Emotion Recognition From Facial Expression

Goru Medha Pranami
medhapranami45@gmail.com

Abstract

A person's facial expressions act as a form of communication in personal communication since they are a demarcates of their emotional states, cognitions, intention, personality, and psychopathology. Natural human-machine interactions may involve a major aspect that automatically interprets facial expressions. Additionally, it can be used in clinical situations and behavioural science. Detecting faces in a confusing situation, identifying facial features, and classifying facial expressions are all tasks that an efficient facial expression recognition system ought to do. Convolution Neural Network (CNN) has been used to create a facial expression recognition system. In this project, a facial expression dataset comprising seven facial expressions labelled as angry, disgusted, fearful, happy, sad, surprised, and neutral.

1. Introduction

The way that people convey their feelings is through emotion. Face expressions, body language, and speaking tones are all ways to communicate it. Since the most potent, natural, and universal signal to indicate humans' emotional state is their facial expression, it plays a significant role in communicating emotion. Human facial expressions can be easily classified into 7 basic emotions: happy, sad, surprise, fear, anger, disgust, and neutral. Our facial expressions may indicate a variety of often subtle, but complicated, signals that provide a wide range of details about our emotional state. In addition to being a key element of natural human-machine interactions, automatic facial expression recognition may also be useful in behavioral science and in clinical practice.

Humans can typically identify emotions by their distinguishing facial features as they are expressed in everyday facial expressions. For instance, it is evident that a smile or an upward movement of the lips' corners indicate happiness. Similarly other emotions are characterized by other deformations typical to a particular expression. The difficulties with representing and categorizing static or dynamic aspects of these deformations of face pigmentation are addressed in research on automatic facial expression recognition. The aim of this project is to develop a real-time emotion detection website based on facial expressions where users may upload an image or video and analyze it for emotions, or even assess real-time video from a camera like a cctv to find emotions. Convolutional Neural Network (CNN), a Deep Learning-based algorithm, is used in this study. In the future, the developed application could be improved by adding more face expression categories.

The concept of convolutional networks was influenced by biological processes. They are widely used in natural language processing, recommender systems, and image and video recognition.

2. Literature Survey

As per various literature surveys it is found that for implementing this project four basic steps are required to be performed.

i. Preprocessing
ii. Face Detection
iii. Facial feature extraction
iv. Emotion classification

There are a lot of studies that have been conducted on this topic, and listing each one of them is difficult. However, in Table 2.1, a small number of these studies have been listed whose motivation has been the same as mine.

No	Title	Method	Advantage	Disadvantage
1.	FER Using Salient Features and Convolutional Neural Network	A profundity of webcam method which is used that technique is proposed here for proficient of outward appearance of the acknowledgment. For the every pixel in a profundity picture, of the eight neighbourhood directional qualities are gotten and positioned. When the position of all pixels is acquired, eight histograms are created for the eight encompassing headings.	In this method would be checked with the other traditional like all the method is the different a little storage where it showing its maximum results over other methods. An achieved the highest minimum rating is 95.20%	The weakness of this approach is CNN lays in the amount of data you provide to them. If we train with less number of images, the output would be poor results only
2.	Multiple Task CNN for Pose Invariant Face Recognition.	This method performs various tasks CNN for face acknowledgment were the personality characterize is the primary errand build up a dynamic-weighting plan to the consequently relegate the misfortune loads to the each side assignment, which takes care of the pivotal issue of adjusting between various undertakings in MTL and propose a vitality based weight of the examination technique to investigate how CNN based MTL functions.	Multi-task CNN actually in this method provides dynamic updating weighting scheme it update weights whenever the weight get lost.	Huge amount training set may reduce the processing speed and take huge amount memory the major disadvantage of multiple task approach for emotion classification. Third, both LFW and IJB-A have large variations other than pose such as the expression and blurring, etc. that cannot be well handled by the proposed method

3.	4D FER by Learning Geometric Deformation	Programmed approach for outward appearance acknowledgment from 3-D video groupings. In the proposed arrangement, the 3-D faces are spoken to by accumulations of outspread bends and a Riemannian shape examination is connected to successfully measure the disfigurements incited by the outward appearances in a given subsequence of 3-D outlines.	This approach it's can be achieved to accurately identifying with very less number of containing very different Three D frames with the maximum accuracy of 92% using HMM-based classifier and 93.21% using mean deformation-based classifier.	One limitation of the approach is, if the face contains hair fully in case of non-frontal views or the presence of occlusions.
4.	Survey RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition	This paper shows a general diagram of programmed RGB, 3D, warm and multimodal outward appearance investigation.	An in depth of the discussion about each step in a AFER pipeline followed, including the comprehensive taxonomy and many examples of techniques used on data captured with different video sensors (RGB, 3D, Thermal).	This paper describes about only the theoretical way gathered from various papers and resources.
5.	Automatic analysis of facial affect reg, representation and recognition	Programmed influence investigation has pulled in extraordinary enthusiasm for different settings and including of the acknowledgment of activity units and essential or Nonfundamental feelings. Dissect the cutting edge arrangements by the disintegrating their pipelines into central segments, in particular face enlistment, portrayal, dimensionality decrease and the acknowledgment.	This paper provides the best approach for the (FER) or emotion recognition method to identify the exact expression that approach is contains various illumination technique's to reduce the back-ground noise and other things.	This paper survey make us to understand the problem of the emotion recognition and the future development to develop more in theoretical way

6.	FER Using Hierarchical Features with the Deep of the Comprehensive MultiPatches Aggregation CNN	This paper has proposed a method, named Profound Complete Multipatches Total Convolutional Neural Systems, to tackle FER issue. The proposed technique is a profound based structure, which predominantly comprises of two parts of Convolutional Neural System.	Pooling method is proposed it's to handle the if the picture in different angle like rotating the image remove the noise etc. to achieve the more pre-processing to reduce noisy data etc.	In this method gives the 89.5% accuracy but sometimes it gives the false results it cannot identify the exact results all the times.
7.	High-Performance and Lightweight Real-Time Deep FER.	This work depicts a progressed prepreparing calculation for facial pictures and an exchange learning system, two potential possibility for loosening up this prerequisite.	This method is performs well with the low end devices accurately.	In this algorithm, it has one limitation of the approach is if the face contains hair fully in case of non-frontal views and/or the presence of occlusions.
8.	Micro-expression Recognition Using the Colour Spaces	A smaller scale articulation shading video cut treat the fourth-request tensor, i.e., a four-measurement exhibit. The initial two measurements are transient data, and the final one is the shading data.	The matching of independent colour and static texture of the spatial data and length of the code is very low	In the higher dimensional huge amount of the number of neighbours provides that the maximum local information

9.	Hierarchical Multi-pose FER	This model gives a bound together answer for multi-present FER, bypassing the different preparing and part of the tuning for each posture, and along these lines is adaptable to countless.	This paper provides the best approach for the (FER) or emotion recognition method to identify the exact expression that approach is contains various illumination of the techniques to the reducing the background noise and other things.	Sometimes it counts the miss detections as wrong expressions, and therefore, it will negatively affect the expression recognition performance.

10.	Multi-view FER and Regression	This paper proposes an efficient method called proficient of the calculation is utilizing the estimated expanded Lagrangian multiplier of approach this works achieves 95% accuracy.	This method is identifying with very less number of containing very different Three Dimension frames with an average accuracy of 90% based classifier and 94.01% using mean deformation-based classifier.	If it’s a non-frontal face images is the most common problem that have to deal with the problems of face occlusions.
11.	Regression-Based Facial Expression Optimization	This paper shows a methodology for repeating ideal 3-D outward the appearances dependent on mix shape relapse.	This method is improves the fidelity of facial expressions, and thus, the user experience while maintaining the efficiency of the blend shape method.	The lack of realistic of expression are represents true facial muscle movements, thus reducing acceptance.

12.	Robust FER for A Comprehensive Neuromuscular Signal Analysis	This paper introduces a far reaching study on the investigation of neuro-muscular flag exercises to perceive eleven outward appearances for Muscle Interfacing applications	Discriminant analysis when the values of the parameter is high this method gives maximum accuracy.	Slow communication rate, detecting small number of facial gestures, being sensitive to position and lightening.
13.	FER from Image Sequence based on LBP and Taylor Expansion	This paper presents programmed video-based outward appearance acknowledgment of the framework is to distinguish and order human outward from picture succession	It handles the illumination changes of the matching of the independent colour and static texture of the spatial data and length of the code is very low.	In this method, it cannot able to be identify the faces if the person have moustache like that.
14.	Human FER using Stepwise Linear Hidden Conditional Random Fields	In technique presents the accurate and strong outward appearance framework. For highlight take the feature or extracting proposes FER frame-work utilizes.	This FER method was achieved 93.5% across the 4 unquiet datasets, which is a significant improves in contrast to the existing FER methods.	This method modify the typical CNN structure with the proposed ETI pooling strategy, which reduces not the impact of nuisance variations in classification tasks
15.	Automatic FER Using Features of Salient Facial Patches	This paper proposed a method called system for demeanor acknowledgment by utilizing appearance highlights of chosen facial patches.	This is the method working with the different resolutions provides an optimum solution for a small size images.	In this method it will give the results well only if the images are facials and doesn’t contains any facial hair

Table 2.1 : Literature Survey

3. Software And Hardware Specifications

3.1 Introduction

To analyzing the software requirements there are two different ways are there one is functional requirements and another one non-functional requirements. In functional requirements administrator is the main actor he should have taken all the responsible of the major tasks. In non-functional requirements the administrator does not plays the main role. But it judges the operation of a system, rather than specific behavior.

3.2 Specific Requirements

3.2.1 Functional Requirements:

- The proposed solution shall be able to implement the modules of mini project in user friendly way.
- System should have varied input formats such as camera input or user input.
- Interface for video/image upload in case of user input.
- The model should have enhanced range of emotions.
- The system should extract the important features from the image.
- The module should be easier to understand.
- It will output a display of emotion as label on the face detected.

3.2.2 Non-functional Requirements:

- The module should be simple and clear.
- The system should be user friendly so that anyone can use without any difficulties.

