

THE THIRD EYE INTERVIEW – EMOTION ANALYSIS

2021-070

Hewage Don Lahiru Lakshan

(IT18110180)

B.Sc. (Hons) Degree in IT specializing in Software Engineering

Department of Computer Science & Software Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

October 2021

THE THIRD EYE INTERVIEW – EMOTION ANALYSIS

2021-070

Hewage Don Lahiru Lakshan

(IT18110180)

Dissertation submitted in partial fulfillment of the requirements for the B.Sc. Special
Honors Degree in IT

Department of Computer Science & Software Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

October 2021

DECLARATION

I declare that this is my own work and this dissertation¹ does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. Also, I hereby grant to Sri Lanka Institute of Information Technology the nonexclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Name	Student ID	Date
Lahiru Lakshan	IT18110180	13/10/2021

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of the supervisor:

Date:

Signature of the co- supervisor:

Date:

ACKNOWLEDGEMENT

The final creation is the result of the group's hard work as well as the encouragement, support, and guidance provided by many others. As a result, I feel indebted of appreciation to everyone who helped me finish this huge effort.

I am grateful to Dr. Pradeepa Samarasinghe, Ms. Madhuka Nadeeshani, and the lecturers at Sri Lanka Institute of Information Technology for their recommendations, ongoing encouragement, and support in the creation of this research, especially for the variety of advanced and educational discussions.

I'd want to express my gratitude to Dulmini Dissanayake, Raveen Dissanayake, and Venuri Amalya, members of my research team, for their support and encouragement in finishing this project. Also, I'd like to thank the Sri Lanka Institute of Information Technology (SLIIT) for identifying the resources that were available at the right time to groom and develop my skill.

Additionally, we would like to extend our thanks to all of our colleagues and friends for their assistance, support, attention, and invaluable advice. Finally, I'd want to express my gratitude to everyone else whose names are not mentioned specifically but who have given their support in several ways and inspired me to make this a success.

ABSTRACT

There is an infinite amount of emotions that the human face may communicate without saying anything. Facial expressions play a significant role in how we communicate and create impressions of those around us. In virtual meetings such as interviews or viva sessions, it is impossible to communicate with others through facial expressions as in a real meeting. There may be many misunderstandings if the interviewee is unable to convey facial expression via their responses and the interviewer is unable to understand what the interviewee is trying to say. To address this issue, the facial emotion identification and analysis component of the behavior and attention detection system was developed. The fundamental seven emotions (anger, disgust, fear, happy, joy, sad, and surprise) as well as arousal, violence, and others are identified and shown to the interviewer in order to get an understanding of the interviewee's emotional state. Individual and group interviewees' emotions are monitored and evaluated in order to determine five personality trait rates. Apart from that, a predicted emotion of interviewee was used to determine the interviewee's most prominent emotion. Emotions detected and data analyzed will help the interviewer in getting a greater understanding of the interview session as a whole, as compared to a physical meeting.

TABLE OF CONTENTS

DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	vi
LIST OF FIGURES	vi
LIST OF ABBREVIATIONS	vii
LIST OF APPENDICES	viii
1 INTRODUCTION	1
1.1 Background Context	3
1.2 Research Gap	9
1.3 Research Problem	10
1.4 Research Objectives	11
1.4.1 System Objective	11
1.4.2 Main Objectives	11
2 METHODOLOGY	13
2.1 Methodology	13
2.1.1 Data gathering	16
2.1.2 Emotion detection model building	17
2.1.3 Training the model	18
2.1.4 Arousal and valence detection model	19
2.1.5 Requirements Gathering and Analysis	20
2.1.6 Design the system	21
2.2 Testing and Implementation	22
2.2.1 Identifying the emotion and detect the most prominent emotion.	22
2.2.2 Tools and technologies	23
2.2.3 Front-end implementation	25
2.2.4 Testing	26

2.2.5	Test Cases	27
2.3	Commercialization aspects of the product	30
3	RESULT.....	32
3.1	Results	32
3.2	Research findings	38
3.3	Discussion	40
4	CONCLUSION	41
5	REFERENCES.....	43
	APPENDICES	46

LIST OF TABLES

Table 1.1: Comparison with similar technologies	8
Table 1.2: Comparison of features of researches and applications	9
Table 2.1: Survey questions	20
Table 2.2: Test case	27
Table 3.1: Dataset comparison of emotion detection model	38

LIST OF FIGURES

Figure 1.1: Emotional analysis of changes over time [4]	5
Figure 1.2: Output of Self-Management Interview App[4]	5
Figure 1.3: Valence and arousal circumplex [6]	7
Figure 2.1: Overall System Diagram	14
Figure 2.2: Emotion analysis system diagram	15
Figure 2.3: Emotion Detection Model Summary	18
Figure 2.4: Model training process	19
Figure 2.5: Dashboard of TIN system	25
Figure 2.6: Emotion analysis user interface	26
Figure 3.1: Accuracy curves of model training by FER2013 Dataset	32
Figure 3.2: Loss curves of model training by FER2013 Dataset	33
Figure 3.3: Accuracy curves of model training by CK+ Dataset	33
Figure 3.4: Loss curves of model training by CK+ Dataset	34
Figure 3.5: Accuracy curves of model training by SPOS Dataset	34
Figure 3.6: Loss curves of model training by SPOS Dataset	35
Figure 3.7: Difference between CK+ dataset and SPOS dataset	35
Figure 3.8: Different between images of original CK+ and modified CK+ dataset ..	36
Figure 3.9: Loss curves of model training by modified CK+ Dataset	37
Figure 3.10: Accuracy curves of model training by modified CK+ Dataset	37
Figure 3.11: Receiver operating characteristic curve	39

LIST OF ABBREVIATIONS

Abbreviations	Description
CNN	Convolutional Neural Network
ReLu	Rectified Linear
CK+	The Extended Cohn-Kanade Dataset
Fer-2013	The Facial Emotion Recognition – 2013
TIN	Third Eye Interview

LIST OF APPENDICES

Appendix 1: Classification report of the emotion detection CNN model.....	46
Appendix 2: Training emotion detection model	46
Appendix 3: Dataset splitting function	47
Appendix 4: Dataset loading function.....	47
Appendix 5: Plagiarism report	48

1 INTROUCTION

The pandemic crisis has affected the world's living patterns dramatically in a short period. Among them, a large number of physical meetings conducted in a variety of sectors have moved to virtual meetings, despite the fact that people have largely adapted to it since it is convenient and simple, as well as the best solution in light of the current global pandemic scenario. Additionally, several user-friendly apps have been developed to facilitate these virtual meetings, which has resulted in these applications being widely used and well-known by everyone. During virtual meetings, we may interact with others by gazing directly into their eyes while exchanging papers, and we can deliver our material as if we were giving a presentation in the real world. Work from home and learn from home concepts have become more popular around the world as a result of these apps and their cutting-edge technologies.

When compared to a physical interview, virtual interviewing provides additional challenges for both the interviewer and the interviewee to go through. An interviewee's psychological state is one of the most important aspects of a successful candidate for a job interview. During the viva/interview, interviewees are asked to provide a brief but thorough description of themselves based on their activities, behaviors, and emotional states. But virtual interviews do not give a 's description of the interviewee's activities or behavior. Interviewers are watching closely their clients' body language, emotions, smiles and smile patterns, eye gazing, and head posture, among other things. However, interviewers will have much more trouble evaluating these factors in virtual interviews than they would in physical meetings, which will impact their ability to get a clear understanding of the interviewee. As a solution to this issue, we propose Third Interviewer on the Net (TIN) - Interviewee analysis in online viva/interview session. We propose to evaluate the interviewee's emotion, smile, eye gaze, and head pose in order to provide the interviewer with information about the interviewee's behavior and the personality traits. Additionally, the system focused on delivering a group analysis in which an individual candidate's performance could be compared to the average performance.

In our proposed system, my responsibility is to detect and analyze the interviewee's emotions. The human face is effective in conveying a broad variety of emotions; facial emotions play a significant role in how we communicate and form perceptions of those around us. While emotions play a significant part in communication, emotional awareness is another skill that helps us in communicating successfully. You'll take notice of other people's emotions and the ways in which they communicate differently based on their state of mind. You'll now have a better understanding of how and why individuals behave toward you. This will be useful during interviews or viva sessions, when facial expressions may provide a clearer idea of the interviewee's attitudes and what they are consciously attempting to communicate.

To achieve my goal, seven fundamental emotions (anger, disgust, fear, happy, neutral, sad, and surprise) were detected with a level of accuracy using a Convolution Neural Network (CNN) model. Additionally, another CNN model was utilized to determine the interviewee's arousal and valence using process interview video. The outcome predicted by these two CNN models were used to identify the most prominent emotion in each question and throughout the interview, to determine the interviewee's rating for personality traits during the interview, and to compare the interviewee with the group's average performance.

1.1 Background Context

Numerous researchers have tried to develop a machine learning-based model for identifying the emotion on a human face using a variety of technologies. They focused on ways to improve the predictive accuracy of the developed machine learning model. They utilized a variety of technologies to face detection, feature extraction and the emotion classification.

A Weighted Mixture Deep Neural Network (WMDNN) Model [1]

Pre-processing instructions such as face recognition, rotation correction, and data augmentation were used to restrict areas. There are rotating benchmarking datasets and real-world settings for images of the same topic. These differences have little to do with facial movements, yet they may affect the accuracy of identification. This problem is resolved by aligning the face area using rotation rectification and verifying that the angle of rotation produced by the line segment connecting one eye center to the other is zero. The face was identified using the Viola–Jones framework and converted to both LBP and grayscale images. Mouths, eyes, and brows pop out more in LBP images than in grayscale images.

The suggested CNN for grayscale image feature extraction is based on the VGG16 network, which has good visual detection and fast convergence. They built a shallow CNN model that could automatically extract face expressions from LBP images. The feature vectors are connected by two cascaded full connect layers. The fused feature vector is utilized to classify the phrase using SoftMax classification. This model was trained and tested on the CK +, JAFFE, and Oulu-CASIA datasets, with the accuracy rate of 97.02 %, 92.21 %, and 92.89 % respectively.

Automatic facial expression recognition based on a deep convolutional-neural-network structure [2]

OpenCV (Open Source Computer Vision Library) is the most widely used computer vision library in the world, providing a wide range of programming functions for computer vision. Face identification in this system utilized OpenCV's Haar Cascades classifier. They utilized the histogram equation (HE), an image processing technique that enhances gray value distribution uniformity and reduces

lighting interference. The suggested CNN-based recognition method has two layers called convolutional layer and subsampling layer. The network is fed directly from 2D images, which are subsequently convoluted with multiple convolution layers that are modified to generate matching feature maps, creating the convolutional layer. The same procedure is repeated, and the two-dimensional pixels are rasterized into one-dimensional data after the features are extracted, and then inputted to the conventional neural network classifier, using SoftMax utilized for final classification. Accuracy in training and identification was 76.74% and 80.30%, respectively, using CK + and JAFFE datasets. Additionally, they evaluated accuracy without using HE, and found that the accuracy rates for CK+ and JAFFE databases were decreased to 67% and 76%, respectively.

