PSYCHOMETRIC ANALYSIS TOOL

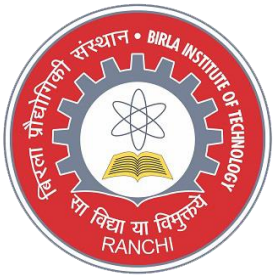
A project report submitted in partial fulfillment of the requirements for the award of the Degree of

BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE & ENGINEERING

BY

APURVA BHARGAVA (BE/25022/15)

PALLAVI JAIN (BE/25001/15)



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

BIRLA INSTITUTE OF TECHNOLOGY, MESRA
JAIPUR CAMPUS, JAIPUR

MO-2018

DECLARATION CERTIFICATE

This is to certify that the work presented in the project entitled "PSYCHOMETRIC ANALYSIS TOOL" in partial fulfillment of the requirements for the award of Degree of Bachelor of Engineering in Computer Science and Engineering of Birla Institute of Technology, Mesra, Ranchi, Extension Center Jaipur is an authentic work carried out under my supervision and guidance.

To the best of my knowledge, the content of this project does not form a basis for the award of any previous degree to anyone else.

Date: 18 December 2018

Dr. Shripal Vijayvargia
Associate Professor
Birla Institute of Technology, Mesra ,Ranchi
Extension Centre,Jaipur

CONTENTS

1. INTRODUCTION

Psychometrics, is a field of study concerned with the theory and technique involved behind psychological measurement. This field is primarily concerned with testing, measurement, and assessment, that is, objective measurement of skills and knowledge, abilities, attitudes, and personality traits. The two areas of research focus are- (i) the construction and validation of assessment instruments such as questionnaires, tests, raters' judgments, and personality tests, and (ii) measurement theory (e.g., item response theory; intra-class correlation).[1] The focus of the project is to employ machine learning models to automate certain aspects of psychometric analysis that usually require human intervention.

1.1 Problem Statement

To design a tool (Psychometric Analysis Tool) for interviewers and/or psychologists that powers judgement by determining the interviewee's psychological characteristics based on the detection and analysis of subjective factors such as facial expressions, speech emotions and written opinion.

1.2 Psychometrics and Machine Learning

Machine learning (ML) is the study of algorithms and statistical models that computer systems use to progressively improve their performance on a specific task. Machine learning algorithms build a mathematical model of sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.[2], [3] Machine learning will facilitate the evaluation of psychometric information outside of a testing context, both from past information (e.g. internet corpus) and from current information (e.g. live cameras, smartphones). This may turn out to be a boon or a dystopian nightmare depending on who uses it and how it is used. (Which side you see is itself an interesting psychometric datum). Take for instance a non-controversial topic like IQ. Machine learning will evaluate how trainable a person is in general. What is the best way to train that person? Would this person do well in this new task? How long will the training last? All this without a test. Of course, psychometrics and psychometricians will still be needed to validate those algorithms.

This age-old domain is largely based upon the work of Charles Spearman, but the technological changes across the entire landscape of instruments, data, applications, and domains associated have brought machine learning techniques into psychometrics. Machine Learning is expected to revolutionize Psychometrics. IRT (Item Response Theory) psychometrics are usually based upon logistic regression techniques. However, the technique fails to promise best class of models for classification anymore. Machine Learning can be utilized to reveal candidate's strengths in the in the social components of collaborative problem solving, such as perspective taking, participation, and social regulation.

This can be achieved by simply analyzing a participant's video/audio recording, or social media data, leveraging Machine Learning techniques. It's already making some waves. Machine learning can be used to match employees with training programs that fit their profiles and career goals, predict responses to certain drugs (particularly neuropsychological disorders), and validation of a wider range of survey tools. Machine Learning techniques can extend to incorporate Virtual Reality technology. This will help towards expanding the scope of problem solving tasks, while enriching the resulting data stream. These techniques can also be applied to study difficult real life scenarios.

1.3 Literature Review

Facial or speech emotion recognition, as well as sentiment classification problem can be reformulated in mathematical terms as a classification task, which can very well be done using statistical machine learning models. Artificial neural networks or connectionist systems are computing systems vaguely inspired by the biological neural networks that constitute animal brains. Computational models of neural networks have been around for a long time, first model proposed was by McCulloch and Pitts.

Neural Network with Backpropagation: Neural networks are made up of a number of layers with each layer connected to the other layers forming the network. A feed-forward neural network or FFNN can be thought of in terms of neural activation and the strength of the connections between each pair of neurons. In FFNN, the neurons are connected in a directed way having clear start and stop place i.e., the input layer and the output layer. The layer between these two layers, are called as the hidden layers. Learning occurs through adjustment of weights and the aim is to try and minimize error between the output obtained from the output layer and the input that goes into the input layer. The weights are adjusted by process of back propagation (in which the partial derivative of the error with respect to last layer of weights is calculated). The process of weight adjustment is repeated in a recursive manner until weight layer connected to input layer is updated.

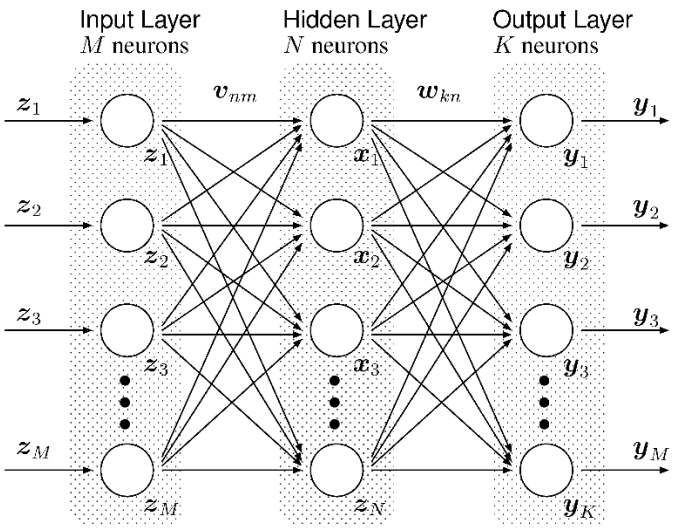


Figure 1

Convolutional Neural Networks (CNN): It is a variant of Multi- Layer Perceptron (MLP) which is inspired from vision, as observed in animals. Convolutional neural networks are designed to process two-dimensional (2-D) image.[4] A simple CNN architecture may consist of three types of layers namely convolution layer, sub sampling layer and the output layer. CNNs exploit spatially local correlation by enforcing a local connectivity pattern between neurons of adjacent layers. In the CNN algorithm, each sparse filter is replicated across the entire visual field. These units then form a feature maps, these share weight vector and bias. The gradient of shared weights is the sum of the gradients of the parameters being shared. Such replication in a way allows features to be detected regardless of their position in visual field. In addition to this, weight sharing also allows to reduce the number of free learning parameters. Due to this control, CNN tends to achieve better generalization on vision problems. CNN also make use of the concept of max-pooling, which is a form of non-linear down-sampling. In this method, the input image is partitioned into non-overlapping rectangles. The output for each sub-region is the maximum value. Recent deep FER systems generally focus on two important issues: overfitting caused by a lack of sufficient training data and expression-unrelated variations, such as illumination, head pose and identity bias.[5], [6]

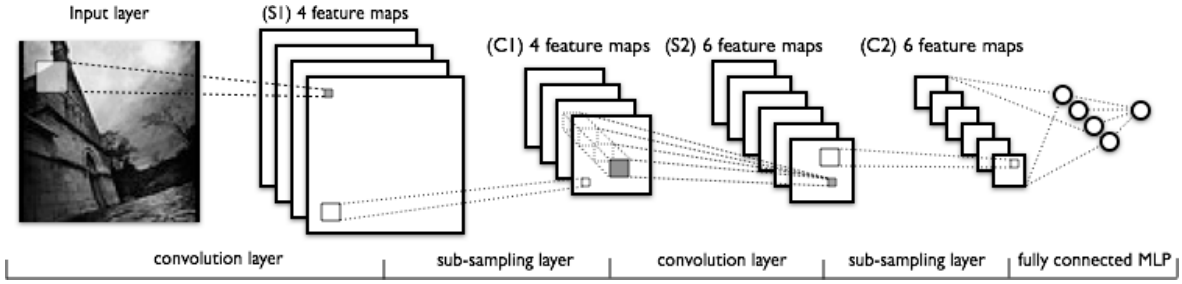


Figure 2

Recurrent Neural Network (RNN): It is a class of neural networks whose connections between neurons form a directed cycle. Unlike feedforward neural networks, RNN can use its internal "memory" to process a sequence of inputs, which makes it popular for processing sequential information. The "memory" means that RNN performs the same task for every element of a sequence with each output being dependent on all previous computations, which is like "remembering" information about what has been processed so far.

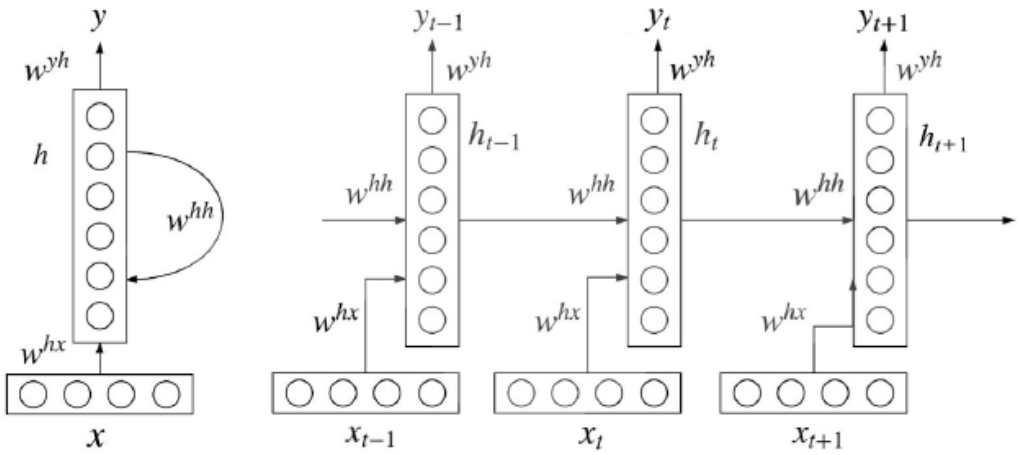


Figure 3

The left graph is an unfolded network with cycles, while the right graph is a folded sequence network with three time steps. The length of time steps is determined by the length of input. x_t is the input vector at time step t. h_t is the hidden state at time step t, which is calculated based on the previous hidden state and the input at the current time step.

$$h_t = f(w^{hh}h_{t-1} + w^{hx}x_t)$$

The activation function f is usually the tanh function or the ReLU function. w^{hx} is the weight matrix used to condition the input x_t . w^{hh} is the weight matrix used to condition the previous hidden state h_{t-1}. y_t is the output probability distribution over at step t.

$$y_t = softmax(w^{yh}h_t)$$

The hidden state h_t is regarded as the memory of the network. It captures information about what happened in all previous time steps. y is calculated solely based on the memory h_t at time t and the corresponding weight matrix w^{yh}.[7], [8]

Long Short Term Memory network (LSTM): It is a special type of RNN, which is capable of learning long-term dependencies. All RNNs have the form of a chain of repeating modules. In standard RNNs, this repeating module normally has a simple structure. However, the

repeating module for LSTM is more complicated. Instead of having a single neural network layer, there are four layers interacting in a special way. Besides, it has two states: hidden state and cell state.[9]

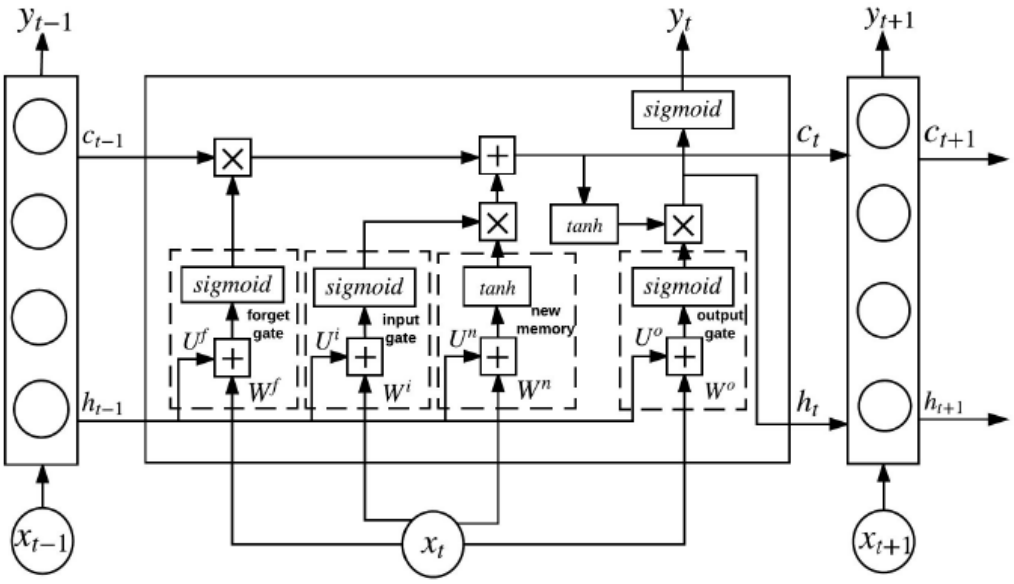


Figure 4

At time step t, LSTM first decides what information to dump from the cell state. This decision is made by a sigmoid function/layer σ, called the "forget gate". The function takes h_{t-1} (output from the previous hidden layer) and x_t (current input), and outputs a number in [0, 1], where 1 means "completely keep" and 0 means "completely dump".

$$f_t = \sigma(W^f x_t + U^f h_{t-1})$$

Then LSTM decides what new information to store in the cell state. This has two steps. First, a sigmoid function/layer, called the "input gate", decides which values LSTM will update. Next, a tanh function/layer creates a vector of new candidate values $\tilde{C}_t$, which will be added to the cell state. LSTM combines these two to create an update to the state.

$$i_t = \sigma(W^i x_t + U^i h_{t-1})$$

$$\tilde{C}_t = \tanh(W^n x_t + U^n h_{t-1})$$

It is now time to update the old cell state C_{t-1} into new cell state C_t. Note that forget gate f_t can control the gradient passes through it and allow for explicit "memory" deletes and

updates, which helps alleviate vanishing gradient or exploding gradient problem in standard RNN.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Finally, LSTM decides the output, which is based on the cell state. LSTM first runs a sigmoid layer, which decides which parts of the cell state to output, called "output gate". Then, LSTM puts the cell state through the tanh function and multiplies it by the output of the sigmoid gate, so that LSTM only outputs the parts it decides to.[9]

$$o_t = \sigma(W^o x_t + U^o h_{t-1})$$

$$h_t = o_t * \tanh(C_t)$$

Cosine Similarity: Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them. The cosine of 0° is 1, and it is less than 1 for any angle in the interval (0,π] radians. It is thus a judgment of orientation and not magnitude: two vectors with the same orientation have a cosine similarity of 1, two vectors oriented at 90° relative to each other have a similarity of 0, and two vectors diametrically opposed have a similarity of -1, independent of their magnitude. The cosine similarity is particularly used in positive space, where the outcome is neatly bounded in [0,1].

The cosine of two non-zero vectors can be derived by using the Euclidean dot product formula:

$$\mathbf{A}.\mathbf{B} = \big||\mathbf{A}|\big|\,\big||\mathbf{B}|\big|\,cos\theta$$

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

Jaccard Similarity: The Jaccard index, also known as Intersection over Union and the Jaccard similarity coefficient, is a statistic used for comparing the similarity and diversity of sample sets. The Jaccard coefficient measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}.$$

2. OBJECTIVES

Psychometric Analysis Tool acts as an interface between the two interacting parties, one of whom evaluates the other's psychological state and competence using the various modules provided by the software as mentioned further. It automates, to a large extent, the task of the interviewer or psychologist. It can be widely used by the companies for recruitment purposes. Also, a virtual psychologist may use it for evaluation of the patient's emotional state. This is achieved by various modules that analyze the responses of the interviewee or patient, which are in the form of text, speech, facial emotion, etc.