3.3 HARDWARE AND SOFTWARE REQUIREMENT

3.3.1 Hardware requirements

- 2.0GHz Intel i5 64bit PROCESSOR or above
- 4GB / 8GB RAM
- WEBCAM
- 5GB HARD DISK SPACE

3.3.2 Software requirements

- Python == 3.7.2
- OS == WINDOWS 8.1 / 10 64BIT
- TENSORFLOW == 1.1.0
- KERAS == 2.0.5
- PANDAS == 0.19.1
- NUMPY == 1.12.1
- OPENCV-PYTHON == 3.2.02.0

4. Problem Statement

Since the time of Aristotle, the behaviour of sentiments through facial feelings has been a subject of research. This became apparent after 1955, when a list of general feelings was established and a few parametrized frameworks were suggested. Building automated recognition frameworks has gained a lot of attention within the field of computer science, encouraged by Deep Learning and Computer Vision. According to Mehran's research, 55% of the time in face-to-face communication, feelings are communicated through outer looks in order to better comprehend communication. It suggests that communication would become increasingly common and appropriate if a computer could detect and compare the user's emotions, especially if we take into account scenarios in which a computer would take the place of a human.

4.1 Objectives

The main objective Build and train a convolutional neural network (CNN) from scratch to recognize facial expressions. Once you have trained, saved, and exported the CNN, you will directly serve the trained model to a web interface and perform real-time facial expression recognition on video and image data.

4.2 Proposed Solution:

Implement the concept of a convolutional neural network as a response to the abovementioned problem statement. However, a lot of neural network methods utilise back propagation, thus we choose convolutional neural networks instead because they have special properties like pooling. The real statistics image is examined for highlights. In order to deliver accurate results, the suggested approach neglects unnecessary background distractions. The highest point for each channel we use was put up by the activation function. A fix to down sample is the largest square shape. Through down sampling, the activation function was shortened.

4.4 Emotion Recognition

Facial recognition is the skill of identifying human emotion from facial expressions and then classifying that emotion into one of the categories listed above. Numerous technological improvements in this field have ensured that attention is maintained and that progress is being made constantly. The next method of communication between a human and a machine is anticipated to be nonverbal communication.

There are several ways to identify emotions, but one popular method is to flatten the image and convert it to ASCII characters. Plain vector points can now be accurately mapped onto the face using ASCII codes, allowing for the interpretation of facial images. However, it is not the sole method of classifying and identifying facial images. The process of identifying emotions is not as simple as it appears as we have seen in one method. When classifying emotions, a variety of characteristics are taken into account. Readers should be aware that each person is different and exhibits a unique range of emotions, and that emotions are not culturally bound.

4.5 CNN?

A convolutional neural network (CNN or convnet) is a subset of machine learning. It is one of the various types of artificial neural networks which are used for different applications and data types. A CNN is a kind of network architecture for deep learning algorithms and is specifically used for image recognition and tasks that involve the processing of pixel data.

A CNN, Convolutional Neural Network is one kind of neural network concept which is mainly used for analysing the image frame by frame purpose and convolutional neural network consists of three layers: a convolutional layer, a pooling layer and a fully connected (FC) layer. The convolutional layer is the first layer while the FC layer is the last.

From the convolutional layer to the FC layer, the complexity of the CNN increases. It is this increasing complexity that allows the CNN to successively identify larger portions and more complex features of an image until it finally identifies the object in its entirety. By using that it performs the operation based on that functions.

Convolutional layer:

The majority of computations happen in the convolutional layer, which is the core building block of a CNN. A second convolutional layer can follow the initial convolutional layer. The process of convolution involves a kernel or filter inside this layer moving across the receptive fields of the image, checking if a feature is present in the image.

Over multiple iterations, the kernel sweeps over the entire image. After each iteration a dot product is calculated between the input pixels and the filter. The final output from the series of dots is known as a feature map or convolved feature. Ultimately, the image is converted into numerical values in this layer, which allows the CNN to interpret the image and extract relevant patterns from it.

Pooling layer:

Like the convolutional layer, the pooling layer also sweeps a kernel or filter across the input image. But unlike the convolutional layer, the pooling layer reduces the number of parameters in the input and also results in some information loss. On the positive side, this layer reduces complexity and improves the efficiency of the CNN.

Fully connected layer:

The FC layer is where image classification happens in the CNN based on the features extracted in the previous layers. Here, fully connected means that all the inputs or nodes from one layer are connected to every activation unit or node of the next layer.

All the layers in the CNN are not fully connected because it would result in an unnecessarily dense network. It also would increase losses and affect the output quality, and it would be computationally expensive.

4.2 Scope And Applications

The scope of this system is to tackle with the problems that can arise in day-to-day life. Facial image recognition is used in BPO's for identifying calls based on emotions. Some of the scopes are:

1. The system can be used to detect and track a user's state of mind.
2.The system can be used in mini-marts, shopping center to view the feedback of the customers to enhance the business,
3.The system can be installed at busy places like airport, railway station or bus station for detecting human faces and facial expressions of each person. If there are any faces that appeared suspicious like angry or fearful, the system might set an internal alarm.
4.The system can also be used for educational purpose such as one can get feedback on how the student is reacting during the class.
5.This system can be used for lie detection amongst criminal suspects during interrogation
6.This system can help people in emotion related -research to improve the processing of emotion data.
7.Clever marketing is feasible using emotional knowledge of a person which can be identified by this system.
8. It can also be used in medical industries and to identify whether a driver is fatigued or not, which can help prevent possible accidents.

Before images are fed into the machine for deep learning, we must be careful about the quality of the images, for example, facial rotation, illumination, facial alignment, etc, and lots of pre-processing techniques are applied so that the images are flattened at an earlier stage to avoid problems and diminishing accuracy result at a later stage of time.

5. Design Of System

5.1 Approach Of The Design

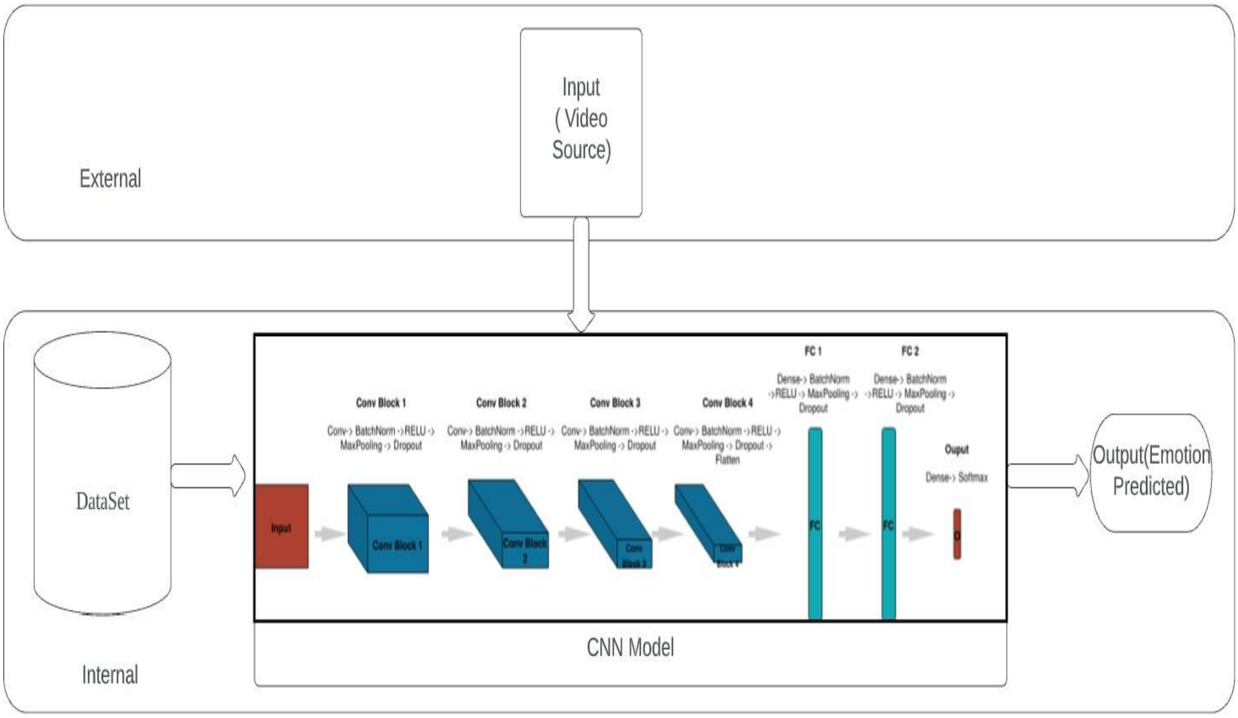

Figure 5.1: Architecture Design

In this fig 3.1 the external process describes about the task which is done externally which the user can able to see and user can control. The internal process describes about the task which is done dynamically without the user involvement.

