KLE Society's

KLE Technological University

A Mini Project ReportOn

EMOTION CLASSIFICATION IN VIDEOS

submitted in partial fulfillment of the requirement for the degree of

Bachelor of EngineeringIn

Computer Science and Engineering

Submitted By

SEEMA AGER	**01FE19BCS097**
ARPITA CHULAKI	**01FE19BCS098**
KHOT SHEJAL	**01FE19BCS090**
SNEHA KASHETTI	**01FE19BCS112**

Under the guidance ofDr.
SUJATHA C

SCHOOL OF COMPUTER SCIENCE & ENGINEERING

HUBLI-580 031 (India).

Academic year 2021-22

KLE Society's

KLE Technological University

2021- 2022

SCHOOL OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that Mini Project entitled Emotion Classification in videos is a bonafied work carried out by the student team Ms. KHOT SHEJAL – 01FE19BCS090, Ms. SEEMA AGER – 01FE19BCS097, Ms. ARPITA CHULAKI – 01FE19BCS098, Ms.SNEHA KASHETTI – 01FE19BCS112, in partial fulfillment of completion of Fifth semester B. E. in Computer Science and Engineering during the year 2021 – 2021. The project report has been approved as it satisfies the academic requirement with respect to the project work prescribed for the above said programme.

Guide
Dr. Sujatha C

Head, SoCSE
Dr. Meena S. M

External Viva:

Name of the Examiners	**Signature with date**
1.	
2.	

ABSTRACT

Human emotion classification plays a vital role in the interpersonal relationship. The automatic classification of emotions has been an active research topic from early eras. Therefore, there are several advances made in this field. Deep learning based emotion detection gives performance better than traditional methods with image processing. Our project presents the design of an artificial intelligence system capable of emotion classification through facial expressions. It discusses about the emotion classification, which includes basically three main steps: face detection, features extraction, and emotion classification. We proposed a convolutional neural networks based deep learning architecture for emotion classification from input images. We have trained our model using dataset from the FER-2013 dataset resources. The accuracy achieved with our proposed model is 85.70%.

ACKNOWLEDGEMENTS

We would like to thank our faculty and management for their professional guidance towards the completion of the project work. We take this opportunity to thank Dr. Ashok Shettar, Vice-Chancellor and Dr. P.G Tewari, Dean Academics, KLE Technological University, Hubballi, for their vision and support. We also thank Dr. Meena S. M, Professor and Head, SoCSE for having provided us direction and facilitated for enhancement of skills and academic growth. We thank our guide Dr Sujatha C for the constant guidance. We extend our acknowledgement to the reviewers for critical suggestions and inputs. We also thank faculty in-charges for their support during the course of completion.

KHOT SHEJAL 01FE19BCS090
ARPITA CHULAKI 01FE19BCS098
SEEMA AGER 01FE19BCS097
SNEHA KASHETTI 01FE19BCS112

Chapter No.	TABLE OF CONTENTS			
1.	**INTRODUCTION**			
	1.1	Preamble		
	1.2	Motivation		
	1.3	Objectives of the project		
	1.4	Literature Survey		
	1.5	Problem Definition		
2.	**PROPOSED SYSTEM**			
	2.1	Description of Proposed System.		
	2.2	Description of Target Users		
	2.3	Advantages/Applications of Proposed System		
	2.4	Scope		
3.	**SOFTWARE REQUIREMENT SPECIFICATION**			
	3.1	Overview of SRS		
	3.2	Requirement Specifications		
		3.2.1	Functional Requirements	
		3.2.2	Use case diagrams	
		3.2.3	Use Case descriptions using scenarios	
		3.2.4	Nonfunctional Requirements	
			3.2.4.1	Performance requirements
			3.2.4.2	Safety requirements
			3.2.4.3	Security Requirements
			3.2.4.4	Usability
	3.3	Software and Hardware requirement specifications		
4	**SYSTEM DESIGN**			
	4.1	Architecture of the system		
	4.1	Class Diagram		
	4.2	Sequence diagram		
	4.3	ER diagram and schema		
	4.4	State transition diagram		
	4.5	Data structure used		

1.INTRODUCTION

1.1 Preamble

- Emotion detection is the process of identifying human emotions. It plays a vital role in interpersonal relationship.