Objectives realised in the project are stated as follows:

2.1 Expression Recognition from Face

While the interviewee speaks or gives a verbal response, images are captured periodically from the webcam or video camera feed (where faces are detected in real time using Haar-cascade based face detection [22]). The facial expression recognition model classifies each image, calculating the class probability for each possible target class (angry, disgust, fear, happy, sad, neutral, or surprise). The two classes with highest probabilities are used for analysis.

2.2 Emotion Recognition from Speech

The interviewee's verbal response is recorded. The speech emotion recognition model is used on the sequences of pressure levels (of successive frames) extracted from the audio recording to assign a probability for each possible target class (angry, disgust, fear, happy, sad, neutral, or surprise). The two classes with highest probabilities are used for analysis.

2.3 Sentiment Analysis of Image-Based Description

The interviewee writes an honest descriptions of the shown image. The text in the description is processed and converted to a feature vector. The sentiment analysis model takes the feature vector as input and calculates the probability for each target class (positive, neutral, or negative). This is used to determine the polarity of the interviewee's opinions, and thus, his/her attitude.

2.4 Automated Scoring System for Question/Answers Based On Similarity Measures

The interviewee writes a response to the given question. The database contains the expected answer to the question. The given and expected responses are compared using similarity measures like cosine similarity and jaccard similarity, and a score is assigned to the interviewee's response.

2.5 Provision for Setting Up Images for Image Based Description Test and Questions for Question-Answer Based Test

A database file is created which stores the filename and picture information in the form of BLOB. Another database records list of questions and their respective answers entered by the administrator. They can be accessed by querying the database by making a connection request to it.

2.6 Result as the Analysis of the Various Inputs to Each Individual Functionality

For each functionality (audio/video analysis, image description, or question/answer), raw results are displayed as well as a simple analysis of the raw results.

2.7 Graphical User Interface (GUI)

A graphical user interface is designed using Python's tkinter library [20] where all the features are integrated. There are four screens for the four functionalities (tests) – audio/video emotion recognition, image-based description, question-answer and insertion into database.

3. SRS (SOFTWARE REQUIREMENTS SPECIFICATIONS)

3.1 Introduction

3.1.1. Purpose

This document describes the software requirements specification (SRS) for the application software that enables the psychometric analysis of an individual. Psychometric Analysis Tool looks at measuring the psychological characteristics of an individual: a tool that powers assessment based on mental and emotional state of an individual. It can be used widely by companies for recruitment purposes. The software aids in faster evaluation, avoiding the necessity of an extra psychologist who assesses the emotional state of interviewee. Also, a virtual psychologist can use it for measuring and quantifying a range of different metrics and personal characteristics through the use of psychometric testing. This is achieved by various modules that analyze the responses of the interviewee in the form of text, speech, facial emotion, etc.

3.1.2 Document Conventions

There are no special highlighting Font sizes 18 for headings and 14 for sub-headings has been used throughout the document for maintaining uniformity. Bullet list is used in case of listing the features and similar.

3.1.3 Intended Audience and Reading Suggestions

- The document is intended for researchers, software developers, advanced practitioners, documentation writers, users, testers and evaluators. The SRS contains the requirements and perspective of the software in an elaborated and organized manner which should be read in the same sequence as it is written.
- In the next section, system features with their functional requirements are presented to highlight the major services provided by the intended product. Then the external interface requirements highlighting the logical characteristics of each interface between the software product and the users are discussed. Finally, this specification is concluded with the reference documents on which this document is based on.

3.1.4 Product Scope

We describe what features are in the scope of the software and what are not in the scope of the software to be developed:

In the scope:
- Users can modify the database of images and questions
- Face detection for facial emotion recognition
- Speech recording for speech emotion recognition
- Response as input by test giver in the form of speech, mouse click and text through keyboard

Out of the scope:
- Registration and login for users or administrators
- Long term storage of responses given by users
- No more than seven major class classification possible for emotions

3.2. Overall Description

3.2.1 Product Perspective

The word 'psychometric' conflates the word 'psyche', which is defined as 'mind', with the word 'meter', which means 'measure'. So, psychometric testing and psychometric assessments are effective tools that quantify a person's psychological characteristics in a way that can be measured and analyzed. For example, psychometric tests and psychometric assessments can be used to measure with a great degree of accuracy the characteristics of a person like their personality, cognitive abilities, behavior patterns as well as a wide range of other factors. Psychometrics is the science of assessment. Psychometric Analysis Tool is a tool that powers judgment based on the mental assessment and emotional state of an individual. It is of similar utility as existing interviewers and/or psychologists, who tend to deduce the state of mind of a person by various observations and analysis. Our software aims at provisioning solutions to administer the same in the absence of a human interviewer and/or psychologist. It acts as the interface between the two interacting parties,

one of which evaluates the other using the various modules provided by the software as mentioned further.

3.2.2 Product Functions

- Emotion recognition from facial expressions
- Emotion recognition from speech
- Sentiment Analysis of the image based description to test the attitude of a person
- Automated Scoring system for question/answers based on similarity measures
- Provision for setting up interview questions that can be answered in speech or text
- Provision for setting up images for image based description test
- Result as the analysis of the various inputs to each individual function

3.2.3 User Classes and Characteristics

The users of the software can be differentiated by using their membership and contributions to the system. There are users who cooperate for resource sharing; some are content providers and others being the end-users. In addition, users can vary based on the purpose, size, scope and duration of interaction. For instance, establishment of a long-term business calls for a user who uses the software for organizing the interviews. The most common among them being the recruiters of various companies, who avail this service of the software. The other prospective users are the psychologists who could take advantage of our image description based sentiment analysis and the emotion recognition features. On the other hand, a short-term interaction of the interviewee is required to react within a time frame. They are the person(s) who are to be evaluated and analyzed, whose audio, facial emotions and responses will be recorded for reaching the result.

3.2.4 Operating Environment

This is a standalone system and hence will require the operating environment for a simple GUI.
A modular implementation approach would be useful to perform testing on modules at different stages to ensure correct implementation. It is anticipated that the software will work seamlessly in any functional environment provided the complete and standard configuration and installation of the software.

3.2.5 Design and Implementation Constraints

This system is provisioned to be built on the python tkinter framework which is highly flexible for a simple graphic user interface. It exploits technologies like machine learning and sentiment analysis which are again implemented in python. Decision regarding the database is taken considering the fact that data being stored is large, and hence an appropriate data management system will yield efficient performance. Sqlite3 is used for creating and maintaining the database for images and questions.

3.2.6 User Documentation

Along with the software product, a user manual would be written to help people understand the working methodology and usage of the developed prototype system. It would be written for nontechnical individuals and the level of content or terminology would differ considerably from, for example, a System Administration Guide, which is more detailed and complex. The user manual would follow common user documentation styles capturing purpose and scope of the product along with key system features and operations; step-by-step instructions for using the system including conventions, messaging structures, quick references, tips for errors and malfunctions; pointers to reference documents; and glossary of terms.

3.2.7 Assumptions and Dependencies

- For face and speech emotion classification, the seven classes are anger, disgust, fear, surprise, sadness, happiness and neutral.
- For text, the sentiment classes are positive, negative and neutral.
- The combined observations from different modules are enough to reach to a conclusion.
- The efficiency of the model used for analysis depends on the CPU/GPU used.
- There are no dependencies that the project has on external factors.

3.3 External Interface Requirements

3.3.1 User Interfaces

This section describes the logical characteristics of each interface between the intended software product and the users. For user interface design, common GUI standards will be followed along with the presence of keyboard shortcuts, error message display standards etc., and standard buttons and functions will appear on every screen. The administrator is expected to be familiar with the interface of the system. The software provides a simple GUI interface that has a separate window for each of the four modules giving an enhanced test experience to both the user and the responder.

3.3.2 Hardware Interfaces

Reliable software device drivers shall be provided for every I/O component used in the Software system. Web camera, keyboard, mouse, microphone for recording responses shall be completely tested to prove the full access to the required software functionality and the correct exploitation of its resources.

3.3.3 Software Interfaces

The Psychometric Analyzer has different types of software interfaces (this term is used in a very broad meaning) to external packages, depending how the interaction is realized:
i. User interface: The various functionalities of the application can be accessed by the means of buttons with simple, self-explanatory labels.

ii. Message interface: Since automation is the goal, methods have been implemented to run the various functionalities and communicate among different objects (instances of different classes).
iii. Database interface: The test or analysis organizer can easily input one's own images and question-answers in the database.

3.4. Functional Requirements

The major services and functional requirements for the product can be illustrated by system features. In the following, necessary description is provided for each module in the system. Each description provides information of the associated actors, triggering condition, preconditions, postconditions, response sequences, exceptions and functional requirements (assumptions). Being a major important section of the SRS, this section is expected to go through iterative improvement to make the most logical sense for the intended product.

3.4.1 Face and Speech Emotion Recognition

Introduction:

This module provides the main emotion recognition of the responder whose facial expressions and speech is captured and recorded in fixed intervals using the GUI and sent to the emotion recognition model for classification. The handling of the web camera is done by another library cv2 that sends the snapshot to our classification model to recognize the emotional state.

Input:

i. Audio feed from microphone

ii. Video feed from webcamera

Output:

The result of the analysis of the same is displayed on the GUI itself. It gives the percentages of the strongest detected emotions among the seven, and a brief analysis. This exploits the CNN model for Facial Emotion Recognition and LSTM model for Speech Emotion Recognition.

3.4.2 Sentiment Analysis of Image Based Description

Introduction:

This analysis requires the responder to input a description for an image that is displayed to him/her within a pre-specified word limit. This text is then sent to the sentiment analysis model for estimation of the positive, negative and neutral perception percentages of an

individual in response to the respective images. The result is displayed on the window of the GUI.

Input:

i. Text input from keyboard

Output:

The percentages of the individual's response being positive or negative is displayed on the GUI, as well as a brief analysis. This exploits NLP (Natural Language Processing) and the ANN Keras model in Python for classification.

3.4.3 Adaptive Questions/Answers Scoring Module

Introduction:

Questions will be displayed which can be answered by the responder and this will be further analyzed for finding the relevance and similarity percentages. Next question to be displayed is selected with respect to the answer of the previous questions, i.e., it is an adaptive model. It will have a Chat type interface.

Input:

i. Text input from keyboard

Output:

The score is generated based on the similarity in the content and length of the expected answer and the answer given by the responder and displayed on the GUI. This exploits the cosine similarity and jaccard similarity for finding the score.

3.4.4 Images and Questions/Answers Database

Introduction:

The users who are given authorized access to the database are provided with the facility of modifying the database. They can add the images and questions/answers to the database for organizing the tests using the fourth window of the GUI.

Input:

i. Questions and their expected answers as text

ii. Image that can be browsed from the system

Output:

Three-tier client server architecture is used because data is stored in database can be accessed and modified by the user, the client with the help of the GUI between them.

3.5. Other Nonfunctional Requirements

3.5.1 Performance Requirements

The system shall be interactive and the delays involved shall be less. So in every action-response of the system, there are no immediate delays. In case of opening different windows of various modules and saving the settings or sessions there is delay much below 1 second, In case of opening databases, sorting questions and evaluation there are no delays and the operation is performed in less than 2 seconds for opening ,sorting, computing, posting more than 95% of the files.

3.5.2 Safety Requirements

Information transmission shall be securely transmitted to database server without any changes in information. As the system provide the right tools for recording and analysis, it must be made sure that the system is reliable in its operations and for securing the sensitive details.

3.5.3 Software Quality Attributes

Availability: The software is standalone and independent of network connections pertaining to the node system on which it is installed.
Usability: The system is easy to handle and navigates in the most expected way with no delays. The system program reacts accordingly and transverses quickly between its states.

4. SDS (SOFTWARE DESIGN SPECIFICATIONS)

4.1 Introduction

This document is designed to be a reference for any person wishing to implement or any person interested in the architecture of the software. This document describes the application's architecture and sub architecture, the associated interfaces, the database and the motivations behind choosing the design. Both high level and low level designs are included in this document. This document should be read by an individual with a technical background and has experience reading data flow diagrams, control flow diagrams, interface designs and development experience in object-oriented programming as well as sequential programming.

The document will provide developers an insight in meeting client's needs efficiently and effectively. It would demonstrate how the design will accomplish the functional and nonfunctional requirements captured in the SRS.

4.1.1 Document Description

It is the "how will we do it" part after we wrote the "what will we do" part, which is the SRS.

The System Architecture section is the main focus of this document. It provides an overview of the system's major components and architecture, as well as specifications on the interaction between the system and the user.

The Detailed System Design description of components section will also be covered in this document. It will describe lower-level classes, components, and functions, as well as the interaction between these internal components. It contains specific information about the expected input, output, classes, and functions. The interactions between the classes to meet the desired requirements are outlined in detailed figures (class diagram) at the end of the document. It supplies a snapshot of the intended system from multiple points of view.

Here is the outline of the proposed software design specifications.

- Introduction
- System Overview
- Design Considerations
    - Assumptions and Dependencies
    - General Constraints
    - Goals and Guidelines
    - Development Methods
- System Architecture
    - Module-1 Graphical User Interface
    - Module-2 Facial Expression Recognition Model
    - Module-3 Speech Emotion Recognition Model
    - Module-4 Image Description Sentiment Analysis Model
    - Module-5 Questions Answers Automated Scoring Model
    - Module-6 Database for Images and Question Answers
- Detailed System Design
    - Classification
    - Definition of components

- Class Diagram for the application
- Use Case Diagrams
- Sequence Diagrams

4.1.2 System Overview

Psychometrics is the science of assessment of mental capacities and processes. Psychometric Analysis Tool powers judgment based on the assessment of emotional state of an individual by using webcam, microphone and keyboard input. It is a tool that can be used for assistance to existing interviewers and/or psychologists which aims at provisioning solutions to administer the same in the absence of a human interviewer and/or psychologist. It acts as the interface between the two interacting parties, one of which evaluates the other using the various modules provided by the software as mentioned further. It can be widely used by the companies for recruitment purposes. Also, a psychologist can use it for evaluation of emotional state. This is achieved by various modules that analyze the responses of the interviewee or patient in the form of text, speech, facial emotion, etc.

4.2. DESIGN CONSIDERATIONS

This section describes many of the issues which need to be addressed or resolved before attempting to devise a complete design solution.

4.2.1 Assumptions and Dependencies

- For face and speech emotion classification, the seven classes are anger, disgust, fear, surprise, sadness, happiness and neutral.
- For text, the sentiment classes are positive, negative and neutral.
- The combined observations from different modules are enough to reach to a conclusion.
- The efficiency of the model used for analysis depends on the CPU/GPU used.
- There are no dependencies that the project has on external factors.

4.2.2 General Constraints

- **Operating Environment:** This is a standalone application software and hence will require the operating environment for a simple GUI, preferably 64-bit system.
- **Performance Requirements:** The system shall be interactive and the delays involved shall be set to a minimum. So, in every action-response of the system, there are no immediate delays. In case of opening different windows of various modules and saving the settings or sessions, there is a delay of duration below 1 second.
- **Safety Requirements:** Information transmission shall be securely transmitted to server/ database without any changes in information. As the system provides the

right tools for recording and analysis, it must be made sure that the system is reliable in its operations and for securing the sensitive details.