Facial emotion recognition in real-time and static images system [3]

The web camera is capturing a real-time video in the format of a sequence of frames rather than a single frame. Each frame of the webcam video will be scanned for faces, and those identified will be processed further. Additionally, they utilized OpenCV to identify the face in real time. After grayscale transformation, the image is adjusted for contrast using adaptive histogram equalization. A corner point detection method was used to extract the necessary corner points from the feature areas. Support Vector Machines classification was used to assess the outcome of the model. They have also evaluated accuracy using Linear SVM, Polynomial SVM, K-Means Clustering, and Random Forest classifier classification algorithms, and the system has achieved 94.1% accuracy in Linear SVM classification algorithm in real time emotion detection.

Self-Management Interview App [4]

A mobile application was created to assist users in preparing for face-to-face interviews. Participants in mock interviews may improve their communication and interviewing abilities. The system will evaluate the user's emotions based on an analysis of the recorded interview, as well as identify the emotion that was expressed the most often. It teaches students how to face the interview with a positive attitude. They utilized multi-block deep learning and various AdaBoost learning methods to

develop this application. For facial recognition and emotion extract, the CAS-PEAL face database and the CK+ database is used.

They then analyze how the emotion changes over time; Fig 1.1 illustrates this in a tabular format. As a result of their system's output, they generated a rate of emotions that was evaluated throughout the video, as shown in Fig 1.2.

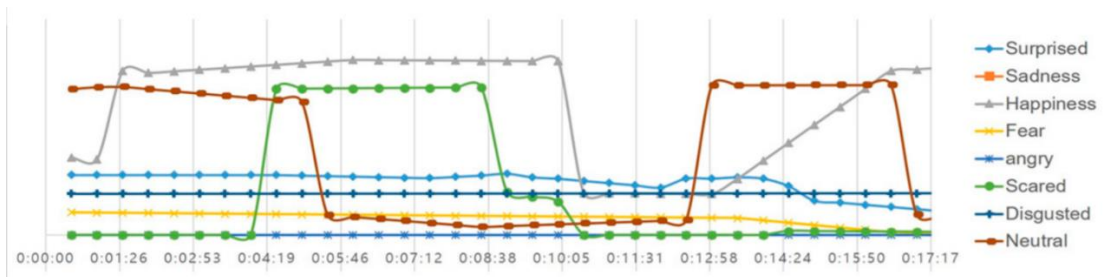


Figure 1.1: Emotional analysis of changes over time [4]

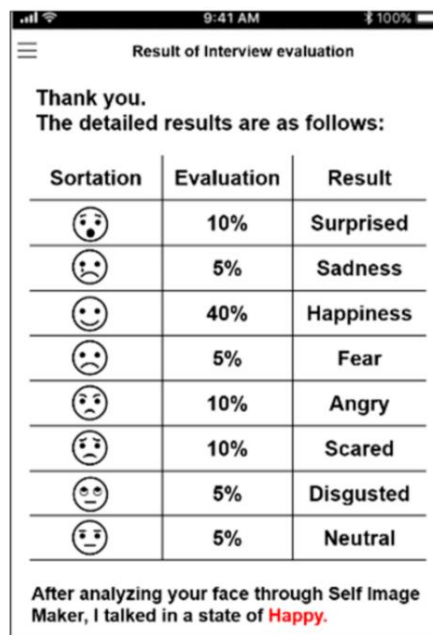


Figure 1.2: Output of Self-Management Interview App[4]

Interviewee Performance Analyzer Using Facial Emotion Recognition and Speech Fluency Recognition [5]

R. Halder et al. suggested a method automates this procedure by constructing two multiclass classification models. The suggested system is provided a video recorded during an interview and it extracts frames and audio from that too. It uses the HaarCascade classifier, Gabor filters, and a convolutional neural network to classify facial emotion as one of seven different emotions: happy, surprise, angry, disgusted, neutral, fear, and sad. The second model is fed audio and utilizes Mel frequency cepstral coefficient characteristics and logistic regression to classify speech into four categories: fluent, stuttering, cluttering, and pauses. Combining the predictions of these two models results in a performance rating for the interviewee. In comparison to using just CNNs or Deep Neural Networks for face emotion detection, the Gabor Filter-based method that they utilized achieved a higher level of accuracy with few hidden layers and less training time.

The FER2013 and ck+ datasets were utilized for Facial Emotion Recognition. For speech fluency recognition, they used data from two datasets called the Speech Accent Archive and the LibriSpeech ASR Corpus. For 24 epochs, the accuracy of the Facial emotion recognizer is 92% and validation accuracy are 86% in the emotion classification model.

Estimation of continuous valence and arousal levels from faces in naturalistic conditions [6]

Facial affect analysis attempts to enable computers to better understand a person's emotional state, thus allowing new kinds of human–computer interactions. Due to the fact that discrete emotional categories (anger, happiness, sadness, and so on) do not properly represent the full range of emotions displayed by humans on a daily basis, psychologists typically rely on dimensional measures, specifically valence (how positive the emotional display is) and arousal (how calming or exciting the emotional display looks like) as shown in the Fig 1.3. While it is easy for humans to estimate these values from a face, it is very difficult for computer-based systems, and automated assessment of valence and arousal in natural conditions is an outstanding problem.

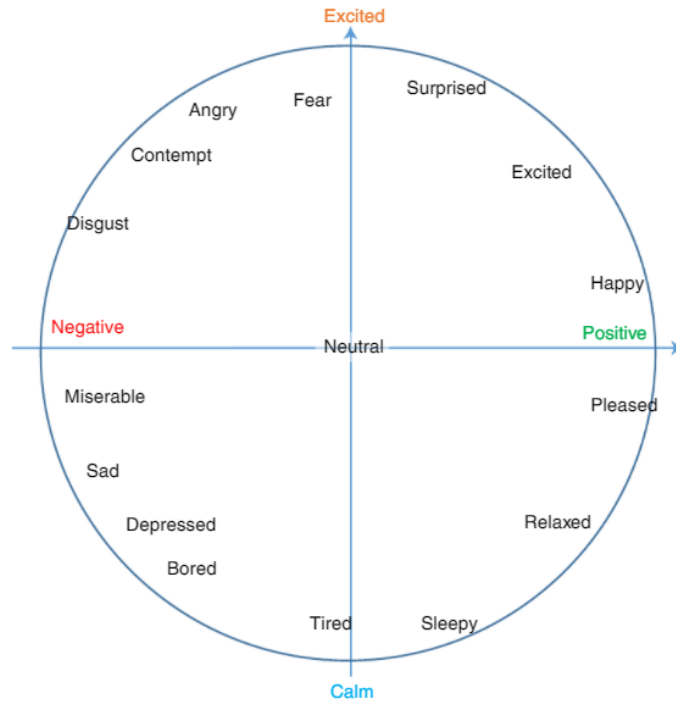


Figure 1.3: Valence and arousal circumplex [6]

They present a new deep neural network architecture for analyzing facial expressions in naturalistic settings and demonstrate that their method beats the other methods on three challenging datasets gathered in naturalistic conditions.

The input image may include one or more faces that have varying orientations, translations, and scaling. A face detector is used to locate and extract each face from the input picture. To begin, the basic process identifies facial landmarks to align each face and then predicts categorical or continuous emotions based on the aligned faces (multi-step approach). Bottom line, their method assesses facial landmarks, discrete and continuous emotions directly using a single deep neural network, enabling the real-time prediction (single-step approach).

Finally, they obtained a 62% accuracy rate using the Original AffectNet dataset. However, they applied additional data pre-processing methods and cleaned the AffectNet dataset and reached a 75% accuracy rate.

Table 1.1 shows the comparison of technology and accuracy of emotion detection which is used by different researchers.

Table 1.1: Comparison with similar technologies

Research	Technology used for			Accuracy
	Face Detection	Feature extraction	Final classification function	
T. Tashu et al. [1]	Viola–Jones framework and DRMF technology	VGG16 network and shallow CNN model	SoftMax	CK+ - 97.02% JAFEEEE - 92.21% Oulu-CASIA - 92.89%
K. Shan et al. [2]	HAAR filter from OpenCV	New proposed CNN-Based Recognition base algorithm with two basic layers	SoftMax	CK+ - 80.303% JAFEEEE - 76.7442%
S. Gupta et al. [3]	HAAR filter from OpenCV	Corner point detection algorithm	SVM	94.1%
R. Halde et al. [5]	Local SMQT features and Split up Snow classifier	Corner point detection algorithm	SVM	83%
J. Kim et al. [7]	fuzzy color filter and VFM based histogram analysis	New Proposed method	Fuzzy classifier	74.0%

1.2 Research Gap

The majority of the studies and applications I described earlier are not utilized to identify emotion for non-emotional purposes. The majority of these studies are conducted in order to create new machine learning models and discover new technologies capable of more correctly detecting the basic seven emotions. The Self-Management Interview App analyzes interview videos for the most prominent emotion. However, according to the interviewer's question, they did not predict the most prominent emotions. There is currently no researcher who has predicted the interviewer's personality characteristics using the average values of seven fundamental emotions arousal and valence. In our proposed method, we identify those emotions with the highest level of accuracy and utilize the results to provide a good opinion of the interviewee to the interviewer by predicting the interviewer's most prominent feelings and personality characteristics. A more detailed comparison of the existing systems and researchers are tabulated in Table 1.2.

Table 1.2: Comparison of features of researches and applications

Product	Identify Seven Basic Emotion in image/video	Identify Arousal	Identify Violence	Identify Most prominent emotion of each question and overall video	Predict the personality trait values
T. Tashu et al. [1]	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
K. Shan et al. [2]	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
H. Siqueira et al. [8]	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Self-Management Interview App	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Emotimeter [9]	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Feely [10]	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Third Eye Interview (Proposed System)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

1.3 Research Problem

Interviews and vivas are an essential part of academic and professional life. When these encounters are held online through different platforms due to a pandemic scenario or for any other reason, it gives both interviewers and interviewees a lot of flexibility, but it has a cost. Our body language, emotions, smile, eye gazing, and other nonverbal cues cannot be used to communicate ourselves in virtual meetings. These factors are essential in fostering collaboration and mutual understanding, which is to say, in order to get the most from a meeting.

Strong emotion may be expressed via the human face without saying anything. Faces play a significant role in how we communicate and create impressions of others. Virtual meetings such as interviews or viva sessions do not allow for facial emotion to be used to communicate. A lack of facial expression and an inability to understand what the interviewee is trying to say may contribute to misunderstandings. Like a physical meeting, an interview may show a person's knowledge, confidence, and skill. It may impact both the interviewer and the interviewee, impeding the development of a connection or understanding among them and so defeating the meeting's goal.