5.2 Flow Chart Of Solution :

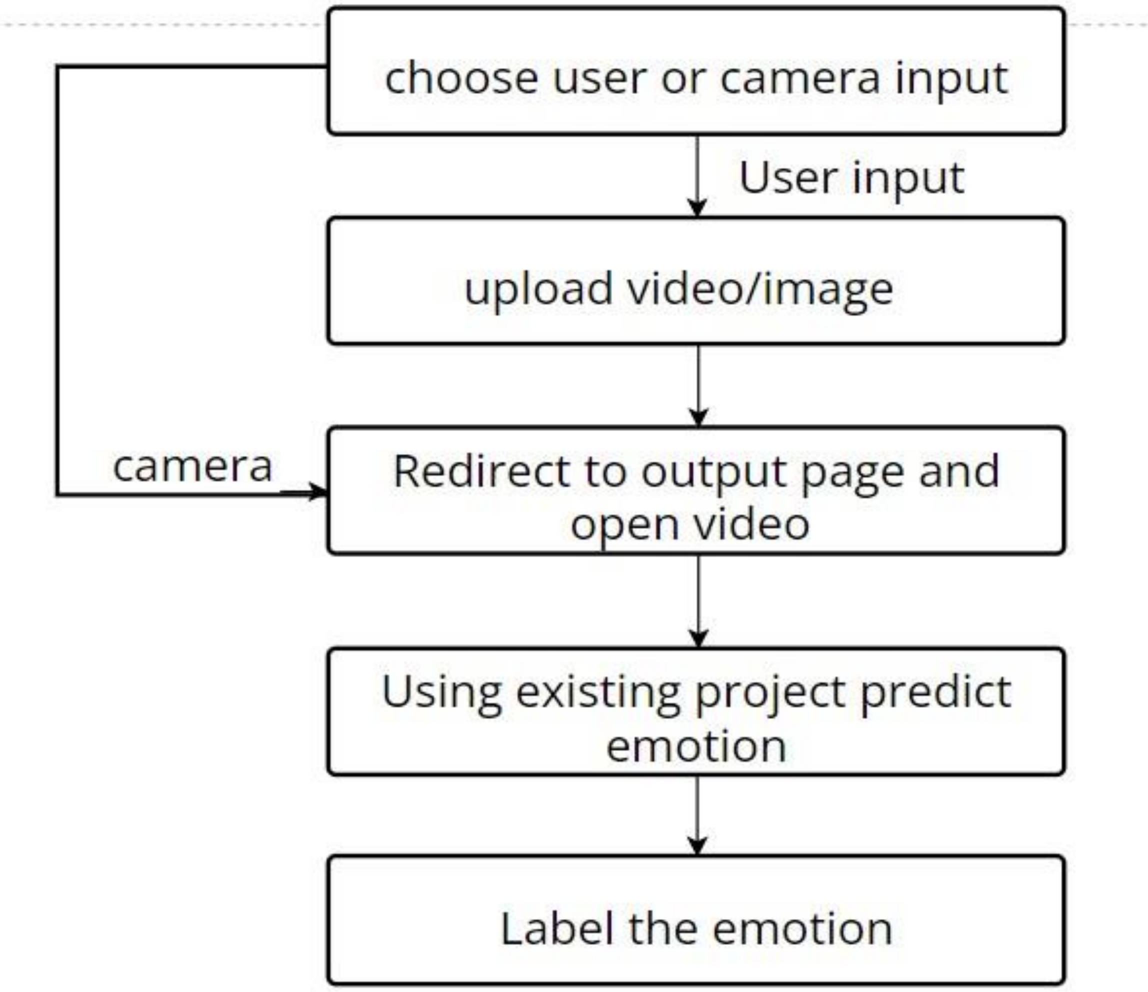

Figure 5.2: Activity Diagram

5.3 Methodology

Build and train a convolutional neural network (CNN) from scratch to recognize facial expressions. The data consists of 48x48 pixel images of faces. The objective is to classify each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

We will use OpenCV to automatically detect faces in images and draw bounding boxes around them. Once trained, saved, and exported the CNN, you will directly serve the trained model predictions to a web interface and perform real-time facial expression recognition on video and image data. We developed CNNs with variable depths to evaluate the performance of these models for facial expression recognition.

The first part of the network refers to M convolutional layers that can possess spatial batch normalization (SBN), dropout, and max-pooling in addition to the convolution layer and ReLU nonlinearity, which always exists in these layers. After M convolution layers, the network is led to N fully connected layers that always have Affine operation and ReLU nonlinearity, and can include batch normalization (BN) and dropout. Finally, the network is followed by the affine layer that computes the scores and soft-max loss function. The developed model gives the user the freedom to decide about the number of convolutional and fully connected layers, as well as the existence of batch normalization, dropout and max-pooling layers. Along with dropout and batch normalization techniques, we included L2 regularization in our implementation. Furthermore, the number of filters, strides, and zero-padding can be specified by user, and if they are not given, the default values are considered. As we will describe in the next section, we proposed the idea of combining HOG features with those extracted by convolutional layers by mean of raw pixel data. To this end, we utilized the same architecture described above, but with this difference that we added the HOG features to those exiting the last convolution layer. The hybrid feature set then enters the fully connected layers for score and loss calculation.

5.3.1 Dataset and Features

In this project, we used a dataset of our own in which we collected images from various sources, which consists of about 400 well structured images of faces. The images are processed in such a way that the faces are almost centered and each face occupies about the same amount of space in each image. Each image has to be categorized into one of the seven classes that express different facial emotions. These facial emotions have been categorized as: 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral.

Figure 5.3 Data set

Figure 1 depicts one example for each facial expression category. In addition to the image class number (a number between 0 and 6), the given images are divided into two different sets which are training, validation sets. There are about 300 training images, 100 validation images.

After reading the raw pixel data, we normalized them by subtracting the mean of the training images from each image including those in the validation sets. For the purpose of data augmentation, we produced mirrored images by flipping images in the training set horizontally.

In order to classify the expressions, mainly we used the features generated by convolution layers using the raw pixel data. As an extra exploration, we developed learning models that concatenate the HOG features with those generated by convolutional layers and give them as input features into Fully Connected (FC) layers.

5.3.2 Face Extraction:

The human face is captured by using PC's web cam or external webcam. From that live steam the face is extracted and all other unwanted components are not considered. To achieve this efficiency and comprehensiveness I have utilized the Open CV library (to be specific classifier).

5.3.3 Feature Extraction:

Pre-processing:

It is a term used to describe processes where the input and output images are intensity images, which represent the lowest level of abstraction. Pre-processing is a step in the processing of images that aims to improve the image data by enhancing or suppressing undesirable distortion or certain image elements that are crucial for subsequent processing.

Haar cascade classifier:

Haar Cascade is a machine learning-based approach where a lot of positive and negative images are used to train the classifier.

Positive images – These images contain the images which we want our classifier to identify.

Negative Images – Images of everything else, which do not contain the object we want to detect.

Region splitting:

For the emotion recognition the main region of face under consideration are eyebrows and mouth. And the splitting of mouth and the eyebrows is named as region splitting.

5.3.4 Emotion Classification:

After the sub task of feature extraction is completed the reaction of the person is produced simultaneously with their percentage level.

5.3.5 Facial Expression Evolution:

Angry:

Anger can be characterized by these characteristics below:

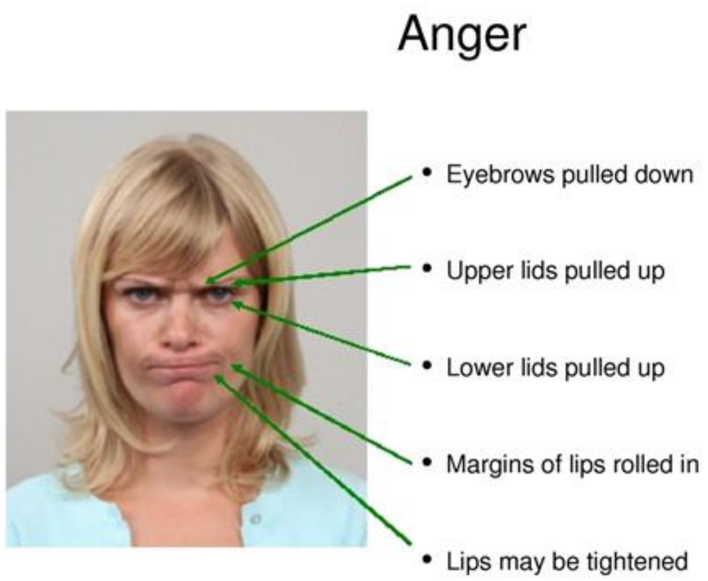

Figure 5.4: The angry face

Disgust:

Disgust can be characterized by these characteristics below:

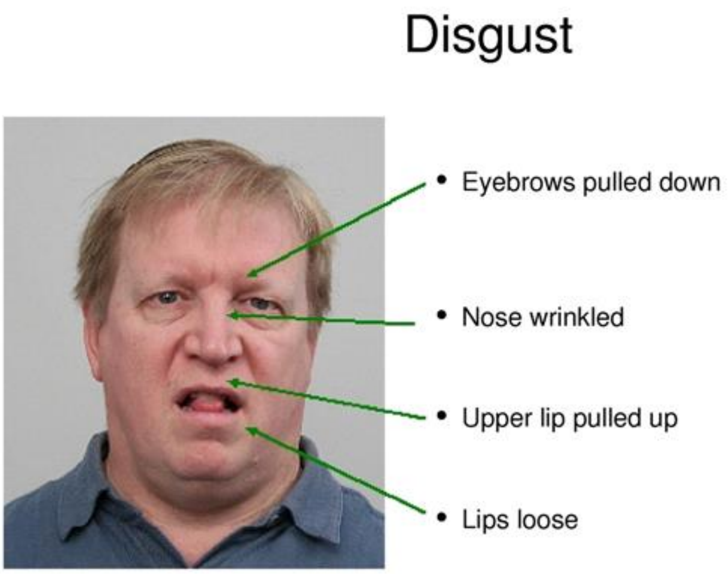

Figure 5.5 The disgust face

Fear:

Fear is recognized all around the world by this facial expression of emotion, with the characteristics which can be seen below. Its important to note that other words describing fear are also expressed by this same face (or portions of this face). Emotions such as scared, mortified, horrified and petrified all have characteristics.

Figure 5.6 The fear face

Happy:

Happiness can be characterized by these characteristics below:

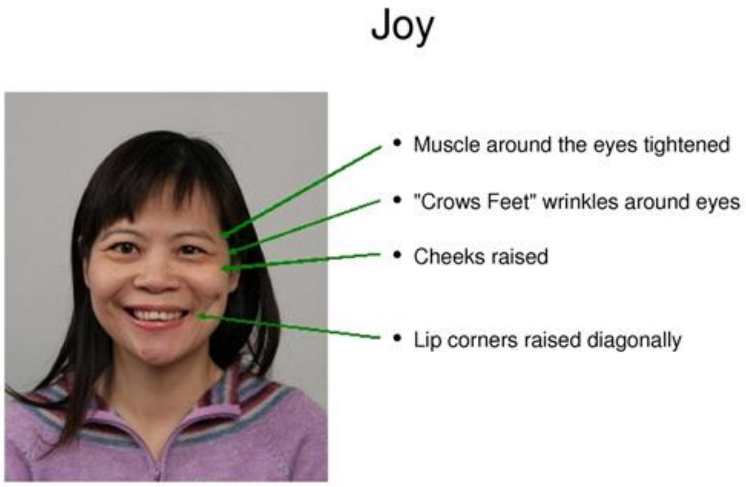

Figure 5.7 The Happy face

Neutral:

It does not include in any of the reaction like happy, shock, sad, disgust, angry etc. this expression is a simple one where the lips and eyes are in normal position. Which indicates that the user is not showing any reaction. The default type of emotion is neutral. The every reaction changes starts from the neutral.