- Understanding emotions and knowing how to react to people's expression greatly enriches the interaction.

- The most frequently used emotion recognition methods include facial expression analysis, audio signal analysis in terms of modulation, textual input analysis and many others.

- In this project we are dealing with emotion classification, based on facial expressions, i.e the system classifies facial expressions of the person into the basic emotions namely disgust, fear, happy, surprise, sad and anger.

- The purpose of this system is to classify the emotion of the person based on their facial expression and hence improvise interaction between human beings and machines.

1.2 Motivation

In today's networked world the need to maintain security of information is becoming both increasingly important and difficult. In several countries the rate of crimes are increasing day by day. No automatic systems are there that can track person's activity. If we will be able to track Facial expressions of persons automatically then we can find the criminal easily since facial expressions changes doing different activities. So we choose to make a Facial Expression Classification System. As a result we are highly motivated to develop a system that recognizes facial expression and track one person's activity.

1.3 Objectives of the Project

- Split the input video into frames/images.
- Detect the face of a person present in the frame.
- Extract the facial expression features from the detected face.
- Classify the emotion of a person using the extracted features.

1.4 Literature Survey

[1]Emotion Detection: A Machine Learning Project by Aarohi Gupta.

- Information about facial landmark model that is 'Dlib's pre-trained Facial Landmark Detection Model' which detects 68 2-Dimensional points on the humanface.
- The comparison between the KNN and CNN algorithms is done, concluding that CNN is more accurate in detecting the emotion than KNN.

[2]Facial Emotion Detection Using Neural Network, by Md. Forhad Ali, Mehenag Khatun, Nakib Aman Turzo

- Emotion Detection system involves face and facial parts detection using the Viola-Jones algorithm, extraction of facialFeature, and emotion Detection usingCNN model.
- Keras is used as an open-source neural network for modeling, preprocessing,evaluating and optimization.

[3]Study of Video based Facial Expression and Emotions Recognition Methods, Husam K Salih

- Spatiotemporal feature extraction for facial detection on video sequences.
- To detect the face in the images Viola–Jones face detector is used.
- SVM and Neural Network with different kernels are used for the emotionclassification.

[4]Facial Emotion Detection Using Deep Learning Publisher: IEEE

- Convolutional neural networks (CNN) based deep learning architecture isused for emotion detection from images.
- The performance given by Deep learning based emotion detection is betterthan traditional methods with image processing.

[5]Facial Emotion Recognition Using Machine Learning, Nitisha Raut San Jose State University

- Viola- Jones technique is used for face detection which uses Ada-boost algorithm. Weak classifier is combined to form a strong classifier by iterations, where a weak classifier reduces the weighted error rate per iteration.
- Local Binary Pattern (LBP) simple and powerful technique for feature extraction.
- Support Vector Machines, Machine Learning Algorithms are used to classifythe emotions.

1.5 Problem Definition

- To develop a system to detect the faces of people from the video and classify theiremotion by extracting their facial expression features.