As the solution of this problem emotion analysis component is developed within our new proposed attention and behavior system 'Third Eye Interview'. Interviewers conducting virtual interviews can gain a better knowledge of their interviewees by calculating the interviewer's personality characteristics, which are calculated based on predicted emotional changes, the percentage of smiles, changes in eye direction, and head nodding and shaking.

1.4 Research Objectives

1.4.1 System Objective

This research aims to inform the interviewer on the interviewee's behavior and attentiveness during a virtual meeting, exactly as it would during a real meeting. A summary of the questions asked during in the interview/viva will be provided as an outcome of the analysis based on the emotion, smile, eye gaze and the head pose. Individual and group data may be displayed separately for each question.

1.4.2 Main Objectives

The emotion detection section is used to determine the interviewee's emotion and how that emotion varies over time and in response to the questions given. The detected data is processed and utilized to determine the interviewee's most prominent emotion and the personality traits as an output to display the interviewer to gain a sense of the interview. In comparison to a physical meeting, detected emotions and processed data will assist the interviewer in gaining a better understanding of the interview, the interviewee's conduct, and expertise.

1.4.2.1 Specific Objectives

1. Gather emotion datasets

Find the most well-known and valid annotated datasets for emotion detection.

2. Create a model to extract features from the interviewee's responses and identify the seven basic emotions exhibited by the interviewee throughout the interview.

Identify happy, anger, disgust, neural, fear, sadness, surprise

3. Determine the interviewee's prominent emotion, as they relate to each interviewer's question and to the whole interview.

Using the predicted emotion of the interviewee, the most prominent emotion of the interviewee in each question and throughout the interview should predict .

4. Determine arousal and violence.

Arousal and violence value of the interviewee should be predicted, and the average values should pass the personality trait detection model.

5. Comparison of the individual and group emotional behavior based on questions.

6. Identify the personality traits rate

2 METHODOLOGY

Third Eye Interviewer is an open-access website that anyone with a valid username and password can visit. Users may submit video interviews of interviewees to our system along with the interviewee's information. The system analyzes the video and determines the interviewee's behavioral state and displays the interviewee's important personality information to the interviewer. This section will explain the techniques, specific approaches, and methods utilized to achieve specific objectives. It also describes the emotion recognition model building process and the designed user-friendly user interfaces, testing process, and tools used to develop the models.

2.1 Methodology

Initially, the user has to login to our system by using user credentials. Those credentials have to be generated using company login credentials which are issued by the system after any organization registered with the product. After logging in to the system the user can create an interview group and they can upload all interview videos which are related to creating the group along with the interviewee details.

Initially, user credentials have to be generated by the system. Those credentials must be generated by using the company credentials issued by the system to any company registered with our products. After logging into the system, the user may create an interview group and upload any relevant interview videos to create an interview group along with details about the interviewee.

The video will be stored in our database after it has been successfully uploaded. Google speech and text API are used to determine the interviewer's time intervals for each question [11]. Dlib is used to determine the identity of faces and their coordinators [12]. Detected faces, face coordinators, and the image frames are then sent to smile detection, eye gaze detection, emotion detection, and head pose detection component to analyze them. Those analyzed data are used to predict the personality trait of the interviewee.

To identify seven basic emotions, arousal, and valence two CNN-based models were developed and tested. The CNN model's predicted data were used to predict the most prominent emotions and to moderate the input to the personality trait prediction model.

Overall System Diagram

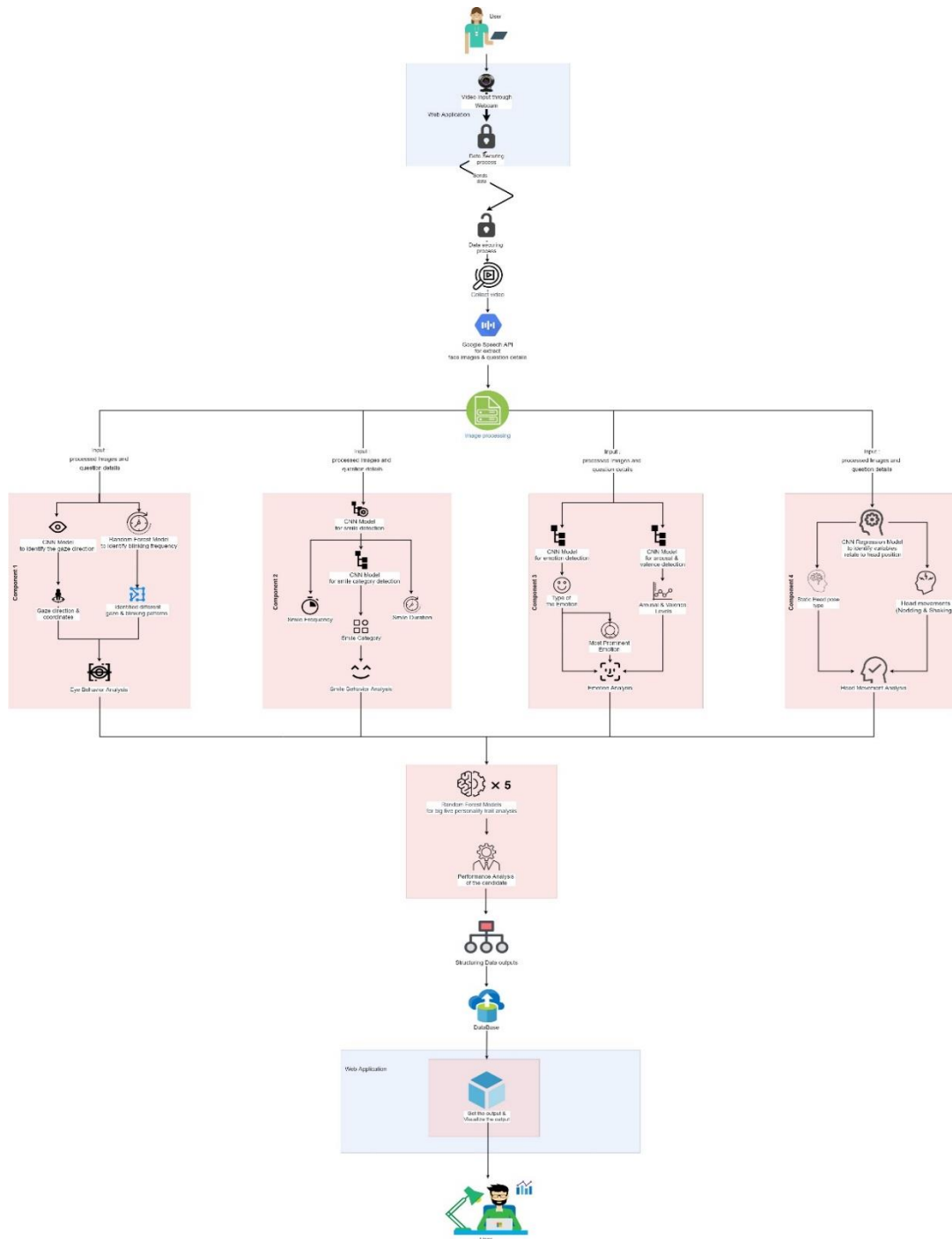


Figure 2.1: Overall System Diagram

Emotion analysis system diagram

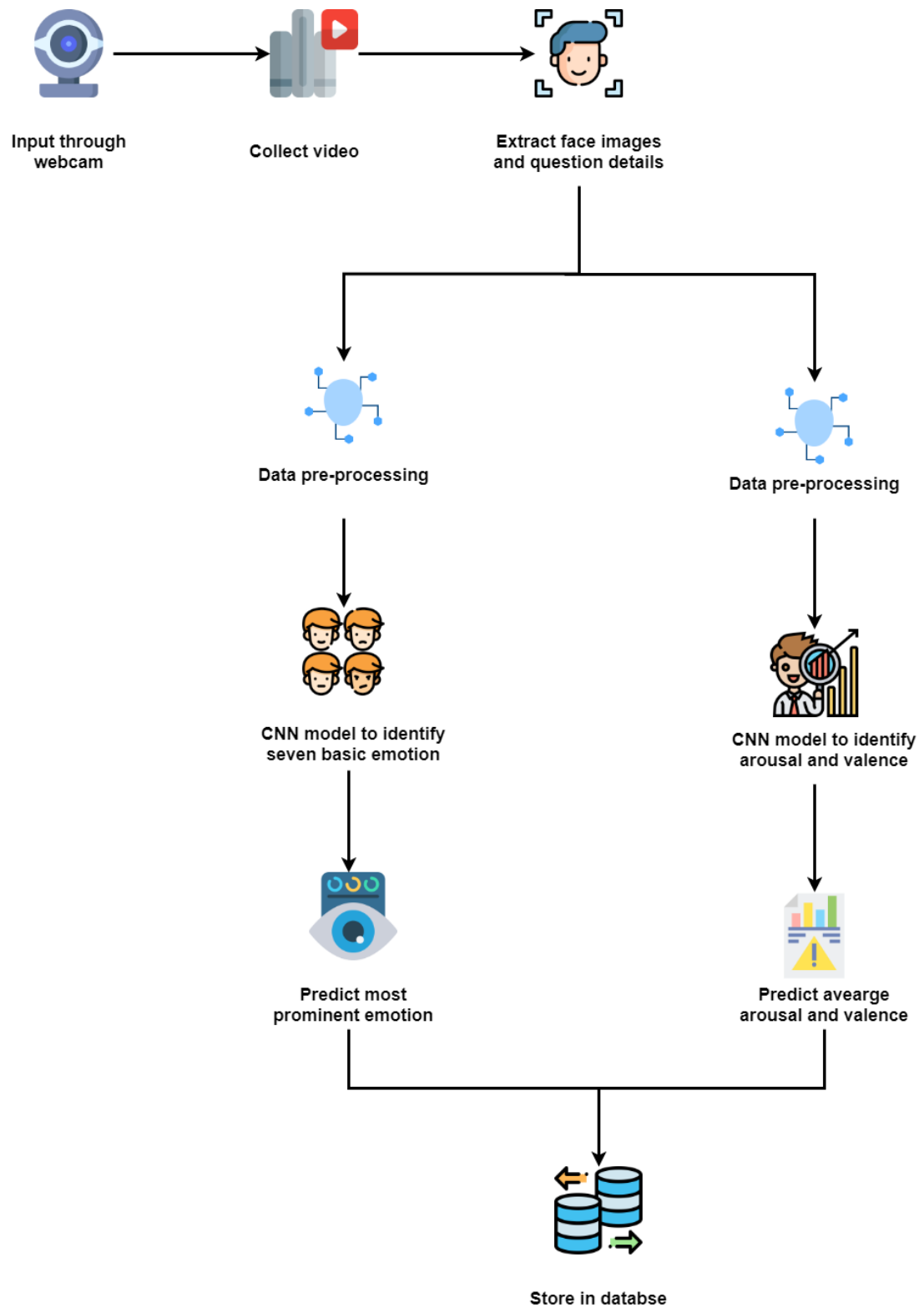


Figure 2.2: Emotion analysis system diagram

2.1.1 Data gathering

When developing a CNN-based model with a high accuracy rate, the dataset is essential. Distinct types of emotion datasets have been created around the world, since emotion detection is a highly popular area of research among researchers. There are primarily two kinds of datasets: annotated and unannotated. However, only annotated, and valid datasets were mostly used by researchers to train models of emotion recognition and arousal and valence detection.

