System Development Project

Human Emotion Detection

By

Mazharul Islam
Roll: 1807102

&

Md Shanjid Arefin
Roll: 1807104

Supervisor:

Md. Ahsan Habib
Lecturer
Department of Computer Science and Engineering
Khulna University of Engineering & Technology

_____
Signature

Department of Computer Science and Engineering
Khulna University of Engineering & Technology
Khulna 9203, Bangladesh

Acknowledgment

With the blessings and limitless mercy of Almighty, we are able to do this. We express our heartiest gratitude to Almighty Allah for this.

A word of special thanks must go to our highly esteemed teacher and this project's supervisor, Md. Ahsan Habib, Lecturer, Department of Computer Science and Engineering, Khulna University of Engineering & Technology (KUET), for his excellent advices, guidance and right directions without which our project may not have reached a state it is in now.

We would like to thank our seniors who inspired us to implement different Ideas throughout this project. We would also like to thank our friends for their association also.

Any constructive comments, suggestions, criticism from teachers as well as seniors will be highly appreciated and gratefully acknowledged.

Abstract

According to the American Psychological Association (APA), emotion is defined as “a complex reaction pattern, involving experiential, behavioral and physiological elements.” Emotion detection is the task of recognizing a person’s emotional state — for example, anger, confusion or deceit across both voice and nonvoice channels. The process of human emotion detection is a complex process. Therefore, an intelligent system is required to detect emotions. In this project, a mobile application has been developed on android platform and a webapp using python flask library which can access device’s camera to detect human faces and recognize the emotions. The performance of the emotion detection of humans using frontal-face depends on classification algorithms. This process requires high level of knowledge about computer vision, convolutional neural network and working process of the algorithm.

In this system, we have used the android studio emulator to build an android application and a webapp to detect human real-time emotions using the camera of the devices. The proposed model was trained in the TensorFlow library with deep neural network would detect different seven types of emotions: happiness, anger, disgust, fear, neutral, sad and surprise. Thus, creating ways for many applications in the future for industrial purpose and many more real time applications like humanoid robot, video gaming, medical diagnosis etc.

Contents **Pages**

1. Introduction 06
2. Objectives 07
3. Previous Work 07-09
4. System Requirements 09-11
5. System Overview 11-12
6. Implementation 12-15
7. Current Prototype 15-16
8. Result Analysis 17-29
9. Limitations 19
10. Future Work 20
11. Conclusion 20
12. References 21

List of Figures Pages

Figure 1. Suicide rate of Bangladesh. 06

Figure 2. Convolutional Neural Network. 10

Figure 3. Block diagram of Emotion detection system in android application. 11

Figure 4. Block diagram of Emotion detection system in web application. 12

Figure 5. Workflow of CNN using TensorFlow. 13

Figure 6. Examples of dataset. 13

Figure 7. Illustration of Accuracy and loss plot. 13

Figure 8. Emotion Detection in live stream and still picture in android application. 14

Figure 9. Emotion Detection in live stream and still picture in webapp. 15

Figure 10. Current prototype of the system (android application) 16

Figure 11. Current prototype of the system (Web application). 16

Figure 12. Emotion detection with real time photo and video. 17

Figure 13. Emotion detection of random image form dataset 18

1. Introduction

According to a survey of World Bank suicidal rate is increasing in Bangladesh as well as in the whole world in every year [1]. Most of them are male and teenagers. At this age range people are so much emotional and their emotions fluctuate very quickly about anything and most of the time parents are not aware of these.