2.PROPOSED SYSTEM

2.1 Description of Proposed System

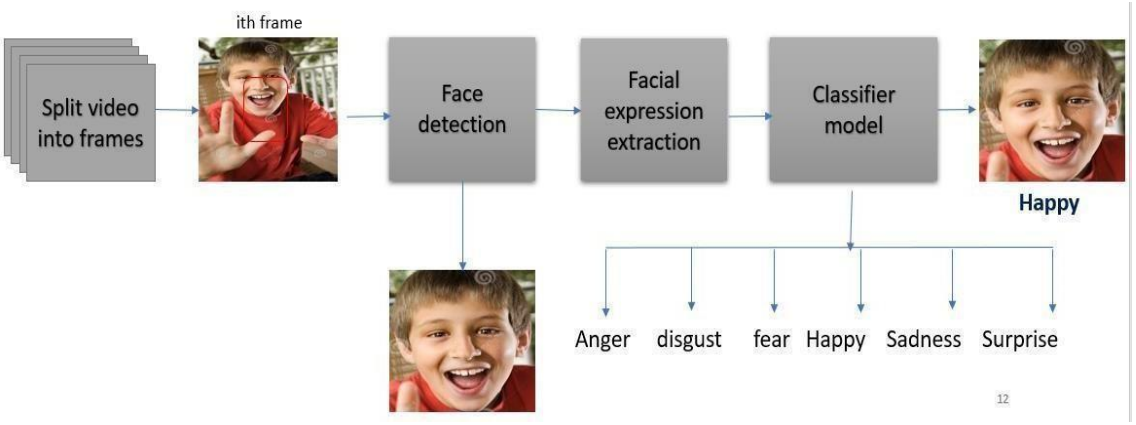

Fig 1. Block Diagram

Split video into frames:
First the user gives the input video, then the system splits the input video into frames.

Face detection:
Viola Jones algorithm is used to detect the faces of people from the frames.
The output of this algorithm is that it plots a bounding box on the detected face with length x width. This image is then cropped and given as an input to the classifier model.

CNN model:
First facial expression features are extracted and probability of each expressed emotion is calculated, then emotion with maximum probability is declared as expressed emotion.

2.2 Description of Target Users

Our project focus on classification of people's emotions. It can be used by malls/shopping marts/super market owners to get customer feedback. Can be used by educational institute to get students feedback. Can be used in driving vehicle to detect driver emotions and pop an alert, hence driving towards safety measures.

2.3 Advantages/Applications of proposed system

- Driver emotion detection in intelligent vehicles.
- The system can be installed at places like railway station, airport, or bus station for detecting human faces and then identifying facial expressions of each person. If there are any faces that appeared suspicious like fearful or angry, the system mightset an alarm.
- Emotion detection in crime investigation field.

- The system can be used at places like super markets, shopping center, malls to get the feedback of the customers to enhance the business.

- The system can also be used for educational purpose like getting the feedback on the student reaction during the class.

2.4 Scope (Boundary of proposed system) and Constraints

- Emotions are subjective, people interpret it differently. The concept of emotions is difficult to define. For example, although emotions may always be the same, different people emerge to express the same feelings in very different ways.
- Since emotions are intertwined in some aspects, for example sometimes there is a fine line between fear, surprise, boredom, sadness, etc.
- People's expressions may be very different. It becomes difficult to gauge the feelings of a less expressive person.
- The intensity of agitation cannot be quantified

3.SOFTWARE REQUIREMENT SPECIFICATION

3.1 Overview of SRS

The purpose of this SRS document is to provide a detailed view of our software product, its parameters and its terms. This document describes the target audience of the project as well as the user interface, hardware and software requirements.

3.2 Requirement Specifications

3.2.1 Functional Requirements

- The system shall be able to convert the video into frames
- The system shall be able to detect face in the frame
- The system shall be able to assign confidence score to the emotionidentified from the face
- The system shall display the identified emotion class based on maximumscore.