- **Availability:** If the file read or video IO service gets disrupted while reading the file, storing to database or accessing the webcam, the system will attempt again on its own.
- **Usability:** As the system is easy to handle and navigates in the most expected way with no delays, the system program reacts accordingly and transverses quickly between its states.
- **User Interfaces:** The administrator is expected to be familiar with the interface of the system. A simple GUI interface that has a separate window for separate tasks, gives an enhanced test experience to both the user and the responder.
- **Hardware Interfaces:** Reliable software device drivers should be provided for every I/O component used in the software system. Web camera, keyboard, mouse, microphone for recording responses are to be completely tested to prove the full access to the required software functionality and the correct exploitation of its resources.
- **Software Interfaces:** The Psychometric Analyzer has different types of software interfaces (this term is used in a very broad meaning) to external packages, depending how the interaction is realized:
    1. User interface: The various functionalities of the application can be accessed by the means of buttons with simple, self-explanatory labels.
    2. Message interface: Since automation is the goal, methods have been implemented to run the various functionalities and communicate among different objects (instances of different classes).
    3. Database interface: The test or analysis organizer can easily input one's own images and question-answers in the database.

4.2.3 Goals and Guidelines

- Creating a software product that assists a human interviewer and/or psychologist and automates a large part of psychometric assessment.
- Classification into emotions/ sentiment and using some heuristic or guidelines to understand their significance.
- Combination of observations from different modules to reach to a conclusion.
- The efficiency of training of the model used for analysis depends on the CPU/GPU used and hence powerful ones are recommended.
- A properly working web camera and microphone are must for this to function appropriately.

4.2.4 Development Methods

Object-oriented design strategy was used for this software design. This design strategy focuses on entities and its characteristics, and using method calls to change characteristics

or communicate among entities. The whole concept of software solution revolves around the engaged entities. Function-oriented design was considered, wherein, each feature or utility of the tool would be implemented as a function, however, it was better suited to treat each functionality as an entity and handle its various aspects using method calls.

4.3 System Architecture

This section provides a high-level overview of how the functionality and responsibilities of the system are partitioned and then assigned to subsystems or components. The main purpose here is to gain a general understanding of how and why the system was decomposed, and how the individual parts work together to provide the desired functionality.
At the top-most level, the description of the major responsibilities that the software undertakes and how the higher-level components collaborate with each other in order to achieve the required results are given.

4.3.1 Graphical User Interface

A simple GUI interface that has a separate window for the major functional modules, giving an enhanced test experience to both the user and the responder. Each window consists of several frames that encapsulate the displayed non-interactive regions like images in canvas, labels, helper information, displayed questions; and the responder area in the form of the editable text boxes as well as buttons for interacting with the tool. A toolbar on the top of every window allows navigation between the four windows (video-audio feed test, image description-based test, question response-based test and user interface to database) provides for a good user experience. Libraries for designing GUI: tkinter, PIL.

4.3.2 Facial Expression Recognition Model

The Facial Expression Recognition Model is a convolutional neural network model with a sequence of convolution layers, followed by a max-pooling layer, and a dense layer with nodes = 1024 and dropout rate = 0.5 succeeding all the convolution and pooling layers. Lastly, a dense layer for 7 outputs with softmax activation follows. This model is trained using the FER2013 dataset and saved for use. The face detection component is implemented using Haar-feature based cascade classifier using cv2 (OpenCV). A rectangular frame is drawn around the detected part in real-time video feed from webcam. Detected sections of images with faces, captured periodically, are resized to 48x48 pixels, improved by increasing brightness and converted to grayscale. These grayscales are fed to the trained neural network model for classification of facial expressions into disgusted,

angry, sad, happy, surprised, fearful and neutral. The results of analysis are displayed on the GUI. Libraries for FER model: keras, numpy, sklearn, cv2.

4.3.3 Speech Emotion Recognition Model

This component records audio from microphone periodically and feeds it to the trained stacked LSTM model for classification of speech emotions into disgust, anger, sadness, happiness, surprise, fear and neutral. The LSTM model is trained using the RAVDESS and SAVEE dataset. Libraries used: scipy for audio extraction and keras for training.

4.3.4 Image Description Sentiment Analysis Model

This analysis requires the responder to input a description of a displayed image in about 100 words or some other specified limit. This text is then pre-processed by removing punctuations, extending contractions, POStagging for selecting sentiment descriptors, lemmatizing and stemming. From a similar pre-processing, followed by CountVectorizer and employing a self-designed variation of TF-IDF, a lexicon was generated. For the response as well, the lexicon is used for extracting numerical features. These features are fed to the trained sentiment analysis model (datasets: Stanford Movie Reviews, Amazon Reviews, Twitter Airline Sentiment) for estimation of the positive, negative and neutral perception percentages of an individual in response to the respective images. The results are displayed on the GUI window. Libraries used: nltk, numpy, pandas, csv for pre-processing, and keras for model.

4.3.5 Automated Questions Answers Scoring Model

Questions displayed on the GUI can be navigated through using Next and Previous buttons. The responder has to answer the question in the answer box and upon submission, the answer will be compared to the response already pre-fed in the database with the question, using the Cosine similarity and Jaccard similarity measures. Depending on the measures and respective lengths of the known and given responses, the answers will be scored. Libraries used: sklearn.

4.3.6 Database for Images and Questions Answers

The users who are given authorized access to the database are provided with the facility of modifying the database. They can add the images and questions/answers to the database for organizing the tests using the fourth window of the GUI. Client server architecture is used because data is stored in database can be accessed and modified by the user, the client with the help of the GUI between them. Libraries used: sqlite3, IPython.

4.4 Detailed System Design

Most components described in the System Architecture section will require a more detailed discussion. Other lower-level components and subcomponents may need to be described as well. Each subsection of this section will refer to or contain a detailed description of a system software component. The discussion provided should cover the following software component attributes:

4.4.1 Classification

We have used four classes for four screens that provide different functionalities. Each of these classes have various functions that are called on some button click or submission or for sending the data to the model for different data mining purposes. These classes are independent of each other.

4.4.2 Definition of Components

4.4.2.1 GUI (Graphical User Interface):

Class: App - A simple GUI interface that has a separate window for the major functional modules, giving an enhanced test experience to both the user and the responder. Each window consists of several frames that encapsulate the displayed non-interactive regions like images in canvas, labels, helper information, displayed questions; and the responder area in the form of the editable text boxes as well as buttons for interacting with the tool. A toolbar on the top of every window allows navigation between the four windows (video-audio feed test, image description-based test, question response-based test and user interface to database) provides for a good user experience. Libraries for designing GUI: tkinter, PIL.

4.4.2.2 Facial and Speech Emotion Recognition Model:

The first module provides the main emotion recognition of the responder whose facial expressions and speech is captured and recorded in fixed intervals using the GUI and transported to the emotion recognition model for classification. The result of the analysis of the same is displayed on the GUI itself. It gives the percentages of the highest two emotions among the seven.

Models:

FER Model - The Facial Expression Recognition Model is a convolutional neural network model with three convolution layers (5x5, 3x3, 3x3), each followed by a pooling layer (5x5, 3x3, 3x3), and two dense layers with nodes = 512 and dropout rate = 0.5 succeeding all the convolution and pooling layers. Lastly, a dense layer for 7 outputs with softmax activation follows. This model is trained using the FER2013 dataset and saved for use. The face detection component is implemented using Haar-feature based cascade classifier using cv2 (OpenCV). A rectangular frame is drawn around the detected part in real-time video feed from webcam. Detected sections of images with faces, captured periodically, are resized to 48x48 pixels, improved by increasing brightness and converted to grayscale. These grayscales are fed to the trained neural network model for classification of facial expressions into disgusted, angry, sad, happy, surprised, fearful and neutral. The results of analysis are displayed on the GUI. Libraries for FER model: keras, numpy, sklearn, cv2.

SER Model – The Speech Emotion Recognition Model is an LSTM (Long Short Term Memory) model which is trained using RAVDESS and SAVEE dataset and saved for use. The speech recorded via microphone is directly fed to the trained model as a test data for classification of speech emotions into sad, neutral, happy, surprise, anger, disgust and fear. The results of analysis are displayed on the GUI. Libraries for FER model: keras, numpy, sklearn, scipy.

4.4.2.3 Image Description Sentiment Analysis Model:

Class: Screen2 - This analysis requires the responder to input a description for an image that is displayed to him/her in about 100 words. This text is then transported to the sentiment analysis model for estimation of the positive, negative and neutral perception percentages of an individual in response to the respective images. The result is displayed on the window of the GUI. The percentages of the individual's response being positive, negative or neutral is displayed on the GUI.

Model: Sentiment Analysis Model - This text is then pre-processed by removing punctuations, extending contractions, POStagging for selecting sentiment descriptors, lemmatizing and stemming. From a similar pre-processing, followed by CountVectorizer and employing a self-designed variation of TF-IDF, a lexicon was generated. For the response as well, the lexicon is used for extracting numerical features. These features are fed to the trained sentiment analysis model (datasets: Amazon Reviews, Twitter Airline Sentiment, Stanford Movie Reviews)

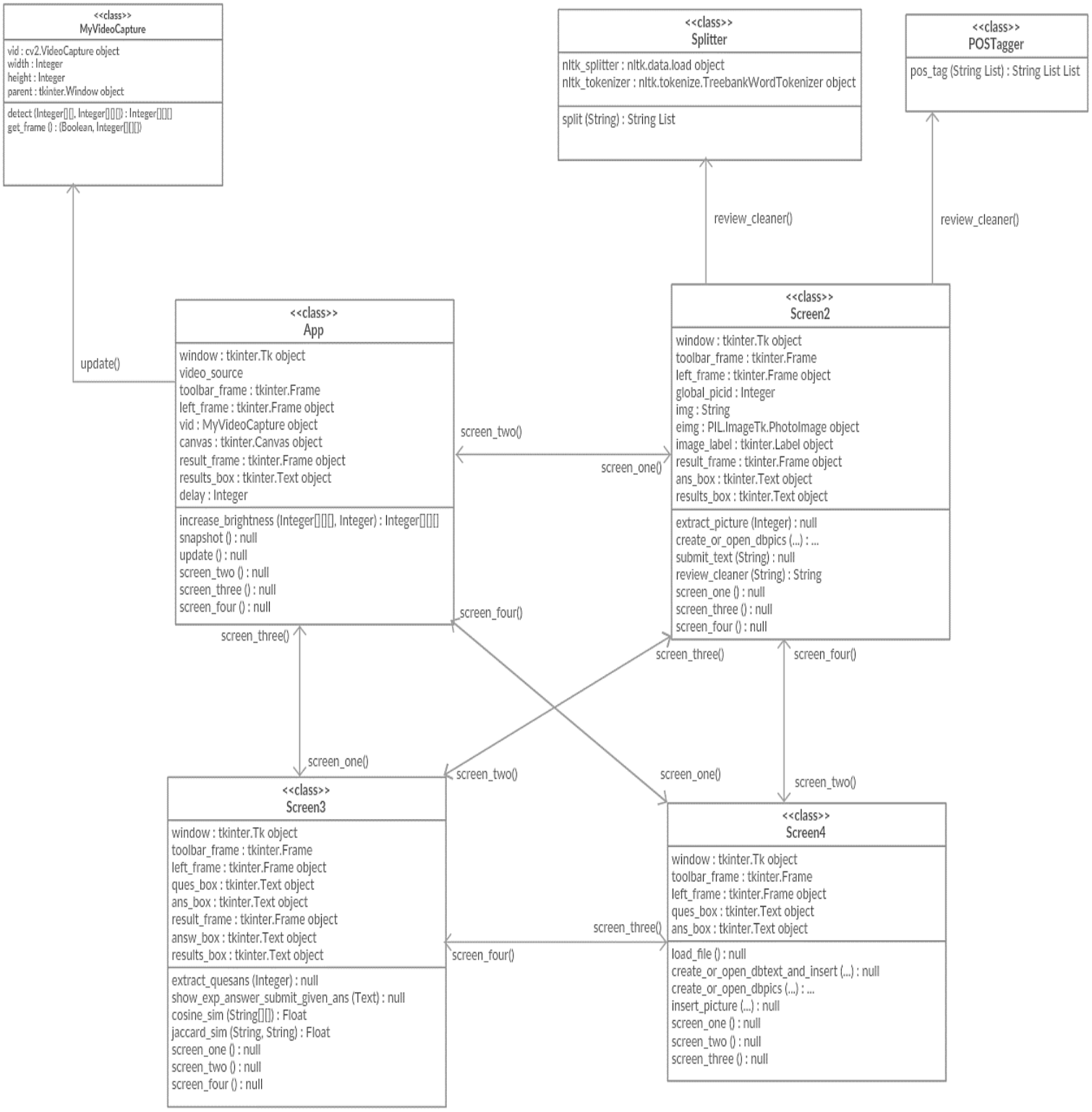
4.4.2.4 Questions Answers Automated Scoring Model:

Class: Screen3 - Questions displayed on the GUI can be navigated through using Next and Previous buttons. The responder has to answer the question in the answer box and upon submission, the answer will be compared to the response already pre-fed in the database with the question, using the Cosine similarity and Jaccard similarity measures. Depending on the measures and respective lengths of the known and given responses, the answers will be scored. Libraries used: sklearn.

4.4.2.5 Database for Images and Questions Answers:

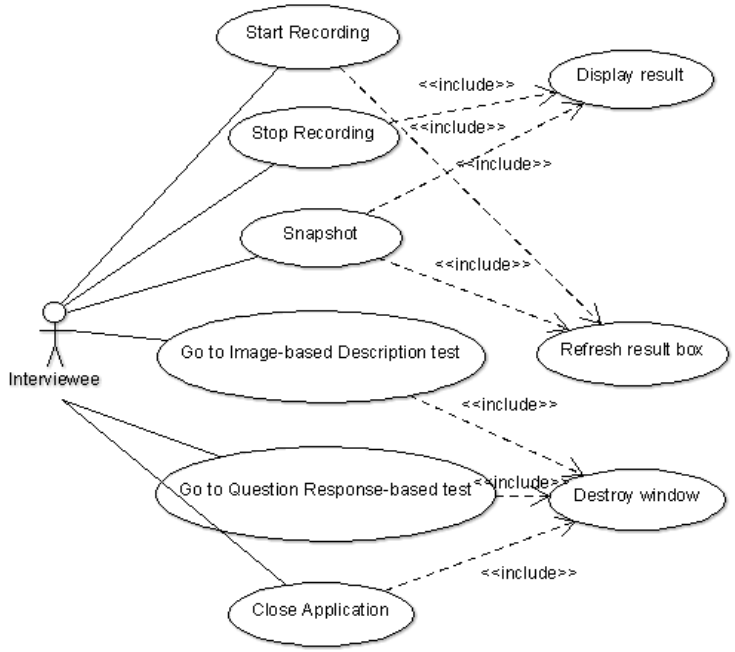
Class: Screen4 - The users who are given authorized access to the database are provided with the facility of modifying the database. They can add the images and questions/answers to the database for organizing the tests using the fourth window of the GUI. Libraries used: IPython, sqlite3, os.