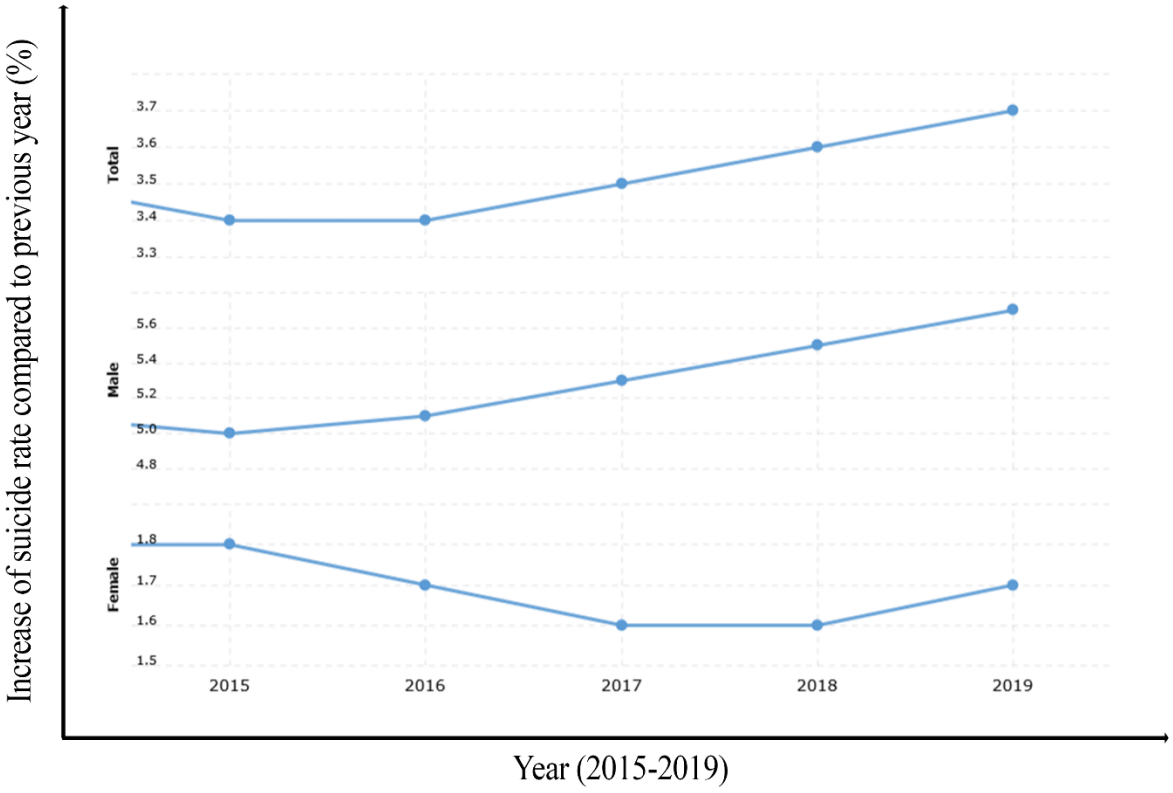

Figure 1. Suicide rate of Bangladesh

In this world of technology artificial intelligence is revolutionizing and humanoid robot is a great example. Humanoid robot can act like a human and taking care of the emotions of the audience and can improve human-robot interaction.

Emotions play an important role in human interactions as they let people articulate themselves without words. Emotions include cognitive appraisal, bodily language, action tendencies, expressions, and feelings [2]. People would not be able to get along with each other without emotions.

Emotion recognition involves considerable information including facial expressions, body language, pitch and tone of voice, and semantics. Facial expressions are crucial because they convey considerable information that can be widely used in various applications in different fields. Furthermore, facial expression can convey the same information across different cultures and countries.

It remains challenging for computers and robots to classify facial expressions under different light conditions, poses, and backgrounds and across people of different ages, genders, and ethnicities. In one study [3], the Facial Action Coding System (FACS) was proposed for quantifying human facial movement. This system is a practical solution for detecting facial movement within the field of behavioral science. Essentially, facial expressions are identified according to several muscle movements. Based on the movements of these facial muscles, the FACS decomposes the facial expressions into their component actions.

2. Objectives

- ✓ Our target is to help to reduce suicide rate of our country as well as the whole world by developing a monitoring mobile app of human emotions.
- ✓ The main priority of the system is to detect human emotions by using mobile app or webapp using a trained model with deep neural network: CNN and TensorFlow library of python.
- ✓ We developed a system that could detect human emotions.
- ✓ The system is a real time mobile app and a webapp which can detect seven types of emotions from pictures and real-time video.

3. Previous Work

In the last decade, there are several works related to emotion recognition, facial expression recognition, deep neural network and transfer learning. And we can discuss them below.

Tzuu-Hseng S. Li and Ting-Nan Tsai [4] developed an emotion detection robot which can detect human emotions. They analyzed static images for facial expression recognition. However facial expressions are produced by the contraction and relaxation of some facial muscles.
They consider six different emotions using static photo for facial emotion detection, but it is not feasible way, because human facial expression can fluctuate rapidly.
Building a model which can take photos of a human continuously and detect emotions with the best frame can give the accurate data.

Jose Maria Garcia-Garcia, Victor M. R. Penichet and Maria D. Lozano [5] developed emotion detection technology for the corporate sector. They developed a survey gathering emotional information from the user of a system on their voice. They focus on that When a person starts talking, they generate information in two different channels: primary and secondary [6]. The primary channel is linked to the syntactic-semantic part of the talking (what the person is literally saying), while the secondary channel is linked to paralinguistic information of the speaker (tone, emotional state, and gestures). E.g., someone says “That’s so funny” (primary channel) with a

serious tone (second channel). By looking at the information of the primary channel the message that the speaker thinks that something is funny, and by looking at the information received by the second channel, we get to know that the real meaning.

Yisi Liu, Olga Sourina, and Minh Khoa Nguyen [7] developed a system on human emotion detection concentrating on recognition of "inner" emotions from electroencephalogram (EEG) signals. They propose real-time fractal dimension -based algorithm of quantification of basic emotions using Arousal-Valence emotion model. Two emotion induction experiments with music stimuli and sound stimuli from International Affective Digitized Sounds (IADS) database were proposed and implemented. Finally, the real-time algorithm was proposed, implemented and tested to recognize six emotions such as fear, frustrated, sad, happy, pleasant and satisfied. Real-time applications were proposed and implemented in 3D virtual environments. The user emotions are recognized and visualized in real time on his/her avatar adding one more so-called "emotion dimension" to human computer interfaces. An EEG-enabled music therapy site was proposed and implemented. The music played to the patients helps them deal with problems such as pain and depression. An EEG-based web-enable music player which can display the music according to the user's current emotion states was designed and implemented.

Egger Maria, Ley Matthias, Hanke Sten [8] worked on research and their aim was to give an overview of methods to recognize emotions and to compare their applicability based on existing studies. They work on smart wearables which provide contact with the skin and physiological parameters such as electrodermal activity and heart related signals can be recorded unobtrusively also during dynamical tasks. Looking forward, heart-related parameters might be an option to measure emotions accurately and unobtrusive with the help of smart wearables. They achieved of 88.86% accurate data based on the real emotions.

Robert Horlings [9] worked on Human emotion detection based on brain activity, measured by EEG signals. They classified the received EEG signals into 5 classed on two emotional dimensions, valence and arousal. That system designed using prior knowledge EEG signals in practice. For that purpose, they gathered a dataset with EEG signals from people that were emotionally stimulated by pictures. That method enabled us to teach our system the relationship between the characteristics of the brain activity and the emotion. They found that the EEG signals contained enough information to separate five different classes on both the valence and arousal dimension. However, using a 3-fold cross validation method for training and testing, we reached classification rates of 32% for recognizing the valence dimension on from EEG signals and 37% for the arousal dimension when. Much better classification rates were achieved when using only the extreme values on both dimensions, the rates were 71% and 81%.