- **The extend Chon-Kanade (CK+) dataset [13]**

The Chon-Kanade dataset was published in 2020 and rapidly grown in popularity as a testbed for algorithm development and evaluation. Due to the discovery of three restrictions, they published a new version called the Extended Chon-Kanade (CK+) dataset. CK+ dataset of 5876 images classified as Happy, Angry, Contempt, Disgust, Fear, Sadness, and Surprise. It consists of photographs of persons aged 18 to 50, with 69% of males and 31% of females.

- **Facial Expression Recognition 2013 dataset (FER2013) [14]**

FER2013 dataset consist around 30000 grayscale face images with 48*48 pixels. All the face images labeled into seven categories as Angry, Disgust, Fear, Happy, Sad and Neutral. All the images and the label were saved in the comma-separated value (CSV) file.

- **SPOS Dataset [15]**

The SPOS database contains 2,337 facial images. All images are gray and only the area of the person's face is in the images. All of the images have been labeled with six different emotions: Anger, Disgust, Fear, Happy, Sad and Surprise.

- **Affectnet Dataset [16]**

AffectNet includes over one million face images collected from the Internet by searching by 1250 emotion-related keywords in six different languages in 3 major search engines. The existence of seven

unique facial expressions (Happy, Sad, Surprise, Fear, Anger, Disgust and Contempt) and the intensity of valence and arousal were manually labeled in images.

2.1.2 Emotion detection model building

The models for emotion identification and arousal and valence detection make use of a deep learning technique called a Convolutional Neural Network, which is often considered the most efficient way for image processing. The emotion detection model is constructed using four convolutional layers to detect emotion when given the face image as input.

A ReLU activation function is used with each hidden convolutional layer. There are two significant advantages to the ReLU activation function. One significant advantage is the decreased probability that the gradient vanishes. Another advantage of ReLUs is its sparsity [17]. However, the output layer employs the Softmax activation function. The Softmax function enables the output layer to be converted into probabilities for each class.

To reduce the dimension, we then apply a pooling step for each hidden layer. Pooling involves down sampling features in order to reduce the number of parameters to learn during training. The most often used kind of pooling is max pooling. For each dimension of the input image, we execute a max-pooling operation that finds the largest value among the four pixels across a specified height and width, usually 2x2. The idea is that when categorizing an image, the maximum value has a greater probability of being more important. Batch normalization enables each hidden layer of the network to learn more freely. It is used to normalize the preceding layers' output. Fig 2.3 shows the model summary of the build emotion detection model.

Model: "sequential"		
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 48, 48, 64)	640
batch_normalization (Batch Normalization)	(None, 48, 48, 64)	256
activation (Activation)	(None, 48, 48, 64)	0
max_pooling2d (MaxPooling2D)	(None, 24, 24, 64)	0
dropout (Dropout)	(None, 24, 24, 64)	0
conv2d_1 (Conv2D)	(None, 24, 24, 128)	73856
batch_normalization_1 (Batch Normalization)	(None, 24, 24, 128)	512
activation_1 (Activation)	(None, 24, 24, 128)	0
max_pooling2d_1 (MaxPooling2D)	(None, 12, 12, 128)	0
dropout_1 (Dropout)	(None, 12, 12, 128)	0
conv2d_2 (Conv2D)	(None, 12, 12, 256)	295168
batch_normalization_2 (Batch Normalization)	(None, 12, 12, 256)	1024
activation_2 (Activation)	(None, 12, 12, 256)	0
max_pooling2d_2 (MaxPooling2D)	(None, 6, 6, 256)	0
dropout_2 (Dropout)	(None, 6, 6, 256)	0
conv2d_3 (Conv2D)	(None, 6, 6, 512)	1188160
batch_normalization_3 (Batch Normalization)	(None, 6, 6, 512)	2048
activation_3 (Activation)	(None, 6, 6, 512)	0
max_pooling2d_3 (MaxPooling2D)	(None, 3, 3, 512)	0
dropout_3 (Dropout)	(None, 3, 3, 512)	0
flatten (Flatten)	(None, 4608)	0
dense (Dense)	(None, 512)	2359808
batch_normalization_4 (Batch Normalization)	(None, 512)	2048
activation_4 (Activation)	(None, 512)	0
dropout_4 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 256)	131328
batch_normalization_5 (Batch Normalization)	(None, 256)	1024
activation_5 (Activation)	(None, 256)	0
dropout_5 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 7)	1798
activation_6 (Activation)	(None, 7)	0
Total params: 4,048,671		
Trainable params: 4,046,215		
Non-trainable params: 3,456		

Figure 2.3: Emotion Detection Model Summary

2.1.3 Training the model

FER2013, CK +, and SPOS datasets were used to train the emotion detection model. Initialization includes importing all of the required libraries such as NumPy, Pandas, Keras, Tensorflow and Matplotlib. Matplotlib is used to visualize the result of the model [18]. Prior to training the model, some reusable methods were created for data pre-processing technologies that will work with any dataset. After importing the dataset, all image inputs will be resized to the appropriate input size of the model. After successfully loading all images and labels, the database is split into a training database and a validation database with a 4:1 aspect ratio, respectively.

Once the number of epochs and batch size have been chosen, the model starts to train using the parameters provided. The length of time required to train the model

is dependent on the size of the database and the number of epochs. Also, at the end of each epoch, if the validation accuracy of the emotion detection model is higher than the validation accuracy of the previous epochs, a callback function is used to identify and save the model that performed best as shown in the Fig 2.4.

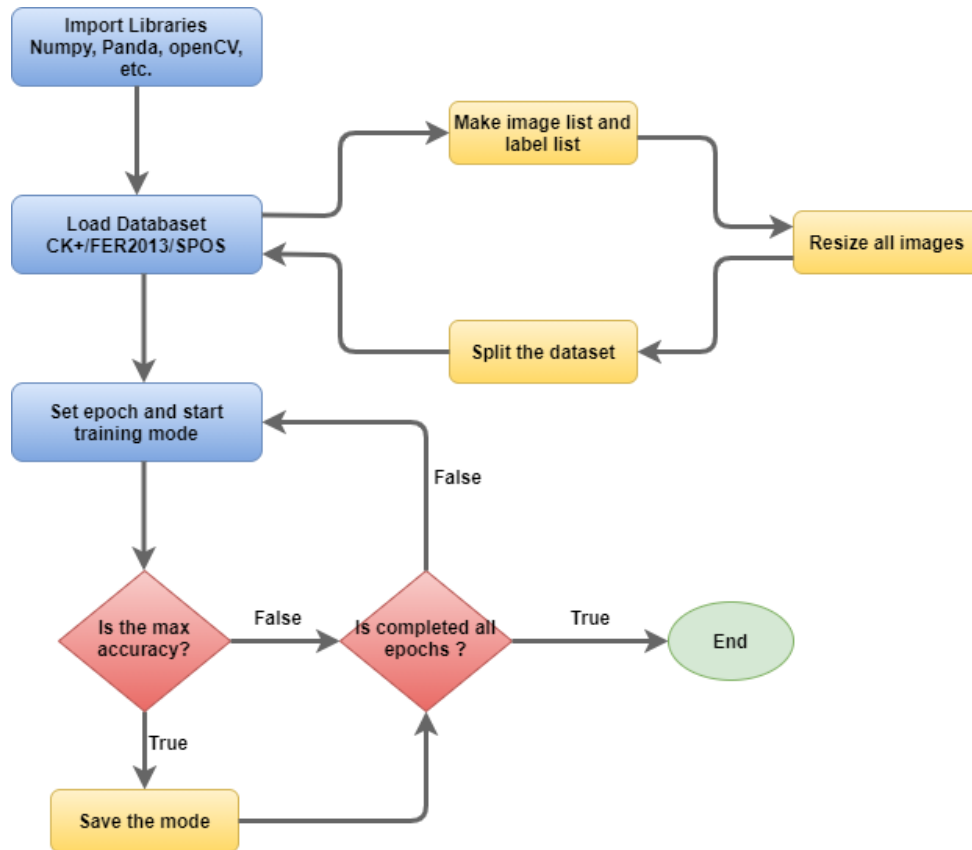


Figure 2.4: Model training process

2.1.4 Arousal and valence detection model

A CNN Regression model was developed using 4 convolutional layers and two dense layers with rectified linear activation function (ReLU) activation function. Because this is a CNN-type model, the ReLU activation function is used.

Initially, image is resized to the 256*256 pixel and converted to grayscale image before sending it to the model. The input image converted to the image type contiguous array using `ascontiguousarray` function in Numpy library. Contiguous array converted to the tensor image using Torchvision transform function. The pre-processed image is used to train the model as well as get prediction the result of the image.

1.1.1 Requirements Gathering and Analysis

We reviewed various aspects of our research idea in the initial stages and started exploring current systems, apps, and researchers that were utilized to evaluate interview behaviors. Additionally, as mentioned in the introduction, the study concentrated on obtaining the most recent emotion detection systems and research, studied how those variables influenced interviewers' conduct, and gained a solid understanding and familiarity with the most recent technology for using it. The majority of studies utilized a CNN-based method to predict emotions with a high degree of accuracy. The goal of the research was to optimize the system design using previously mentioned techniques and technologies. Finally, all functional and non-functional requirements are determined, and a system requirement specification was created.

A survey was conducted to determine the interviewer's impression of changes in the interviewee's emotions and behaviors. The survey was attended by all 30 participants from major companies and lecturers that regularly conduct virtual interviews / viva. The survey's results suggested that the system would be very beneficial for interviewers. Table 2.1 highlights some of the questions contained in the Google survey's questionnaire.

Table 2.1: Survey questions

Question	Yes (%)	No (%)
Do you prefer traditional interview over virtual interview?	70	30
Can you determine the interviewee genuineness through a virtual interview?	10	90
Can you track emotion of interviewee?	71.5	28.5

As a beginning point for the study project, a set of specific requirements are clearly defined, evaluated, and properly documented using the above information.

