

Real-time Emotion Detection



Group Members –

- 1) Ashish Borate, Roll no - 3
- 2) Atharva Bharekar, Roll no – 21

INDEX

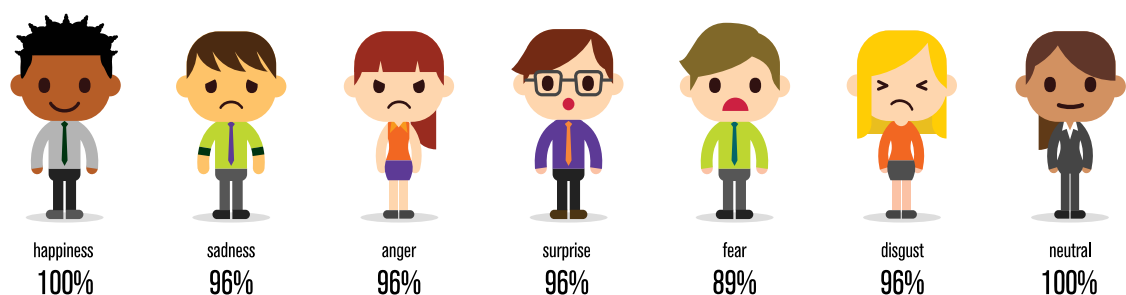
Sr. No.	Particulars	Page No.
1.	Abstract	3
2.	Introduction	3
3.	Motivation	4
4.	Problem Statement	4
5.	Purpose/Objectives & Goals	5
6.	Literature survey	6 -12
7.	Flowchart	13
8.	UML Diagrams	14-17
9.	User Interface	18-19
10.	Hardware/Software Requirements	20
11.	Future Scope	21-22
12.	Acknowledgment and References	23

Abstract –

Humans express their emotions through their facial expressions. They are crucial in interpersonal communication. Every human mood can be documented using this project's real-time emotion recognition system. In this project, various models generated using machine learning and deep learning algorithms are used. To identify human expressions in real time, application software is created that makes use of a few powerful Python packages. This project makes use of a number of libraries, including OpenCV, and Matplotlib. Training those emotions can detect various human emotions. This real-time emotion detection can be used across multiple platforms.

Introduction -

Humans express their emotions through their facial expressions. They are crucial in interpersonal communication. Every human mood can be documented using this project's real-time emotion recognition system. In this project, various models generated using machine learning and deep learning algorithms are used. To identify human expressions in real time, application software is created that makes use of a few powerful Python packages. This project makes use of a number of libraries, including Keras, OpenCV, and Matplotlib. Training those emotions can detect various human emotions. This real-time emotion detection can be used across multiple platforms.



Motivation –

Emotions play an essential role in identifying the mood of a human being. There are generally six raw emotions: happy, sad, anger, fear, surprise, disgust, contempt. It is seen that research work focuses on the four main emotions named happy, sad, angry and neutral. Also, recognising the emotions plays an important role in camera surveillance to capture the suspects; for example, in the case of a feared person, a system raising the alarm can help. Emotion recognition systems can be used as a sub-module of various applications like recommending music and various camera surveillance systems.

Problem Statement –

Develop an emotion detection system capable of real-time analysis of human facial expressions and speech intonation. The system should accurately classify a range of emotions including but not limited to happiness, sadness, anger, surprise, fear, and neutral states. It should process live video input from a webcam and provide instantaneous, multi-class emotion predictions. The project should employ state-of-the-art machine learning techniques for efficient real-time processing while maintaining high accuracy, and it should be user-friendly, with a clear and intuitive interface that visually displays the detected emotions in real-time. It should operate in real-time, providing low-latency emotion predictions and visualizing the detected emotions on the video feed in a user-friendly manner.

Purpose/Objectives & Goals –

- Facial expression recognition system has turn into a most emphasizing research area since it plays a most important part in human-computer-interaction.
- The face can articulate emotion sooner than people verbalize or even understand their posture.
- To identify the key facial features that are associated with different emotions.
- Modelling the facial emotion and its intensity.
- To recognize emotional state / expression using mouth information.
- Tool for building systems which can identify, interpret, process, and simulate human emotions.
- To optimize the algorithm for real-time performance.
- To evaluate the accuracy and reliability of the system on a variety of datasets.
- The final deliverable will be a functional real-time emotion detection system that accurately detects and classifies emotions from live video feeds. This will include the source code, trained machine learning models, and a user manual.

Literature survey

The process of emotion recognition involves the processing images and detecting the face then extracting the facial feature. Facial Expression Recognition consists of three main steps. In first step face image is acquired and detect the face region from the images and pre-processed the input image to obtain image that have a normalized size or intensity. Next is expression features are extracted from the observed facial image or image sequence. Then extracted features are given to the classifier and classifier provides the recognized expression as output

A. Face Detection and Pre-processing

The face detection is the process of extracting the face region from the background. It means to determine the position of the face in the image. This step is required because images having a different scale. Input image having a complex backgrounds and variety of lightning conditions can be also quite confusing in tracking. Face expression recognition tends to fail if the test image has a different lighting condition than that of the training images. For that facial point can be detected inaccurately for that pre-processing step is required.

B. Feature Extraction And Classification

Selecting a set of feature points which represent the important characteristics of the human face. After the face has been located in the image, it can be analysed in terms of facial features. The features measure the certain parts of the face such as eyebrows or mouth corners. Various methods exist which can extract feature for expression based on motion of the feature such Active Appearance Model [4] which is statistical model of shape and gray scale information. The Features describe the change in face texture when particular action is performed such as wrinkles, bulges, forefront, regions surrounding the mouth and eyes. Image filters are used, applied to either the whole-face or specific regions in a face image to extract a feature vector. Principal Component Analysis [1], Local Binary Pattern

(LBP)[2],Fisher's Linear Discriminator[3] based approaches are the main categories of the approaches available. After the set of features are extracted from the face region are used in classification stage. The set of features are used to describe the facial expression. Classification requires supervised training, so the training set should consist of labelled data. Once the classifier is trained, it can recognize input images by assigning them a particular class label. The most commonly used facial expressions classification is done both in terms of Action Units, proposed in Facial Action Coding System (FACS) [5] and in terms of six universal emotions: happy, sad, anger, surprise, disgust and fear.