Table 3.1: Summary table for the previous works on Emotion Detection

Authors	Model	Input	Detected Emotions
Li and Tsai [4]	CNN	Facial Expression Recognition	6
Garcia et al. [5]	GNN	Speech and Voice	7
Liu et al. [7]	VAD	Electroencephalogram (EEG) signals	7
Maria et al. [8]	VAD	Smart wearables and Electrodermal activity and heart related signals	6
Robert Horlings [9]	VAD	Brain activity measured by (EEG) signals	7

4. System Requirements

The necessary requirements for the proposed system are CNN Model, LSTM and knowledge about combination of CNN and LSTM. These required elements are given below:

4.1. CNN Model

Deep Learning has proved to be a very powerful tool because of its ability to handle large amounts of data [10]. The interest to use hidden layers has surpassed traditional techniques, especially in pattern recognition. One of the most popular deep neural networks is Convolutional Neural Networks in deep learning. Since the 1950s, the early days of AI, researchers have struggled to make a system that can understand visual data.

The convolutional layer encompasses a set of kernels for determining a tensor of feature maps. [11] These kernels convolve an entire input using "stride(s)" so that the dimensions of an output volume become integers. The dimensions of an input volume decrease after the convolutional layer is used to execute the striding process. Therefore, zero padding is required to pad an input volume with zeros and maintain the dimensions of an input volume with low-level features.

$$F(i,j) = (I * K)(i,j) = \sum\sum I(i+m, j+n)\, K(m,n) \tag{1}$$

where I refer to the input matrix, K denotes a 2D filter of size m × n, and F represents the output of a 2D feature map. The operation of the convolutional layer is denoted by I*K. To increase nonlinearity in feature maps, the rectified linear unit (ReLU) layer is used . ReLU computes activation by keeping the threshold input at zero. It is mathematically expressed as follows:

$$f(x) = \max(0, x) \tag{2}$$

The pooling layer performs a down sampling of a given input dimension to reduce the number of parameters. Max pooling is the most common method, which produces the maximum value in an input region. The FC layer is used as a that makes a decision on the basis of features obtained from the convolutional and pooling layers.

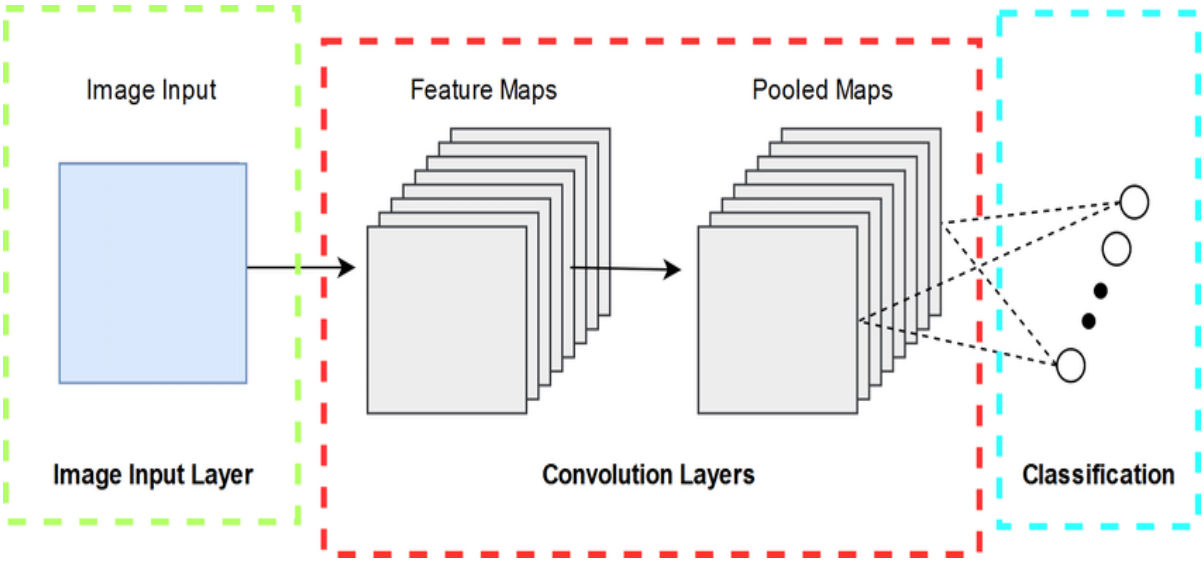

Figure 2. Convolutional Neural Network [12]

4.2. Dataset

To train the proposed model a dataset was collected from Kaggle contributed by Manas Sambare. The famous dataset is known as FER-2013. The data consists of 48 × 48 Pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. It consists of seven classes: angry, disgust, fear, happy, neutral, sad and surprised. The training set consists of 28,709 examples and the public test set consists of 3,589 examples.

4.3 Haar-cascade frontal face

The Haar-Cascade Face Detection Algorithm is a sliding-window type of algorithm that detects objects based upon its features like size and location of certain facial features: nose bridge, mouth line and eyes. Eye region being darker than upper-check region, nose bridge region being brighter than eye region.

5. System Overview

The system has two parts with each part detecting emotions in its own way. Firstly, the trained model was deployed in an android application and then it was deployed in a web application.

5.1. Android Application

This application has 5 working layers. In first layer we open the application and there appear 3 different options: Live camera, Image capture and local drives. User can choice any of these options. After getting images from the options our model can detect emotions. The block diagram of the system is illustrated in following Figure 5.