4.5 Class Diagram for the Application

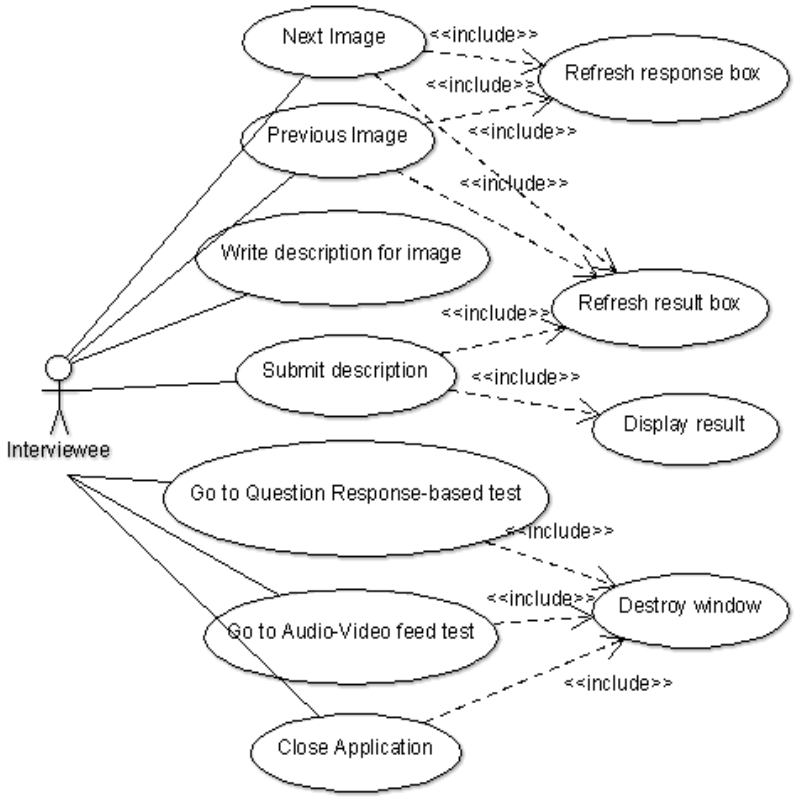


4.6 Use Case Diagrams

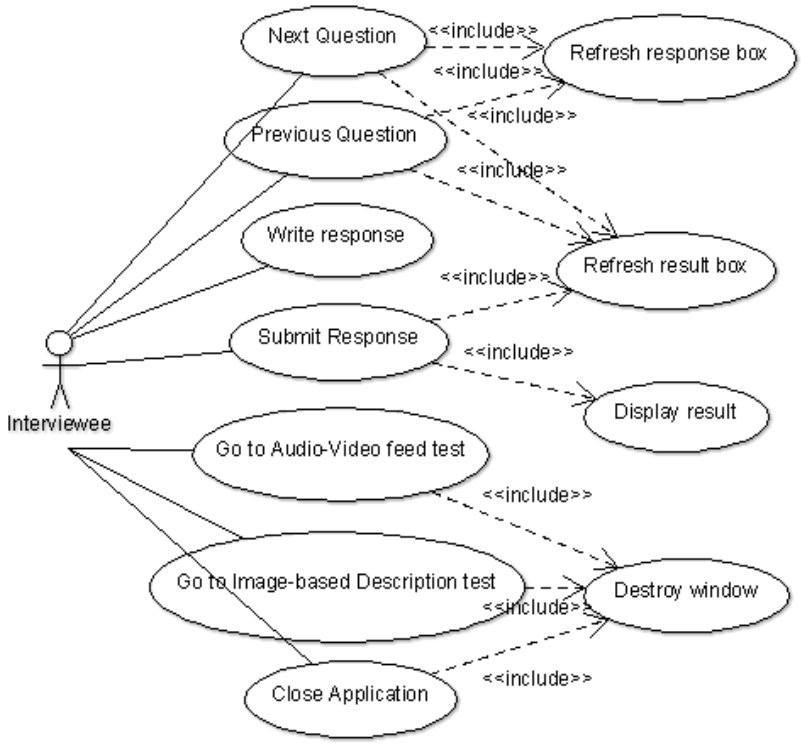
4.6.1 Screen 1 (Audio/Video Feed Based Test)



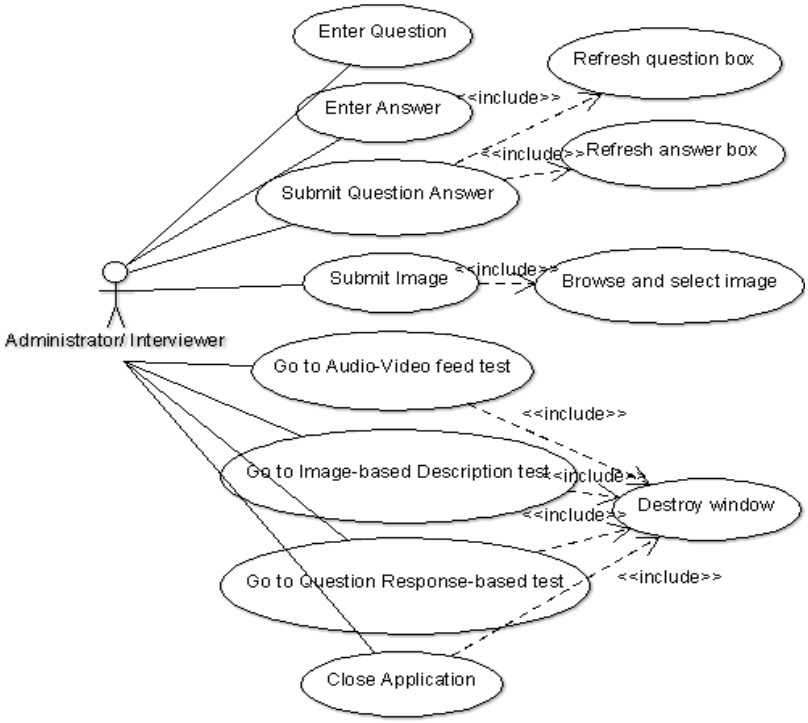
4.6.2 Screen2 (Image Description Sentiment Analysis Model)



4.6.3 Screen3 (Questions Answers Automated Scoring Model)

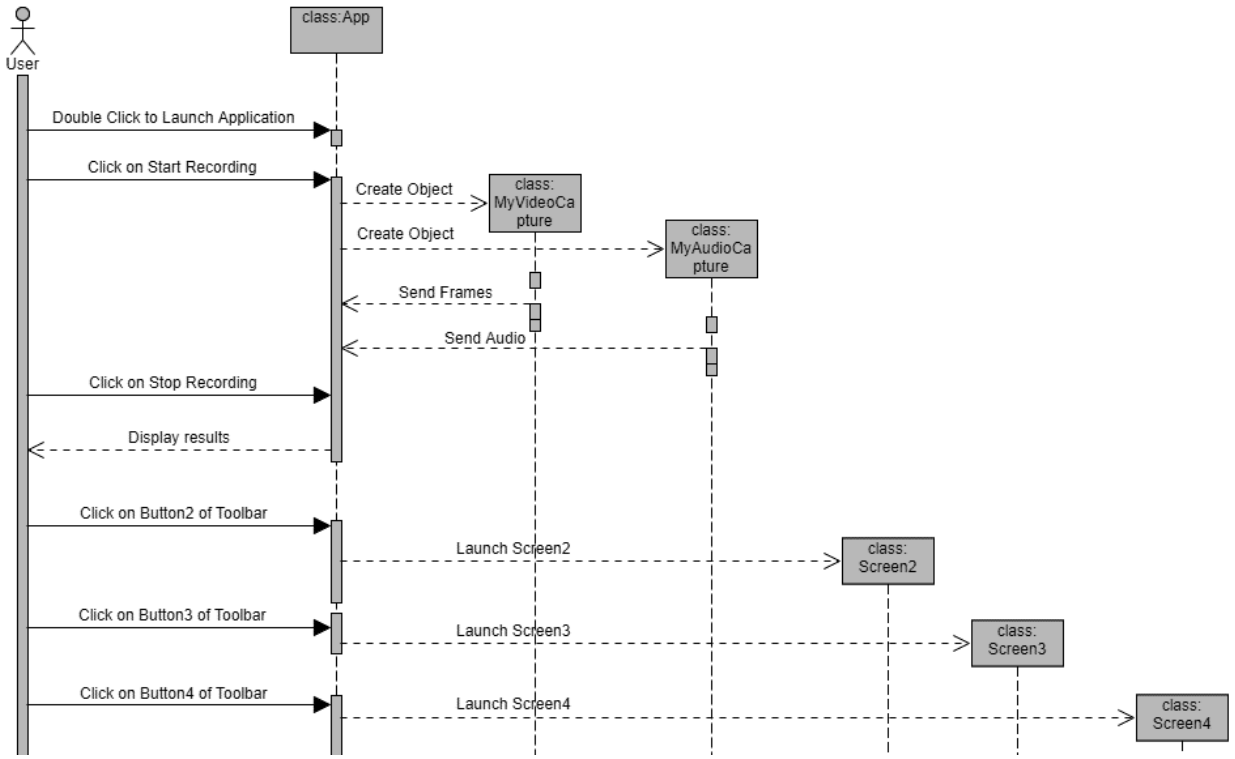


4.6.4 Screen4 (Database for Images and Questions Answers)

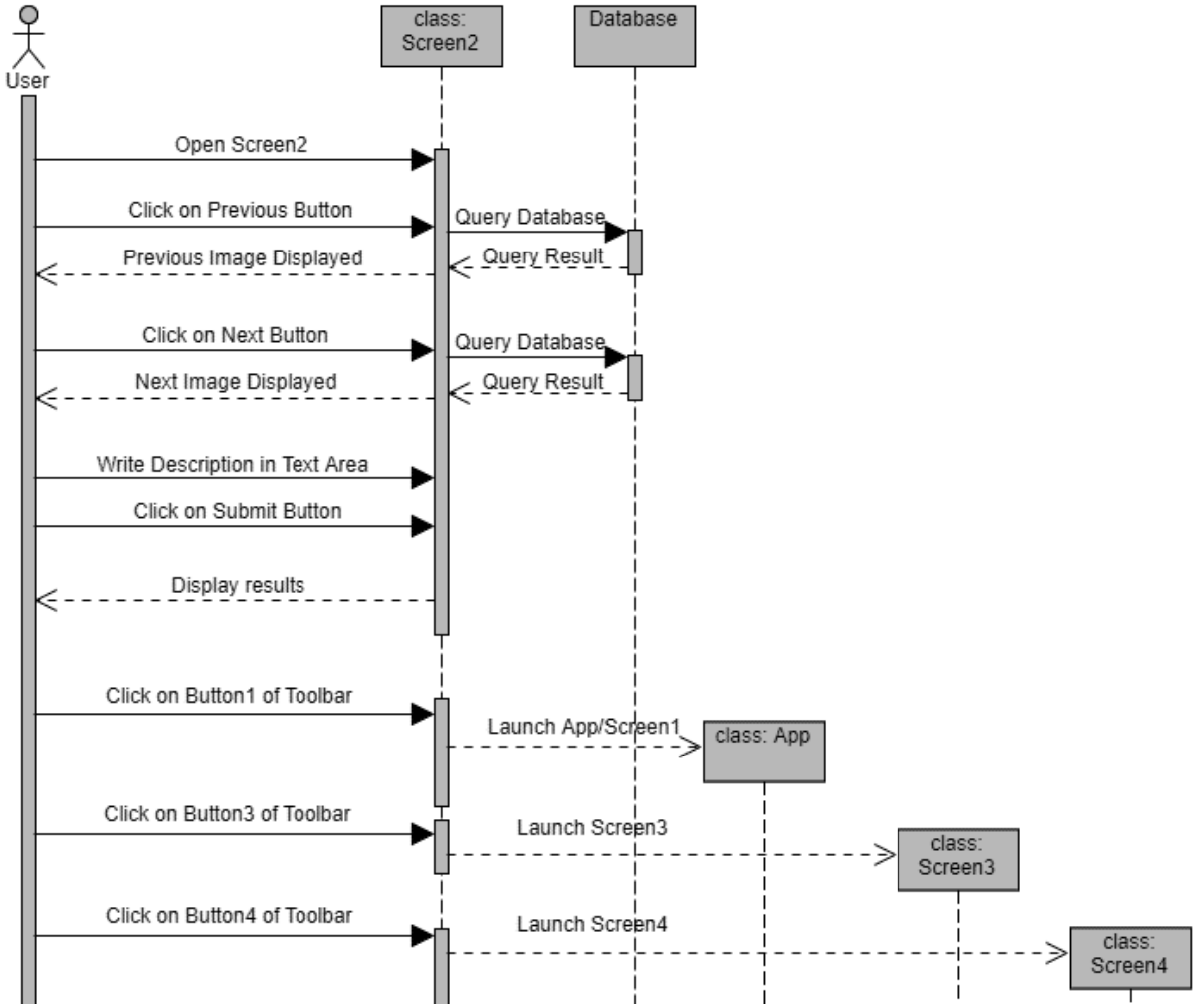


4.7 Sequence Diagrams

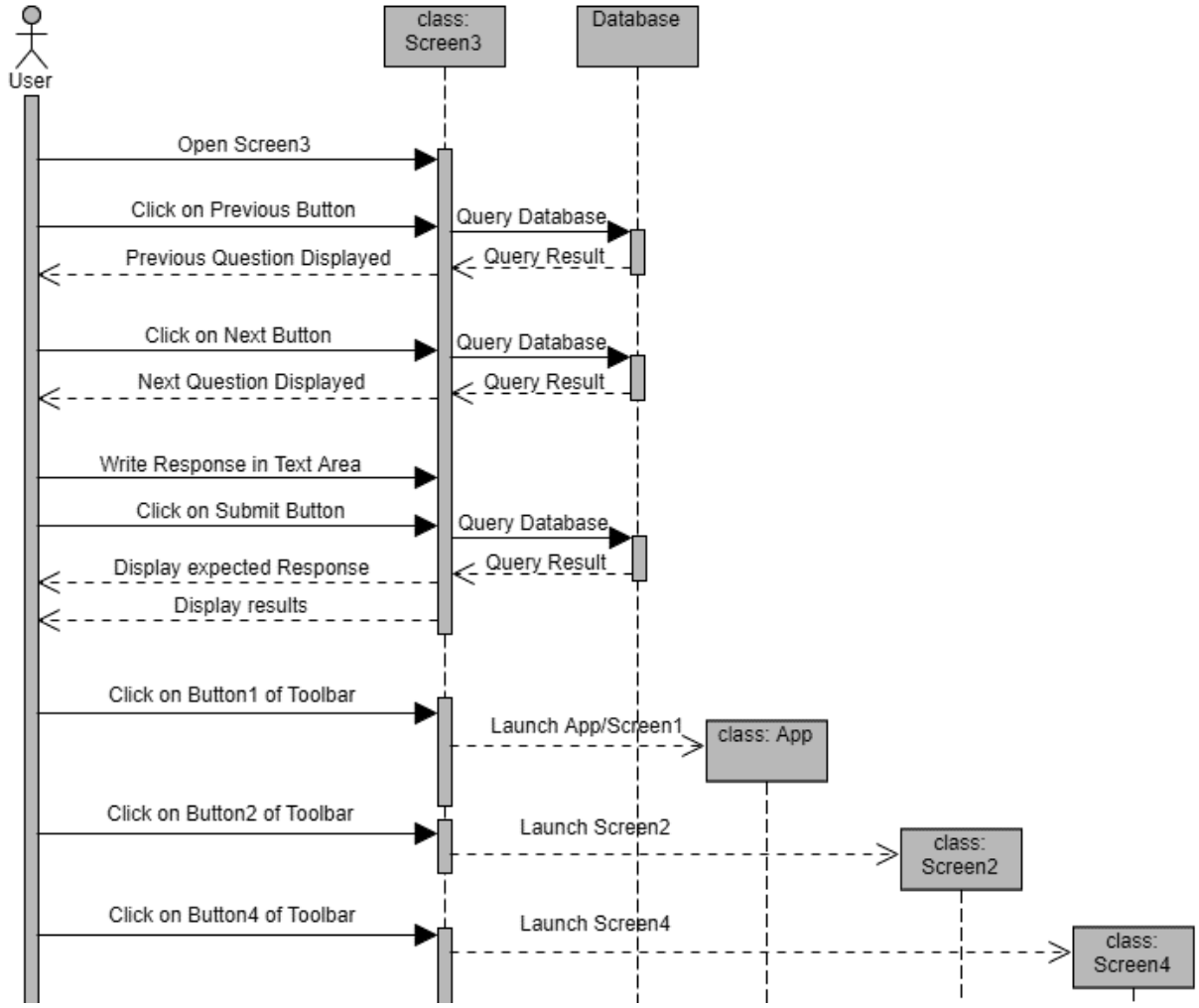
4.7.1 Screen 1 (Audio/Video Feed Based Test)