Random Forest

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning**, which is a process of *combining multiple classifiers to solve a complex problem and to improve the performance of the model*.

As the name suggests, *"Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset."* Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a class of deep neural networks that have proven to be very effective in image classification tasks. Here's a basic overview of how CNNs work for classification:

1. **Convolutional Layers:** CNNs use convolutional layers to automatically and adaptively learn spatial hierarchies of features from input images. These layers consist of filters (also called kernels) that slide over the input image to detect patterns like edges, textures, and more complex structures.
2. **Pooling Layers:** After each convolutional layer, pooling layers are often added to reduce the spatial dimensions of the representation and the number of parameters and computations in the network. Max pooling is a common type of pooling layer that takes the maximum value from a group of neighboring pixels.
3. **Flattening:** The output from the convolutional and pooling layers is flattened into a one-dimensional vector. This vector is then fed into one or more fully connected layers.
4. **Fully Connected Layers:** These layers are traditional neural network layers where each neuron is connected to every neuron in the previous and next layers. The final fully connected layer produces the output for classification.
5. **Activation Functions:** Rectified Linear Unit (ReLU) is commonly used as the activation function in convolutional and fully connected layers.
6. **Softmax Activation:** In the output layer, a softmax activation function is often used for multi-class classification problems. It converts the raw output scores into probabilities, assigning a probability to each class.
7. **Loss Function:** The categorical crossentropy loss is commonly used for classification problems with softmax activation.

Haar Cascade

A Haar Cascade is a machine learning object detection method used to identify objects in images or video. It's based on the Haar wavelet technique, and the most popular application is face detection. The concept was introduced by Viola and Jones in their 2001 paper, "Rapid Object Detection using a Boosted Cascade of Simple Features."

Here's a basic overview of how Haar cascades work:

1. **Haar Features:** Haar features are simple, rectangular filters that can be applied to an image. These features are used to capture different characteristics of objects, such as edges, corners, and textures.
2. **Integral Image:** To efficiently compute Haar features, an integral image is used. The integral image allows the rapid calculation of the sum of pixel values in any rectangular region of the image.
3. **Training:** The Haar Cascade is trained using a machine learning algorithm, often AdaBoost. During training, the algorithm selects the most informative Haar features and assigns weights to them, creating a strong classifier.
4. **Cascade Classifiers:** The trained model is organized into multiple stages, forming a cascade. Each stage consists of a subset of the selected Haar features. The cascade structure allows the classifier to quickly reject regions of the image that are unlikely to contain the object of interest, leading to faster processing.
5. **Sliding Window:** The Haar Cascade uses a sliding window approach to scan the image at multiple scales and positions. At each step, the cascade evaluates the presence of the object based on the selected Haar features.

6. **Thresholding:** Each stage in the cascade has a threshold, and if a region fails to meet the criteria at any stage, it is quickly rejected. This helps in reducing false positives.

Haar cascades are computationally efficient and have been widely used for real-time object detection, especially in applications like face detection in images or video streams. OpenCV, a popular computer vision library, provides pre-trained Haar Cascade classifiers for various objects, making it easy to implement object detection in your applications.

Adaboost

AdaBoost, short for Adaptive Boosting, is an ensemble machine learning algorithm that can be used in a wide variety of classification and regression tasks. It is a supervised learning algorithm that is used to classify data by combining multiple weak or base learners (e.g., decision trees) into a strong learner. AdaBoost works by weighting the instances in the training dataset based on the accuracy of previous classifications

AdaBoost Algorithm

Freund and Schapire first presented boosting as an ensemble modelling approach in 1997. Boosting has now become a popular strategy for dealing with binary classification issues. These algorithms boost prediction power by transforming a large number of weak learners into strong learners.

Boosting algorithms work on the idea of first building a model on the training dataset and then building a second model to correct the faults in the first model. This technique is repeated until the mistakes are reduced and the dataset is accurately predicted. Boosting algorithms function similarly in that they combine numerous models (weak learners) to produce the final result (strong learners).

Project Scope -

The seven basic emotions into which human facial expressions can be arranged are happy, sad, surprise, fear, anger, disgust, and neutral. Our facial expressions are caused by the enhancement of certain facial muscles. In this project, we will design a deep learning neural network to allow machines to make references about our emotional states.

- Real-time video capture: The system will capture live video feeds from a camera source.
- Facial detection: It will detect faces within the video frames using computer vision techniques.
- Facial feature extraction: The system will extract key facial features such as eye movements, mouth shape, and eyebrow positions.
- Emotion classification: Using machine learning algorithms, the extracted facial features will be used to classify emotions such as happiness, sadness, anger, surprise, etc.
- Real-time visualization: The detected emotions will be visually displayed in real time, possibly as overlays on the video feed.
- User interface: A user-friendly interface will allow users to start, stop, and interact with the emotion detection system.

Limitations –

1. Accuracy of Emotion Detection:

The accuracy of emotion detection is dependent on the quality of the training data and the chosen machine learning algorithms. Achieving high accuracy may require a large and diverse dataset and complex model architectures.

2. Limited Expressions:

The system might struggle with recognizing subtle or nuanced emotions, as well as emotions that are expressed in atypical ways. Some individuals might express emotions differently due to cultural factors or personal habits.

