Facial Emotional Recognession

Ataa Shaqour

Hiba Habash

Ahmad Droobi

Clemson University<–>An-Najah National University

A thesis presented for the degree of

Undergraduate Research Fellowship

Dec 2021

Acknowledgements

The authors would like to express their gratitude to the university of Clemson (CU), which has provided Research guiding through the Fundamental Research Grant. The authors would also like to thank the Computer Engineering department in Both An-Najah National University(NNU) and CU for providing facilities to support the research work and (Future Computing Technologies Lab) for their Technical support that enabled us to propose this simple Research.

Abstract

This proposes the PCA (Principal Component Analysis) facial recognition system. The PCA aims to reduce a large amount of data storage to the feature space required to represent the data economically. The purpose of this paper is to make a study on recent works on automatic (facial emotion recognition) FER via deep learning.

This Paper underline these contributions treated, the architecture, and the datasets used, and we present the progress made by comparing the proposed methods and the results obtained. The interest of this paper is to serve the idea of emotional Recognition in real-time, to be obtained in business analysis, customer satisfaction, and interaction between machines and machine learning algorithms and the real-time world. We tried to come through recent works and provide insights to improve.

Keywords— Face detection, Face emotional recognition, PCA, Tensorflow, CNN, Deep learning, OpenCV

Table of Contents

1 | Introduction

1.1 Overview

The face gives us a significant amount of information about human social interaction. It provides us the visual information that allows us to determine the age or sex and determine information about the feelings and thoughts of its owner. People generally infer other people's emotional states, such as happy, sad, angry, disgusting, fear, surprised, and neutral. This paper highlights that banks, stores, and organizations evaluate the services they provide to customers by calling them or sending questionnaires to fill out. And due to its ineffectiveness, the client may ignore it because it might be way too long.

To make the evaluation process easy, effective, productive, and at a daily rate. This proposal aims to track clients' facial emotions by installing cameras in these organizations or banks, or stores so that we can see the clients' satisfaction with this service.

1.2 Thesis Outline

The remainder of this paper is organized as follows:

Chapter 2 Methodology — This chapter shows the Methodology of this paper. The research methodology contains the procedure step by step and how the research reached an optimization of the results and briefly talks about the tools used in this paper.

Chapter 3 Literature Review — Discuss the Model's Architecture and split the data into training and testing sets and fitting sets the module with this data. Then by using real-time facial emotional recognition, and adding the recognized faces to convolutions neural network (CNN) and obtain the results in real-time.

Chapter 4 Results and Discussion — Show the results of the published questioner and analyze the results.

Chapter 5 Conclusion and Future Works — Describe how the research could be embedded with other techniques to make it more Augmented with any service we want to apply the system on and make our architecture more optimal.

2 | Methodology

2.1 GOOGLE COLAB IMPLEMENTATION AND RESULTS

2.1.1 Experimental setup

The experiment was carried on the fer2013 dataset. The data consists of 48x48-pixel grayscale images of faces. This dataset was prepared by Pierre-Luc Carrier, and Aaron Courville as a part of the Kaggle Challenge named Challenges in Representation Learning: Facial Expression Recognition Challenge in 2013. It consists of 28709 train samples and 3589 test samples. It included seven emotions, namely Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral.

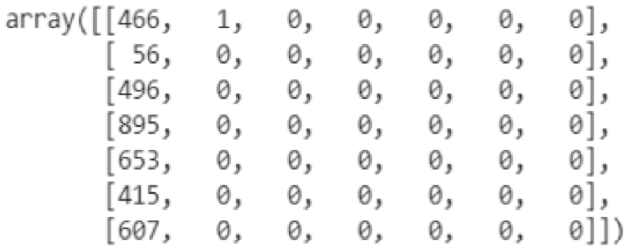
The integrated development environment (IDE) used for the process was Google Colaboratory or Google Colab, in short. Google Colab is a free Jupyter notebook environment that requires no setup and runs entirely in the cloud. With Google Colab, it is possible to write and execute code, save and share our analyses, and access powerful computing resources, all for free from the browser. The fer2013 dataset was mounted to the Google Colab using Google Drive in a CSV file. After Loading the dataset, the batch size was set to 32, and the training epochs set to 12. The software used was Python 3 with machine learning libraries, including Keras 2.1.6 and Tensorflow 2. After initializing the training and the testing instances, the data was given to the convolutional neural network (CNN), which consisted of 3 convolutional layers and one fully connected neural network. The model trained for about 1.30 hours.