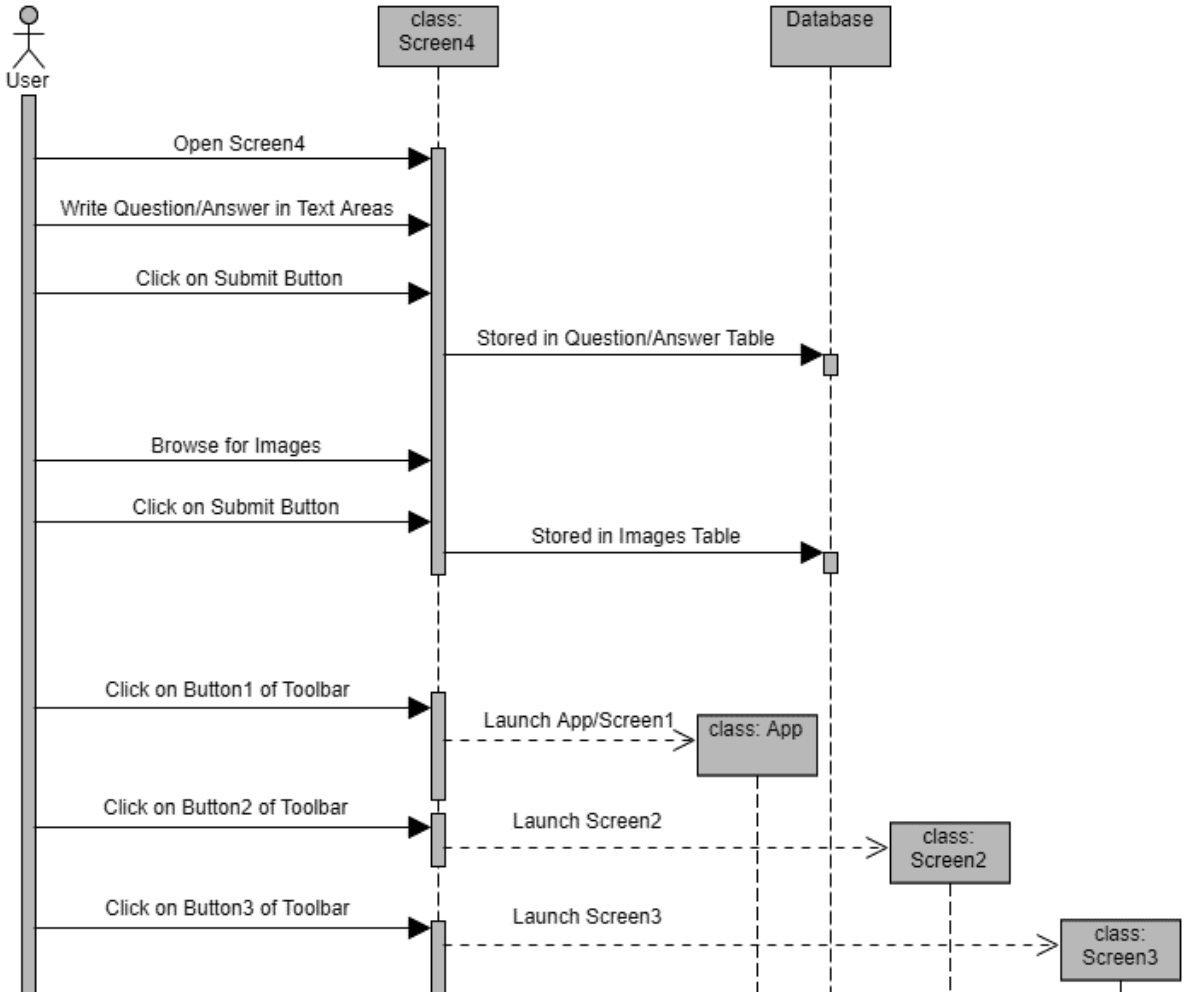
4.7.2 Screen2 (Image Description Sentiment Analysis Model)



4.7.3 Screen3 (Questions Answers Automated Scoring Model)



4.7.4 Screen4 (Database for Images and Questions Answers)



5. IMPLEMENTATION

5.1 Facial Expression Recognition Model

The images captured from webcam are classified using the Facial Expression Recognition or FER model. The FER model is trained using the publicly available FER-2013 dataset [24]. The dataset is challenging as the depicted faces vary significantly in terms of person age, face pose, and other factors, reflecting realistic conditions. The dataset is split into training, validation, and test sets with 28,709, 3,589, and 3,589 samples, respectively. Basic expression labels are provided for all samples. All images are grayscale and have a resolution of 48 by 48 pixels. The human accuracy on this dataset is around 65.5%.

The training has been performed on the images without any preprocessing in order to preserve the variations for realistic usage. The process for model design involved training the various models and different hyperparameters combinations on training set of 28709 samples, and testing on the validation set (3589 samples). The final performance data was gathered by training on the combined training and validation data and testing on the test dataset. The model with the highest accuracy (~63%) was saved for use. The model consists of a sequence of convolution and subsampling (max-pooling) layers, as shown in Figure 5, with subsequently increasing filters for better feature representation using more feature maps. The last of the feature maps are flattened to a vector which is passed to the dense layer (1024 nodes), followed by an output layer with softmax activation.[5], [6]

The model was built using keras library's Sequential model [18], [19] and trained on Google Colaboratory's GPU (1xTesla K80, compute 3.7, having 2496 CUDA cores, 12GB GDDR5 VRAM).

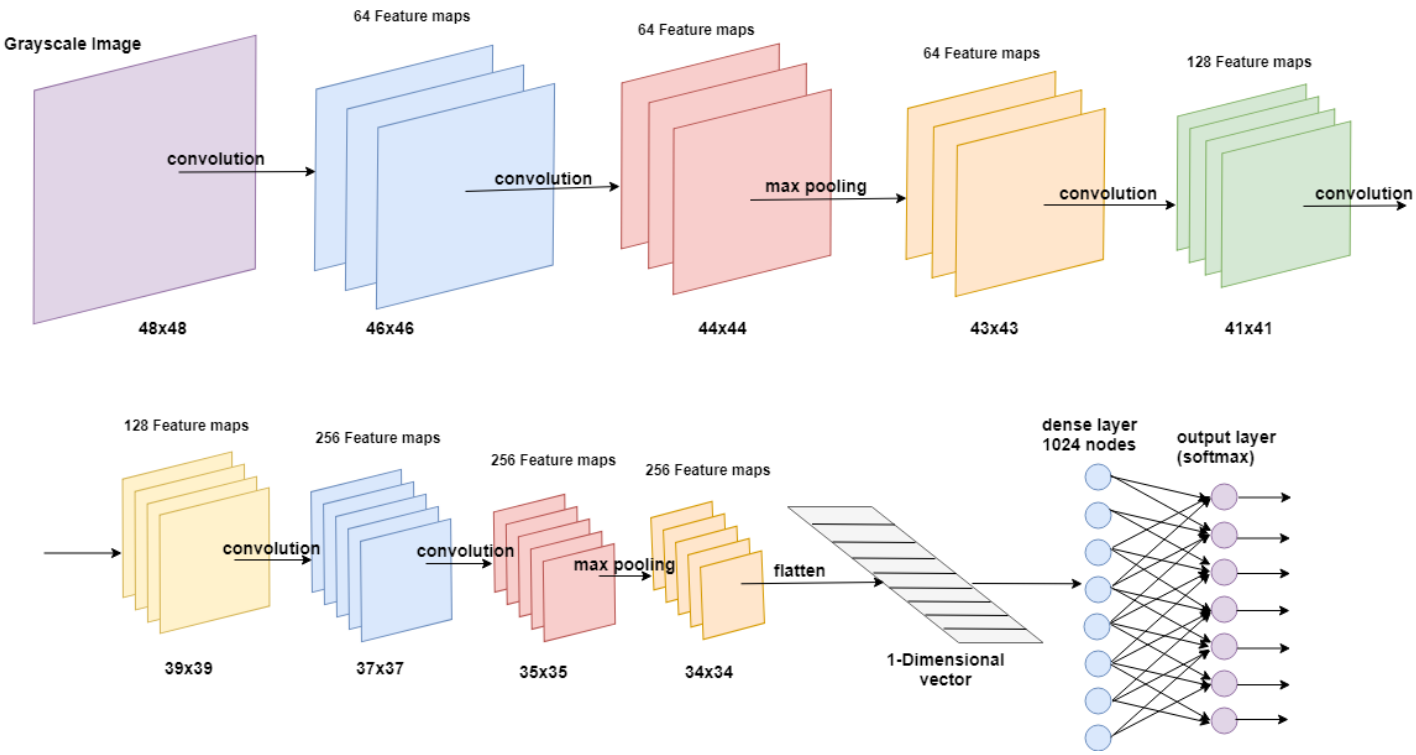


Figure 5

Training and testing accuracy vs epoch and loss vs epoch plots for batch_size = 256 are given in Figure 6 and 7 respectively. The confusion matrix for the same is shown in Figure 8.

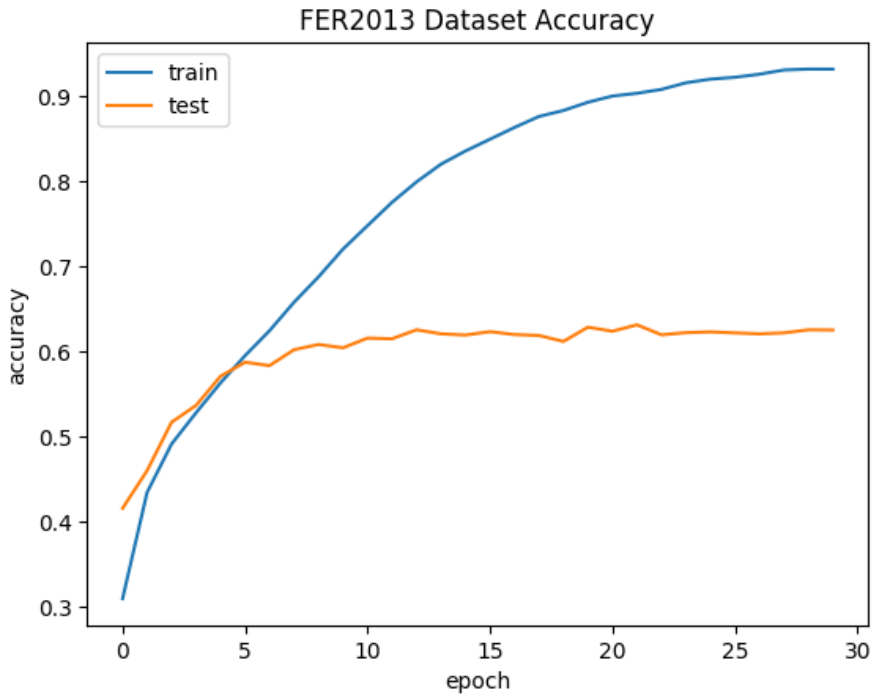


Figure 6

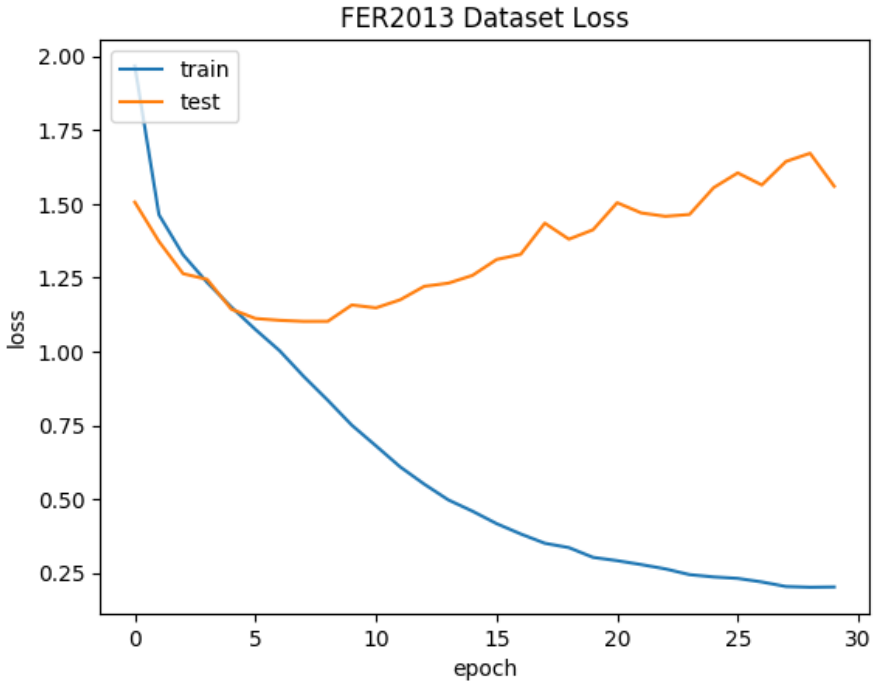


Figure 7

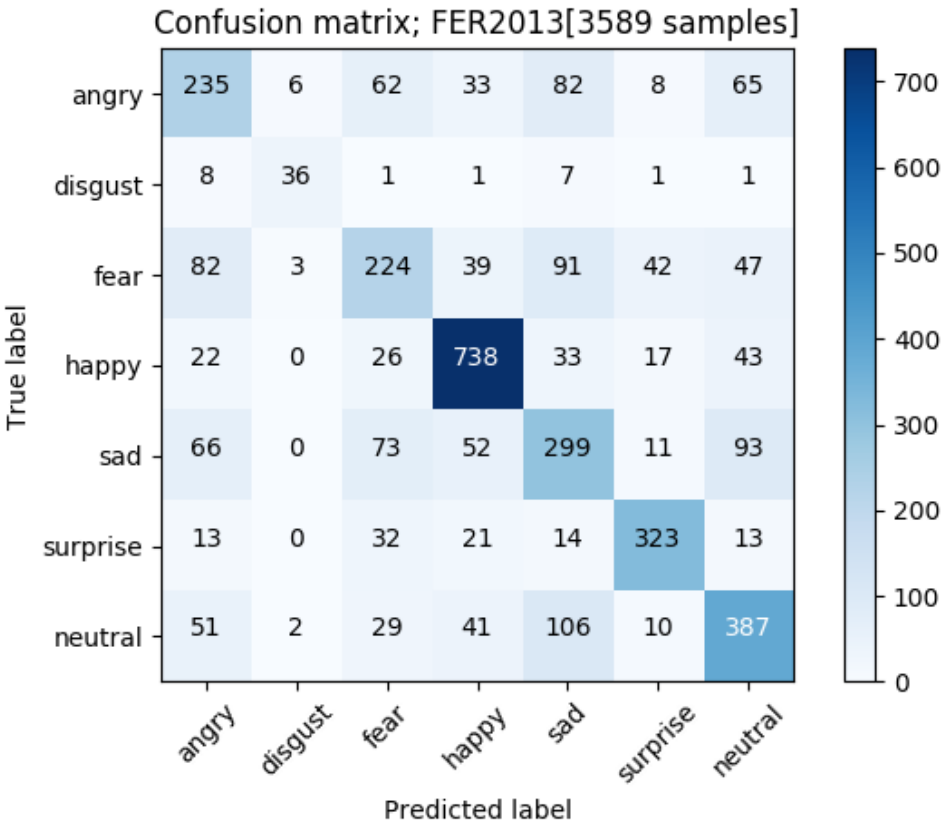


Figure 8

For using the model, preprocessing is required. Preprocessing entails operations that are applied once to each image. This includes face detection, smoothing and means for correcting for illumination variations. The libraries used are keras.preprocessing, cv2 (OpenCV), and PIL.