3. Environmental Factors:

The system's accuracy could be affected by environmental conditions such as lighting, background, and camera quality. Poor lighting or cluttered backgrounds might lead to false positives or negatives.

4. Gender and Age Bias:

If the training data is not balanced across genders, ages, and ethnicities, the system might be biased in its recognition of emotions across different demographics.

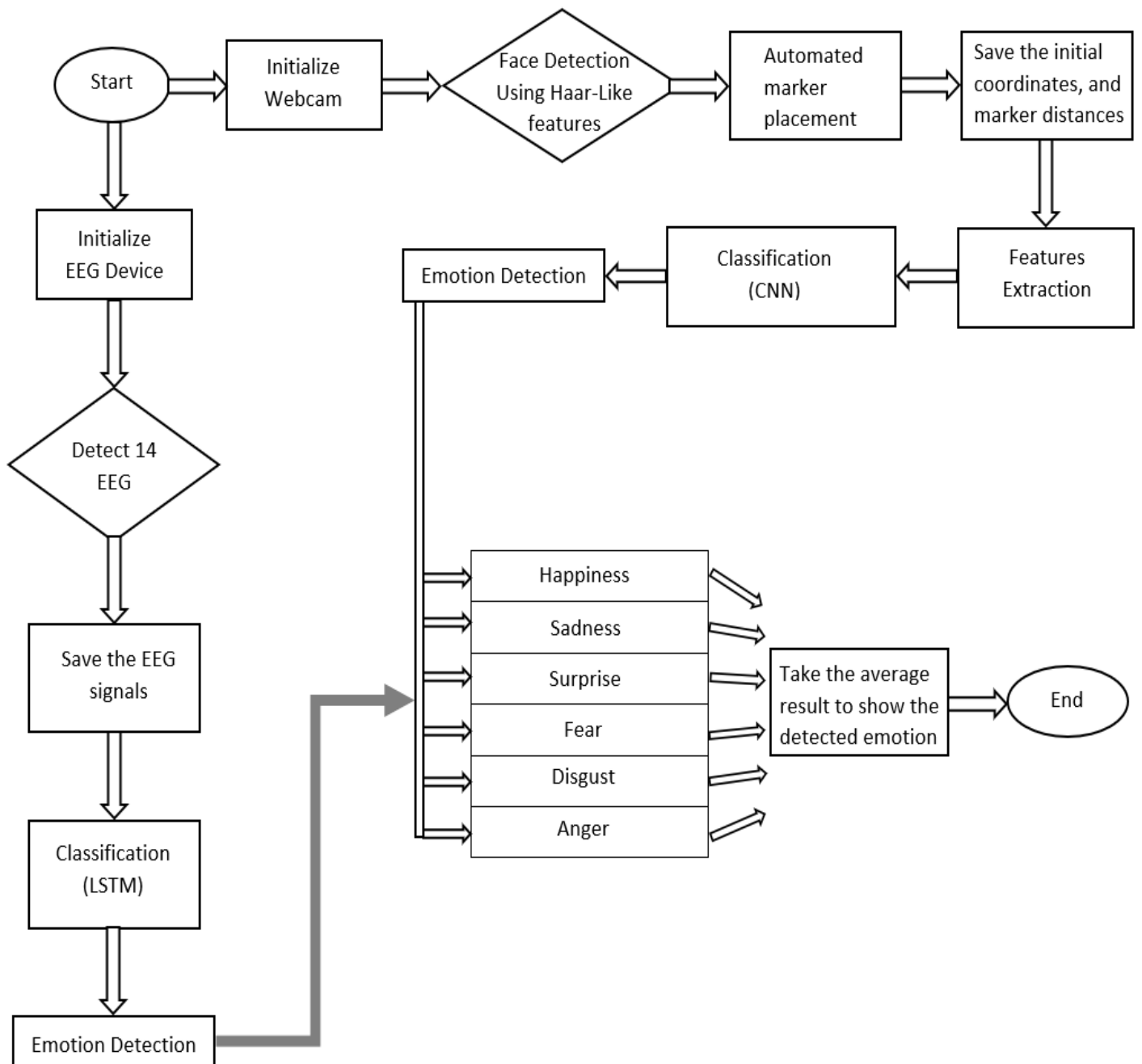
5. Continuous Learning:

The model might require periodic updates and retraining to adapt to changing facial expression trends and improve over time.