2.1.5 Design the system

Prior to the system's implementation, the whole system and its subsystems were designed. The connections, inputs, and outputs of the whole system and subsystems were identified and designed. System design helps in specifying hardware and system requirements and also helps in defining the overall system architecture. Under Designing the proposed system is mainly focused on developing a CNN based prediction model, user-friendly User Interfaces and the database design.

The CNN based model's accuracy is more significant to predict values more accurately to predict personality traits of the interviewee. Therefore, model building with the latest technologies were considered. Also, more consideration was given to the database and the database design because of the handling of very sensitive data of the interviewee. Also, a user-friendly user interface was developed to show the outcomes to an interviewer in a straightforward way.

2.2 Testing and Implementation

2.2.1 Identifying the emotion and detect the most prominent emotion.

After the emotion detection component retrieved the frames and the face coordinators along with the question numbers which detected using google speech to text API, prediction starts. Seven basic emotion, arousal and valence are identified in each frame retrieved by the emotion detection component.

After predicting the emotion, arousal rate and valence rate of all frames of one question. The average predicted result of each emotion, arousal and valence are saved in the database. Also, those data are used to predict the most prominent emotion of the interviewee. The most prominent emotion is calculated as follows.

The most prominent emotion calculation

Let us take,

predicted value for happy emotion of i^{th} frame as PH_i

predicted value for sad emotion of i^{th} frame as PS_i

predicted value for contempt emotion of i^{th} frame as PC_i

predicted value for surprise emotion of i^{th} frame as PR_i

predicted value for disgust emotion of i^{th} frame as PD_i

predicted value for angry emotion of i^{th} frame as PA_i

predicted value for fear emotion of i^{th} frame as PF_i

N = Total Number of frames of a whole video or a question

Average of predicted values of Happy = $\frac{1}{N} \sum_{i=1}^N PH_i$

Average of predicted values of Sad = $\frac{1}{N} \sum_{i=1}^N PS_i$

$$\text{Average of predicted values of Contempt} = \frac{1}{N} \sum_{i=1}^{i=N} PC_i$$

$$\text{Average of predicted values of Surprise} = \frac{1}{N} \sum_{i=1}^{i=N} PR_i$$

$$\text{Average of predicted values of Disgust} = \frac{1}{N} \sum_{i=1}^{i=N} PD_i$$

$$\text{Average of predicted values of Angry} = \frac{1}{N} \sum_{i=1}^{i=N} PA_i$$

$$\text{Average of predicted values of Fear} = \frac{1}{N} \sum_{i=1}^{i=N} PF_i$$

After predicting the average values of a frameset of a question, the most prominent emotion has been calculated by identifying max values among from the average values of the happy, sad, contempt, surprise, disgust, angry and fear.

Using the same method, average arousal and the valence rate of each question is calculated. Average values of each seven emotion, arousal and valence in each question along with the question number is saved in the database.

2.2.2 Tools and technologies.

Frontend-technologies

React/Redux

Facebook and Instagram developed and maintain this feature. React is a JavaScript library for the development of user interfaces. Providing an overview of MVC architecture. Appropriate for large-scale web applications that access and change data in real time without reloading the whole page.

NPM

The Node Package Manager is a package manager written in JavaScript for the Node platform. It builds modules so that node can find them and resolves dependency issues intelligently. It may be configured to handle a broad variety of situations. It is the most often used programming language for publishing, finding, installing, and creating node applications.

Backend Technologies

Python and other Libraries

Python is a high-level programming language with a strong resemblance to the English language's syntax. Python is a very productive programming language that is completely free to use and share. Python's standard library is extensive and contains functions that are helpful for almost everything.

- NumPy

NumPy is a Python third-party module for executing complicated mathematical calculations.

- PyTorch

PyTorch has two main features: Tensor computing with high GPU acceleration (like NumPy) and Automated distinction for neural network development and training.

- Flask

Flask is a Python web application framework. It was developed by Armin Ronacher, who led a team of international Python enthusiasts called Poocco. This framework is lightweight; there are few dependencies to maintain and scan for security vulnerabilities.

- Matplotlib

Matplotlib is a cross-platform data visualization and graphical plotting library written in Python for use with NumPy's numerical extension. It helps in the creation of accuracy and loss curves.

- TensorFlow

Tools

Jupyter Notebook

Jupyter is a free, open-source, interactive web tool known to as a programming notebook that allows researchers to combine software code, numerical output, explanatory prose, and visual features into a single document.

Database handling

MongoDB is an open-source database management system (DBMS) that uses a document-oriented database model that supports various forms of data. As a document database, MongoDB makes it easy for developers to store structured or unstructured data. It uses a JSON-like format to store documents.

2.2.3 Front-end implementation

Dashboard

The TIN system dashboard displays the overall analysis of the interview video. The interviewer displays the most prominent emotion and its percentage ratio in the emotion analysis section.

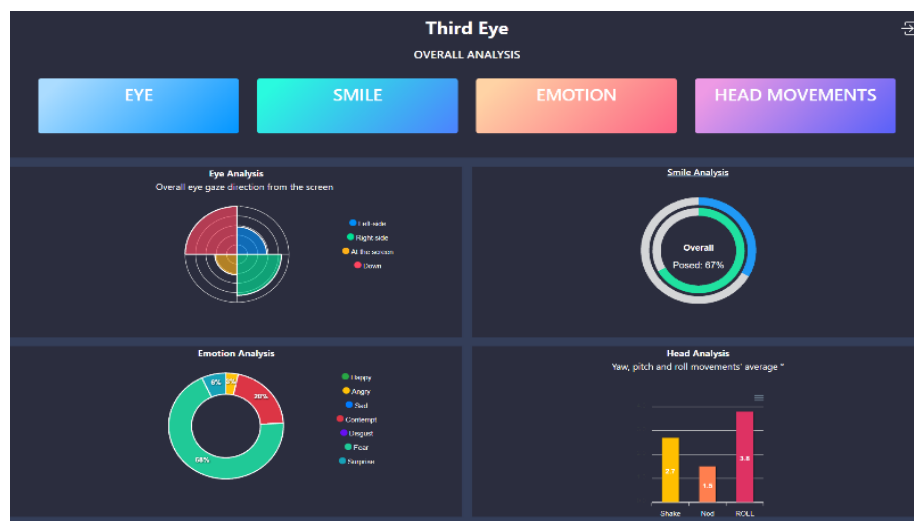


Figure 2.5: Dashboard of TIN system

Emotion Analysis Component

As illustrated in Fig 2.6, the predicted emotional status, arousal rate, and valence rate, as well as the question number for each question, are displayed in the emotion analysis user interface.

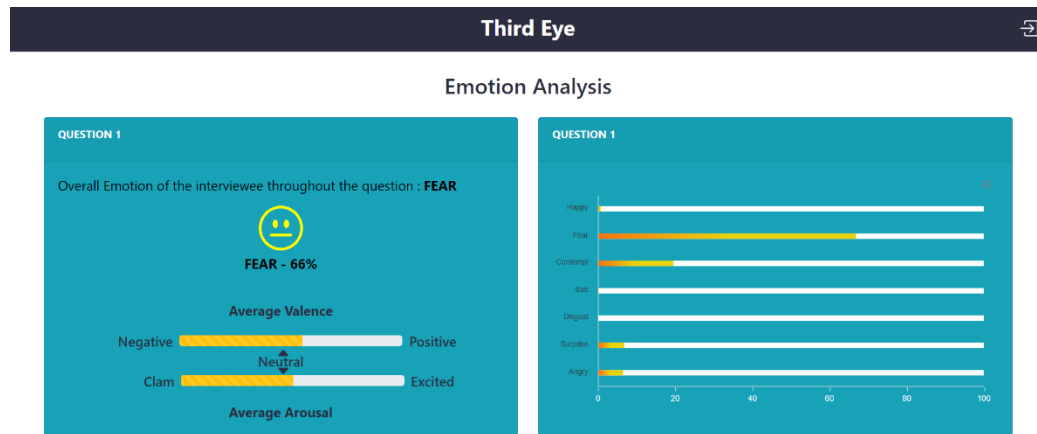


Figure 2.6: Emotion analysis user interface

2.2.4 Testing

Several evaluations were conducted at this phase to test the implemented system.

Unit Testing –

Each component that has been implemented under relief optimization has been subjected to unit testing.

Component Testing –

Component testing has been done by combining several units, component testing.

Integration Testing –

Integration Testing has been conducted to ensure that each component communicates properly with the others.



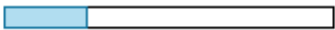
System Testing –







After connecting the components, the entire system (Web application) is tested to ensure appropriate operation.




After developing the emotion detection, arousal and valence detection models, unit testing was done to test the prediction accuracies of the system. Using several sample datasets images, the most prominent emotion with its percentage was identified and compared whether it matches the expected output. Furthermore, arousal and valence also follow the same procedure.

2.2.5 Test Cases

Table 2.2: Test case

Test ID: 001	
1) Test Description: An image of a person with contempt emotion	
Test Data: 	Expected Outputs: Emotion – Contempt Arousal: 0.597 Valence: -0.524 (Low than 0.05 error)
Actual Output:	
<div> <div> Emotion - Contempt - 85.4% Valence= -0.526 Arousal = 0.585 </div> <div> <p>Valence</p> <p>Negative  Positive</p> <p>Clam  Excited</p> <p>Arousal</p> </div> </div>	
Test Status: Pass	
2) Test Description: An image of a person with happy emotion	

<p>Test Data:</p> 	<p>Expected Output:</p> <p>Emotion – Happy</p> <p>Arousal: 0.171</p> <p>Valence: 0.815</p> <p>(Low than 0.05 error)</p>
<p>Actual Output:</p> <div data-bbox="322 595 1399 855"> <p>Emotion - Haapy- 82.0%</p> <p>Valence= 0.820</p> <p>Arousal = 0.175</p> <div> <p>Valence</p> <p>Negative  Positive</p> <p>Clam  Excited</p> <p>Arousal</p> </div> </div> <p>Test Status: Pass</p>	
<p>3) Test Description: An image of a person with sad emotion</p>	
<p>Test Data:</p> 	<p>Expected Output:</p> <p>Emotion – Sad</p> <p>Arousal: -0.424</p> <p>Valence: -0.723</p> <p>(Low than 0.05 error)</p>
<p>Actual Output:</p> <div data-bbox="322 1507 1369 1767"> <p>Emotion - Surprise- 81.5%</p> <p>Valence= 0.325</p> <p>Arousal = 0.785</p> <div> <p>Valence</p> <p>Negative  Positive</p> <p>Clam  Excited</p> <p>Arousal</p> </div> </div> <p>Test Status: Pass</p>	
<p>4) Test Description: An image of a person with surprise emotion</p>	

<p>Test Data:</p> 	<p>Expected Output:</p> <p>Emotion – Surprise</p> <p>Arousal: 0.783</p> <p>Valence: 0.329</p> <p>(Low than 0.05 error)</p>
<p>Actual Output:</p> <div data-bbox="330 607 1378 869"> <p>Emotion - Surprise- 81.5%</p> <p>Valence= 0.325</p> <p>Arousal = 0.785</p> <div> <p>Valence</p> <p>Negative  Positive</p> <p>Arousal</p> <p>Clam  Excited</p> </div> </div> <p>Test Status: Pass</p>	

2.3 Commercialization aspects of the product

Target Audience

Third Eye Interviewer Analyzer is intended to provide a pleasant user experience, particularly for human resource managers in any business and lecturers at any institution or institute that conducts virtual viva/interviews. The primary target users will be university instructors and lecturers, as well as interviewers who need to monitor interviewees for human resource management purposes.