5.2 Speech Emotion Recognition Model

The Speech Emotion Recognition or SER model classifies the segments of a recording of a speaker (extracted using pyaudio [23]) into various categories of emotions. The SER model is trained on Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). This dataset contains the complete set of 7356 RAVDESS files (total size: 24.8 GB). Speech file (Audio_Speech_Actors_01-24.zip, 215 MB) contains 1440 files: 60 trials per actor x 24 actors = 1440. Song file (Audio_Song_Actors_01-24.zip, 198 MB) contains 1012 files: 44 trials per actor x 23 actors = 1012.

The other dataset used for training was Surrey Audio-Visual Expressed Emotion (SAVEE). The SAVEE database was recorded by four native English male speakers (identified as DC, JE, JK, KL), aged from 27 to 31 years. Emotion has been described in discrete categories: anger, disgust, fear, happiness, sadness, surprise and neutral. The text material consisted of 15 TIMIT sentences per emotion: 3 common, 2 emotion-specific and 10 generic sentences that were different for each emotion and phonetically-balanced. The 3 common and 2 × 6 = 12 emotion-specific sentences were recorded as neutral to give 30 neutral sentences. There is a total of 120 utterances per speaker- 480 utterances in all. [25], [26], [27].

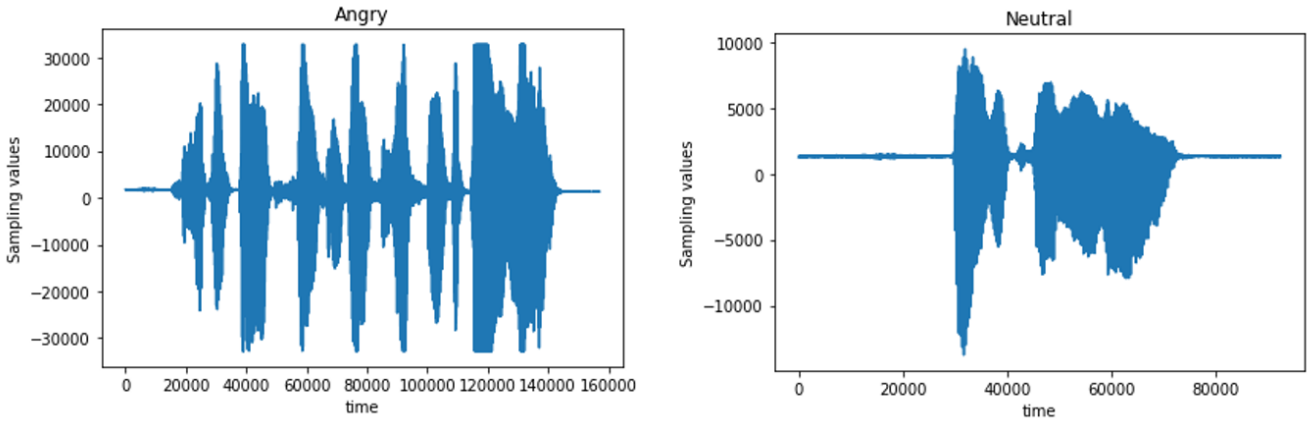


Figure 10 (contrasting the pressure values versus time plot for two samples labeled Angry and Neutral from SAVEE Dataset

The SER model is implemented as a keras Sequential model [19] with a stack of three LSTM layers (256, 128 and 64 units respectively) followed by a dense layer and an output layer. The timesteps for an input sequence are set to 10.[10]

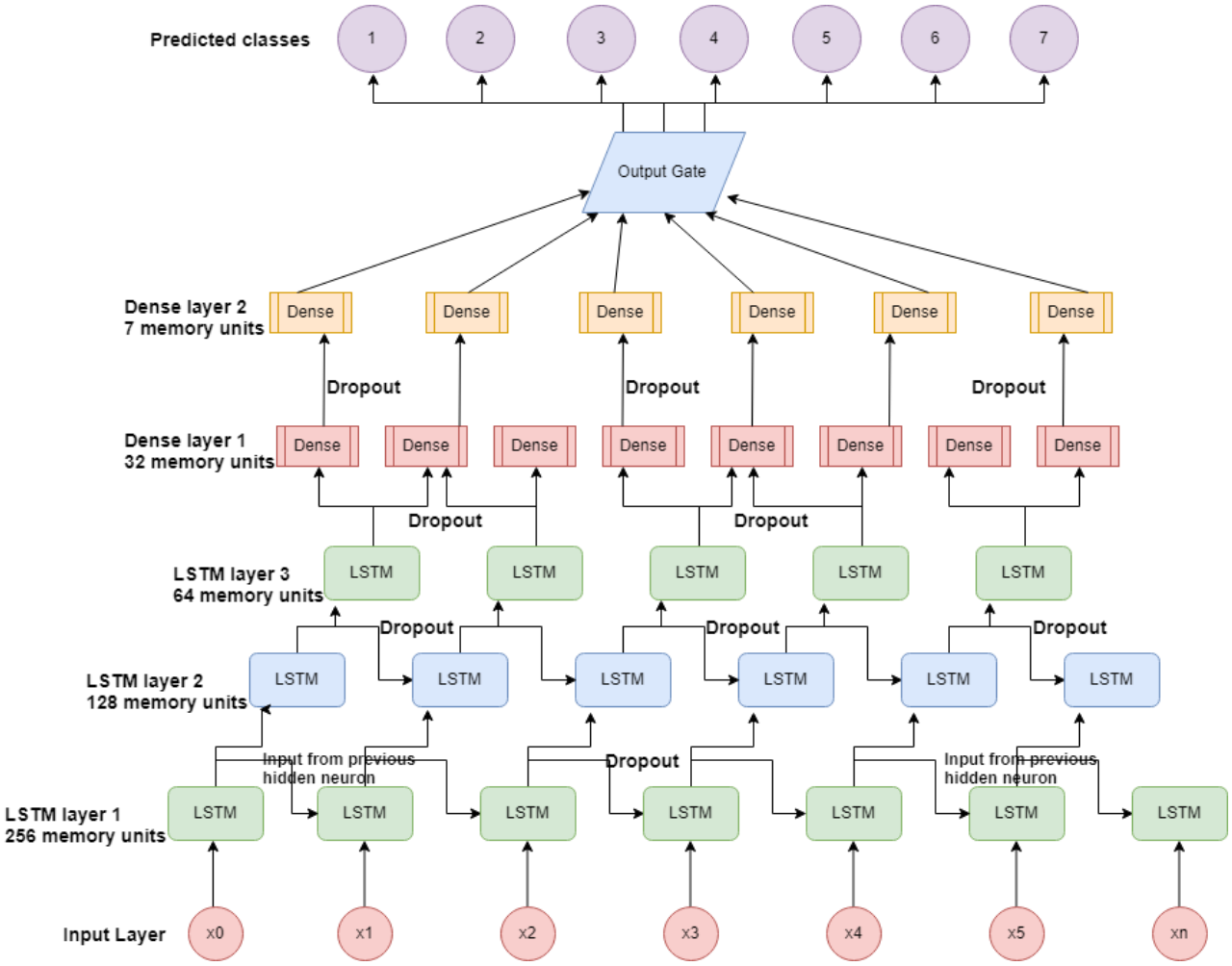


Figure 9

The SAVEE dataset has been augmented by doubling the samples. The training and testing accuracy vs epoch and loss vs epoch plots for RAVDESS Songs dataset are shown in Figure 10. The training and testing accuracy vs epoch and loss vs epoch plots for SAVEE dataset are shown in Figure 11. Figure 12 shows the confusion matrix for the entire dataset on RAVDESS Songs dataset, containing 1012 samples. Figure 13 shows the confusion matrix for augmented set on SAVEE dataset, with 960 samples.

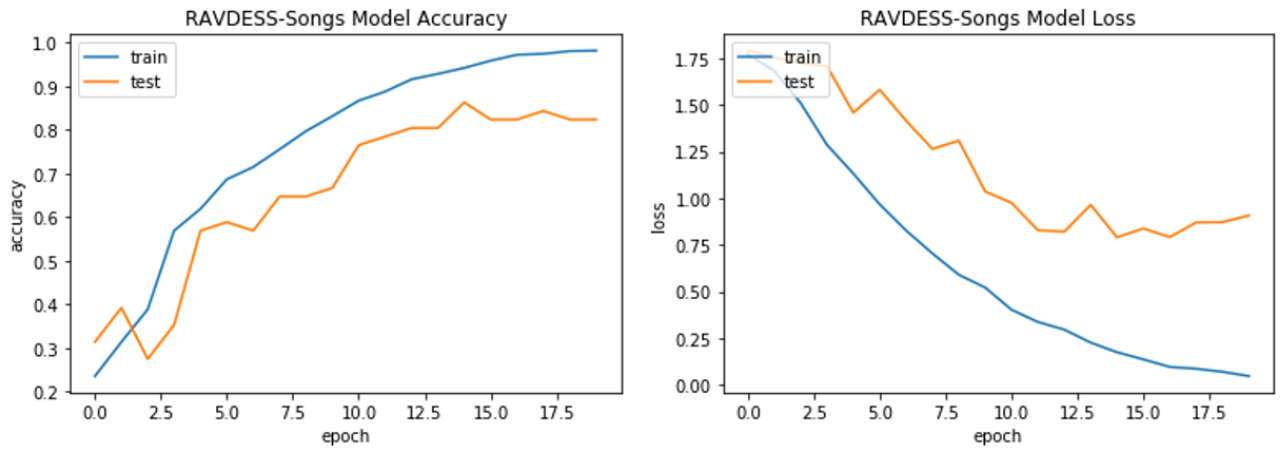


Figure 10

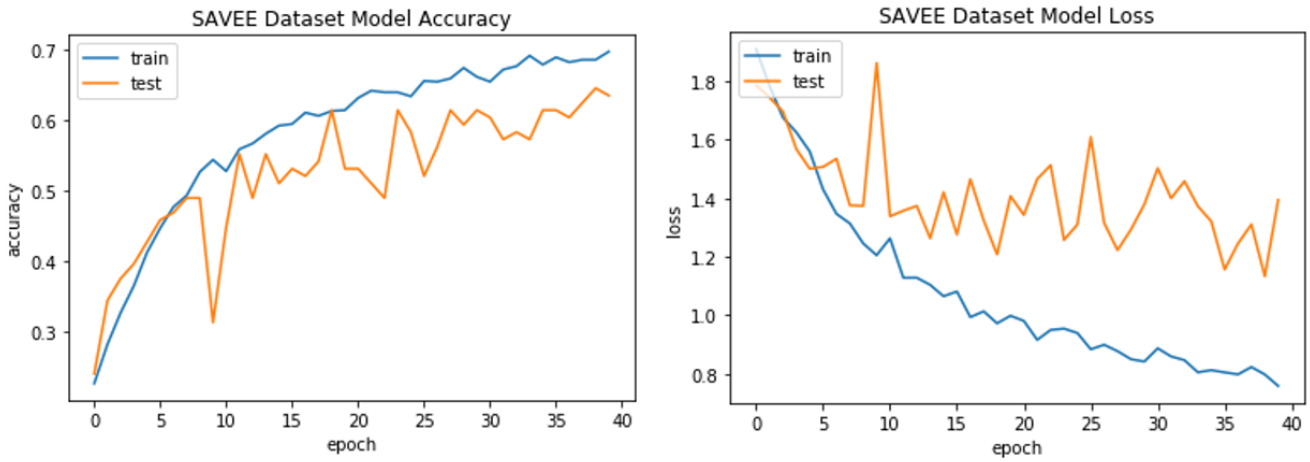


Figure 11

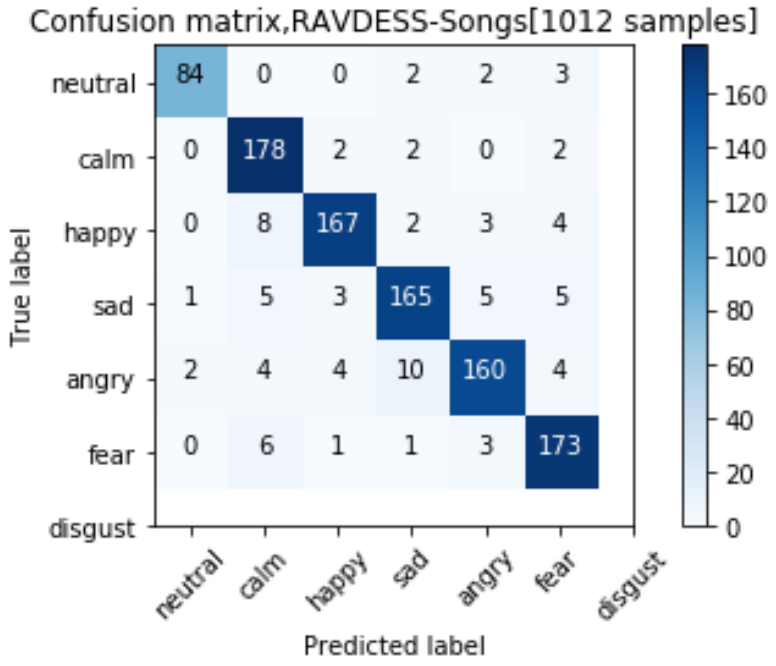


Figure 12

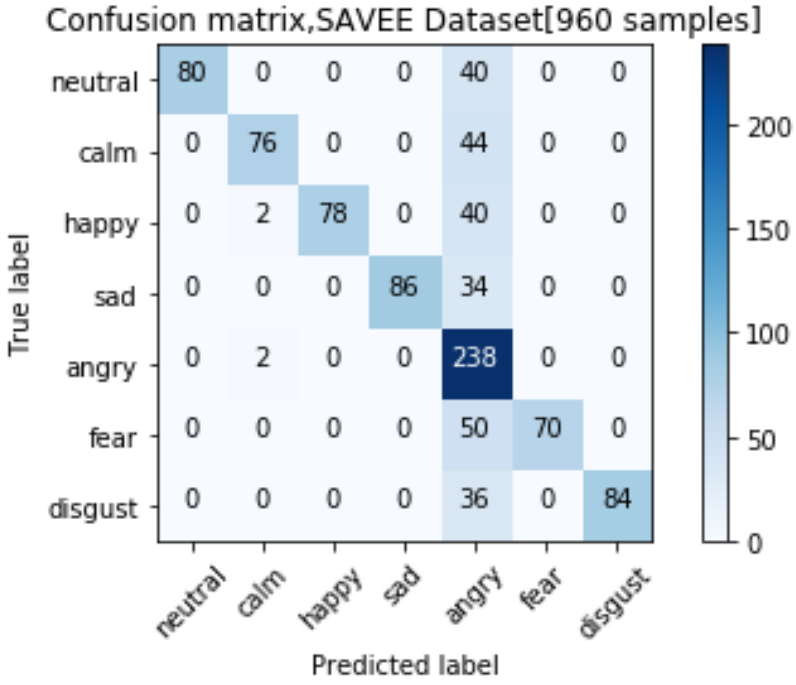


Figure 13

5.3 Sentiment Classification Model

The Sentiment Classification model's lexicon is built using Amazon review dataset, Twitter Airline Sentiment Dataset and Stanford Movie Review dataset [28] [29] [30]. The model is

trained on the combined text and label inputs from the three datasets. There are 36003 samples in training and 6354 samples in validation set. After preprocessing (according to the steps given in Figure 14) [11], [12], a file containing the occurrence count of words in each class is generated. Following that, a lexicon file is generated with each word having 3 representative values for each class (rv(w, X), where w is word and X is class), as calculated by:

```
for each word w:

if (count of w in class X = 0)
      rv(w, X) = 0
else if (count of w in class X = total count of w in all classes)
  rv(w, X) = 2.5 * count of w in class X
else
  rv(w, X) = 1/log10(total count of w in all classes/ count of w in
class X)
```