3.2.2 Use case diagrams

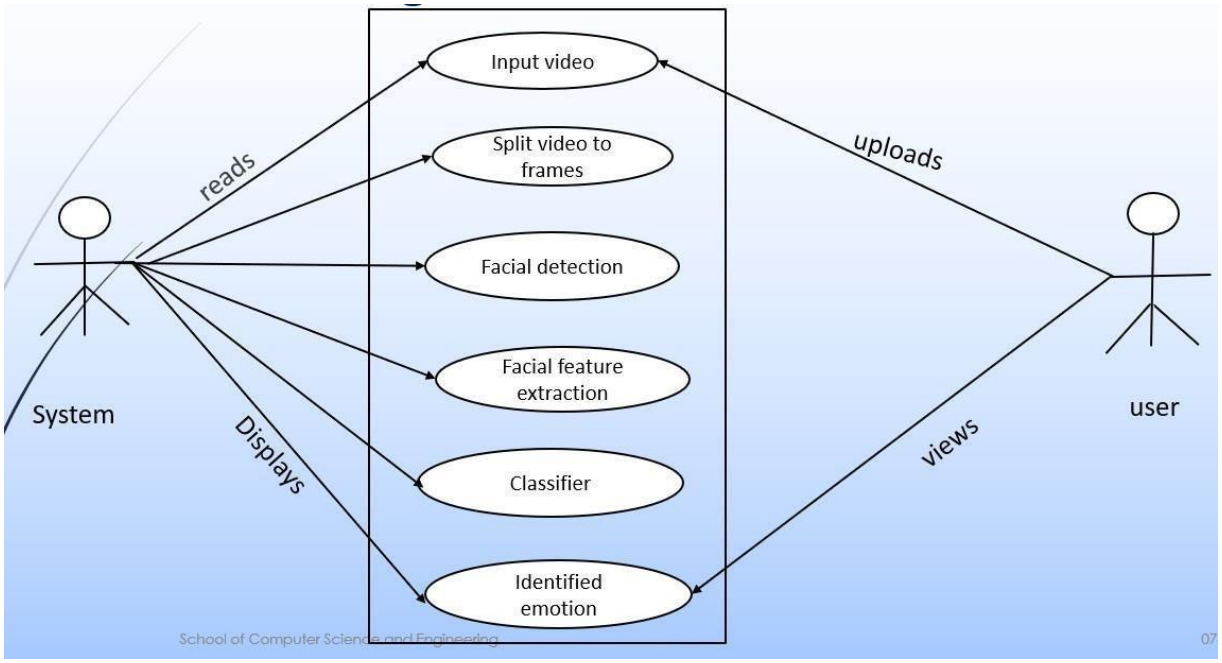

Fig 2. Use case diagram

3.2.3 Use Case descriptions using scenarios.

Use case: classification of emotions in videos.

Actors: system,user
Pre-condition:
1) system must be in proper working condition.
2) video must be uploaded in correct format and available.

Post condition:
Based on confidence score calculated for each emotion, the maximal expressed emotion must be displayed as an expressed emotion.

Main success scenario:
1) begins when user uploads the video
2) The system reads the input video.
3) The system spilt the video into number of individual frames.
4) From an individual ith frame the face is detected.
5) System must extract the facial expression features
6) Based on classifier model the system must calculate the confidence score of all the emotions.
7) System must identify the maximum expressed emotion.
8) System must display identified expressed emotion.

Exception case scenario:
1a Display an error message of video not available.
4a Display a message, face not detected if face isn't available

3.2.4 Non Functional Requirements

- The system shall be able to detect face in the frame within 30msec per frame
- The system should calculate the confidence score of emotions with an accuracy of about 80%
- The system should display the identified emotion class based on maximum score within 30msec per frame

3.2.4.1 Performance requirements

- Reliability based on this system defines the evaluation result of the system, correct
- identification of the facial expressions and maximum evaluation rate of the facial
- expression recognition of any input images.

3.2.4.2 Usability

The system is simple, user friendly, graphics user interface implemented so any can use this system without any difficulties.

3.3 Software and Hardware requirement specifications

Software Requirement

- Python programming language
- Jupyter notebook
- OpenCV framework
- Linux platform Ubuntu OS 2.7.2 or Windows OS

Hardware Requirement

- Fluently working Laptops
- RAM minimum 4G

4.SYSTEM DESIGN

4.1 Architecture of the system

- CNN architectural Model is a class of artificial neural network, commonly applied to analyse visual imagery. We have used CNN model to classify the emotion from facial expression features.
- Computer vision gives the computer the ability to see as people see.
- OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library that is easy to import in Python.
- Haarcascade algorithm is used in the model which is a machine-based learning method in which the cascade function is trained using a large number of positive and negative images. This is then used to find objects in any given input images.