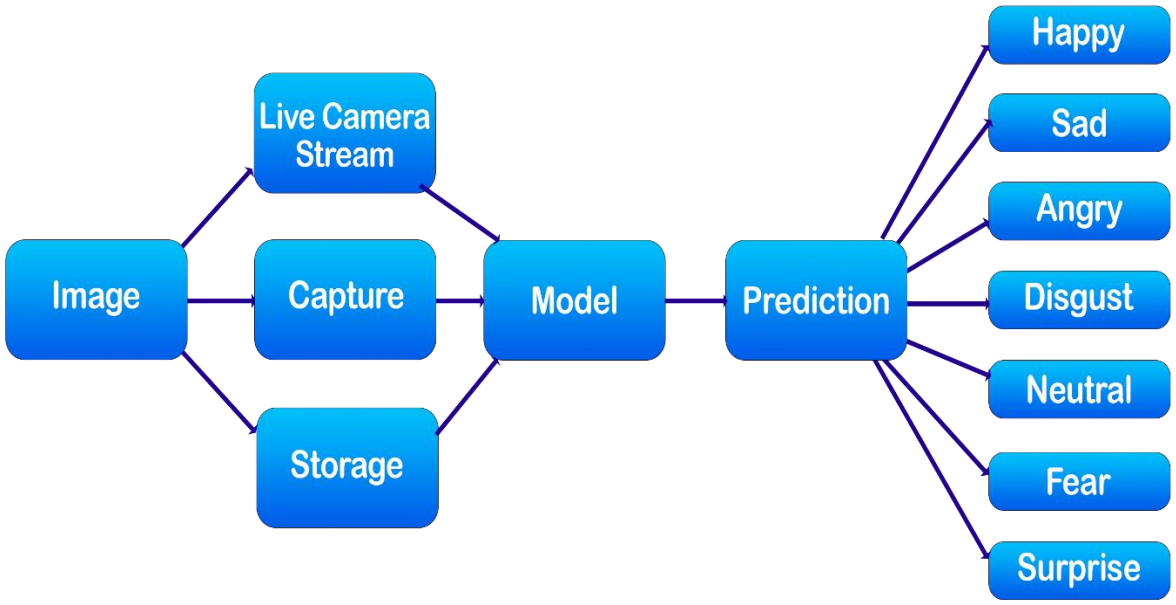

Figure 3. Block diagram of Emotion detection system in android application.

5.2. Web Application

A web application was developed where it can detect emotions from live streaming video and also from the pictures. Figure. 6 shows us the system overview.

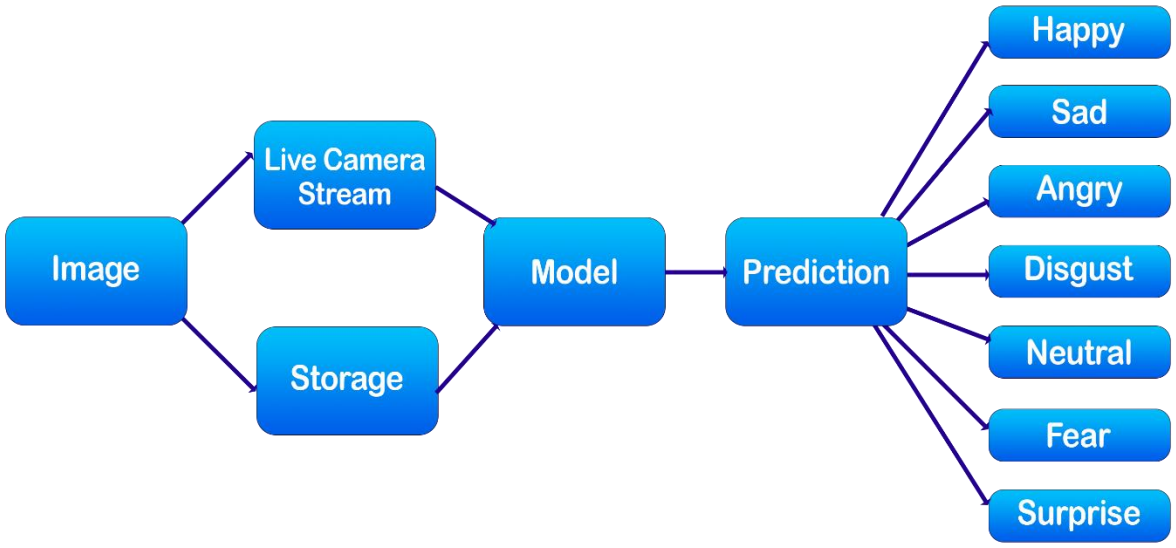

Figure 4. Block diagram of Emotion detection system in web application.

6. Implementation

The developed system has two parts. Firstly, training a model using CNN and LSTM and secondly deploy the model in an android mobile application and a webapp.

6.1. Training a model using CNN model

The CPU has a pre-installed "TensorFlow" library in the python module. We used the CNN (Convolutional Neural Network). First, CNN is used to capture appearance features because it provides state-of-the-art performance for several vision tasks. Fig. 6 shows the network structure. The inputs are the cropped region of interest of the image, which is also the region of the detected face. The cropped region is converted to gray scale and resized to 128 × 128 pixels. Color information is considered less relevant in facial expressions; therefore, it is not necessary to use RGB images. To reduce the memory usage, grayscale images are used in this study. The width and length of the input are 128 pixels, which is large enough for facial expressions because the face region in a 640 × 480 frame is approximately 128 × 128 pixels

These images are pre-trained and provides results when the real time image is provided in the model. For this a high-speed processor was required for faster output. Here is the workflow CNN using TensorFlow library.

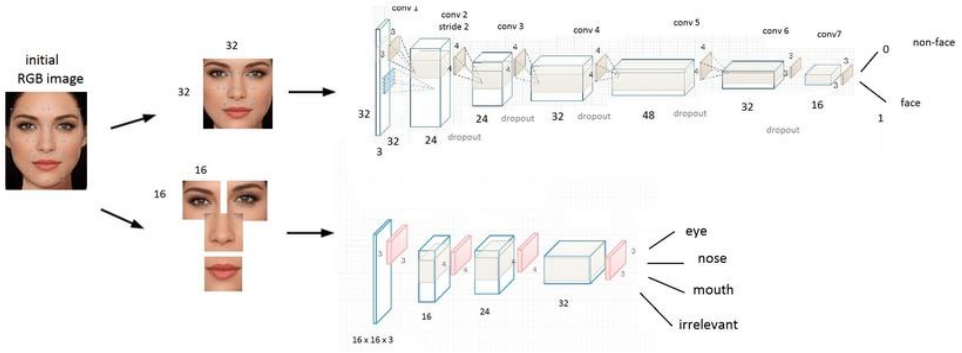

Figure 5. Workflow CNN using TensorFlow .