Figure 14

For each sample, the words in the text are looked up in the lexicon files and their corresponding rv(w, X) values are summed to get a three-valued final feature vector, used as input to the model. The model consists of an input layer followed by a dense layer with 6

nodes and ReLU activation and a softmax output layer (Figure 15). The classes are negative, neutral and positive. The validation accuracy of the model is 76.24%. [13], [14]

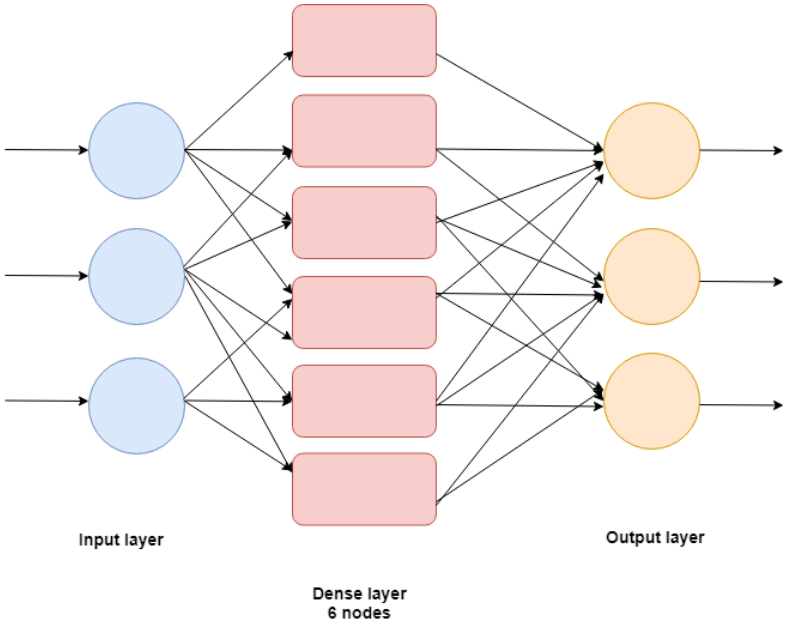


Figure 15

For batch size of 256 and 40 epochs, the training and validation accuracies and losses are plotted in Figure 16 and 17, respectively. The confusion matrix for the same is shown in Figure 18.

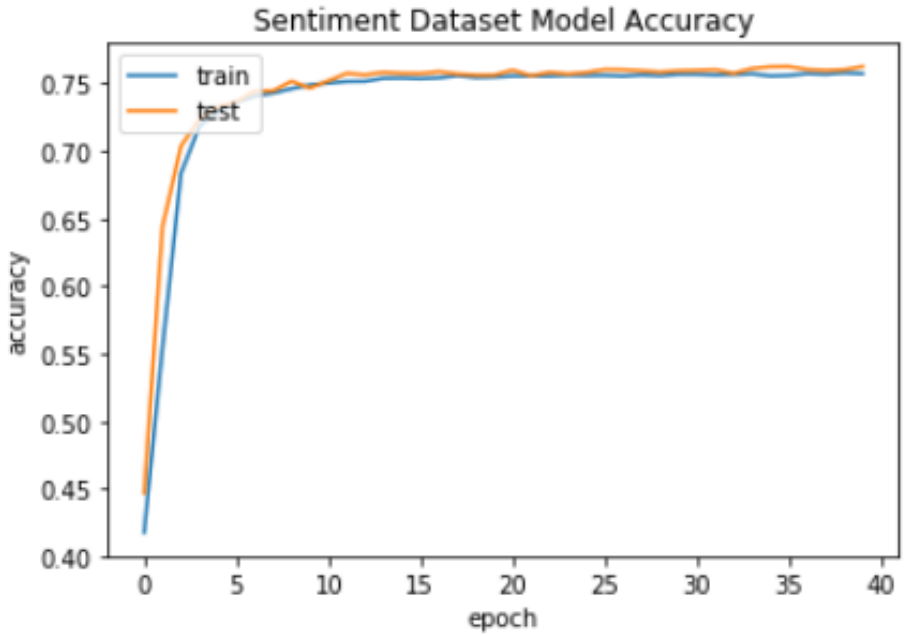


Figure 16

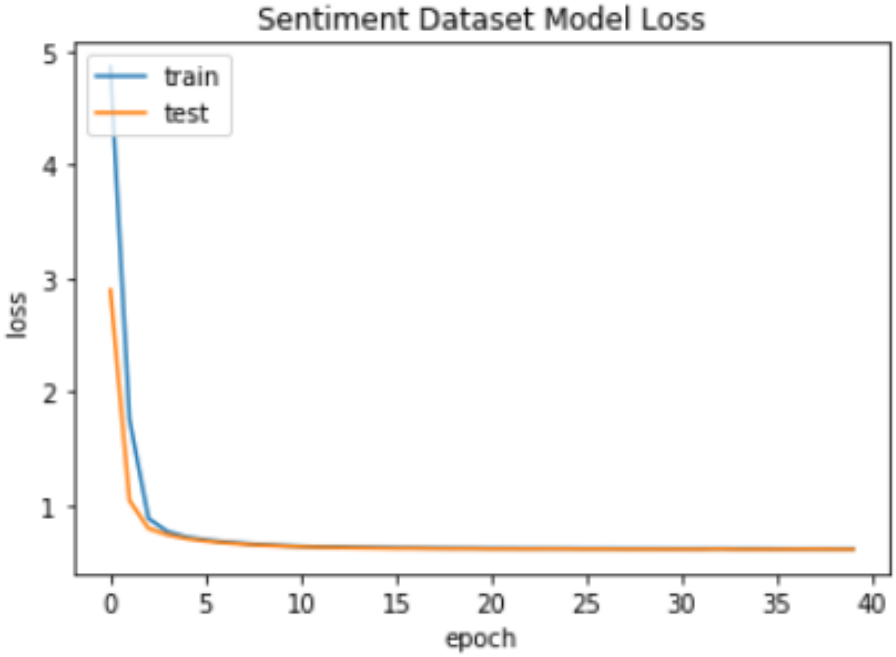


Figure 17

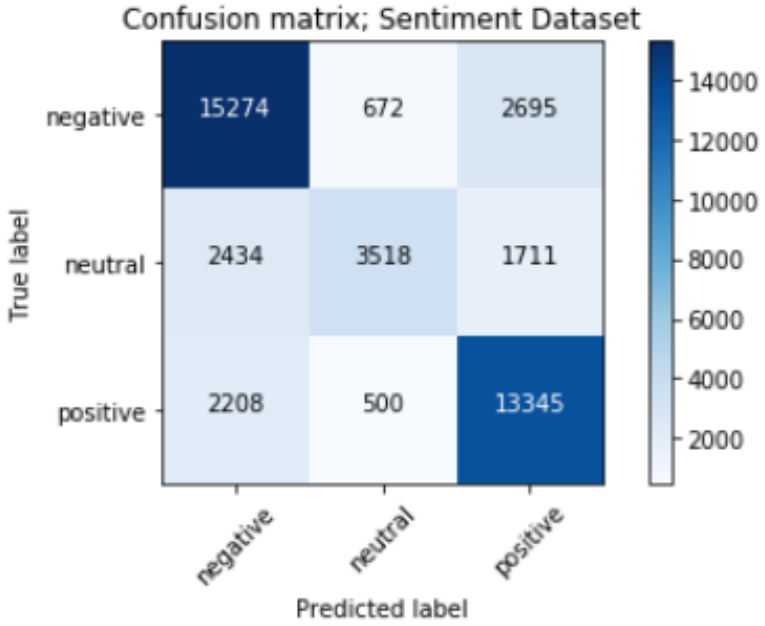


Figure 18

5.4 Similarity Based Scoring Function

In the question/answer functionality, the interviewee's response to a question is checked against the answer in the database, and using some similarity measures[15] and scaling, a score out of 100 is assigned to the interviewee's response. The interpretation of the score is entirely subjective, however, and requires the interviewer's understanding of the various ways of answering the question. This also implies the condition of the question being limited in terms of interpretations and kinds of responses it can generate.

The two answers are treated as two separate strings. Each string is tokenized using nltk library's TreebankWordTokenizer and each token is converted to its stem using nltk library's Snowball Stemmer. The words are detokenized back into a string. For cosine similarity, a list of terms, L occurring in the two strings are created. Count vectors are created for each string, that is, each string is represented by a vector containing occurrence count for each term in the list of terms L. Thus, we have two vectors A and B. Given n terms, there are n values in each of A and B, written as A_i and B_i where $1 \leq i \leq n$. The cosine similarity is calculated as follows:

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

For cosine similarity considering only two grams, the above process is repeated using two consecutive terms as a single entity for count vector creation. For jaccard similarity, the count of words common to both strings is found and divided by the number of terms in list of terms L. [16], [17]

$$Jaccard\ Index = \frac{count\ of\ common\ terms}{total\ number\ of\ terms}$$

The above scores are all in range [0, 1]. While cosine similarity calculates a good enough score, Jaccard similarity is given some weight to reduce the score in case the sizes of the strings differ vastly despite there being many common terms. The pseudocode for final score out of 100 is given as follows:

```
function final_score(string1, string2) {
     cosine1 = cosine_similarity_1gram(string1, string2)
     cosine2 = cosine_similarity_2gram(string1, string2)
     jaccard = jaccard_similarity(string1, string2)
     score = (cosine1 x 60) + (cosine2 x 16) + (jaccard x 24)
     if (0.1 ≤ cosine2 ≤ 0.2)
```

```
            score += 3
    else if (0.2 ≤ cosine2 ≤ 0.4)
            score += 6
    else if (0.4 ≤ cosine2 ≤ 0.6)
            score += 6
    else if (0.6 ≤ cosine2 ≤ 0.8)
            score += 3
    else if (0.8 ≤ cosine2)
            score += 16 * (1 - cosine2)
    return score
}
```

5.5 Database

Provision for two separate databases is provided in the tool that are to be maintained by the client or administrator. A sqlite database file is created which stores the filename and picture information in the form of BLOB. This can be modified by the administrator at any point of time. The previous and next buttons in the GUI window initiates the query in the database by maintaining a connection to it and returns respective images that are served on the GUI screen for image based description. Another database records list of questions and their respective answers entered by the administrator. They can be accessed and modified by querying the database by making a connection request to it. Questions are fetched from the database and displayed on the screen3 of the GUI window that can be answered by the user in the space provided. Further, the expected answer stored in the database is also retrieved along with the question and stored for similarity measure. [21]

6. RESULTS

The **FER model** trained on FER-2013 dataset shows 63% accuracy on validation set. The model was tested on a few images outside of the dataset. The images and results are given in Figures 19 to 22.