2.1.2 Experimental results

This experiment used 12 epochs for training the data samples. With each epoch, the training accuracy increased while reducing the loss. Performance evaluation is shown in Table 4, while the confusion matrix is shown in Figure 4. The testing results were made more accurate if we could include the Haar cascade face detection process. It is a machine learning object detection algorithm used to identify objects in an image or video. It detected the face from the image to reduce the additional noise.

Performance Parameters	Performance Metrics
Training Loss	0.10639
Training Accuracy	0.5986
Testing Loss	2.3
Testing Accuracy	0.6242685914039612

Table 2.1: Performance evaluation based on training and testing accuracy and loss.



```
array([[466,   1,   0,   0,   0,   0,   0],
       [ 56,   0,   0,   0,   0,   0,   0],
       [496,   0,   0,   0,   0,   0,   0],
       [895,   0,   0,   0,   0,   0,   0],
       [653,   0,   0,   0,   0,   0,   0],
       [415,   0,   0,   0,   0,   0,   0],
       [607,   0,   0,   0,   0,   0,   0]])
```

Figure 2.1: Confusion matrix

3 | Literature Review

3.1 Introduction

To make a well-fitted model for Real-Time emotion recognition, there must be proper feature frames(C.H.Wu & Wei, 2014) of the facial expression within the scope. Instead of using conventional techniques, deep learning provides various accuracy, learning rate, and prediction.

Conventional neural networks (CNN) is one of the deep learning techniques which have provided support and a platform for analyzing visual imagery. Convolution is the fundamental use of a filter to an input that outcome in an actuation. Rehashed use of a similar filter to an input brings about a map of enactments called a feature map, showing the areas and quality of a recognized element in input, for example, a picture. The development of convolution neural systems is the capacity to consequently gain proficiency with an enormous number of filters in equal explicit to a training dataset under the requirements of a particular prescient displaying issue, for example, picture characterization. The outcome is profoundly explicit highlights that can be distinguished anywhere on input pictures. (Z.Zeng & T.S.Huang, 2014)Deep learning has achieved great success in recognizing emotions, and CNN is the well-known deep learning method that has achieved remarkable performance in image processing. There has been a great deal of work in visual pattern recognition for facial emotional expression recognition, just as in signal processing for sound-based recognition of feelings. Numerous multimodal approaches are joining these prompts [2]. Over the past decades, there has been extensive research in computer vision on facial expression analysis [3].

The objective of this paper is to develop video-based emotion recognition using deep learning with Google collab, Tensorflow, OpenCV.

3.2 REAL-TIME EMOTION RECOGNITION USING DEEP LEARNING ALGORITHMS

general architecture for building a Real-Time emotion recognition model using a deep learning algorithm are described. Moreover, the architectural diagram and various pre-and post-processing processes are briefly described. The overview of the system using CNN is shown in Figure 1. Before CNN comes into action, the input Real-Time has to go through several processes.