The input image has size of 128 × 128 pixels, and it is vectorized to a size of 16384 × 1

Dataset used which is contributed by Manas Sambare named “FER-2013” contains seven classes of emotions: angry, disgust, happy, fear, neutral, sad, surprised. Some examples of the dataset given below:

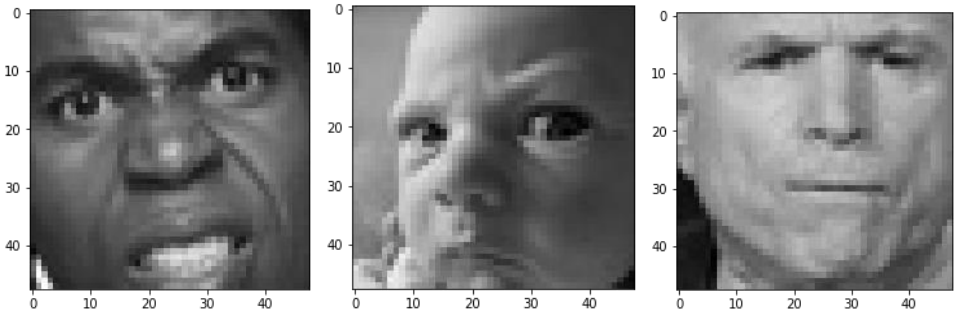

Figure 6. Examples of dataset [15]

Using CNN, a great training accuracy and a decent testing accuracy has been resulted and the result is fair enough which can detect actual emotions almost all the time.

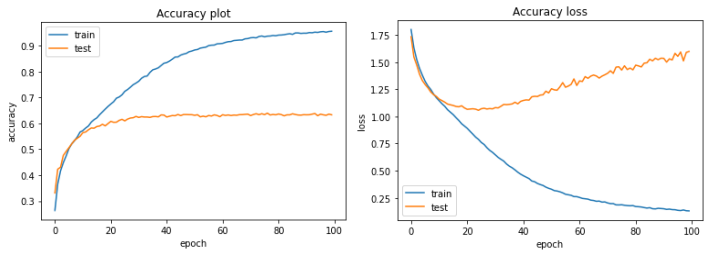

Figure 7. Illustration of Accuracy and loss plot

6.2. Android Mobile Application

Android studio is used to build the application. Android Studio is the official Integrated Development Environment (IDE) for Android app development, based on IntelliJ IDEA. Java and kotlin programing language is used to implement the android project.AT First the proposed model was converted into tf-lite file to embed it into the applications. Using OpenCV library which allow us to perform image processing and computer vision tasks, we extract data from the image into matrices to calculate it via our model. Haar-cascade frontal_face.xml file helps to detect human face from still pictures or videos. By this xml file detection of human face from pictures id done and OpenCV library processes it for the trained model.

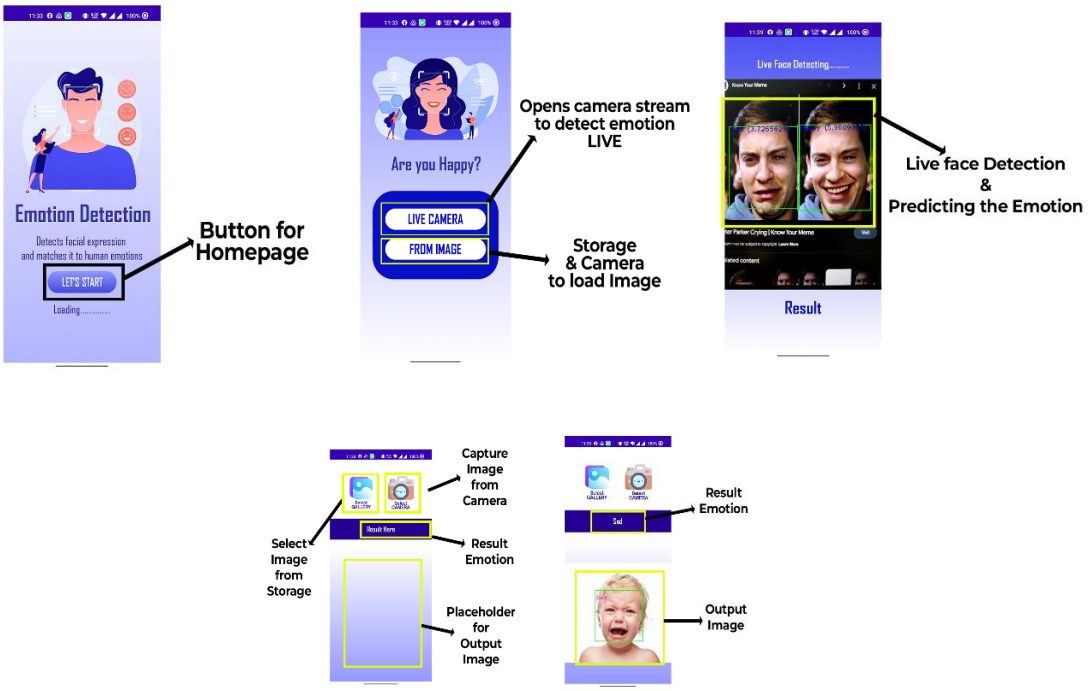

Figure 8: Emotion Detection in live stream and still picture in android application.