4.2 Class Diagram

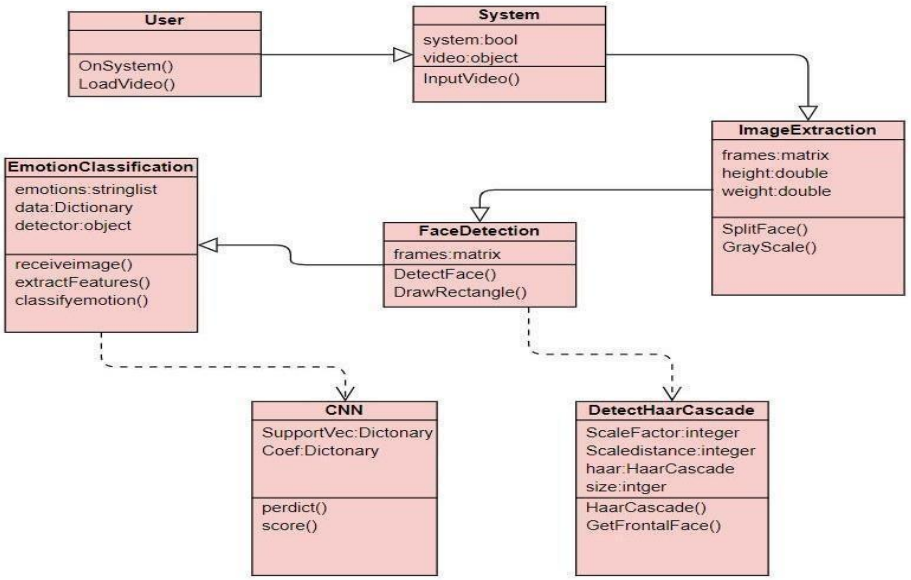

Fig 3. Class Diagram

4.3 Sequence Diagram

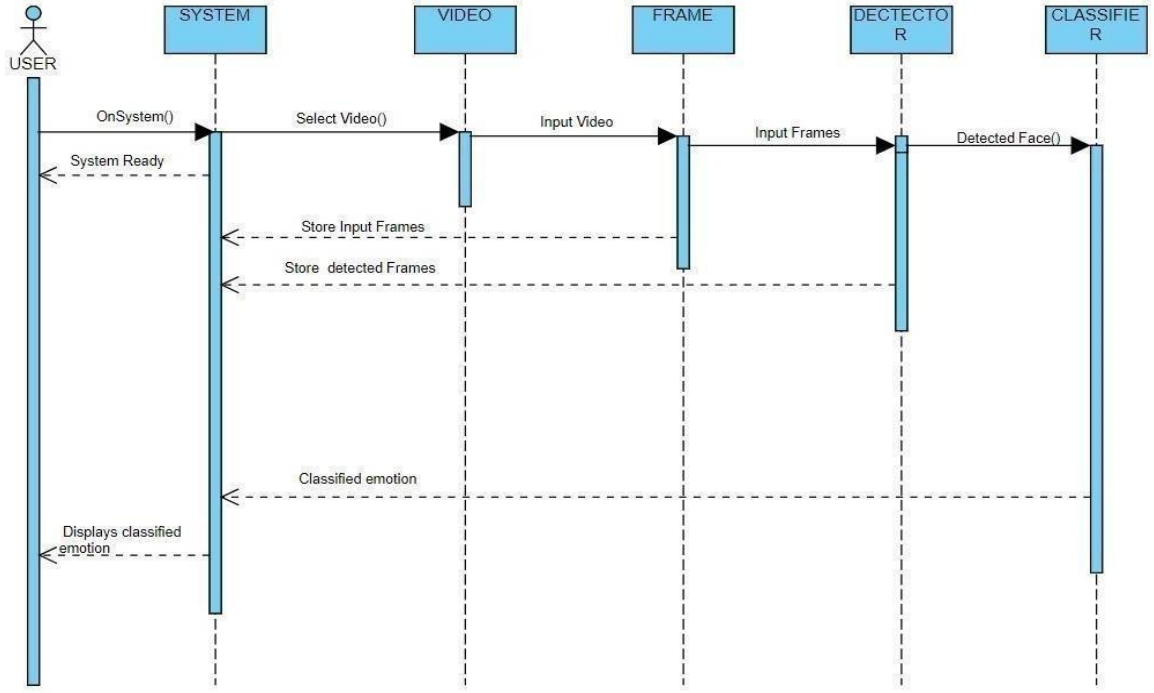

Fig 4. Sequence Diagram

4.4 Data set Description

Dataset Source: The FER-2013 dataset
Dataset Analysis:

- The dataset has of 28,709 images in the training set and 7178 images in the test set.
- The dataset consists of seven different categories of images, each category having 4000 images in training set.