6. Computational Complexity:

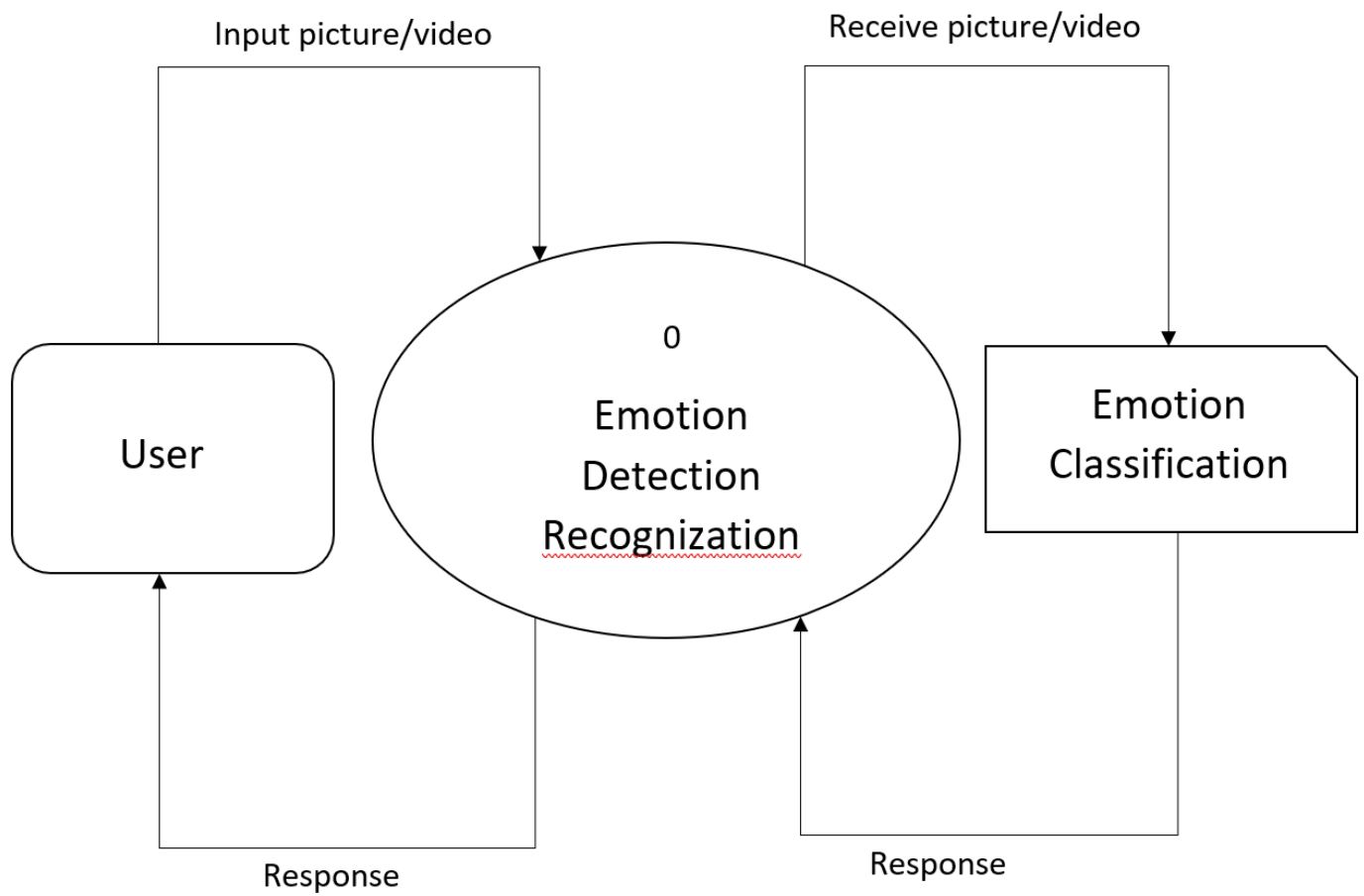
Real-time processing of video and facial analysis can be computationally intensive, potentially limiting its performance on certain devices.