Figure 5.8 The neutral face

Sad:

Sadness can be characterized by these characteristics below:

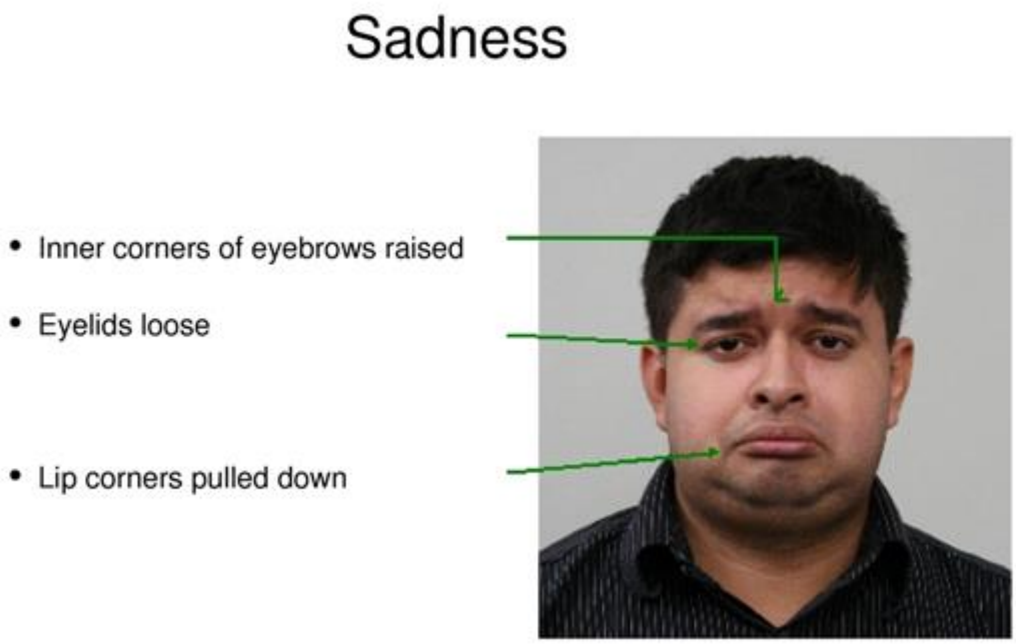

Figure 5.9 The sad face

Surprise:

Surprise can be characterized by these characteristics below:

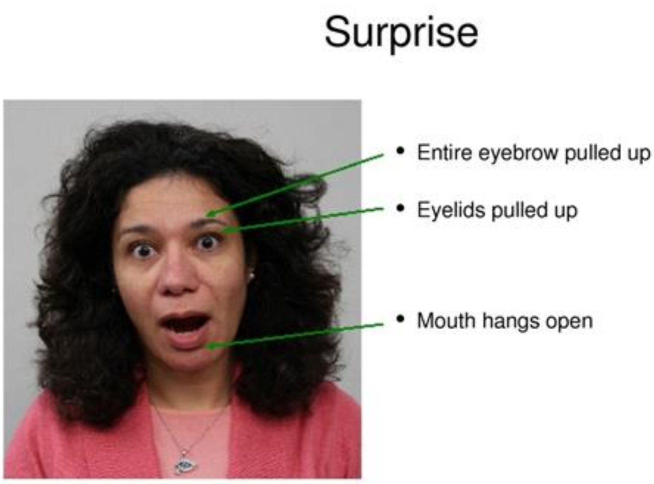

Figure 5.10 The surprise face

5.3.6 WORKING

How face extraction works?

OpenCV its commonly well-known library for facial extraction. OpenCV utilizes AI calculations to scan for countenances inside an image. Since countenances are so confused, there isn't one basic test that would be understand it found a face or not. Rather, there are a great many little examples and highlights that must be coordinated. The calculations break the assignment of recognizing the face into a great many littler, nibble estimated undertakings, every one of which is anything but difficult to tackle. These undertakings are likewise called classifiers. For something like a face, you may have at least 7,000 classifiers, all of which must counterpart for a face to be identified inside blunder limits, obviously. Be that as it may, in that lies the issue: for face recognition, the calculation begins at the upper left of an image and moves down crosswise over little squares of information, taking a gander at each square. Like a progression of cascades, the OpenCV course breaks the issue of distinguishing faces into various stages. For each square, it completes an extremely harsh and speedy test. On the off chance that that passes, it completes a somewhat progressively itemized test, etc.

The calculation might had 20 to 40 this stages of falls, and it will possibly recognize the facial should clear. The preferred standpoint is that most of the image will restore a negative amid the initial couple of stages, which implies the calculation won't sit around idly testing every one of the 6,000 highlights on it. Rather than taking hours, face identification should now be possible progressively. Since face recognition is such a typical case, OpenCV accompanies various inherent falls for recognizing everything from countenances to eyes to hands to legs.

How face detection works?

Understanding the human outward appearances and the investigation of articulations has numerous perspectives, from PC examination, feeling acknowledgment, lie indicators, airplane terminal security, nonverbal correspondence and even the job of articulations in craftsmanship. Improving the abilities of perusing articulations is an essential advance towards fruitful relations. Articulations and feelings go inseparably, for example exceptional blends of face strong activities mirror a specific feeling. For specific feelings, it is exceptionally hard, and perhaps inconceivable, to stay away from it's fitting outward appearance. For instance, an individual who is endeavoring to disregard his supervisor's irritating hostile remark by keeping an unbiased articulation may in any case demonstrate. This marvel of a short, automatic outward appearance appeared on the essence of people as per feelings experienced is called 'micro expression'. The direct opposite marvel alludes to the manner in which that some muscle developments speak to a feeling, and the contrary muscle developments speak to the contrary feeling. A great clarification for the outward appearance speaks to 'weakness' should be possible utilizing direct opposite. Weakness body motion includes hands spreading to the sides, fingers spreading and shoulders shrugging. Its outward appearances include pulling down the base lip and raising eyebrows. Darwin clarified the highlights of this articulation utilizing the direct opposite standard. He found that those developments restricting to the developments of a man who is prepared to confront something. The developments of an individual who is setting himself up for something will resemble that: shut hands and fingers (as though he is getting ready for a battle, for instance), hands near the body for insurance and the neck is raised and tight. At a weakness circumstance the shrugging of the shoulders discharges the neck. Concerning the face: eyebrows are low (like in a method of assault or solidness), upper lip may uncover teeth. The practical wellspring of the absolute opposite can be clarified with the examination of muscles, and to be exact the rival's muscles. Each muscle has an adversary muscle that plays out the contrary development. Spreading fingers is a development done by a few muscles, and shutting the fingers is finished by the foe muscles. For a few articulations we can't generally tell just by taking a gander at it, what is the contrary articulation, so the choice that took the gander the muscles including all the while, at that point it turns out to be exceptionally clear. A fascinating clarification to the direct opposite practical source depends on hindrance. An individual or a creature is attempting to utilize the hostile muscles. Indeed, when an improvements flag is send to a muscle, an inhibitory flag is send naturally to the rival muscle. Outward appearances that can be clarified with direct opposite all the parts of identity with hostility and maintaining a strategic distance.

How CNN works?

How a human figures out how to perceive objects, we have to demonstrate a calculation on a huge number of pictures before it is have the capacity to sum up the information and make expectations for pictures it has never observed. PCs 'see' uniquely in contrast to we do. Their reality comprises of just numbers. Each picture can be spoken to as 2- dimensional varieties of numbers, known as pixels. Any of the case have thing that they see pictures in an unexpected way, doesn't mean we can't prepare them to perceive designs, as we do. Its need to be consider what a picture is in an unexpected way.

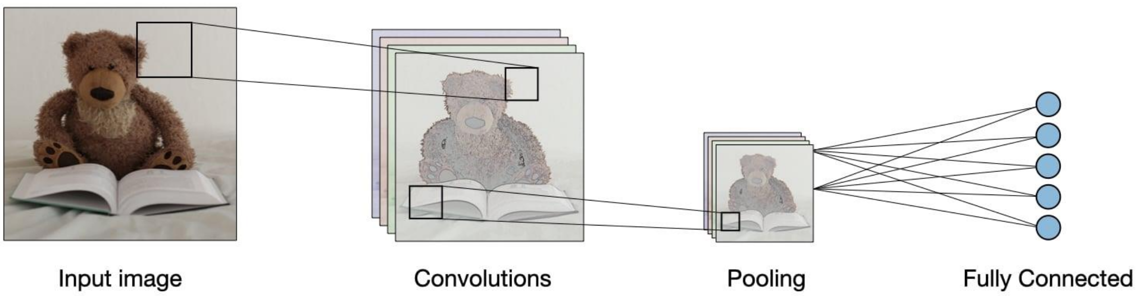

Figure 5.11 The Image Identification And Classification

Fig 5.11 shows the calculation how to perceive protests in pictures, we utilize a particular kind of Artificial Neural Network, a Convolutional Neural Network (CNN). The name comes from it’s the most important tasks in the system called convolution. The straightforward cells enact, for instance, when they recognize fundamental fixed by lines as the shapes territory and a particular point. The unpredictable cells have bigger open fields and their yield isn't touchy to the particular position in the field. The complex cells keep on reacting to a specific improvement, despite the fact that its total direction of the eyes will change. Complex alludes to increasingly adaptable, for this situation. In vision, a responsive field of a solitary tangible area of the retina in which something will influence the terminating of that neuron (that is, will dynamic the neuron). Each tangible neuron cell has comparable open fields, and their fields are overlying.

