in several situations where sunlight, temperature, social restrictions, and physical access to a mobile device will preclude a person from talking to their phone or utilizing a fingerprint reader. Once the technology is fully developed, implemented, and introduced, facial recognition and face authentication can be taken to new heights.

In day-to-day business uses other than law enforcement and surveillance, where there is a need for additional encryption than is needed to open a phone or buy a cup of coffee, the most efficient approach is authentication through liveness detection. When a program may be assured that the person requiring access is not only a legal consumer but is genuinely alive at the moment of the

request for access, high-value accounts such as banking/finance, medical history, automated vehicle access, and blockchain-based applications such as cryptocurrency wallets and secure contract signing can rapidly build trust, adding a variety of advanced resources to an eagerly anticipated customer.

We're now starting to see face-to-face security systems deployed in sectors such as identity protection and banking, but also mainly in combination with other existing solutions such as fingerprint or SMS verification. Yet over the year or two, we'll see big multinational companies adopting increasingly sophisticated face biometrics use state-of-the-art, AI-driven technology to enhance their protection capabilities and better defend customers from identity fraud and data loss. Face authentication with recognition of liveliness would not only have an improved, frictionless user interface, but anybody with a traditional smart app will enjoy unprecedented levels of real-world security.

CHAPTER 6

CODING

6.1 Software Requirements

- Operating system: Windows 10 and above.
- Coding Language: Python
- IDE: Python IDLE
- Editor: Visual Studio Code
- Database: Microsoft SQL (if required)

6.2 Hardware Requirements

- System: i3 Processor and above.
- Hard Disk: 40 GB.
- Monitor: 15 VGA Color.

6.3. Project Resources

- <u>Hardware Resources Required:</u>

1. Processor: Intel i3 and above
2. Hard Disk: Minimum 100GB

3. RAM: 4GB

• Software Resources Required:

1. Platform: Windows 10 and above.
2. Backend: python 3.10
3. Front End: python, PyQt 6

CHAPTER 7

RESULTS AND DISCUSSION

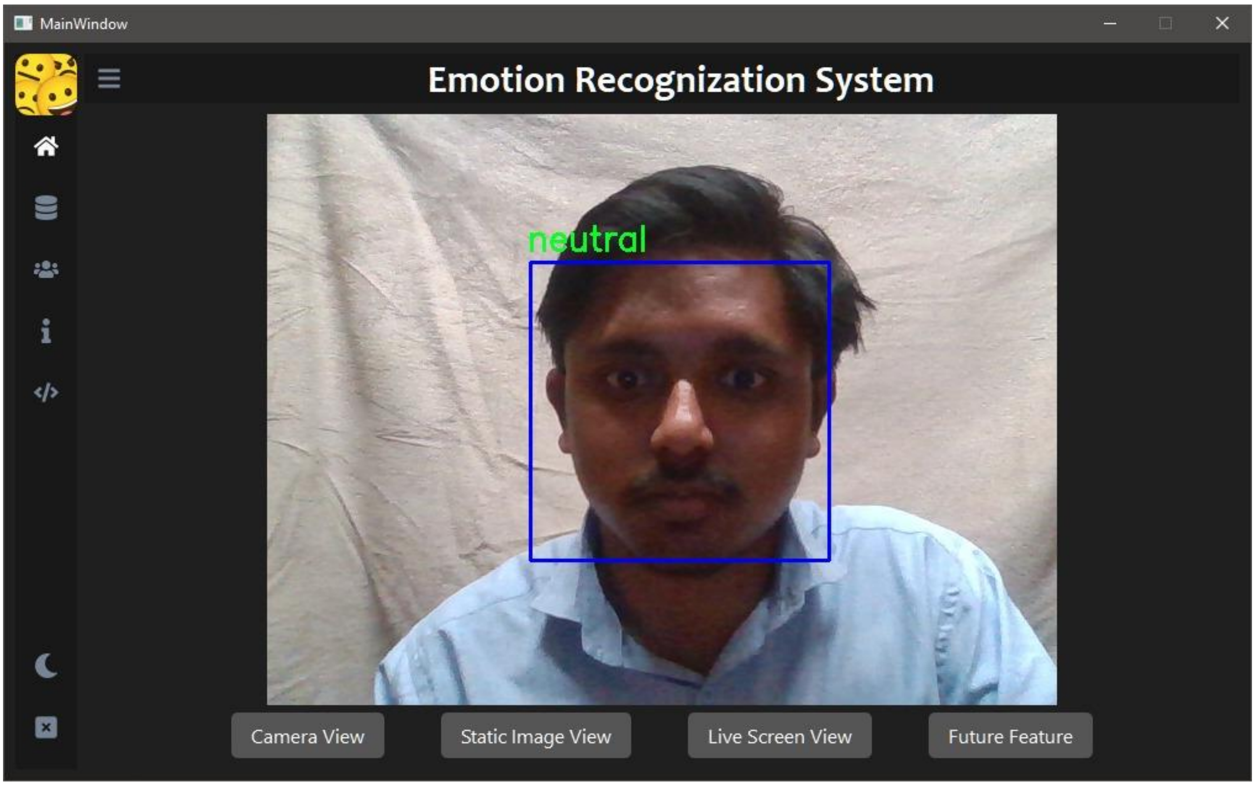


Fig 7.1 Emoji Generated of the emotions -Neutral

In the above figures, output is the window where the user expressions are captured by the webcam and the respective emotions are detected. On detecting the emotion, the respective emotion is shown on the left side of the screen. This emoticon changes with the change in the expression of the person in front of the webcam. Hence changes with the change in the expression of the person in front of the webcam. Hence this real time application is very beneficial in various fields like psychology, computer science, linguistics, neuroscience and related disciplines.

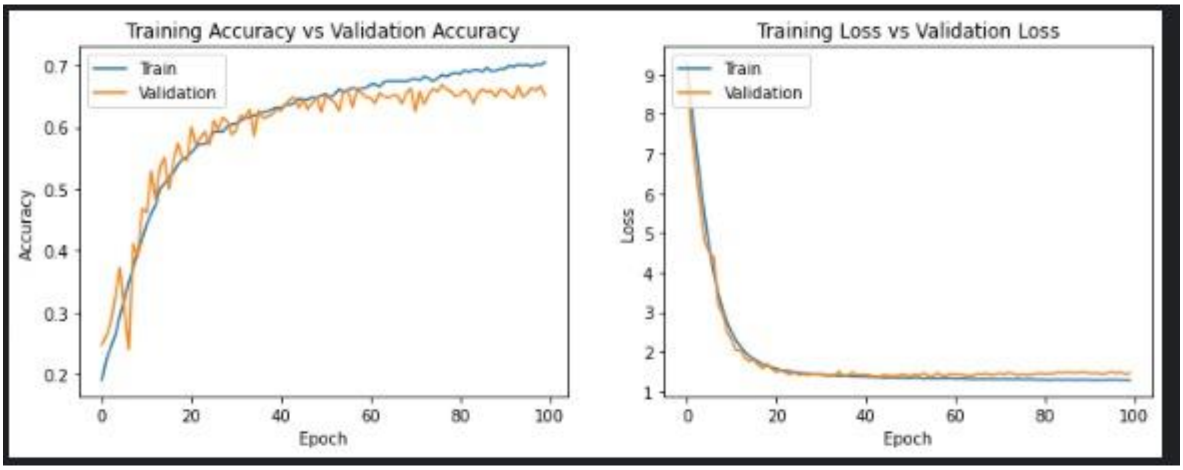
Results were obtained by experimenting with the CNN algorithm. It was observed that the loss over training and test set decreased with each epoch. The batch size was 256, which was kept constant over all experiments.

The following changes were made in the neural network architecture to achieve good results:

- Number of epochs: It was observed that the accuracy of the model increased with increasing number of epochs. However, a high number of epochs resulted in overfitting. It was concluded that eight epochs resulted in minimum overfitting and high accuracy.
- Number of layers: The neural network architecture consists of three hidden layers and a single fully connected layer. A total of six convolution layers were built, using 'relu' as the activation function.
- Filters: The neural network accuracy on the dataset varied on the number of filters applied to the image. The number of filters for the first two layers of the network was 64, and it was kept 128 for the third layers of the network.
- Accuracy: The final, state-of-the-art-model gave a **training accuracy of 79.89%** and a **test accuracy 65.22%** as shown in the table. The architecture used could correctly classify 22936 out of 28709 images from the train set and 2158 out of 3589 images from the test set.

Loss and accuracy over time

It can be observed that the loss decreases, and the accuracy increases with each epoch. The training versus testing curve for accuracy remains ideal over the first five epochs, after which it begins to deviate from the ideal values. The training and test accuracy along with the training and validation loss obtained for the FER2013 dataset using CNN are given in fig.5.2.



Confusion Matrix

The confusion matrix generated over the test data is shown in figure 7. The dark blocks along the diagonal show that the test data has been classified well. It can be observed that the number of correct classifications is low for disgust, followed by fear. The numbers on either side of the diagonal represent the number of wrongly classified images. As

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

these numbers are lower compared to the numbers on the diagonal, it can be concluded that the algorithm has worked correctly and achieved state of the art results.