Data preprocessing:

- Converting the images to greyscale.
- Resizing the detected face to 48x48 pixels**.**

5. IMPLEMENTATION

5.1 Proposed Methodology

- CNN architectural Model is a class of artificial neural network, that is most commonly applied to analyze visual imagery. We have used CNN model to classify the emotion from facial expression features.

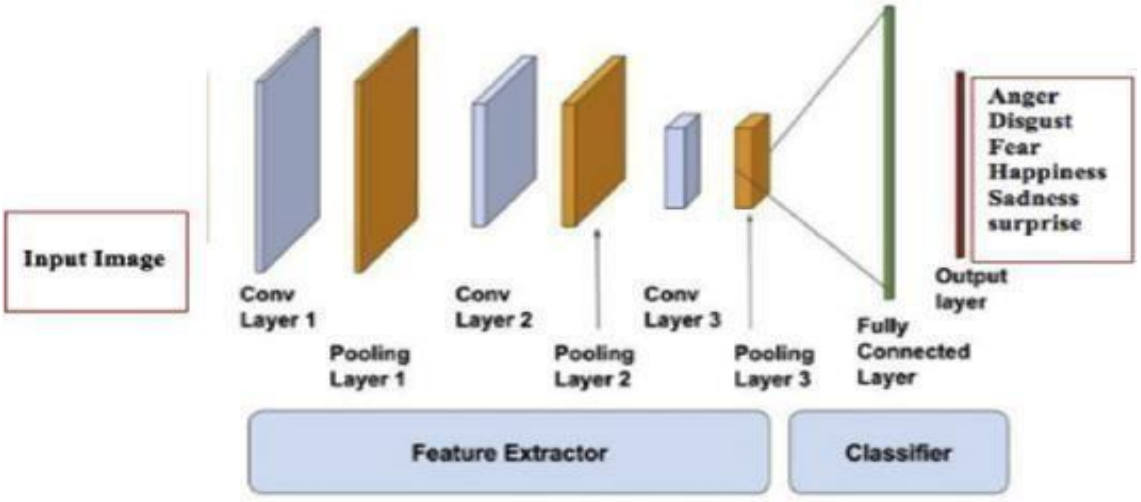

Fig 5. CNN Model

CNN Model Description:

We are using CNN model to classify the emotion expressed by the people.
The Convolutional neural network architecture consists of the following layers.
1.Convolution layer:
Here grey scale cropped image is sent to the first convolution layer of a kernel size 3X3 and activation function relu. A filter is used to recognize the special features or patterns present in the image.
2.Max pooling:
The activated frame is passed to max pooling layer, here the maximum parameter is taken out at each step and the rest is lowered along with this the network is simply down-sampled. Similarly, we have applied three convolution layers and three max pooling layers. 3.droupout layer:
To avoid overfitting we used dropout layer.
4.dense layer:
Dense Layer is used to classify the emotion from the image based on output from convolutional layers.
5. Fully connected layer:
Fully Connected Layer is added to classify the expressed emotion based on the highest probability distribution assigned by the softmax function into seven different categories.

Model Implementation:

First the input video is split into number of frames. Then we have used viola Jones object detection algorithm to detect the people face from the frame. This algorithm uses edge or line detection features from the frontal facial frame. After detecting the frontal face of the people a bounding box is drawn on the detected face. Then this cropped frame is converted to grey scale and sent to the first CNN layer of kernel size 3x3 and activation layer relu. In this layer a set of

frames are applied.

Then this activated frame is passed to pooling layer, here image are shortened. Similarly we have applied 3 convolution layers and 3 pooling layers. To avoid overfitting we used dropout layer. At the end of the network the fully connected layer is added to classify the expressed emotion based on the highest probability distribution assigned by software function into seven different categories like anger, disgust, fear, happiness, sadness, surprise, and neutral.

5.2 Descriptions Of Modules

Module 1 : Collection Module

- Selection of data
- Data formatting

Module 2 : Processing Module

- Resizing
- Conversion

Module 3 : Model Generation

- Selecting the best suited model.
- Importing the required libraries.

Module 4 :

- Training the model.
- Testing the model.