Demand for the system

Virtual interviews were becoming more popular even before the pandemic. However, as a result of the country's lockdown, businesses have altered their health policies. Virtual interviews are used by almost all businesses to interview individuals while they are in various places. They are now used by every organization to hire new employees.

This method can be used to assess the applicants' performance and rank them in order to find the best candidate for a job. In university viva circumstances, students and participants may be less real in their viva sessions, prompting inspectors and lecturers to doubt their honesty and sincerity.

Marketing Plan

Our initially the marketing strategy is to introduce the system to IT recruiting companies giving the opportunity to improve the system with more features. As the second and third step we will introduce the system other local companies and universities in Sri Lanka. After building the local market, we will introduce our system globally to foreign companies and universities.

Pricing

The system will consist of 3 categories. The initial category will be determined as the Pro category which will provide 20 interviews per month. Followed by the enterprise category which will let the user analyze 30 interviews per month. And finally, the Community category will contain 100 interviews per month.

Budget

It costs \$1.25 per month to register the systems' domain name, and around \$30 per month to assign a server to deploy the model. A database needs be managed to store data, and the database costs \$35 per month. Thus, the final and total cost for the entire system is approximately \$66.25 each month.

3 RESULT

3.1 Results

The CNN model was built to detect emotion of interviewees as a main objective in the system. After developing the model, the main goal was identified to be determining and increasing the model's accuracy rate. The model was trained using several datasets and was evaluated comparing all datasets and accuracy rates obtained from each dataset.

Trained using FER2013 dataset

The FER2013 dataset was used to train the basic model. It includes 32,398 face pictures, which are split into a training set of 28,709 photos and a validation set of 3,589 images. When trained on the FER2013 database, the model achieved validation and training accuracy rates of 60% and 80%, respectively with 30 epochs as shown in the Fig 3.1. Fig 3.2 shows the loss curve obtained for a emotion detection CNN model.

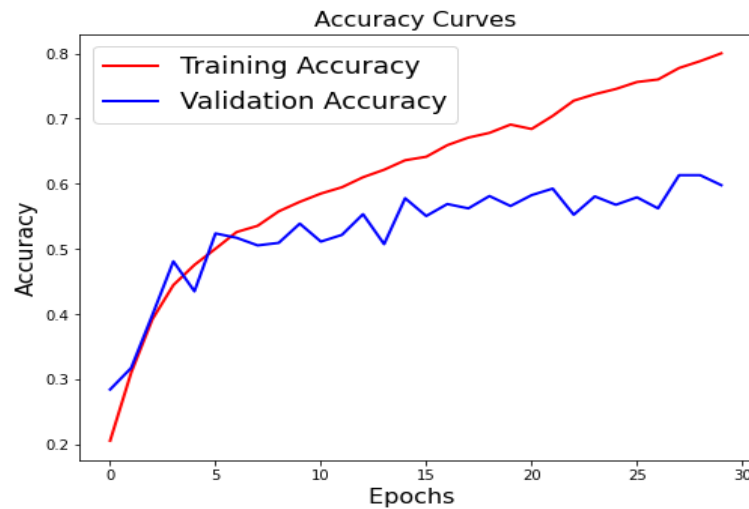


Figure 3.1: Accuracy curves of model training by FER2013 Dataset

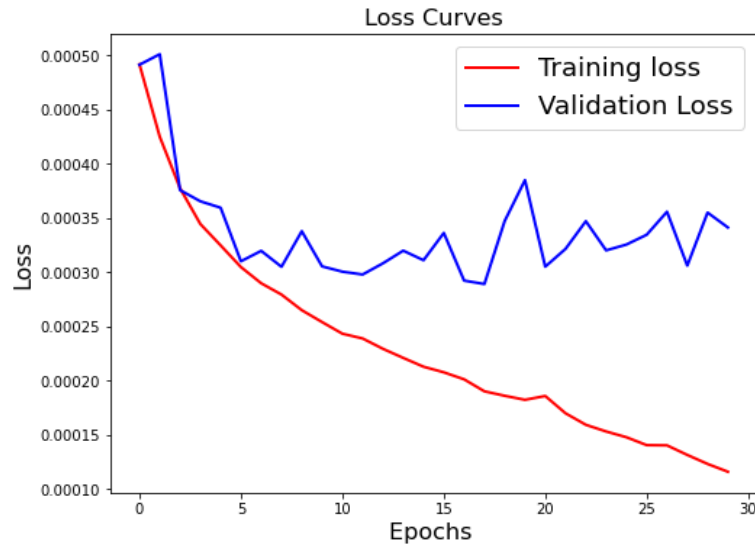


Figure 3.2: Loss curves of model training by FER2013 Dataset

Additionally, this model was retrained using enhanced pictures from the FER2013 collection. However, when compared with the previous result of the model, the accuracy was identified to be less.

Trained the model using the CK+ dataset.

The model was trained on the CK+ dataset, which contains 5,876 facial images categorized into seven emotional states. After splitting the dataset, the model began training and validating on 4701 and 1175 faces, respectively. As shown in Fig 3.3 training and validation had a maximum accuracy rate of about 62%.

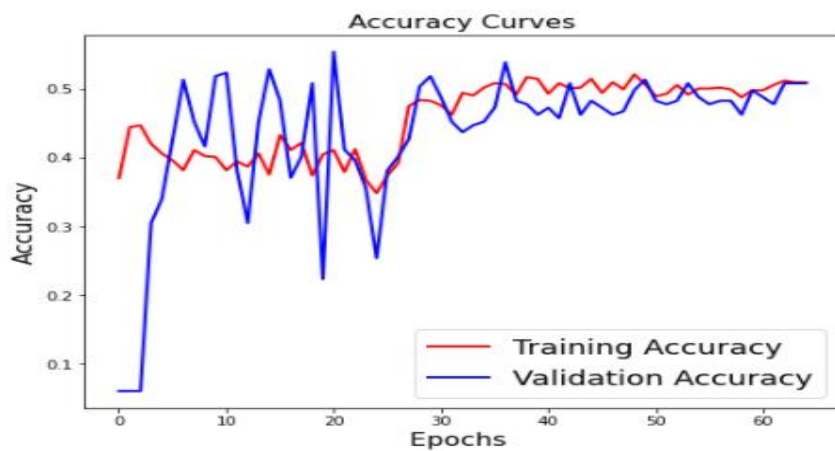


Figure 3.3: Accuracy curves of model training by CK+ Dataset

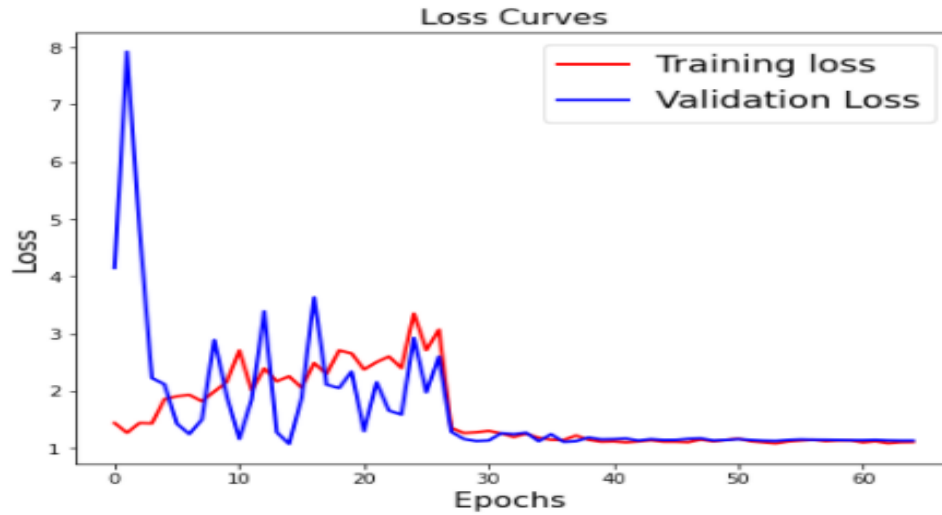


Figure 3.4: Loss curves of model training by CK+ Dataset

Trained the model using the SPOS dataset.

After that, the emotion identification model was trained using the SPOS database, which includes about 2,000 pictures classified into six emotion categories. When trained on the SPOS dataset, the model's training and validation accuracy was increased to be over 96%. The accuracy and loss curves for an emotion detection CNN model trained on the SPOS dataset are shown in Figures 3.5 and 3.6 respectively.

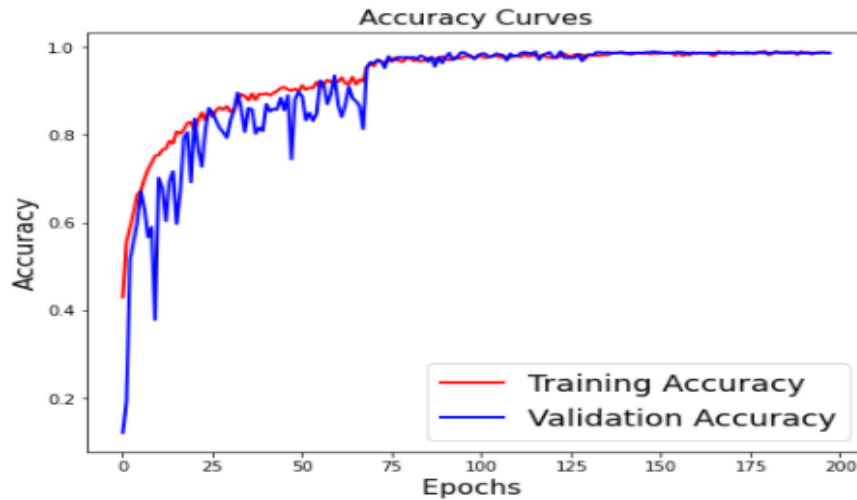


Figure 3.5: Accuracy curves of model training by SPOS Dataset

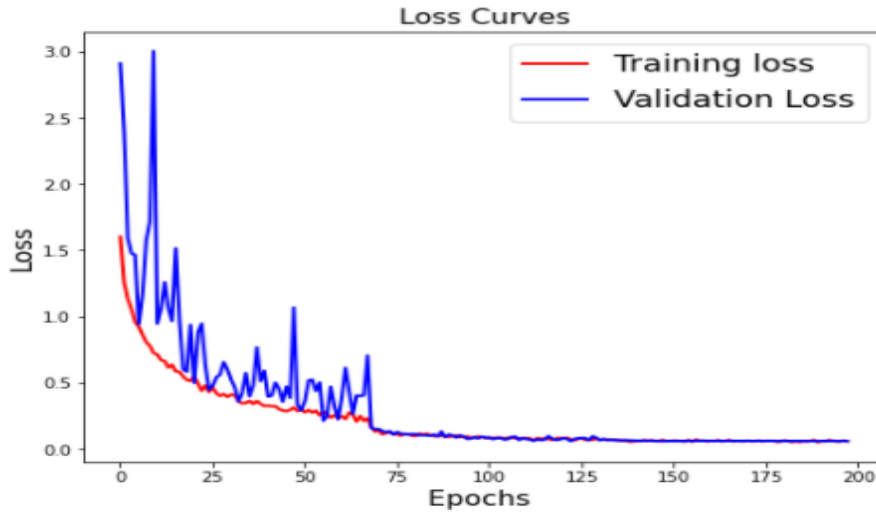


Figure 3.6: Loss curves of model training by SPOS Dataset

The accuracies obtained by training the model with the 3 datasets mentioned were analyzed and compared to differentiate the most suitable dataset for the model.