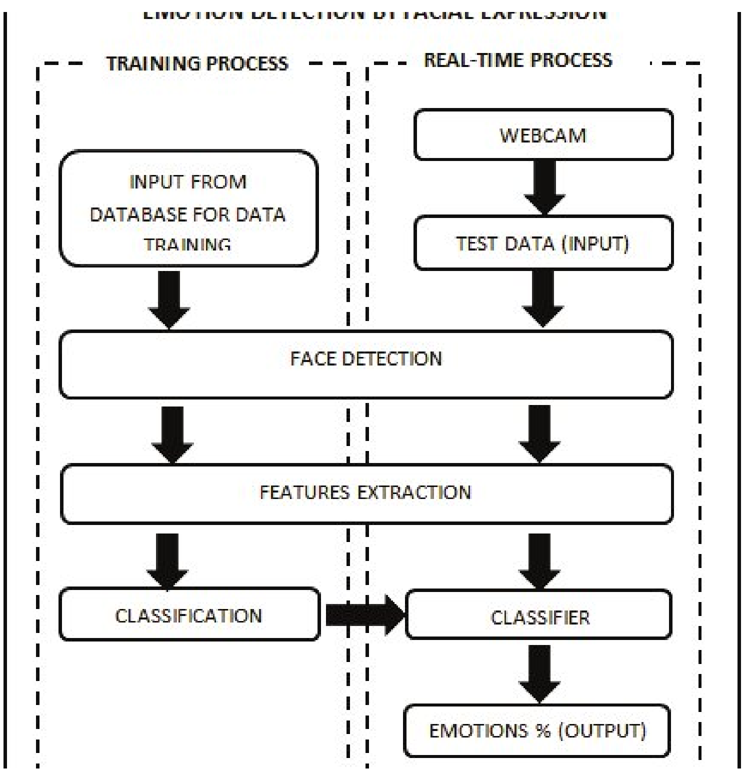


Figure 3.1: Emotion recognition using deep learning algorithms

3.2.1 Pre-processing

This is the first process that is applied to the real-time input sample. Emotions are usually categorized as happy, sad, anger, disgusted, fear, surprise; hence, frames must be extracted from the real-time input stream [4]. Different researchers vary the number of frames based on complexity and computational time. The frames are then converted to grayscale. The frame obtained after gray scaling is somewhat black and white or gray monochrome. The contrast with low-intensity results in grey and that with solid intensity results in white [5]. This step is followed by the histogram equalization of the frames. Histogram equalization is a computer picture handling strategy used to improve contrast in pictures. It achieves this by viably spreading out the most successive intensity esteems, for example, loosening up the intensity scope of the picture. A histogram is a graphical portrayal of the intensity dissemination of a picture. In straightforward terms, it represents the number of pixels for every intensity value considered [6].

3.2.2 Face detection

Emotions are featured mainly from the face. Therefore, it is crucial to detect the face to obtain facial features for further processing and recognition. Many face detection algorithms are used by many researchers like OpenCV, DLIB, Eigenfaces, local binary patterns histograms (LBPH), and Viola-Jones (VJ) [7]. Conventional algorithms included face acknowledgment work by distinguishing facial highlights by extricating highlights, or milestones, from the face picture. For instance, to extricate facial highlights, a calculation may examine the shape and size of the eyes, the size of the nose, and its relative situation with the eyes. It might likewise dissect the cheekbones and jaw. These extracted highlights would then be utilized for looking through different pictures that have matching features. Throughout the years, the industry has moved towards deep learning. CNN has been utilized recently to improve the exactness of face acknowledgment calculations. These calculations accept a picture as information and concentrate a profoundly intricate arrangement of features out of the picture. These incorporate features like the width of the face, the stature of the face, the width of the nose, lips, eyes, proportion of widths, skin shading tone, and surface. Essentially, a convolutional neural network separates many highlights from a picture. These highlights are then coordinated with the ones put away in the database.

3.2.3 Image cropping and resizing

In this phase, the face detected by the face detection algorithm is cropped to obtain a broader and clearer look at the facial image. Cropping is the expulsion of undesirable external regions from a photographic or illustrated picture. The procedure, as a rule, comprises the expulsion of a portion of the fringe regions of a picture to expel incidental rubbish from the image, improve its surrounding, change the perspective proportion, or highlight or disengage topic from its background. After performing cropping operations on the frames, the size of the images varies. Therefore, to attain uniformity, these cropped images are subjected to resizing, say for our example, 80Ã80 pixels. A digital image is just information numbers showing varieties of red, green, and blue at a specific area on a framework of pixels. More often than not, we see these pixels as smaller than normal square shapes sandwiched together on a PC screen. With a little inventive reasoning and some lower-level control of pixels with code, in any case, we can show that data in a horde of ways. The size of the frame determines its processing time. Hence, resizing is very important to shorten the processing time. Moreover, better resizing techniques should be used to preserve image attributes after resizing [8]. The accuracy of the classification depends on whether the features are well representing the expression or not. Therefore, the optimization of the selected features will automatically improve classification accuracy [9].

3.2.4 CNN structure with ConvNet

A CNN is a deep learning algorithm that can take in an info picture, allocate significance (learnable loads and bias) to different viewpoints in the image and can separate one from the other. The pre-preparing required in a ConvNet is a lot lower when contrasted with other algorithms. While in simple techniques, filters are hand-designed, with enough preparation, ConvNets can get familiar with these qualities. The engineering of a ConvNet is pretty much similar to that of neurons in the human brain and was enlivened by the association of the Visual Cortex. Singular neurons react to improvements just in a confined locale of the visual field known as the receptive field. An assortment of such fields overlaps to cover the whole visual zone [10].