Flowchart -

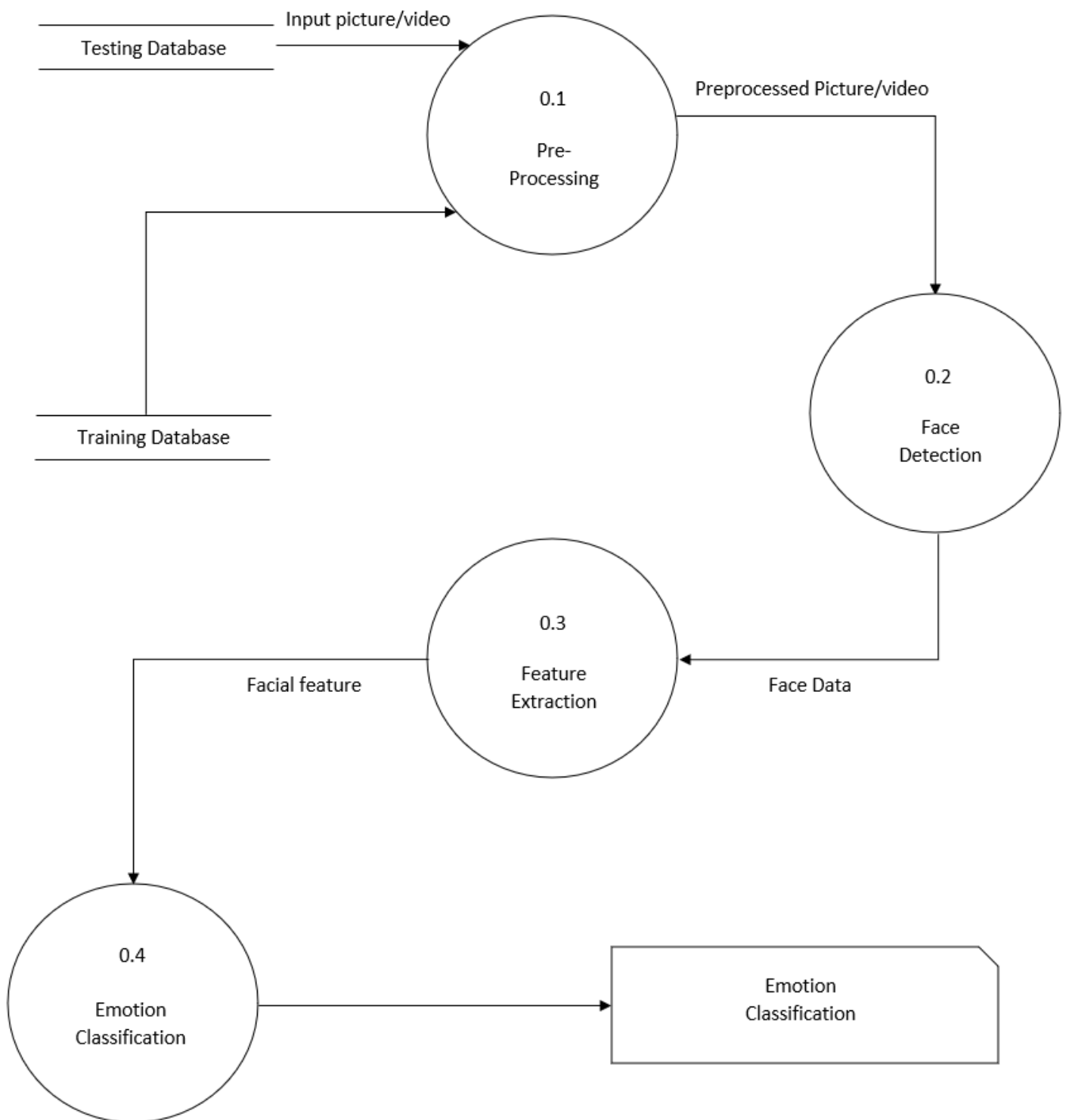


DFD

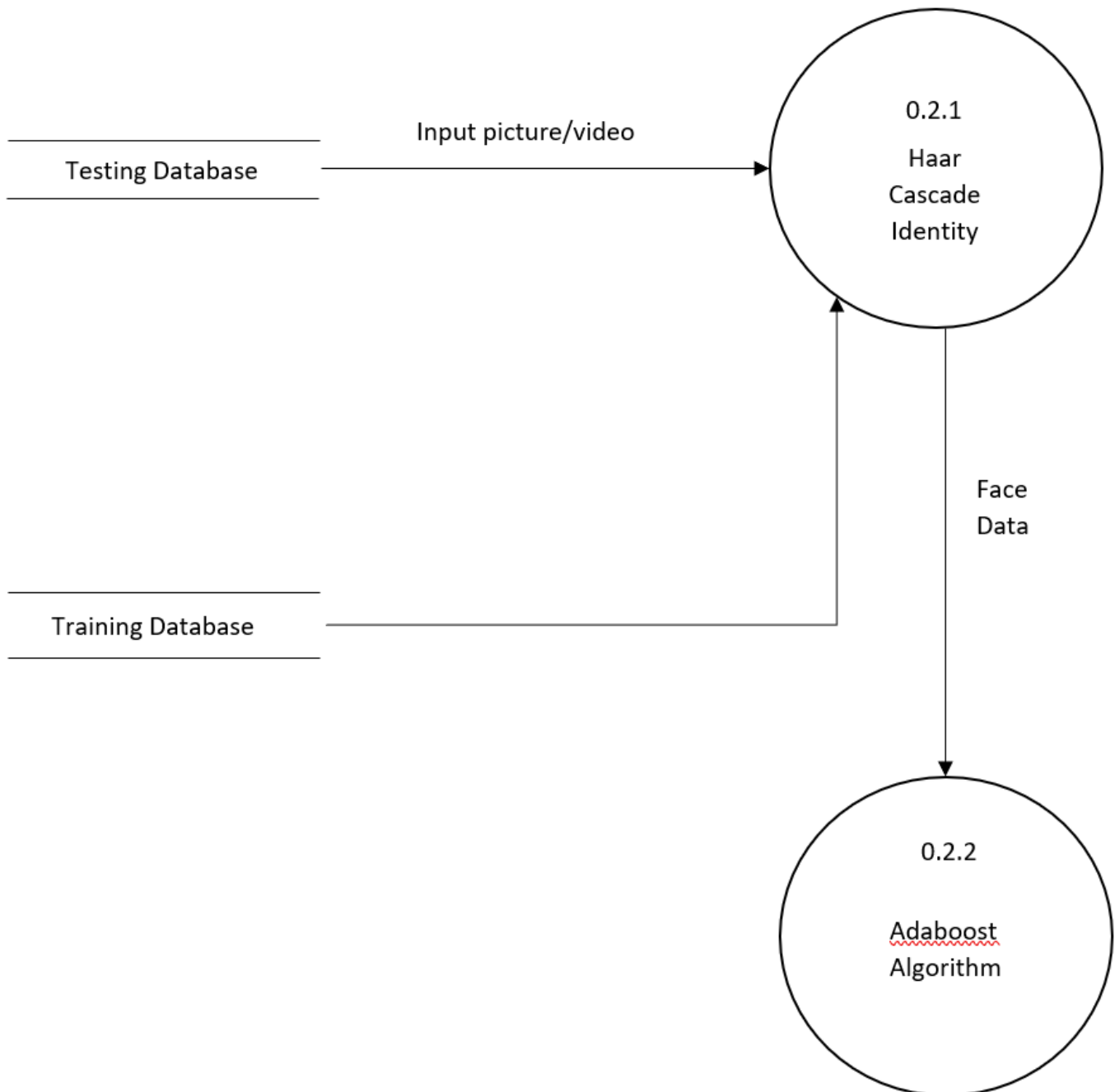
Context level or Zero level