Fig 3.7 shows the main difference of the images in the SPOS database which contributed to the highest model performance had cropped images. But in CK + images, the database contained an unwanted region. In order to observe more, CK + dataset was modified and was used to train the model again.

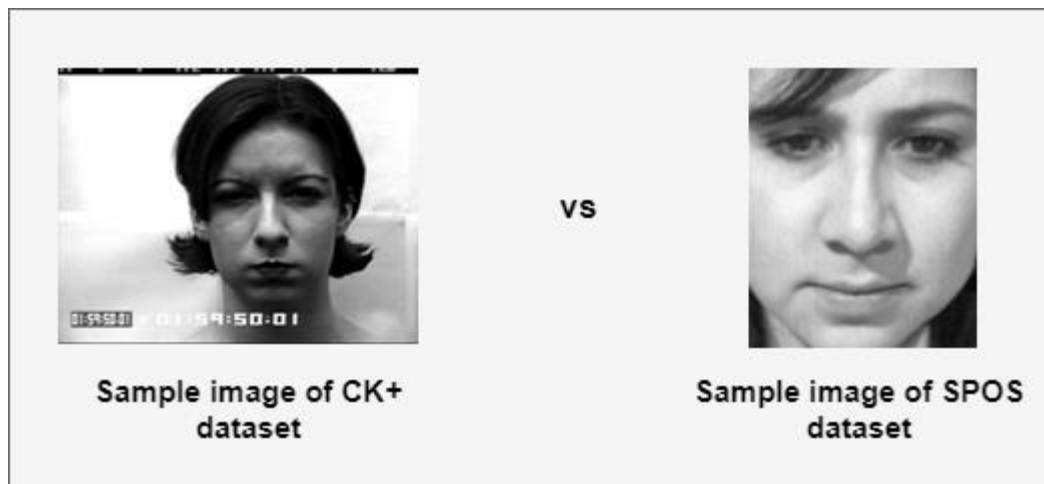


Figure 3.7: Difference between CK+ dataset and SPOS dataset

Trained using modified CK+ dataset

Before starting the training, all the images in the CK + dataset, the face region, were cropped. The face coordinates in the images were identified using the Dlib face detector. Using the detected face coordinates and OpenCV, the face area was cropped as shown in the Fig 3.8. Cropped faces were stored in the appropriate folders based on the image's emotional category.



Figure 3.8: Different between images of original CK+ and modified CK+ dataset

The modified dataset was expanded by augmenting the images.

After training the model with the enhanced CK + database, the validation and training accuracy of the emotion detection model were improved to 92% and 96% respectively as shown in Fig 3.9.

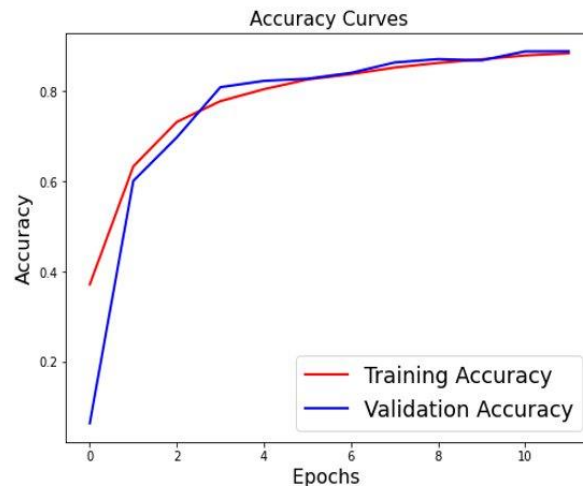


Figure 3.10: Accuracy curves of model training by modified CK+ Dataset

The loss curves for an emotion detection CNN model trained on the modified CK+ dataset is shown in Figures 3.10

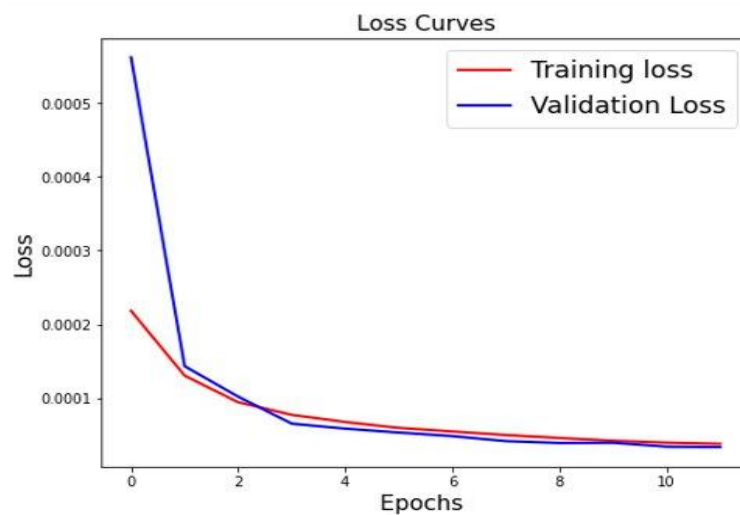


Figure 3.9: Loss curves of model training by modified CK+ Dataset

3.2 Research findings

When building a multi-class classification model for the system using several CNN models, accuracy becomes a critical factor. When attempting to improve the accuracy of a CNN-based emotion detection model, it was identified that the dataset used to train the model had a significant impact on the model's performance.

Prior to training the model, the database should be cleaned by removing parameters that are not required for detection. This enables the CNN model to readily absorb features, thus increasing the model's accuracy.

Additionally, training using augmented images is important for model performance. If the model was practiced with augmented images, the model will be able to study and detect facial emotions from various angles resulting in a significant impact on the system's performance.

Furthermore, increment of the number of classes was identified to be affecting the accuracy level of the CNN model. As an example, the model trained by the SPOS dataset gave more than 94% accuracy rate. But it contains only six emotion categories.

Table 3.2.1 shows a comparison of training and validation accuracy between the datasets FER, CK+ and SPOS datasets.

Table 3.1: Dataset comparison of emotion detection model

Dataset Name	Number of Labels	Training Accuracy (%)	Validation Accuracy (%)
FER	7	80	80
Original CK+	7	65	62
Cropped & Augmented CK+	7	96	92
SPOS	6	96	94

Fig 3.2.1 shows the Receiver operating characteristic curve obtained for emotion detection CNN model.

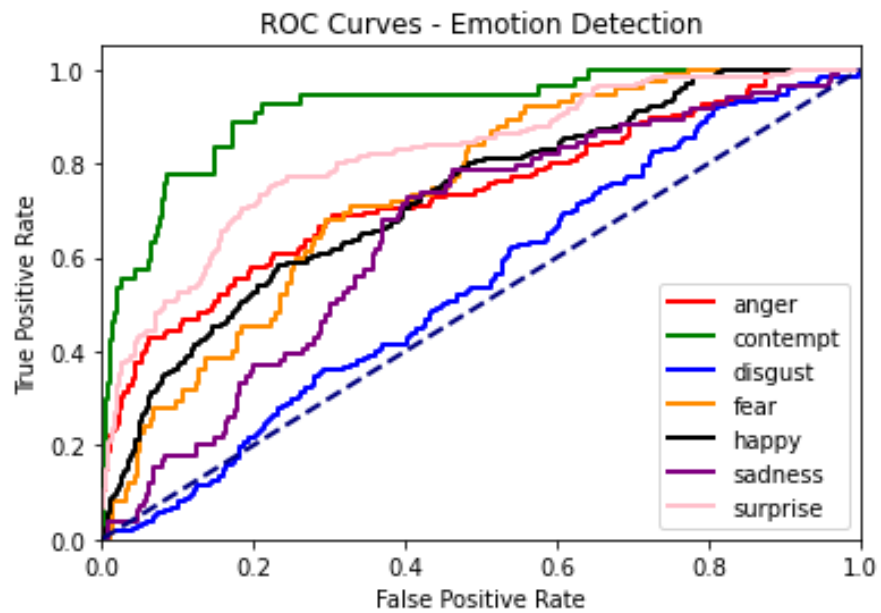


Figure 3.11: Receiver operating characteristic curve

3.3 Discussion

In the interviewee emotion detection scenario, the main responsibility is to identify and analyze the basic seven emotions, arousal rate and the valence rate of the interviewee. To achieve this task two CNN models were build, trained and tested using valid datasets.

After building the emotion detection CNN model with four convolution layers, model was trained with three different datasets. Initially, the model gave nearly about 60% of the accuracy rate for CK+ and FER 2013 datasets. Model achieved 96% accuracy rate when training using the SPOS dataset. After analyzing those results, CK+ dataset was cleaned by cropping the facial region of the image using data pre-processing techniques. Unexpectedly model accuracy was increased from 30%.

The highest accuracy rate was given by the SPOS dataset. But it contains only six emotion categories. Cleaned CK+ dataset gave a rather less accuracy but provided seven emotion categories. The system used the model trained using modified CK+ dataset along with arousal and valence detection regression CNN model which obtained 67% accuracy rate to predict and analysis emotional state of the interviewee.

The developed models may be utilized to detect the interviewee's emotion, arousal, and valence in each frame of the interview video. These predicted results were utilized to determine the interviewee's prominent emotion, as well as his or her average arousal and valence.

4 CONCLUSION

In this paper, two detection models were proposed, which can be used to predict the basic seven emotions and the arousal and valence of a facial images. The training algorithm is a Convolutional Neural Network, one of the well-known deep learning algorithms for image processing.