After that the converted tf-lite file of the trained model detects the emotion of humans and the result showed on the application interface by leveling on the live stream and still images. The target area of the detection work was separated by axis and a square was drawn to that area. Detected result was leveled at the age of that square with a much visible font color. The system is designed for detect multiple faces at a single time with the emotions of those faces.

6.3. WebApp using Python Flask library

In this web application we use flask library of Python. Flask is a web framework, it's a Python module that lets us develop web applications easily. It has a small and easy-to-extend core: it's a microframework that doesn't include an ORM (Object Relational Manager) or such features. It does have many cool features like URL routing, template engine.

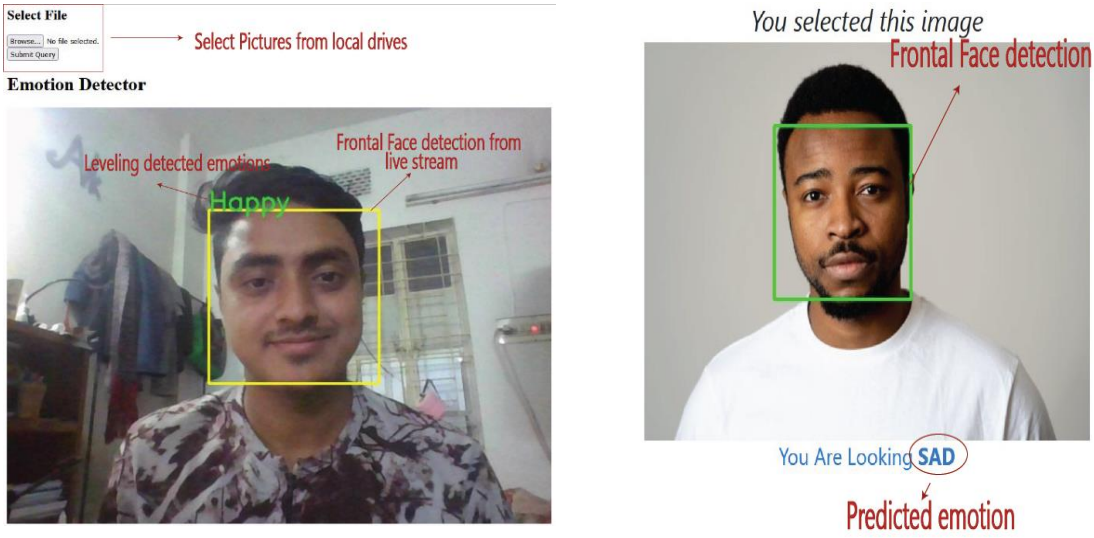

Figure 9. Emotion Detection in live stream and still picture in webapp

Here we also use haarcascade_frontal-face.xml file to detect human frontal face from human body and mark the crroped use OpenCV and NumPy python function to calculate the facial co-ordinates and crop it from the body part so that our model can detect emotions more precisely and more accurately. The we call our model as pass the data of the pictures or frame captures from video stream.

7. Current prototype

With the trained model a prototype was designed. The prototype contains two individual applications. One is an android application for the smartphone of android OS and another one is a web Application. The current prototype of the system (android version) given at figure 13.

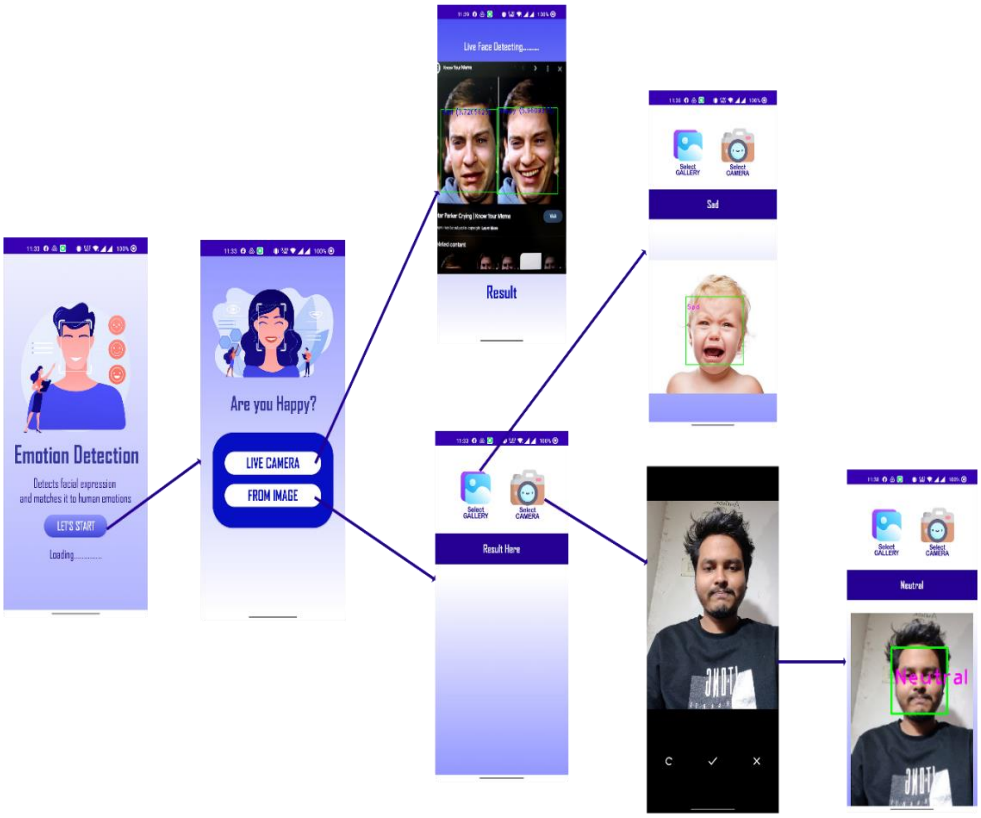

Figure 10. Schematic diagram and working flow of the system (android application).