ConvNet is an arrangement of layers, and each layer of a ConvNet changes one volume of initiations to another through a differentiable function. There are three primary kinds of layers to construct ConvNet models: convolutional layer, pooling layer, and fully-connected layer, as shown in Figure 2. The general architecture of the ConvNet consists of the following [11]:

* Input [80x80x2] will hold the raw pixel estimations of the picture, right now a picture of width 80, height 80.

* The convolutional layer will evaluate the yield of neurons that are associated with nearby locales in the info, each processing a dot product between their loads and a little area they are associated with the input volume. This may bring about volume, for example, [80x80x12] if we chose to utilize 12 filters.

* RELU layer will be applied for an element-wise actuation work, for example, the max (0, x) thresholding at zero. This leaves the size of the volume unaltered [80x80x12].

* POOL layer will play out a downsampling activity along with the spatial measurements, bringing about volume, for example, [40x40x12].
* Fully connected layer (FC) will process the class scores. The input to this layer is all the outputs from the previous layer to all the individual neurons.

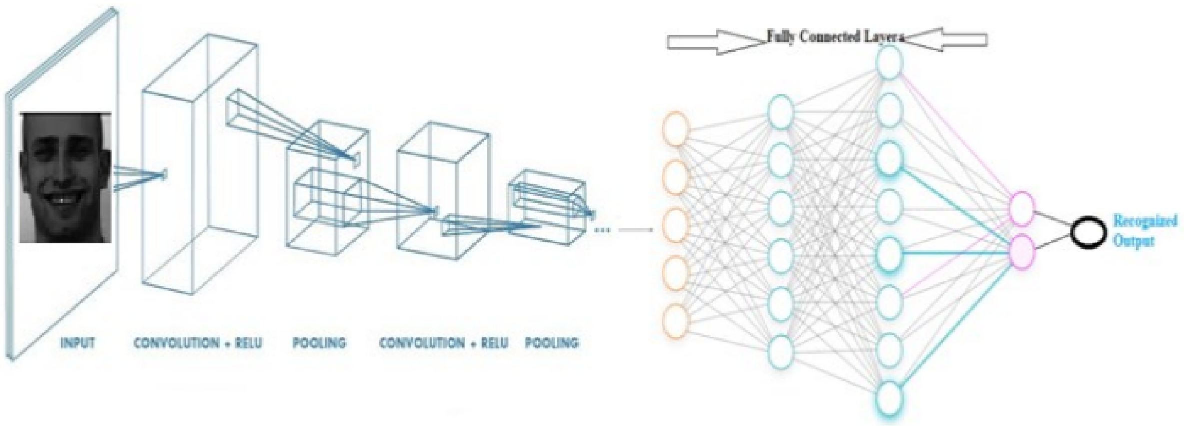


Figure 3.2: General Architectural structure of convolutional neural networks

4 | Results and Discussion

After implementing the facial recognition software, a questioner was built and shared to help this research understand the importance of facial emotion detection and the most efficient way to use this software. The questioner contained many different Questions aiming to collect the most valuable data to help this research improve. Here are the statistics and analysis of the survey.

The survey was answered by 42% Male and 58% Female and people who took the survey were 88.4% at the age from 18 to 24

The survey asked what is the most commonly used service in people life and the results are as shown below :