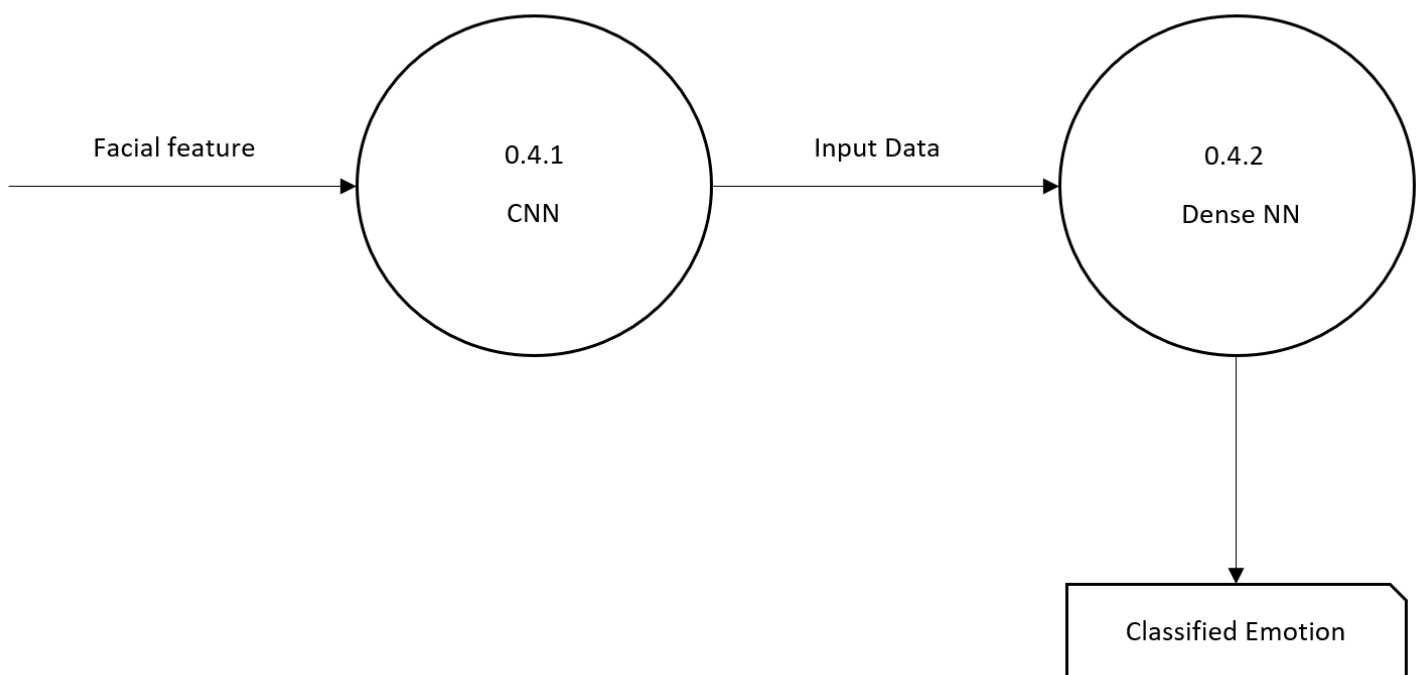
Level 1



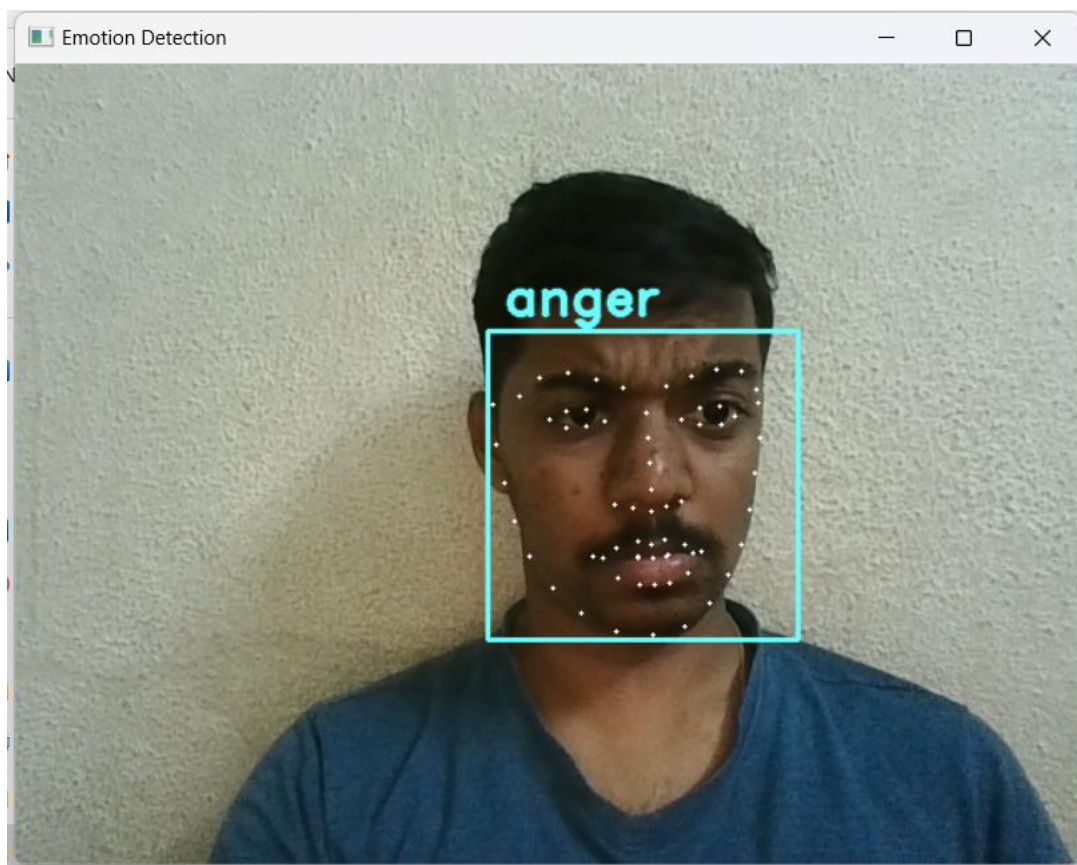
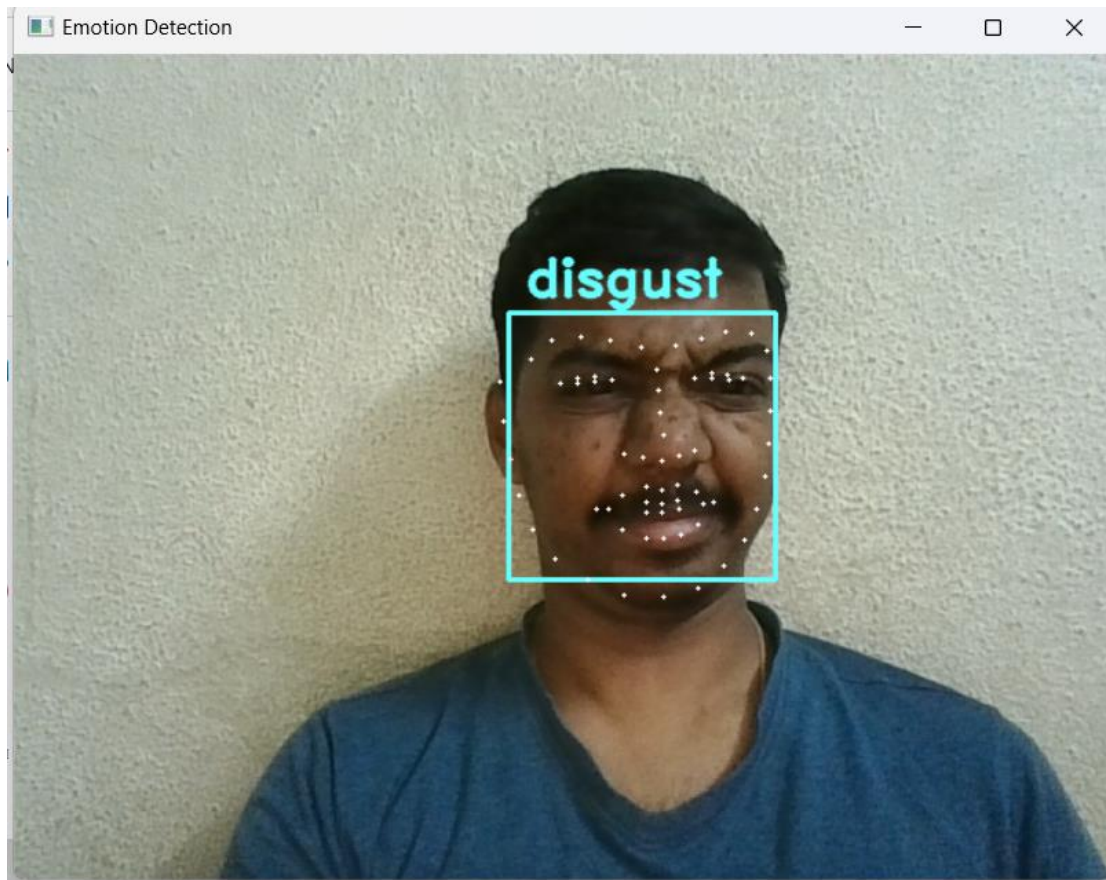
Level 2