The web version of the system prototype given at figure 11.

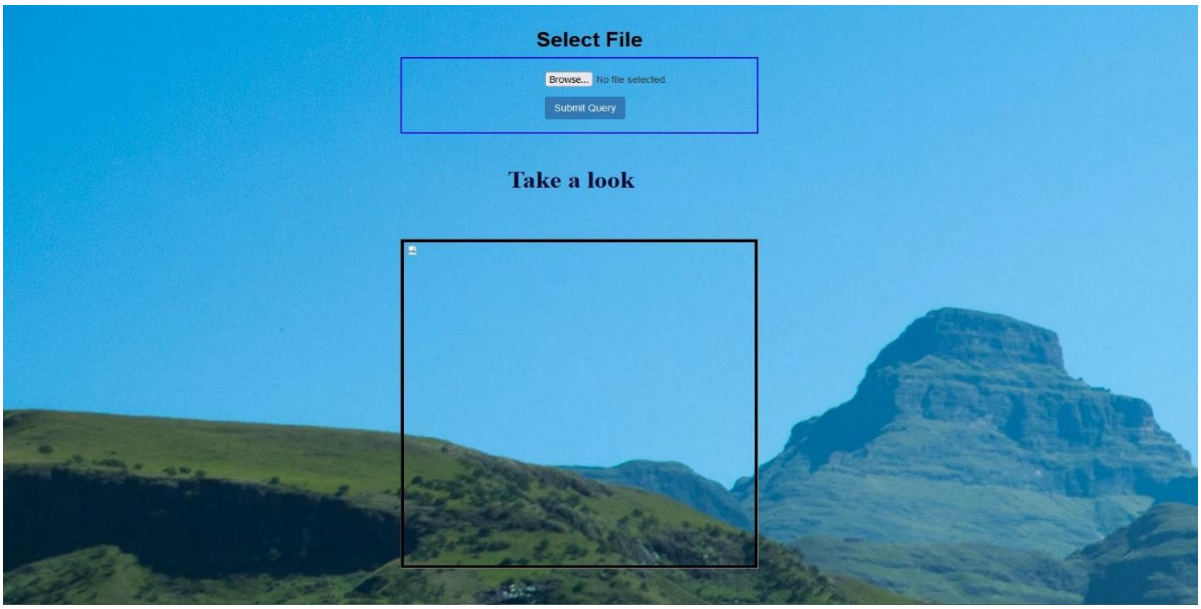

Figure 11. Current prototype of web Application

8. Results Analysis

The system can detect seven types of different emotions: surprise, happy, fear, disgust, angry, sad and neutral by the trained model. We have tested above 200 images and severe live stream on our system. From all of our images we took 200 images randomly and we found 171 images with our expected result. We get an accuracy of 85.5%. The result of our system given here with all classes of emotions detected respectively.