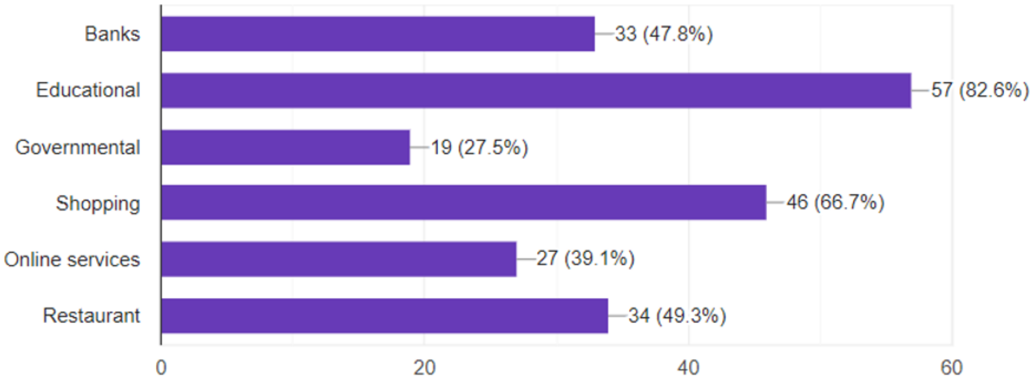


Figure 4.1: The most common services among people

We can conclude from that Question that the most used services among people that took the survey are educational service and shopping comes in number 2. That helps the research to determine what services to implement the software into.

The next question asked people what the reasons that make you prefer one service over another are, and the result was as shown below:

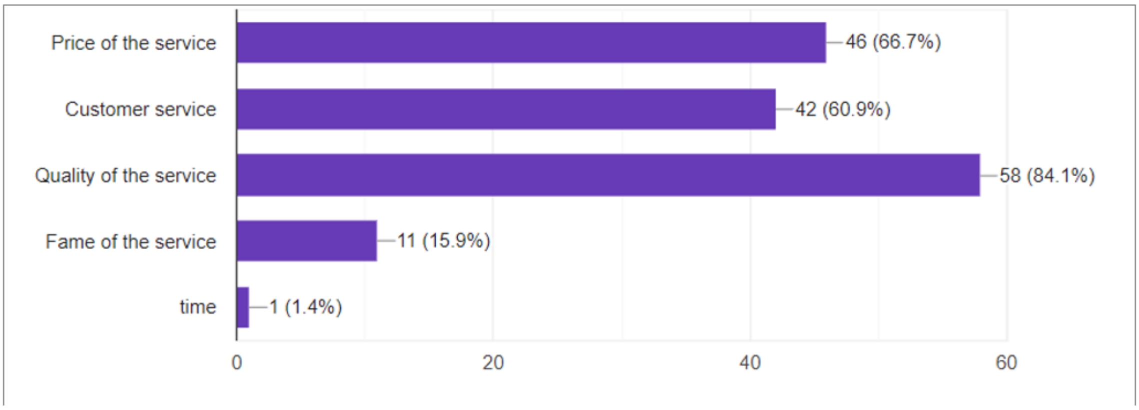


Figure 4.2: what makes a service better than other

84.1% of people prefer service over another due to its quality and 60.9% due to its customer service team. That was a crucial indicator of the Necessity of such software.

The following Question asked whether people have heard about facial emotional recognition and the result were as shown below :

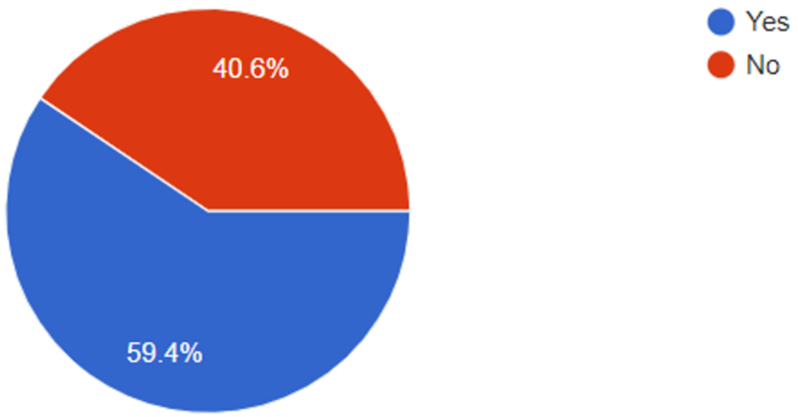


Figure 4.3: percentage of people who have heard about facial emotional recognition

The research concluded that facial recognition is a well-known concept but not enough since 40.6% still haven't heard about it.

The next question asked if feelings appear on people's faces. Hence, it helps us see the software's efficiency since it is useless to apply software to a non-emotional person. Still, the results were promising since 36.2% strongly agreed that their feelings appear on their faces and only 8.6% disagreed.

This was an indicator that the software we built would perform well on real-time implementation.