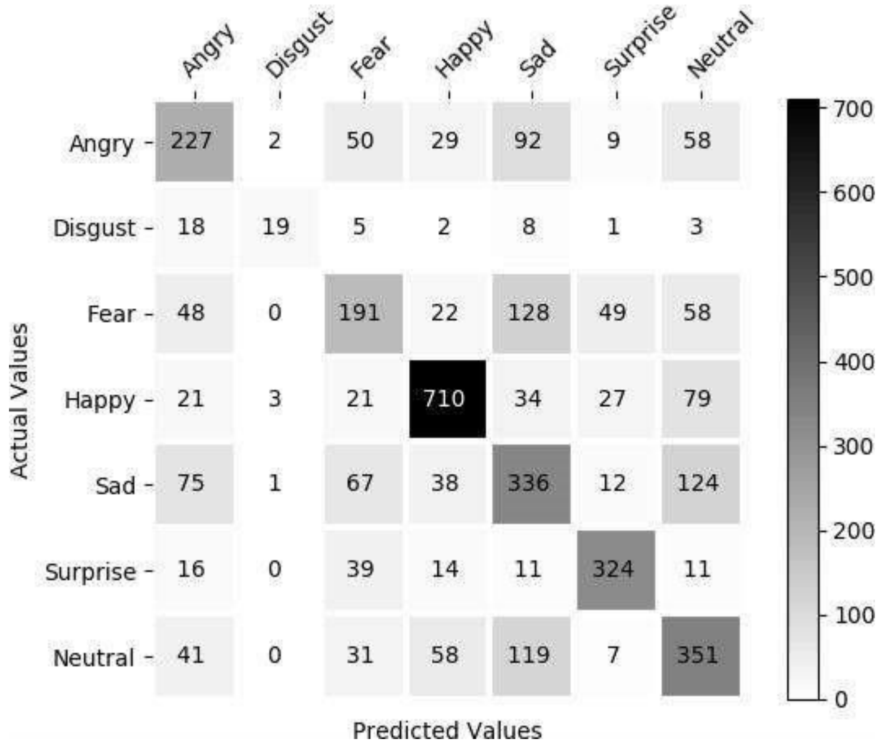


Fig 5.2. Confusion matri

APPENDIX

CONCLUSION

In this project, an approach for FER using CNN has been discussed. A CNN model on the FER2013 dataset was created and experiments with the architecture were conducted to achieve a test accuracy of 76.89 and a validation accuracy of 65.22. This state-of-the-art model has been used for classifying emotions of users in real time using a webcam. The webcam captures a sequence of images and uses the model to classify emotions and generate the corresponding emoji. Proposed is a human emotion detector using emoticons using machine learning, python to predict emotions of the people and represent them using emoticons. These include image acquisition, preprocessing of an image, face detection, feature extraction, classification and then when the emotions are classified the system assigns the user particular music according to his emotion. The main aim of this project is to develop an automatic facial emotion recognition system in which an emoticon is used for giving the output for individuals thus assigning them various therapies or solutions to relieve them from stress. The emotions used for the experiments include happiness, Sadness, Surprise, Fear, Disgust, and Anger that are universally accepted.

REFERENCES

- H.-D. Nguyen, S. Yeom, G.-S. Lee, H.-J. Yang, I. Na, and S. H. Kim, "Facial Emotion Recognition Using an Ensemble of MultiLevel Convolutional Neural Networks," International Journal of Pattern Recognition and Artificial Intelligence, 2018.

- T. Cao and M. Li, "Facial Expression Recognition Algorithm Based on the Combination of CNN and K-Means," presented at the Proceedings of the 2019 11th International Conference on Machine Learning and Computing, Zhuhai, China, 2019.

- N. Christou and N. Kanojiya, "Human Facial Expression Recognition with Convolutional Neural Networks," Singapore, 2019, pp. 539-545: Springer Singapore

- A. Sajjanhar, Z. Wu, and Q. Wen, "Deep learning models for facial expression recognition," in 2018 Digital Image Computing: Techniques and Applications (DICTA), 2018, pp. 1-6: I EEE.

- J. Chen, Y. Lv, R. Xu, and C. Xu, "Automatic social signal analysis: Facial expression recognition using difference convolution neural network," Journal of Parallel and Distributed Computing, vol. 131, pp. 97-102, 2019.

- Al-Sumaidaee, Saadoon AM, et al, Multi-gradient features and elongated quinary pattern encoding for image-based facial expression recognition, Pattern Recognition, 2017, pp. 249—263.

- Barsoum, Emad, et al, Training deep networks for facial expression recognition with crowd-sourced label distribution, ACM International Conference on Multimodal Interaction ACM, 2016, pp. 279—283.

- Martinez, Brais, et al, Automatic analysis of facial actions: A survey, IEEE Transactions on Affective Computing, 2017.

- J. M. B. Fugate, A. J. O'Hare, and W. S. Emmanuel, "Emotion words: Facing change," Journal of Experimental Social Psychology, vol. 79, pp. 264-274, 2018.

- K. Clawson, L. Delicato, and C. Bowerman, "Human Centric Facial Expression Recognition," 2018