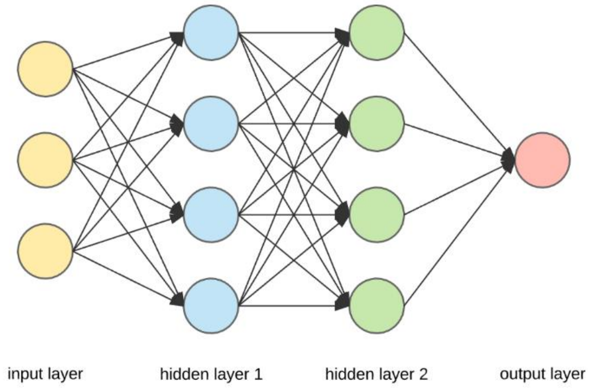

Figure 5.12 Neural Network Diagram

The hidden layer computes the weight and the bias for the purpose of activation purpose. The bias and weights were adjusted to get the accurate output.

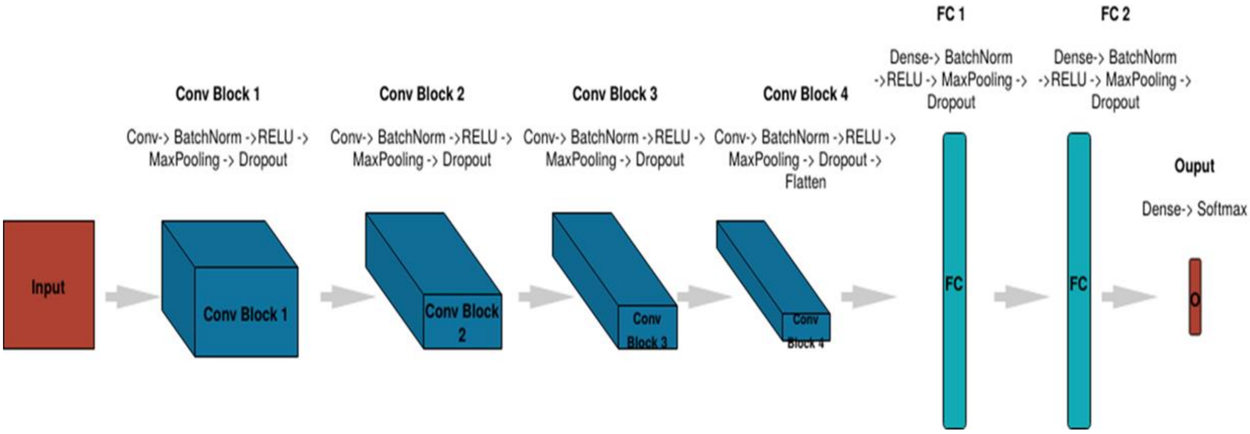

Figure 5.13 Extraction of needed parts from face (Fig 5.12 ref: convolutional neural network introduction part 2)

In fig 5.13 it shows that the huge pictures(input) is diminished in the upcoming layers of the neural network and at last the unwanted parts are eliminated in every layer and the output is produced for our task. In this paper, the face is the input for the system. The input is sent to the next layer and at last the eyes and mouth is extracted and produced as the output.

6. Implementation

- Task 1 : Collect the Dataset
- Task 2: Generate Training and Validation Batches
- Task 3: Create a Convolutional Neural Network (CNN) Model
- Task 4: Train and Evaluate Model
- Task 5: Represent Model as JSON String
- Task 6: Create a Flask App to Serve Predictions
- Task 7: Design an HTML Template for the Flask App
- Task 8: Use Model to Recognize Facial Expressions in Videos

Collect the Dataset

- Display some images from every expression type in the Emotion FER dataset.
- Check for class imbalance problems in the training data.

Generate Training and Validation Batches

- Generate batches of tensor image data with real-time data augmentation.
- Specify paths to training and validation image directories and generates batches of augmented data.
- Generate batches of tensor image data with real-time data augmentation.
- Specify paths to training and validation image directories and generates batches of augmented data.
- We give group of images as a batch for model to train at once.
  (Here our batch size is 32).
- train_datagen here we will generate ‘n’ no of images(5-6) from one image by rescaling, rotating, zooming, flipping, rotating vertical and horizontal.
- After training we will cross check it by validation.
- Images from train_datagen are given to train_generator and images from validation_datagen are given to validation_generator.
- This train_generator and validation_generator images will be given to model.

Create a Convolutional Neural Network (CNN) Model

- Design a convolutional neural network with 4 convolution layers and 2 fully connected layers to predict 7 types of facial expressions.
- Use Adam as the optimizer, categorical crossentropy as the loss function, and accuracy as the evaluation metric.

Train and Evaluate Model

- Train the CNN by invoking the **model.fit()** method.
- Use **ModelCheckpoint()** to save the weights associated with the higher validation accuracy.

Save and Serialize Model as JSON String

- Sometimes, you are only interested in the architecture of the model, and you don't need to save the weight values or the optimizer.
- Use **to_json()**, which uses a JSON string, to store the model architecture.

Create a Flask App to Serve Predictions

- create a flask app to serve the model's prediction images directly to a web interface.

Create a Class to Output Model Predictions

- Create a FacialExpressionModel class to load the model from the JSON file, load the trained weights into the model, and predict facial expressions.

Design an HTML Template for the Flask App

- Design a basic template in HTML to create the layout for the Flask app.

Use Model to Recognize Facial Expressions in Videos

- Run the **main.py** script to create the Flask app and serve the model's predictions to a web interface.
- Apply the model to saved videos / images on disk or web cam video.

7. Testing:

Now we will test the model that we build for emotion detection in real-time using OpenCV and webcam/ input file. To do so we will write a python script. We will use the visual studio in our local system to make use of a webcam. First, we will install a few libraries that are required.

After importing all the required libraries we will load the model weights that we saved earlier after training. After importing the model weights we have imported a haar cascade file that is designed by open cv to detect the frontal face.

After importing the haar cascade file we will have written a code to detect faces and classify the desired emotions. We have assigned the labels that will be different emotions like angry, happy, sad, surprise, neutral. As soon as you run the code a new window will open and your webcam will turn on / video will be opened. It will then detect the face of the person, draw a bounding box over the detected person, and then convert the RGB image into grayscale & classify it in real-time.

8. Results:

Version 1:

It has only one input method where we can upload images or video for detecting emotion.

Version 2:

It has 2 input methods where we can choose between web-cam input or user input.
Errors: The video which we have given as input is not been deleted in server until we restart the server.
We would like to modify code in further versions such that the above error would be rectified.

Ouput Screen:

1.Main screen of version 1:
This will be the main screen of our project.
In the header part on clicking gitam logo it navigates to gitam.edu page.
On clicking contact us it will navigate to contact us page of gitam. And About Us will navigate to about us page of gitam.
On Clicking the user input we will be redirected to page where we can upload our file for detecting emotions.
In the footer part the social media icons navigate to facebook Instagram and twitter pages of gitam.

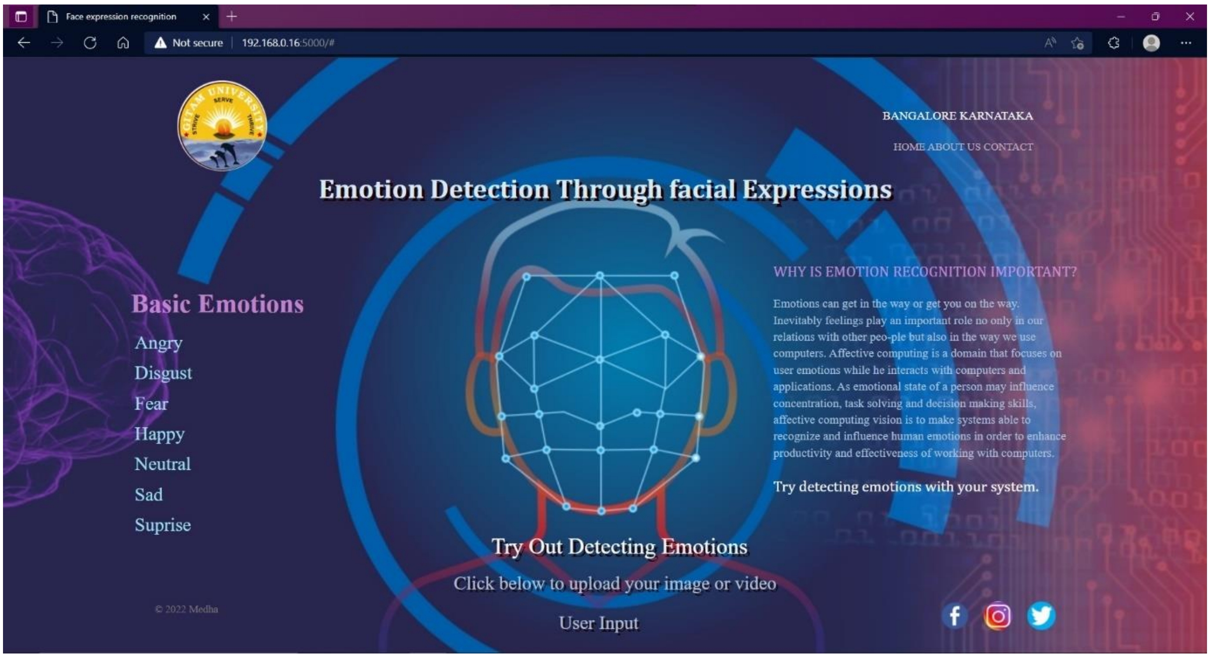

Figure 8.1 Main Screen

2.Main screen when hovered on emotions.