The research also wanted to see how would people accept and react to such software are going to reject or not, and the result was as shown below:

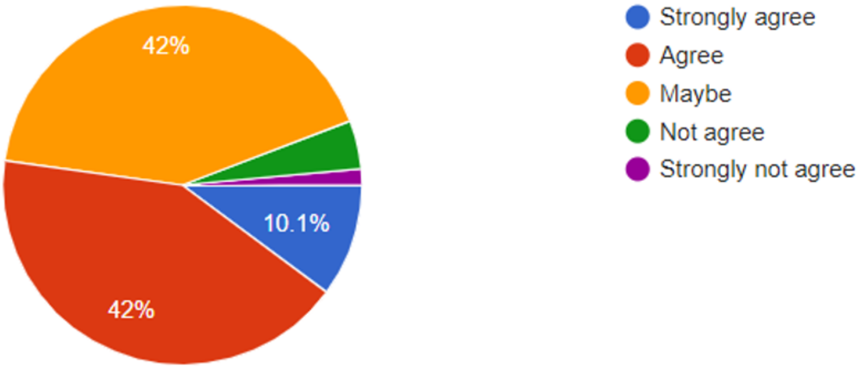


Figure 4.4: People emotions appear on their faces all the time

10.1% strongly agreed that a machine could detect and interact with humanâs emotions while 42% agreed on that and 42% said it was possible to achieve it, which leaves only 5.7% that donât agree with the concept.

On the other hand, 15.9% of people strongly agree that applying such software in service-oriented businesses could violate their privacy. 27.5%agree on that, and 43.5% think that there is a possibility. 13% don't agree.

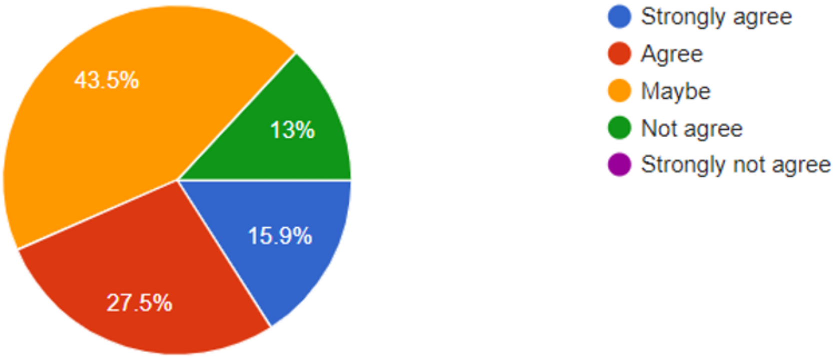


Figure 4.5: Detecting client emotions violate its privacy

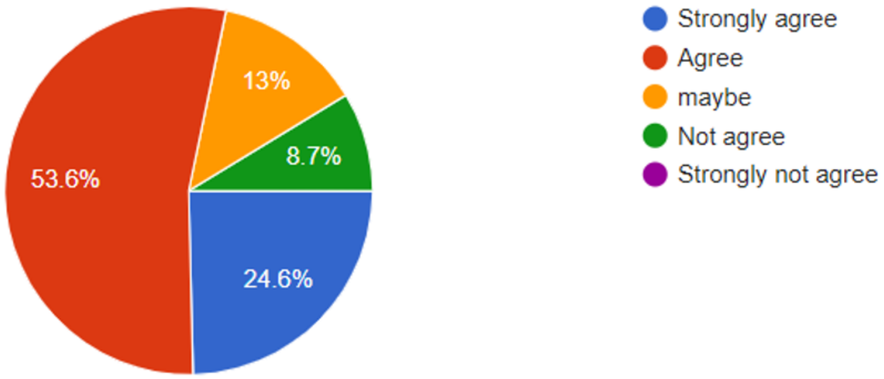


Figure 4.6: knowing what clients feel about a service could help improve that service

But on the bright side, 53.6% of people think if a service-oriented business knows how the clients and customer feels about the service, it could improve that service.14.6 strongly agree with that. 13% believe there is a possibility to do such a thing while only 8.7%don't agree.