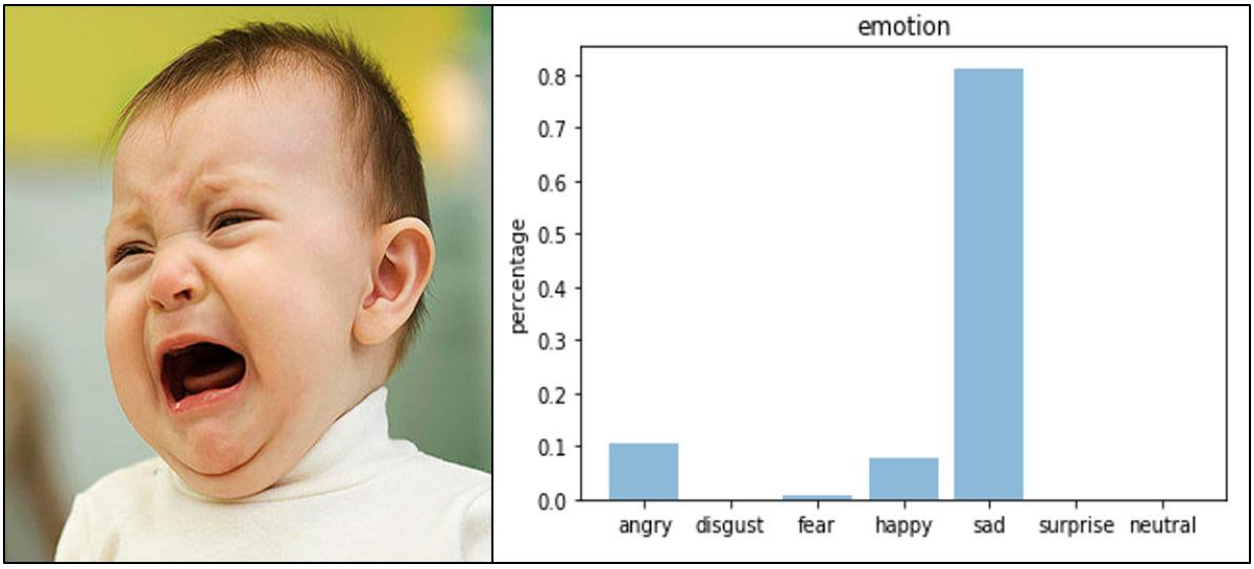


Figure 19

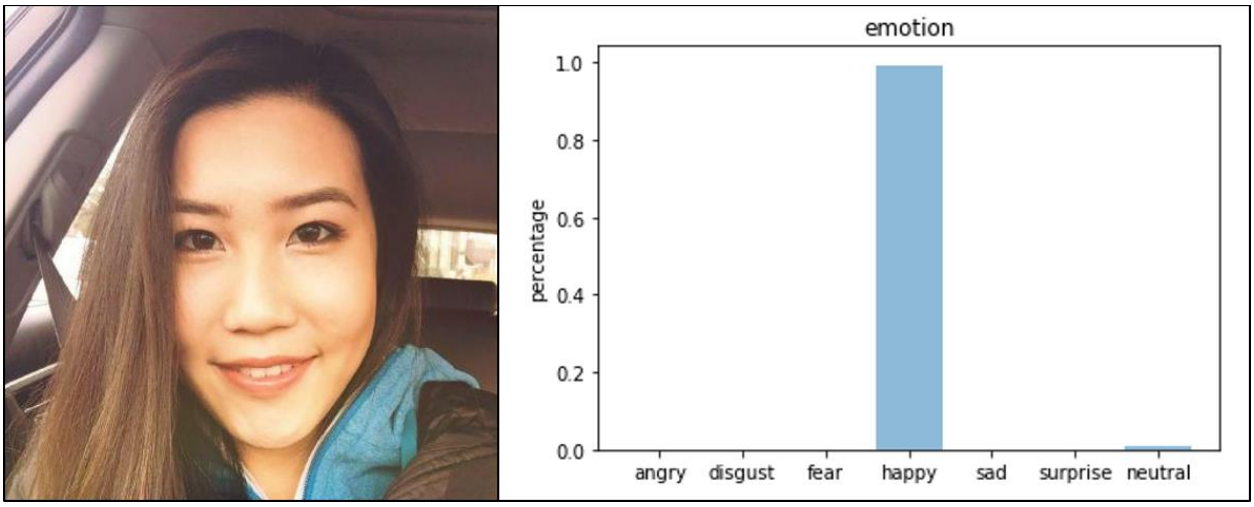


Figure 20

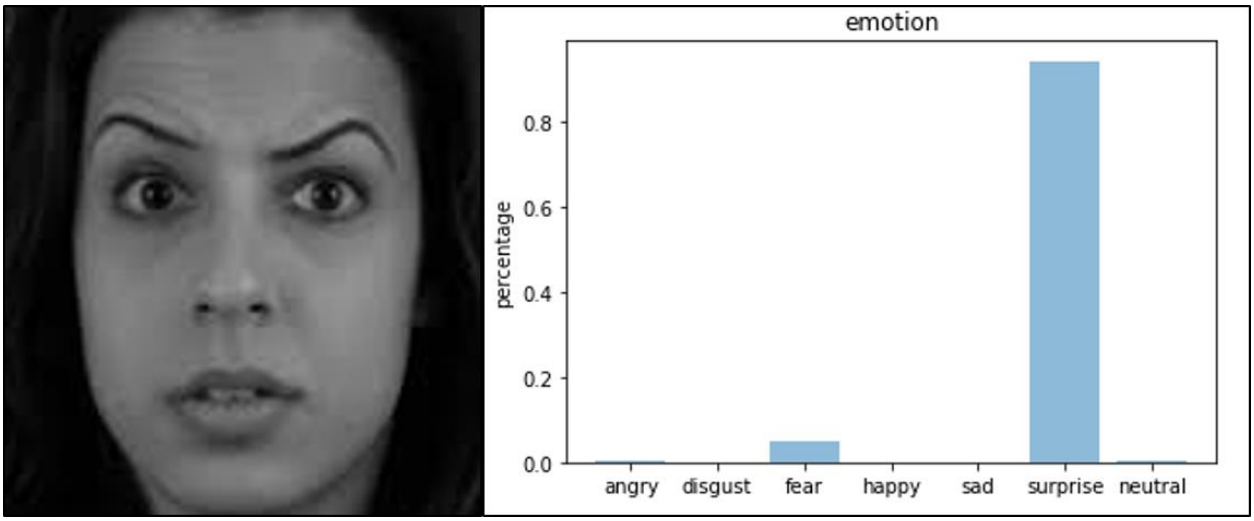


Figure 21

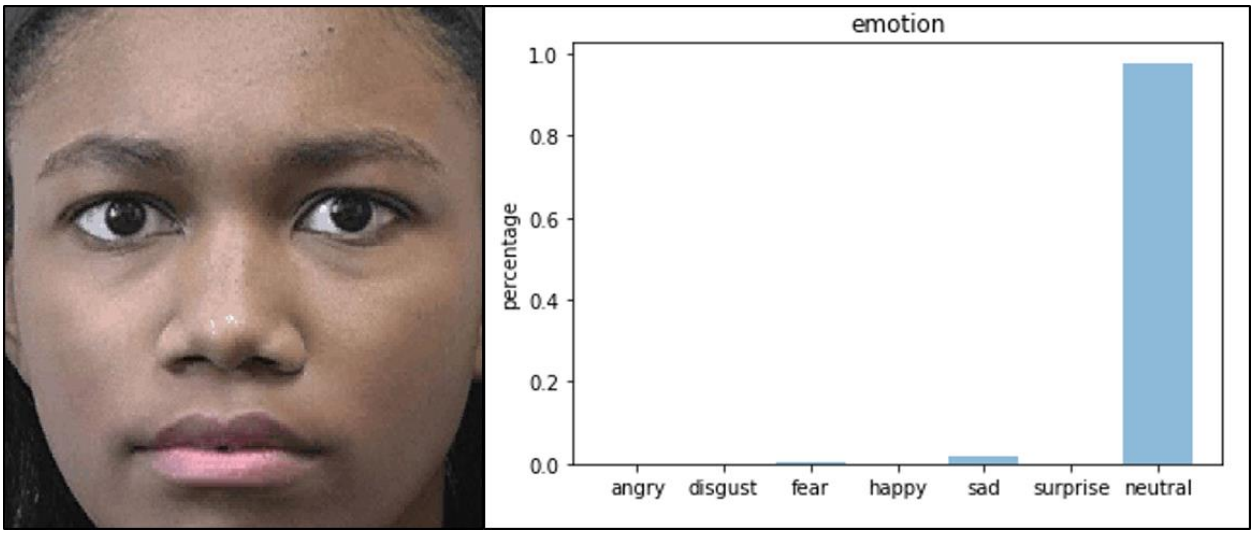
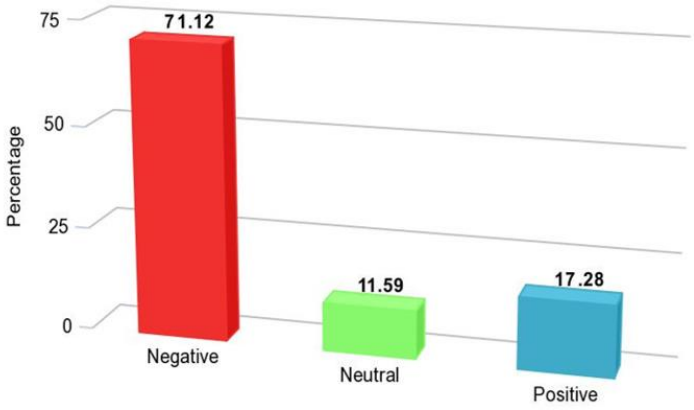


Figure 22

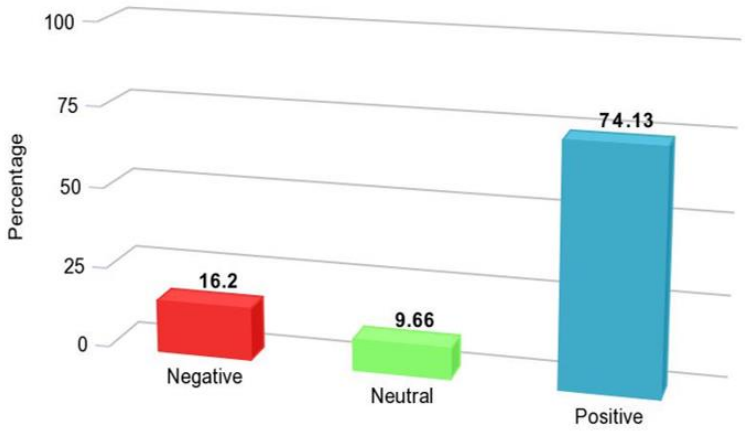
The **SER model** shows validation accuracy of 82.35% on RAVDESS Emotional Songs and 63.54% on SAVEE dataset.

The **Sentiment Classification model's** lexicon is built using Amazon review dataset, Twitter Airline Sentiment Dataset and Stanford Movie Review dataset. The validation accuracy is 76.24%. Figure 23 and 24 show the result on two unseen text inputs (not in dataset).



Text example 1
The lion is cowering before the hyenas, tired, weak and afraid.

Figure 23



Text example 2
A cheerful bird, sitting in the rain and enjoying every minute of it.

Figure 24

The four screens of the Graphical User Interface (GUI) are as follows:

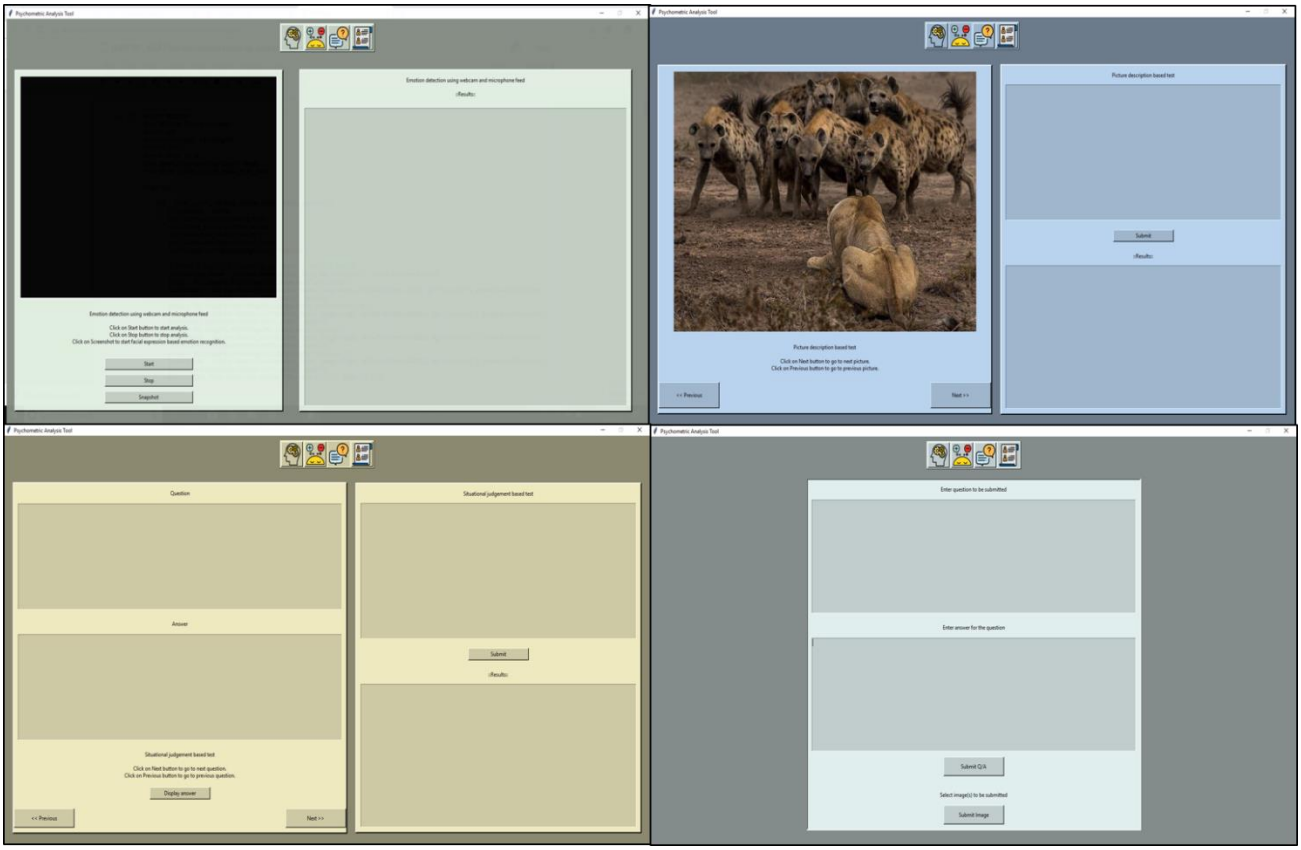


Figure 25

7. FUTURE WORK

The goals of the project were kept within what was believed to be attainable within the given timeframe and with the given resources and datasets. However, there is a significant scope for improvement and refinement by implementing certain changes and enhancements, which are listed as follows:

Registration and login: Currently, the tool can be used indiscriminately by any user. Registration and login functionality for test administrator and test giver can be added for security and restricted use. Using an API, a server can be created to store the registration information (written in the form by users) in a database table. For logging in, the username and password will be validated by the server.

Interaction in Audio/Video Emotion Detection: In the component that performs emotion detection using video and audio input, some visual cue (such as a question or image) could be provided to the user that prompts response. This will require simple modifications in the GUI.

Flexibility in setting tests: Restrictions on the test could be imposed by fixing number and order of questions and images given to the user. Time constraint can be added by fixing the response time per question. User responses can also be stored.

Gelling of components and results: The GUI displays raw results and a simplistic analysis for each individual test. The user's responses for each test could be stored and a comprehensive, all-inclusive analysis can be done for reaching a better conclusion.

Adaptive, interactive chat-bot for questioning and answering: The current navigate-through-questions process can be replaced by an interactive chat-bot that simulates conversational interview by adapting its own responses and deciding questions according to user's responses. This will require context detection and selective context memorization with respect to user's responses. The chat-bot will be trained to understand sentiment and context, and consult the database for shooting more questions.

CRUD Operations in database: Facility for modification and deletion from database contents can be added. Currently, only additions can be made.

8. REFERENCES

[1] National Council on Measurement in Education- http://www.ncme.org/ncme/NCME/ Resource_ Center/Glossary/NCME/Resource_Center/Glossary1.aspx?hkey=4bb87415-44dc-4088-9ed9-e8515326a061#anchorPArchived 2017-07-22 at the Wayback Machine.

[2] Koza, John R.; Bennett, Forrest H.; Andre, David; Keane, Martin A. (1996). Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. Artificial Intelligence in Design '96. Springer, Dordrecht. pp. 151–170

[3] Bishop, C.M. (2006), Pattern Recognition and Machine Learning, Springer, ISBN 978-0-387-31073-2 Zeiler, M.D. & Fergus, R.(2013). Visualizing and understanding convolutional networks. European Conference on Computer Vision(ECCV). 8689. 818-833.

[4] Jaswal, Deepika & Vishvanathan, Sowmya & Kp, Soman. (2014). Image Classification Using Convolutional Neural Networks. International Journal of Scientific and Engineering Research. 5. 1661-1668. 10.14299/ijser.2014.06.002.

[5] Huynh, Phung & Tran, Tien-Duc & Kim, Yong-Guk. (2016). Convolutional Neural Network Models for Facial Expression Recognition Using 3D-BUFE Database. 10.1007/978-981-10-0557-2_44.

[6] Zhang, Ting. (2018). Facial Expression Recognition Based on Deep Learning: A Survey. 345-352. 10.1007/978-3-319-69096-4_48.

[7] Schmidhuber, Juergen. (2014). Deep Learning in Neural Networks: An Overview. Neural Networks. 61. 10.1016/j.neunet.2014.09.003.

[8] Chernykh, Vladimir & Sterling, Grigoriy & Prihodko, Pavel. (2017). Emotion Recognition From Speech With Recurrent Neural Networks.

[9] Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735.

[10] Cen, Ling & Dong, Minghui & Li, Haizhou & Chan, Paul. (2010). Machine Learning Methods in the Application of Speech Emotion Recognition. 10.5772/8613.

[11] Desai, Mitali & Mehta, Mayuri. (2016). Techniques for sentiment analysis of Twitter data: A comprehensive survey. 149-154. 10.1109/CCAA.2016.7813707.

[12] Das, Bijoyan & Chakraborty, Sarit. (2018). An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation.

[13] Effrosynidis, Dimitrios & Symeonidis, Symeon & Arampatzis, Avi. (2017). A Comparison of Pre-processing Techniques for Twitter Sentiment Analysis. 21st International Conference on Theory and Practice of Digital Libraries (TPDL 2017). 10.1007/978-3-319-67008-9_31.

[14] Yun-tao, Zhang & Ling, Gong & Yong-cheng, Wang. (2005). An improved TF-IDF approach for text classification. Journal of Zhejiang University - Science A: Applied Physics & Engineering. 6. 49-55. 10.1007/BF02842477.

[15] Khuat, Tung & Duc Hung, Nguyen & Thi My Hanh, Le. (2015). A Comparison of Algorithms used to measure the Similarity between two documents. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET). 4. 1117-1121.

[16] Shoaib, Muhammad & Daud, Ali & Khiyal, Malik. (2014). An improved Similarity Measure for Text Doocuments.. Journal of Basic and Applied Scientific Research. 4. 215-223. 10.13140/2.1.4814.4006.

[17] Computer Engineering and Applications Vol. 5, No. 1, February 2016. Comparison Jaccard similarity, Cosine Similarity and Combined Both of the Data Clustering With Shared Nearest Neighbor Method Lisna Zahrotun.

[18] https://keras.io/optimizers/

[19] https://keras.io/getting-started/sequential-model-guide/

[20] https://docs.python.org/2/library/tkinter.html

[21] https://www.sqlite.org/docs.html

[22] https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_objdetect/py_face_detection/py_face_detection.html

[23] https://people.csail.mit.edu/hubert/pyaudio/docs/

[24] https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition -challenge/data

[25] https://tspace.library.utoronto.ca/handle/1807/24487

[26] https://smartlaboratory.org/ravdess

[27] http://kahlan.eps.surrey.ac.uk/savee/

[28] http://jmcauley.ucsd.edu/data/amazon/

[29] https://nlp.stanford.edu/sentiment/index.html

[30] https://www.kaggle.com/crowdflower/twitter-airline-sentiment