The proposed emotion detection model is used for predicting the emotion of the interviewee which provides a useful result which can be used for optimizing the personality trait of the interviewee. Also, another CNN model was used to identify the rate of valence and arousal where valence is the degree to which an emotional expression is positive, and arousal represents the calming or exciting nature of the emotional display with the highest accuracy rate.

The model is trained, built, and validated using image processing methods, and the Convolutional Neural Network is a machine learning approach that is mostly used for image processing.

The results of the models help to the user (interviewer) to get an idea about the variation of the emotion of the interviewee. Using the predicted result, the system calculated the most prominent emotion of the interviewee throughout the interview and each question asked by the interviewer. Those data are shown in the user interfaces to give an idea about interviewee's emotional state to interviewer.

The predicted result of the interviewee are used to analysis the group behavior with average emotional state of other interviewees who faced the same interview.

Furthermore, identified average values of arousal and valence rates and the most prominent emotion and the most second prominent emotion of the interviewee are sent to predict the rate of personality traits of the interviewee.

Considering the limitations of the system, system get too much time to predict the output result. Dlib face detector which are used to detect the faces and face coordinators of the interviewee is take too much time predict the outcomes.

As for future directions, due to the lack of time and resources, various adaptations, tests, and experiments are left for the future. Future works will include

the detected the emotion and predict the personality traits of interviewee in live interview sessions.

5 REFERENCES

- [1] T. Tashu, S. Hajiyeva and T. Horvath, "Multimodal Emotion Recognition from Art Using Sequential Co-Attention", *Journal of Imaging*, vol. 7, no. 8, p. 157, 2021. Available: 10.3390/jimaging7080157..
- [2] K. Shan, J. Guo, W. You, D. Lu and R. Bie, "Automatic facial expression recognition based on a deep convolutional-neural-network structure," 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), 2017, pp. 123-128, doi: 10.1109/SERA.2017.7965717.
- [3] S. Gupta, "Facial emotion recognition in real-time and static images," 2018 2nd International Conference on Inventive Systems and Control (ICISC), 2018, pp. 553-560, doi: 10.1109/ICISC.2018.8398861.
- [4] Shin, Chung and Park, "Detection of Emotion Using Multi-Block Deep Learning in a Self-Management Interview App", *Applied Sciences*, vol. 9, no. 22, p. 4830, 2019. Available: 10.3390/app9224830.
- [5] R. Halder, S. Sengupta, A. Pal, S. Ghosh and D. Kundu, "Real Time Facial Emotion Recognition based on Image Processing and Machine Learning", *International Journal of Computer Applications*, vol. 139, no. 11, pp. 16-19, 2016. Available: 10.5120/ijca2016908707.
- [6] R. Halder, S. Sengupta, A. Pal, S. Ghosh and D. Kundu, "Real Time Facial Emotion Recognition based on Image Processing and Machine Learning", *International Journal of Computer Applications*, vol. 139, no. 11, pp. 16-19, 2016. Available: 10.5120/ijca2016908707.
- [7] J. . Kim, M.W and Hoon, Y.,Joo, and Park, “‘Emotion Detection Algorithm Using Frontal Face Image’, ICCAS2005, June 2-5, KINTEX, Gyeonggi-Do, Korea,” no. February 2014, pp. 2373–2378, 2015.
- [8] H. Siqueira, S. Magg and S. Wermter, "Efficient Facial Feature Learning with Wide Ensemble-Based Convolutional Neural Networks", *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, pp. 5800-5809, 2020. Available: 10.1609/aaai.v34i04.6037.

- [9] "Emotimeter Emotion detector For Pc 2021 (Windows 7/8/10 And Mac) - APK Flyer", APK Flyer, 2021. [Online]. Available: <https://apkflyer.com/download/emotimeter-emotion-detector-for-pc-2021-windows-7-8-10-and-mac/>. [Accessed: 09- Oct- 2021].
- [10] "Feely - An emotion detector on Windows PC Download Free - 1.1.1 - com.vladimir.feely", Appsonwindows.com, 2021. [Online]. Available: <https://appsonwindows.com/apk/4939332/>. [Accessed: 09- Oct- 2021].
- [11] "Release notes | Cloud Speech-to-Text Documentation | Google Cloud." <https://cloud.google.com/speech-to-text/docs/release-notes> (accessed Aug. 15, 2021).
- [12] D. E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, 2009.
- [13] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," *2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Work. CVPRW 2010*, no. July, pp. 94–101, 2010, doi: 10.1109/CVPRW.2010.5543262.
- [14] S. Wang et al., "A natural visible and infrared facial expression database for expression recognition and emotion inference," *IEEE Trans. Multimed.*, vol. 12, no. 7, pp. 682–691, 2010, doi: 10.1109/TMM.2010.2060716
- [15] T. Pfister, X. Li, G. Zhao, and M. Pietikainen, "Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 868–875, 2011, doi: 10.1109/ICCVW.2011.6130343.
- [16] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, pp. 18–31, 2019, doi: 10.1109/TAFFC.2017.2740923.

[17] A. De Brack, “Multimodal Emotion Recognition Table of Contents,” 2019.

[18] "Matplotlib: Python plotting — Matplotlib 3.4.3 documentation", Matplotlib.org, 2021. [Online]. Available: <https://matplotlib.org/>. [Accessed: 09- Oct- 2021].

APPENDICES

Appendix 1: Classification report of the emotion detection CNN model

	precision	recall	f1-score	support
anger	0.63	0.41	0.49	32
contempt	0.60	1.00	0.75	8
disgust	0.58	0.78	0.66	35
fear	0.85	0.90	0.87	21
happy	0.80	0.80	0.80	43
sad	0.75	0.78	0.76	10
surprise	1.00	0.90	0.95	47
accuracy			0.70	196
macro avg	0.61	0.79	0.75	196
weighted avg	0.61	0.80	0.74	196

Appendix 2: Training emotion detection model

```

curacy: 0.6166
Epoch 39/50
225/225 [=====] - 478s 2s/step - loss: 2.2481e-04 - accuracy: 0.6227 - val_loss: 3.4376e-04 - val_ac
curacy: 0.5294
Epoch 40/50
225/225 [=====] - 480s 2s/step - loss: 2.2401e-04 - accuracy: 0.6238 - val_loss: 4.1676e-04 - val_ac
curacy: 0.4667
Epoch 41/50
225/225 [=====] - 479s 2s/step - loss: 2.2609e-04 - accuracy: 0.6262 - val_loss: 3.6405e-04 - val_ac
curacy: 0.5322
Epoch 42/50
225/225 [=====] - 474s 2s/step - loss: 2.2072e-04 - accuracy: 0.6294 - val_loss: 3.0295e-04 - val_ac
curacy: 0.5862
Epoch 43/50
225/225 [=====] - 482s 2s/step - loss: 2.2743e-04 - accuracy: 0.6236 - val_loss: 2.7442e-04 - val_ac
curacy: 0.6085
Epoch 44/50
225/225 [=====] - 477s 2s/step - loss: 2.1985e-04 - accuracy: 0.6279 - val_loss: 3.0156e-04 - val_ac
curacy: 0.5759
Epoch 45/50
225/225 [=====] - 477s 2s/step - loss: 2.2122e-04 - accuracy: 0.6291 - val_loss: 2.5498e-04 - val_ac
curacy: 0.6289
Epoch 46/50
225/225 [=====] - 475s 2s/step - loss: 2.1816e-04 - accuracy: 0.6331 - val_loss: 2.8756e-04 - val_ac
curacy: 0.5996
Epoch 47/50
225/225 [=====] - 474s 2s/step - loss: 2.1816e-04 - accuracy: 0.6344 - val_loss: 3.5088e-04 - val_ac
curacy: 0.5520
Epoch 48/50
225/225 [=====] - 481s 2s/step - loss: 2.1707e-04 - accuracy: 0.6355 - val_loss: 2.8473e-04 - val_ac
curacy: 0.6149
Epoch 49/50
225/225 [=====] - 481s 2s/step - loss: 2.1581e-04 - accuracy: 0.6367 - val_loss: 2.6967e-04 - val_ac
curacy: 0.6266
Epoch 50/50
225/225 [=====] - 543s 2s/step - loss: 2.1474e-04 - accuracy: 0.6353 - val_loss: 4.1370e-04 - val_ac

```


Appendix 3: Dataset splitting function

```
def split_data(x, y, validation_split=.2):
    num_samples = len(x)
    num_train_samples = int((1 - validation_split)*num_samples)
    train_x = x[:num_train_samples]
    train_y = y[:num_train_samples]
    val_x = x[num_train_samples:]
    val_y = y[num_train_samples:]
    train_data = (train_x, train_y)
    val_data = (val_x, val_y)
    return train_data, val_data
```

Appendix 4: Dataset loading function

```
datasets.py
import random
import os
import cv2

#add if condition and path
class DataManager(object):
    """Class for loading fer2013 emotion classification dataset or
    imdb gender classification dataset."""
    def __init__(self, dataset_name='imdb',
                 dataset_path=None, image_size=(48, 48)):
        self.dataset_name = dataset_name
        self.dataset_path = dataset_path
        self.image_size = image_size
        if self.dataset_path is not None:
            self.dataset_path = dataset_path

        elif self.dataset_name == 'fer2013':
            self.dataset_path = '../datasets/fer2013/fer2013.csv'

        elif self.dataset_name == 'CK+':
            self.dataset_path = 'G:/1 Sem/RP/Datasets/aug/'
        elif self.dataset_name == 'spos':
            self.dataset_path = 'datasets/spos/'
        else:
            raise Exception(
                'Incorrect dataset name!')

> def get_data(self): ...
> def _load_fer2013(self): ...
> def _load_CK(self): ...
> def _load_spos(self): ...
```

Appendix 5: Plagiarism report

THE THIRD EYE INTERVIEW – EMOTION ANALYSIS

ORIGINALITY REPORT

18%

SIMILARITY INDEX

6%

INTERNET SOURCES

6%

PUBLICATIONS

12%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Sri Lanka Institute of
Information Technology

Student Paper

10%

2

Antoine Toisoul, Jean Kossaifi, Adrian Bulat,
Georgios Tzimiropoulos, Maja Pantic.
"Estimation of continuous valence and
arousal levels from faces in naturalistic
conditions", Nature Machine Intelligence,
2021

Publication

1%

3

Yashwanth Adepu, Vishwanath R Boga,
Sairam U. "Interviewee Performance Analyzer
Using Facial Emotion Recognition and Speech
Fluency Recognition", 2020 IEEE International
Conference for Innovation in Technology
(INOCON), 2020

Publication

1%

4

Biao Yang, Jinmeng Cao, Rongrong Ni, Yuyu
Zhang. "Facial Expression Recognition using
Weighted Mixture Deep Neural Network

<1%