When an emotion is clicked or the cursor is moved on to any of the 7 emotion names which our application detects, we can see that particular picture with that emotion.
The screen looks as follows:

Figure 8.2 Hovered over Angry

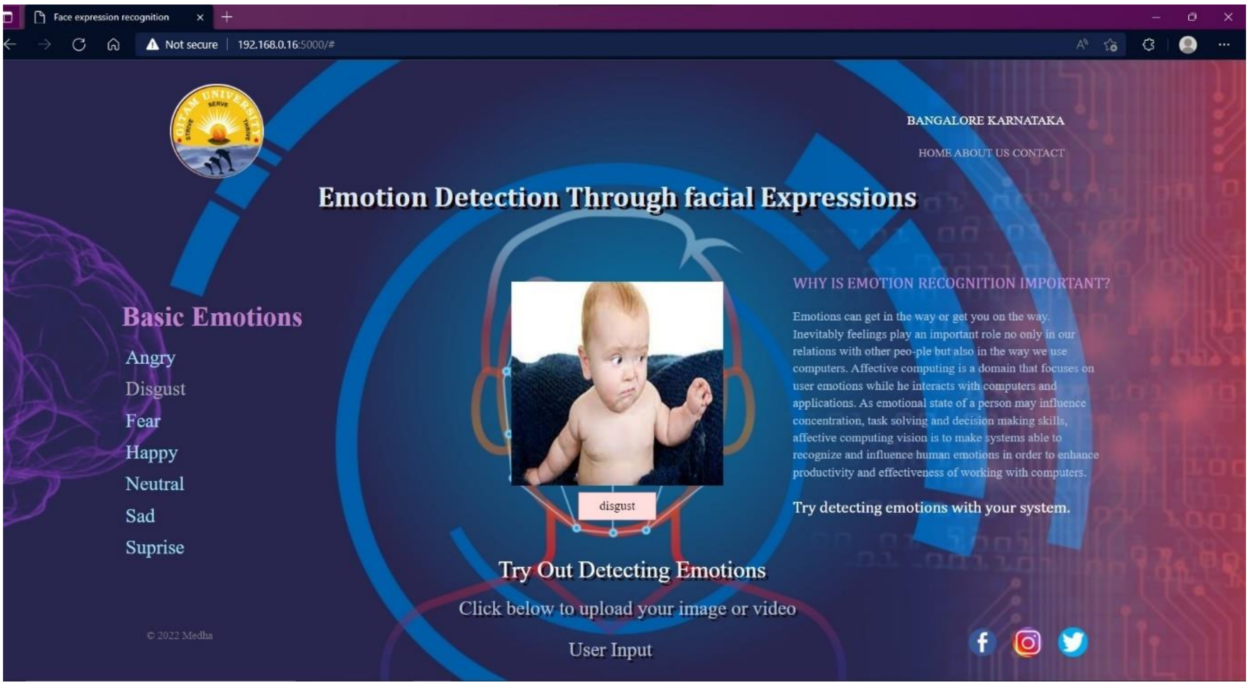

Figure 8.3 Hovered over Disgust

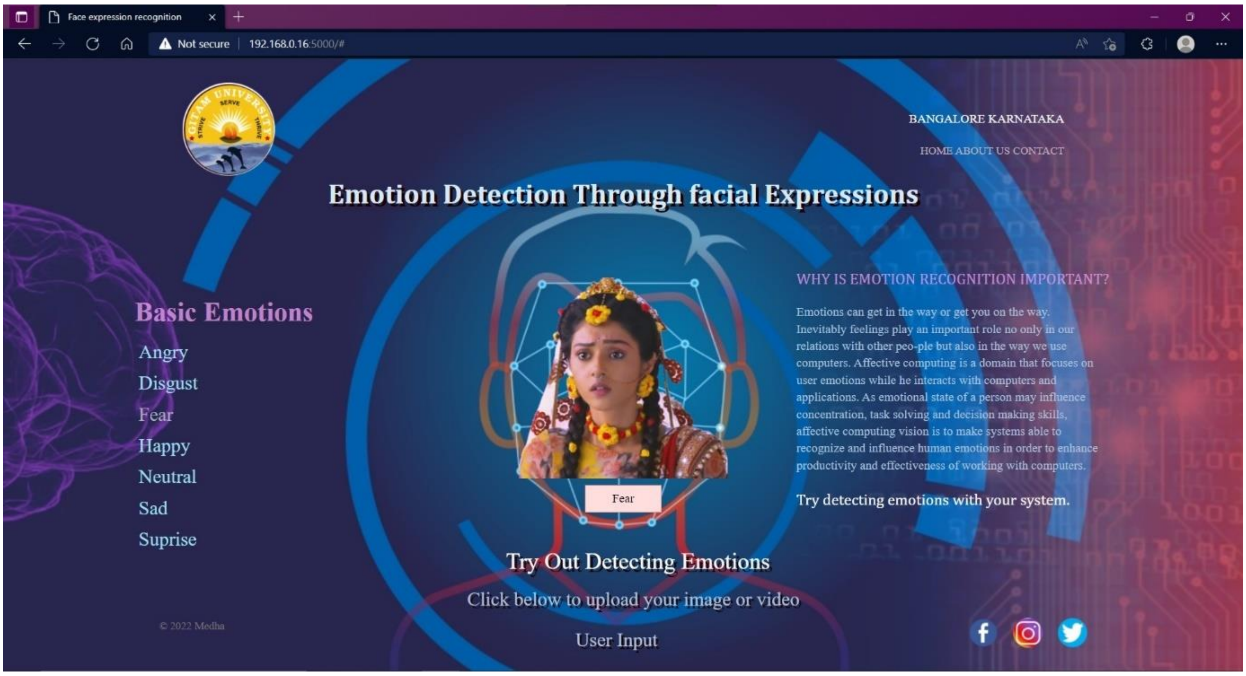

Figure 8.4 Hovered over Fear

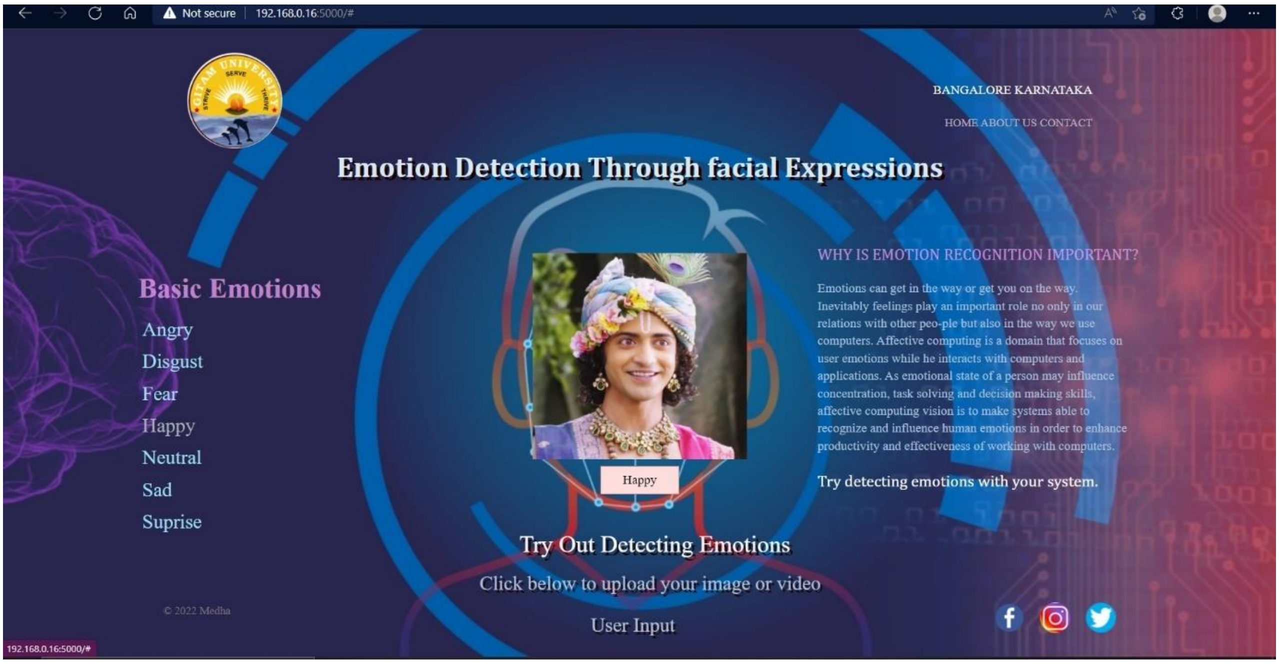

Figure 8.5 Hovered over Happy

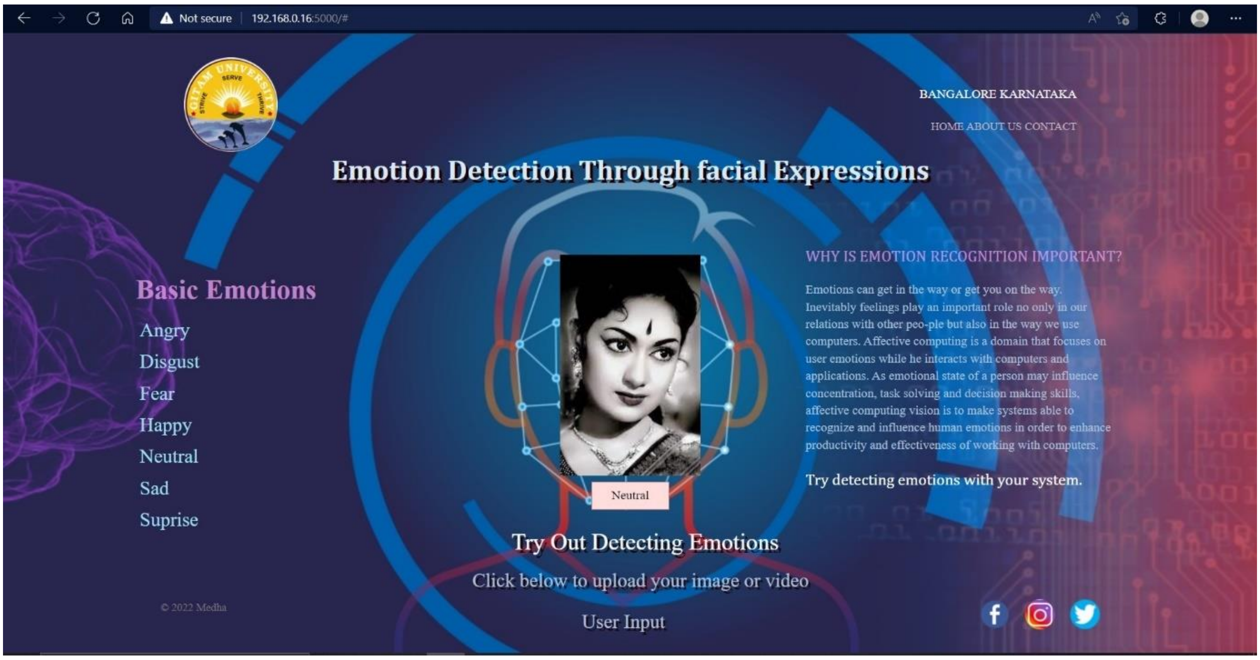

Figure 8.6 Hovered over Neutral