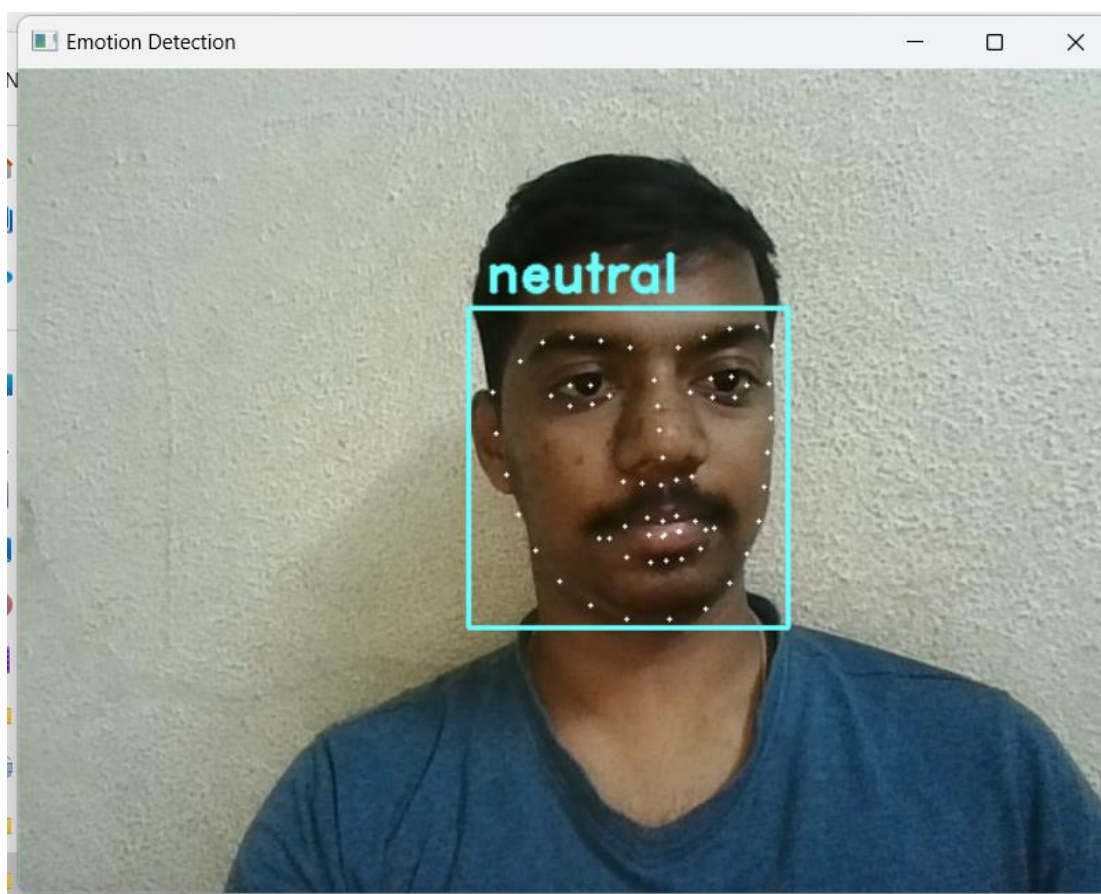
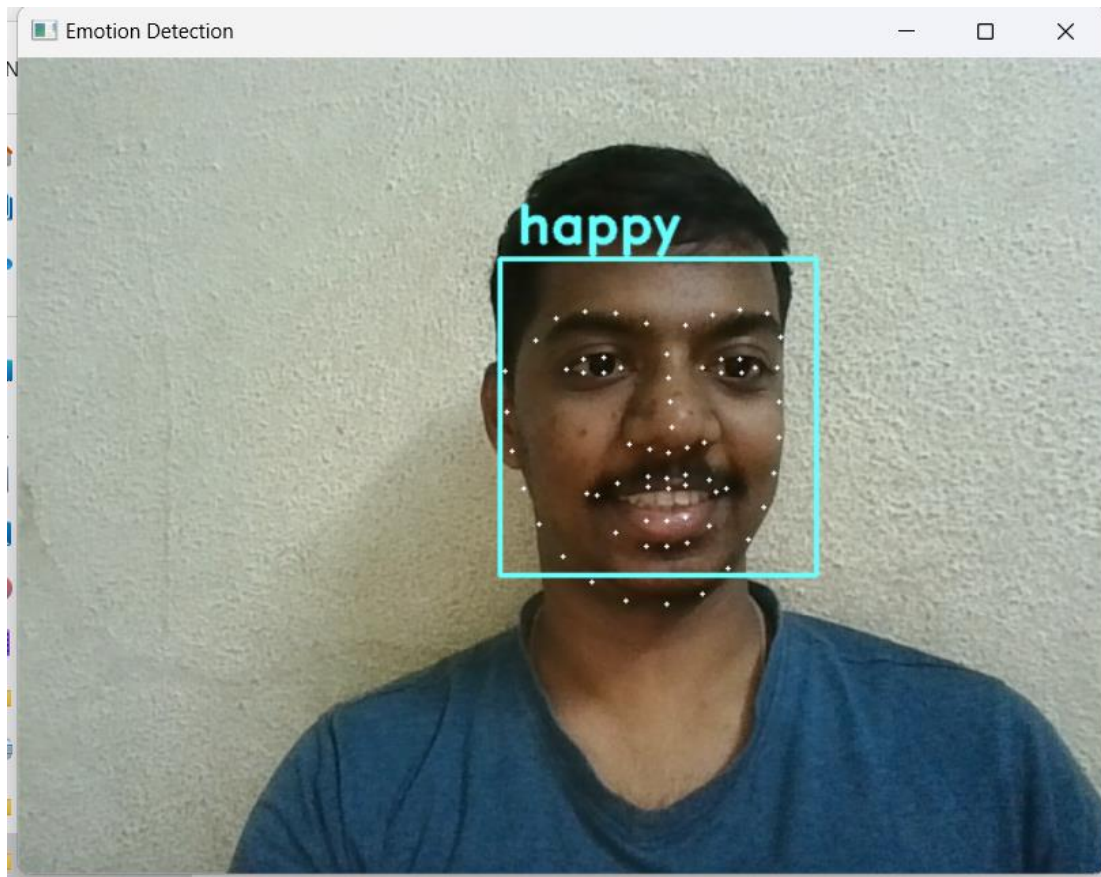