The research also asked people whether when they feel angry about a service does it show on their face or not, and the results are shown below:

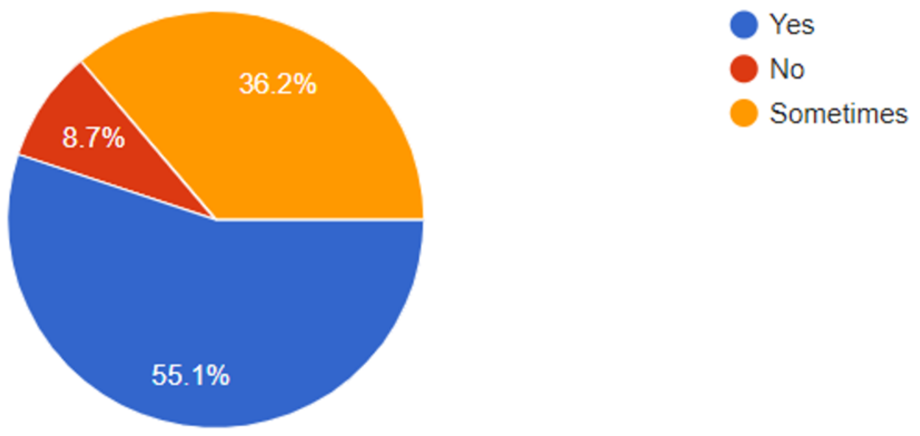


Figure 4.7: Does feeling angry about a service reflect on people's faces

65.1% of people agree that when they are angry, it shows on the face while 36.2% think that there is a possibility of that happening only 8.7% of people Don't show emotions of anger while receiving bad service.

The final question was to give us an idea of whether it is good to implement this software in social media. The question was, while you are scrolling in the social media, does your facial expression indicates whether you like a post or not and the answers were as shown below:

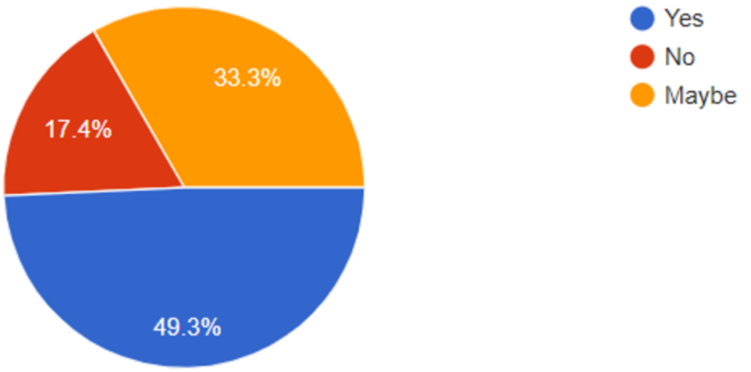


Figure 4.8: while you are scrolling in the social media does your facial expression indicates whether you like a post or not

5 | Conclusion and Future Works

5.1 Conclusion

This paper presented the development of Real-Time emotion recognition using deep learning with Google Colab, Tensorflow, OpenCV, Dlib. The success of this approach in recognizing emotions has been tremendously improving over time. Introducing deep learning techniques like CNN, DNN, or other multimodal methods has also boosted the pace of recognition accuracy. Our work demonstrated the general architectural model for building a recognition system using deep learning (CNN). The aim was to analyze pre and post processes involved in the model's methodology. There is extensive work done on image or video as input (Real-Time System) to recognize the emotion. This paper also covered the datasets available for the researchers to contribute to this field. Different performance parameters were benchmarked on different research to show this sphere's progress. The experimental observation was carried out on the Kaggle dataset involving seven emotions, namely angry, disgust, fear, happy, sad, surprise, neutral, which yielded in the be 0.97 accuracies on the training set and 0.574 accuracy on the testing set when the Haar cascade technique is applied.

5.2 Future Works

The scope for future improvements is very appealing in this field. Different multimodel deep learning techniques can be used along with different architectures to improve the performance parameters [20-27]. Apart from recognizing the emotions only, there can be further addition of intensity scale. This might help to predict the intensity of the recognized emotion. Also, multi modals can be used in future works; for example, video and speech can both be used to design a model along with the use of multi-datasets.

References

C.H.Wu, J., J.C.Lin, & Wei, W. L. (2014). Survey on audiovisual emotion recognition: databases, features, and data fusion strategies. *APSIPA Transactions on Signal and Information Processing*, *3*.

Z.Zeng, G., M.Pantic, & T.S.Huang. (2014). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(1), 39–58.