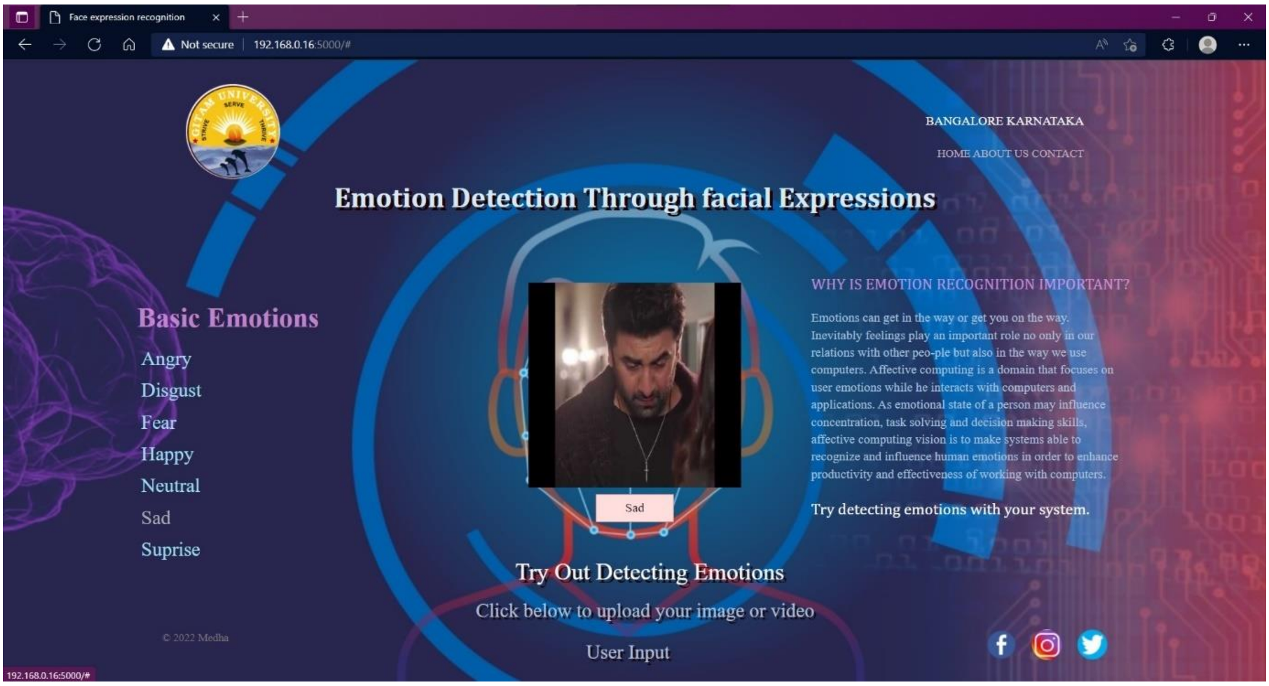

Figure 8.7 Hovered over Sad

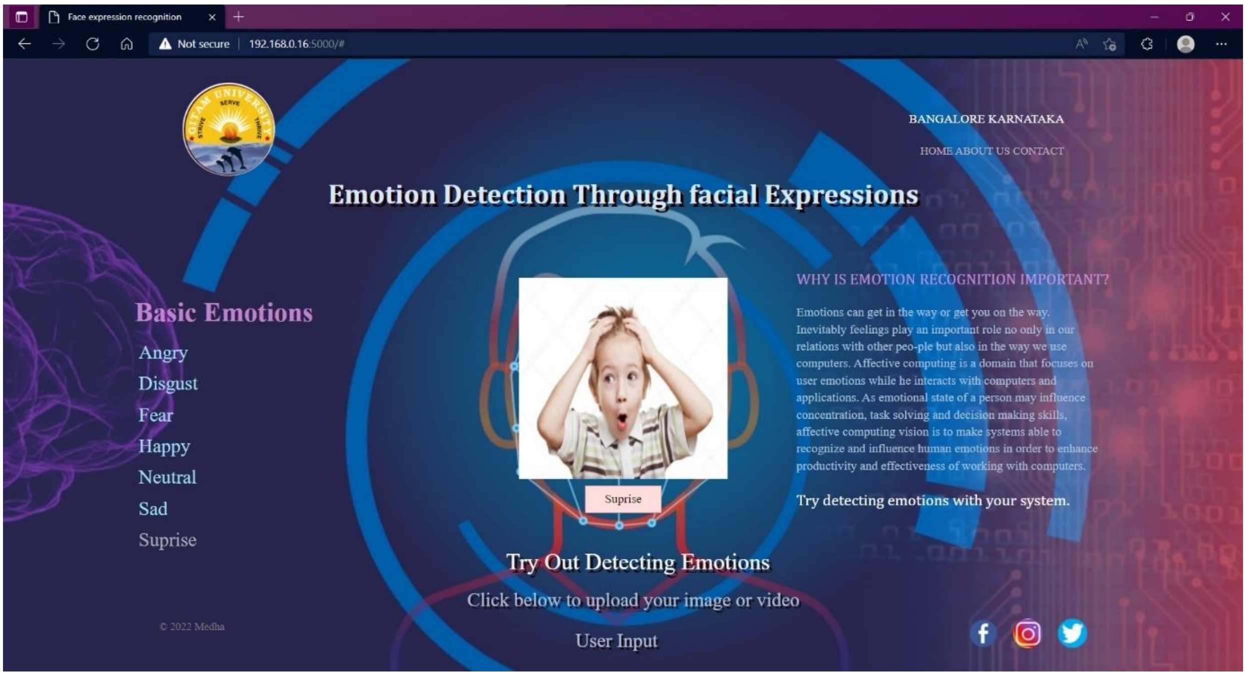

Figure 8.8 Hovered over Surprise

3. Page where we can upload our file :

The header and footer part functions same as that of main page. On clicking choose file, we can choose file from our system for which we want to detect the emotion. Then on clicking submit we will be redirected to file output page.

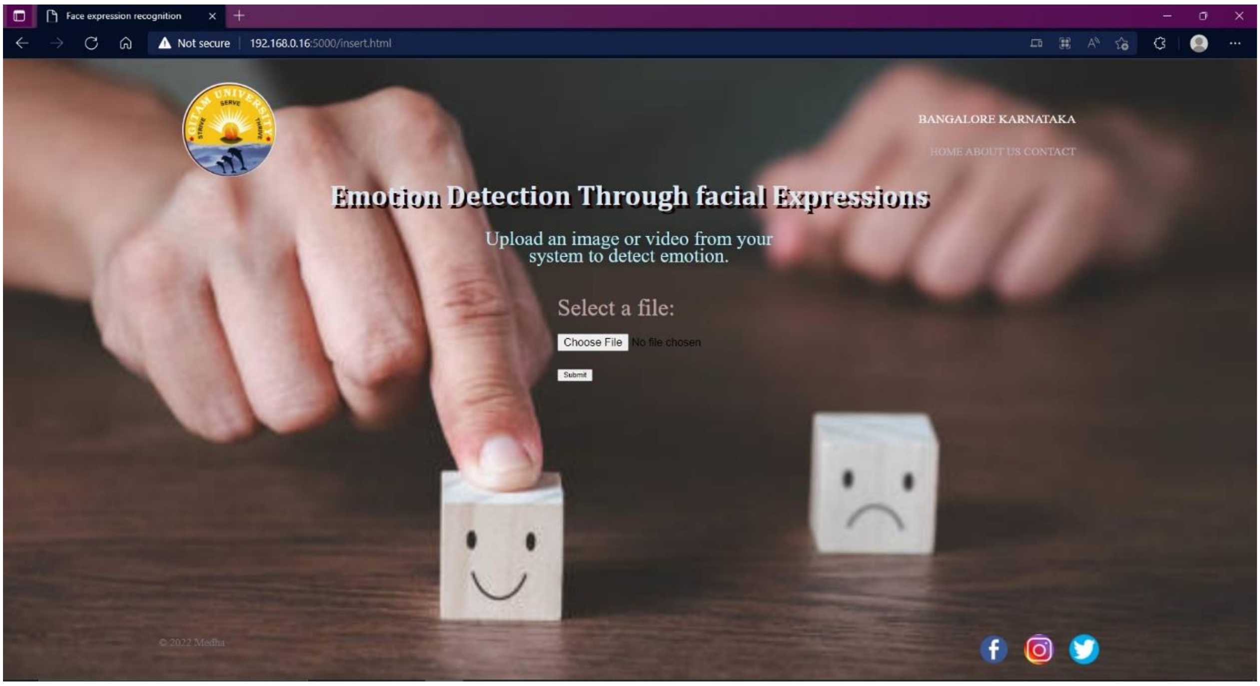

Figure 8.9 Screen before uploading image

4.After uploading file:

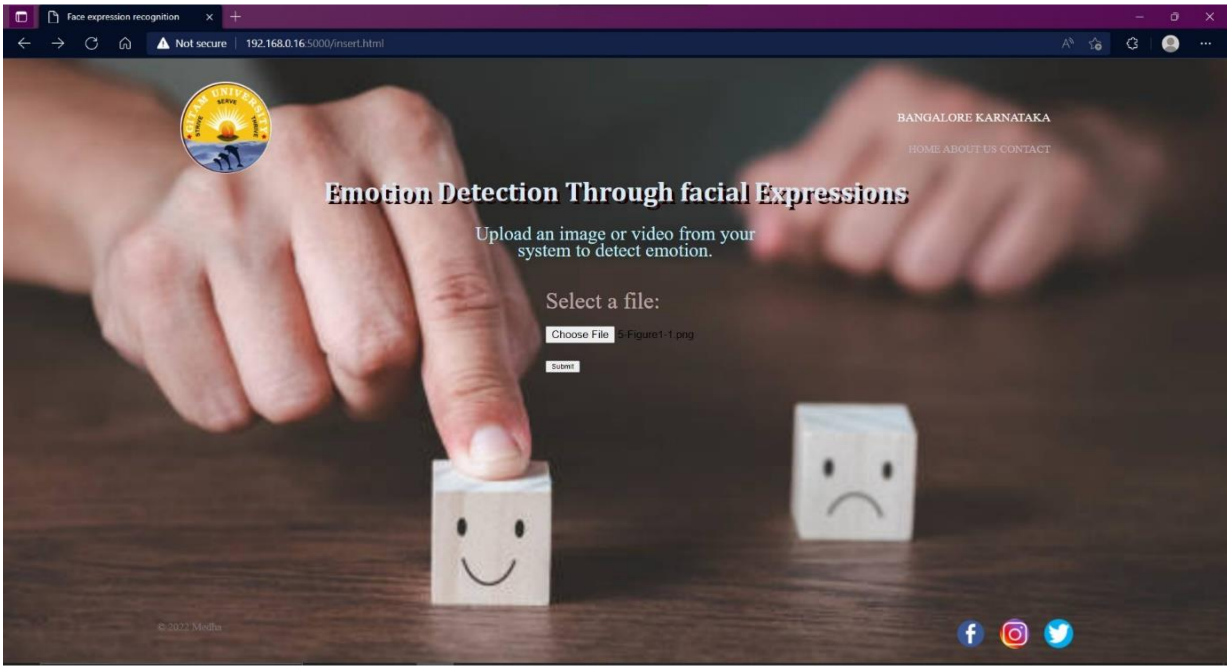

Figure 8.10 Screen after uploading image

5. Final output page:

Here the file uploaded will be opened and emotions are detected using our model.

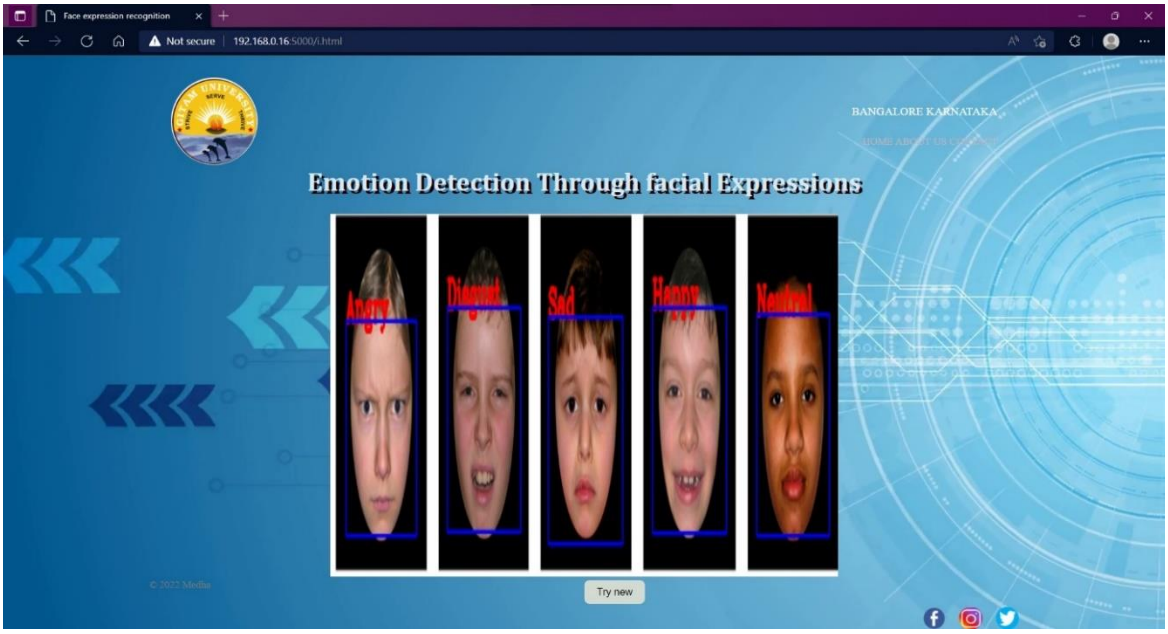

Figure 8.11 Final output

6. Main page of version 2:

All elements work same as version 1. But Here we will have option of choosing input between Video input or camera input. If it is a video input we would be directed to page where we can upload a video file. After submitting we would be redirected to output page. If it is a camera input we would directly be directed to final output page and through web cam video source is taken and output is predicted.

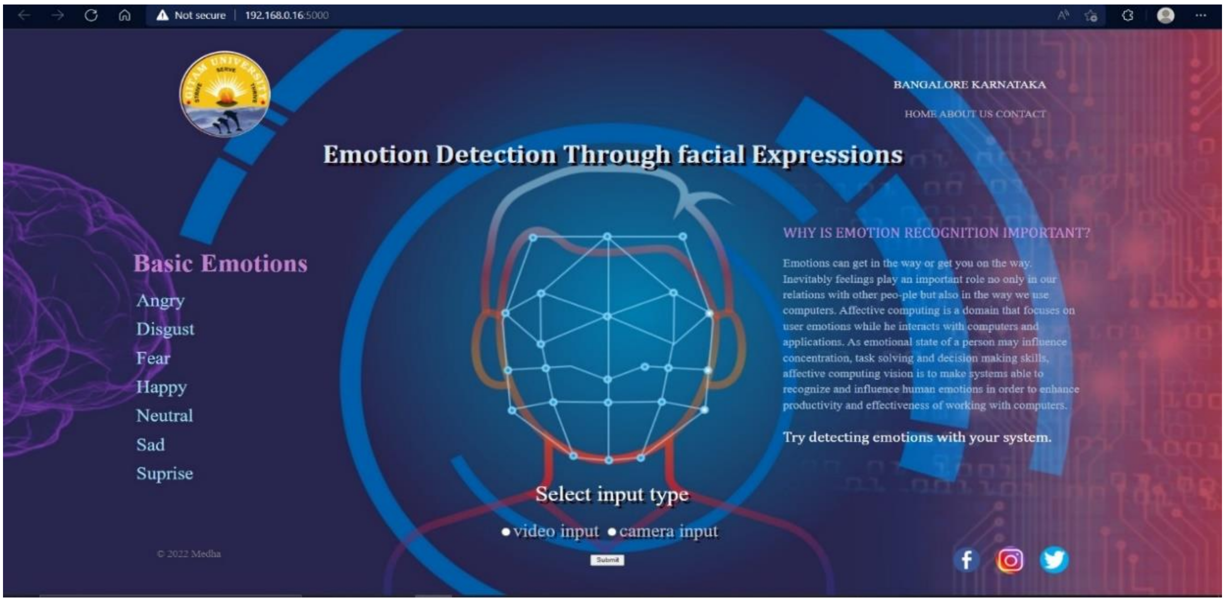

Figure 8.12 Main Screen of version 2

9. Learnings From Project

- The most interesting thing to explore was to learn that every human emotion is different from the rest.

Subject Learnings:

- Learnt open cv methods for image processing and performing computer vision tasks.
- Practical Experience how to train model using machine learning algorithm.
- Learnt Creating a web application of emotion recognition using flask app which makes our project unique from existing projects.

10.Conclusion

Facial expression recognition is a boon to mankind and we can see many enthusiasts experimenting and making the experience better, be it in medical treatment, or in contributing to the Global Happiness Index. In this experiment, various databases had been explored and at last with comprehensive analysis, we used our own database by collecting various pictures from daily life. The challenges faced in this project were at a par level. The most interesting thing to explore was to learn that every human emotion is different from the rest. For example, person A can show him/her being happy in a different way than person B. There is a wide spectrum to it, and detecting each expression, and categorizing them can be hard. In this experiment, the accuracy rate was satisfactory. The rate of detecting happiness was 100% and that was good, however as the facial expression got more complex, the accuracy rate reduced. At a later stage, the accuracy rate hovered at a rate of 70%, which is neither too bad nor too good. As mentioned earlier, human expressions can vary a lot, and concretely categorizing them takes a lot of effort and intelligence. As future enhancements, the facial images can be flattened before training and be converted to ASCII codes. The ASCII codes can then be mapped to the vector points so that facial expression categorization is comprehensive. We also need to work on image clarity, alignment, illumination, facial rotation. All 20 of these factors can play a vital role in deviating the accuracy and recognition rate. These website can be used in daily life such as health care, automotive industry, emotion analysis in video game testing since it has features such as option to upload varied input formats such as camera input or user input (video / image) which makes this project unique from rest of the others.

11.Future Work

The proposed system is highly useful to the society for different applications where emotion recognition plays a major role. In future work, different algorithm can be implemented to improve recognition accuracy. Robots can also be made to recognize emotion by neurological inspiration. Other modality like speech can be combined along with image for emotion recognition.

12.References

[1] B.-F. Wu and C.-H. Lin, "Adaptive feature mapping for customizing deep learning-based facial expression recognition model," IEEE Access, 2018.

[2] B.-K. Kim, J. Roh, S.-Y. Dong, and S.-Y. Lee, "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition," Journal on Multimodal User Interfaces, vol. 10, no. 2, pp. 173–189, 2016.

[3] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in Computer vision and pattern recognition (CVPR), 2012 IEEE conference on. IEEE, 2012, pp. 3642– 3649.

[4] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 815– 823.

[5] G. Pons and D. Masip, "Multi-task, multi-label and multi-domain learning with residual convolutional networks for emotion recognition," arXiv preprint arXiv:1802.06664, 2018.

[6] K. Liu, M. Zhang, and Z. Pan, "Facial expression recognition with cnn ensemble," in Cyberworlds (CW), 2016 International Conference on. IEEE, 2016, pp. 163–166.