Emotion Classification-



User-Interface





HARDWARE REQUIREMENTS:

1. Processor: Intel i3
2. RAM: (recommended) 4GB or higher
3. HD-CAM

SOFTWARE REQUIREMENTS:

1. Operating System: Windows 10
2. Technologies: Python-3.9.5, OpenCV, Modules- Pandas, Matplotlib
3. Development tool: Visual Studio Code, cmd

Future Scope

1. **Human-Computer Interaction:** Real-time emotion detection can enhance human-computer interaction, making interfaces more intuitive and responsive. Applications could include virtual assistants, customer service bots, and gaming systems that adapt to players' emotions.
2. **Mental Health Support:** Emotion detection systems can be integrated into mental health applications to provide real-time feedback to users and therapists. These systems can help individuals monitor their emotional states and provide timely interventions and suggestions for managing stress and anxiety.
3. **Market Research and Advertising:** Emotion analysis can be applied in market research to gauge consumers' emotional responses to advertisements and products. Advertisers can use this data to create more emotionally resonant and effective campaigns.
4. **Security and Surveillance:** Emotion detection systems can be integrated into security systems to identify suspicious behavior or individuals displaying signs of distress, potentially preventing crimes or accidents in public places.
5. **Autonomous Vehicles:** Real-time emotion detection can be used in autonomous vehicles to monitor the emotional state of passengers. This information can be utilized to create a more comfortable and secure driving experience, adapting the vehicle's environment based on passengers' emotions.
6. **Healthcare:** Emotion detection technology can be applied in healthcare settings for patient monitoring. For example, it can be used to monitor the

emotional well-being of patients in hospitals or elderly individuals in assisted living facilities.

7. **Criminal Justice:** Emotion analysis can be used in legal contexts, such as detecting deception or assessing the emotional state of individuals during police interrogations or court proceedings.

Acknowledgement -

<https://www.javatpoint.com>

<https://www.geeksforgeeks.org>

<https://www.mygreatlearning.com/blog/adaboost-algorithm>

References -

1. Principal Component Analysis Book by I. T. Jolliffe
2. Local Binary Patterns: New Variants and Applications (Studies in Computational Intelligence Book 506) 2014 Edition by Sheryl Brahnam (Editor), Lakhmi.C.Jain (Editor), Loris Nanni (Editor), Alessandra Lumini (Editor)
3. The Fisher's Linear Discriminant by Iuliana F. Iatan (Conference Paper)
Part of the Advances in Intelligent and Soft Computing book series (AINSC, volume 77)
4. Visual Computing for Medicine (Second Edition), 2014 by Bernhard Preim, Charl Botha.
5. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS) (Series in Affective Science) by Paul Ekman (Editor), Erika L. Rosenberg (Editor)