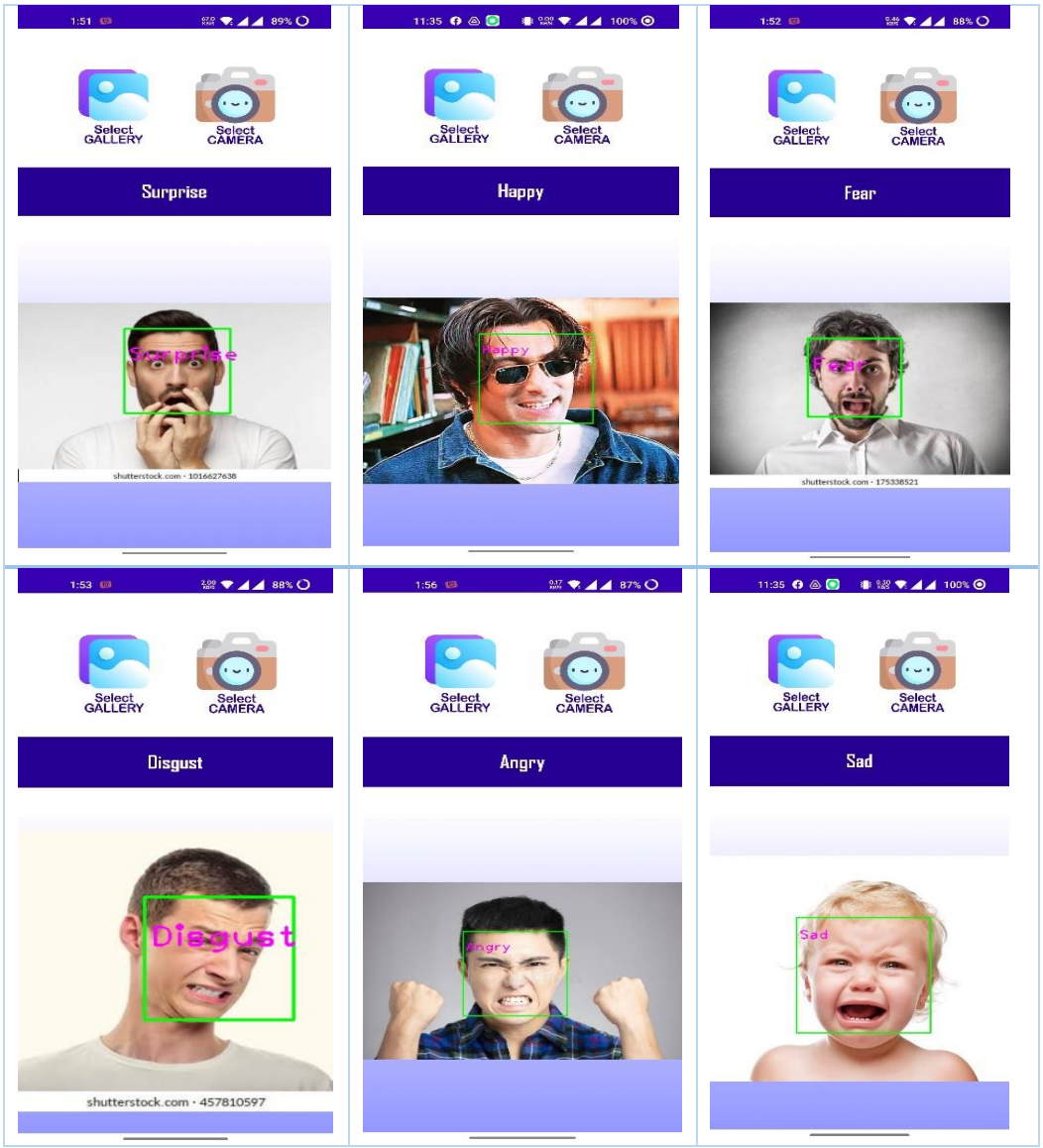

Figure 12. Emotion detection with real time photo and video

Then 10 random images were checked from dataset used to train the model. Some changes were detected in validation accuracy, but a good precision of 97.83% and recall of 97.40% with training accuracy of 97.60% was measured. Here the 10 random testing given below:

Figure 13. Emotion detection of random image form dataset

From the 10 random guess we found 8 were correct, so we can come to a decision that our trained model can give an accuracy of around 80%.

Table 8.1: Value accuracy analytics table

Emotions	Lighting Condition	Test Images	Image Matched	Accuracy
Happy	Bright Light	18	17	94.4%
	Low Light	12	10	83.3%
Sad	Bright Light	15	14	93.3%
	Low Light	11	9	81.8%
Neutral	Bright Light	19	18	94.7%
	Low Light	13	11	84.6%
Disgust	Bright Light	14	12	85.7%
	Low Light	13	10	76.9%
Surprise	Bright Light	16	15	93.8%
	Low Light	15	12	80.0%
Fear	Bright Light	18	16	88.8%
	Low Light	12	9	75.0%
Angry	Bright Light	13	12	92.3%
	Low Light	11	8	72.7%

9. Limitations

- ✓ The dataset we use to train our model has data imbalanced error. So, we don't get great accuracy.
- ✓ As the system detect emotions with camera, having enough light is must. In low light situation camera can't work properly.
- ✓ Some miss calculations between two of the classes.

10. Future Work

- ✓ To make a firebase a model where the user's data will be uploaded at server with classification and we want to make our own dataset with these data.
- ✓ To make the webapp more user friendly and dynamic.
- ✓ Develop a system on suicide prevention using mobile alert application accessing remote camera.
- ✓ Develop a system filtering news according to user's emotion.
- ✓ Want to work with humanoid robot to increase the human-robot interaction.

11. Conclusion

We propose an emotion recognition approach based on a CNN model that effectively extracts facial features. The suggested method uses training sample image data to directly input the picture pixel value. After investigating various face detection, facial expression, emotion recognition, classification methods and techniques, we conclude that the effective result can be achieved by well-trained datasets and algorithms. The ability to determine emotions is very satisfactory. As it gives better result when the face is in the front view and near to the camera view. Expressions of sunglass or other glass wearing faces are also detected quite successfully. In the near future, the study of the emotion detection may provide improved feedback to society as well as the Human-Robot interfaces (HRI). Furthermore, the concept of emotion recognition could be expanded to include emotion detection from speech or body motions in order to address emerging industrial applications.

References

[1] https://www.macrotrends.net/countries/BGD/bangladesh/suicide-rate , Accessed on Dec 25, 2022.

[2] L. E. Parker, E Schneider, and A. C. Schultz, Multi-Robot Systems: From Swarms to Intelligent Automata. Dordrecht, The Netherlands: Springer, 2005.

[3] P. Ekman and W. V. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement, vol. 3. Palo Alto, CA, USA: Consulting Psychologists Press, 1978

[4] Tzuu-Hseng S. Li, Ting-Nan Tsai et al: CNN and LSTM-Based Facial Expression Analysis Model for a Humaniod Robot, vol. 2. Received June 19, 2019, accepted July 4, 2019, date of publication July 11, 2019, date of current version July 30, 2019.

[5] Jose Maria Garcia-Garcia, Victor M. R. Penichet and Maria D. Lozano, Emotion Detection: A Technology review, Interacción '17, September 25–27, 2017, Cancun, Mexico © 2017 Association for Computing Machinery. https://doi.org/10.1145/3123818.3123852

[6] S. Casale, A. Russo, G. Scebba and S. Serrano. 2008. Speech Emotion Classification Using Machine Learning Algorithms. In IEEE International Conference on Semantic Computing. Santa Clara, CA, 158-165. DOI: 10.1109/ICSC.2008.43

[7] Yisi Liu, Olga Sourina, and Minh Khoa Nguyen, Real-time EEG-based Emotion Recognition and its Applications. Nanyang Technological University, Singapore. @ntu.edu.sg

[8] Egger Maria, Ley matthias amd Hanke Sten, Emotion Recognition from Physiological Signal Analysis: A Review. AIT Austrian Institute of Technology GmbH, Vienna, Austria.

[9] Robert Horlings, Emotional recognition using brain activity. Delft University of Technology. March 2008

[10] https://www.analyticsvidhya.com

[11] https://www.sciencedirect.com

[12] https://www.semanticscholar.org

[13] https://www.researchgate.net/figure/

[14] https://intellipaat.com

[15] https://www.kaggle.com/datasets/msambare/fer2013

[16] TensorFlow Models, https://github.com/tensorflow/models