Module 5: Result and discussions

- Predict
- Visualizing the outputs

6.TESTING

6.1 Test Plan and Test Cases

Requirement Id	Test Id	Input	Expected Output	Actual Output
1	1.1	Dataset images of random size	All images should be resized to given dimension	All images were resized to 48x48 pixels
2	2.1	Training the dataset for specified number of epochs	Should give good training and testing accuracy	87.70 % accuracy
3	3.1	Input the image for testing	Plot the bounding box on identified face	Plotted correctly
4	4.1	Input the image for testing	Predict which emotion is expressed	Predicted correctly
5	5.1	Input the video	Determine the percentage of each emotion and identify the maximum	Displayed the correctly identified emotion
6	6.1	Turn on webcam	Identify people present in the frame and display their expressed emotion	Done Correctly

Table 1. Test Plan and Test Cases

7. RESULTS & DISCUSSIONS

Sample results for a single person in a frame

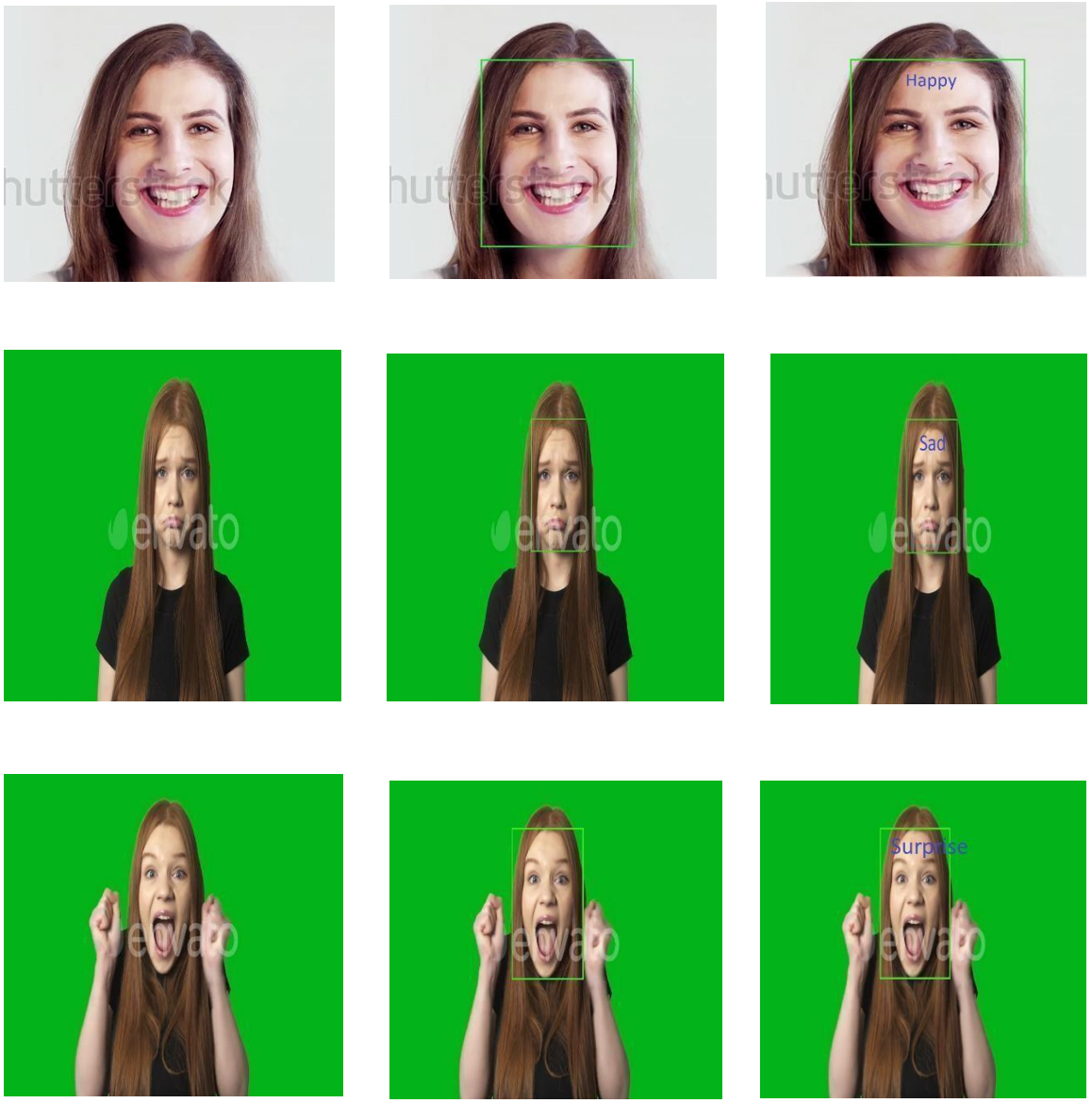

Fig 6. Sample input and output for single person in a frame

First image depicts input image to the model. The second image depicts the output of facial detection, where a bounding box is plotted on detected face. The third face depicts the identified emotion of the person.

The figures and depicts the identification of multiple people in the frame and their expressed emotion

Fig 7. Sample input and output for multiple people in a frame

	Accuracy	Precision	Recall
Video 1	85.91	93.47	92.01
Video 2	83.77	93.04	92.70
Video 3	87.34	93.85	92.76
Video 4	85.76	90.33	91.15
Video 5	84.27	89.23	88.81
Video 6	87.16	92.31	93.97
Average	85.70	92.04	91.89

Fig 8. Final Result

To find test accuracy of our model we used confusion matrix. A confusion matrix is a table that is used to define the performance of a classification algorithm.

The parameters used in confusion matrix are:

Accuracy - Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations.
Accuracy = TP+TN/TP+FP+FN+TN

Precision - Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

Precision = TP/TP+FP

Recall (Sensitivity) - Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes.

Recall = TP/TP+FN

F1 score - F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account.

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

These parameters are calculated for several videos individually and then their average is calculated to give final values of the parameters.

Final test accuracy of the model- 85.70%.

8.CONCLUSION AND FUTURE SCOPE

CONCLUSION

We have used harrcascading to detect the human face and CNN model to classify the emotion from facial expression features. The accuracy obtained from the model used is 87.70%. Further goal is to increase the accuracy.

FUTURE SCOPE

In future this system can be used to classify the emotion of a person. Can be used for Driver emotion detection in intelligent vehicles. The system can be installed in areas such as the airport, train station or bus station to detect the human face and extracts the facial expressions. If there is any face that looks suspicious, such as, angry or nervous, the system may set an alarm. It can be used for Emotion detection in the field of criminal investigation. The system can be used in mini-marts, a shopping center to view customer feedback to improve business. The system can also be used for educational fields such as one can get feedback on how a student responds during class.

9.REFERENCES

[1] Ali, Md. Forhad & Khatun, Mehenag & Turzo, Nakib. (2020). Facial Emotion Detection Using Neural Network.

[2] A. Jaiswal, A. Krishnama Raju and S. Deb, "Facial Emotion Detection Using Deep Learning," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154121.

[3] Aarohi Gupta.(2019). a Machine Learning Project .A computer vision project about emotion detection

[4] Nitisha Raut San. San Jose State University.(2018). Facial Emotion Recognition Using Machine Learning

[5]: Salih, Husam & Baghdad, Iraq & Lalit, & Kulkarni, Lalit. (2017). Study of Video based Facial Expression and Emotions Recognition Methods. 10.1109/I- SMAC.2017.8058267.

10. APPENDIX

A) Gantt Chart

DURATION	LEARNING AND IMPLEMENTATION
Week 1 to Week 2	Understanding the Dataset and Preprocessing.
Week 3 to Week 5	Implementing Classification models.
Week 6 to Week 7	Implementation using Webcam as an input video.
Week 8 to Week 10	Improvising the accuracy of the model.

Table 2: Gantt Chart

B) Description of tools and technology used:

Google Colab for implementation

Google Collaboratory or Google Colab is a free site for Jupyter Notebook. It's a Google's free cloud-based service. One of the great things about Colab is that you do not need to install anything before. We can mount our google drive and use our database.

Team Number	M31
Guide	Dr.Sujatha.C
Project title	Emotion classification in videos
Industry name	Data Science
Department vertical (either of Data/ Network/System)	Data Science
University/ Department Research group	Data Science