VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
Department of Computer Engineering

Project Report on

Pantomath 'Bo- Discover Cross Modal Human Behavior Analysis

In partial fulfillment of the Fourth Year (Semester–VII), Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai Academic Year 2017-2018

Submitted by

Ankit Shaw (D17 - C, 55)
Swastika Thakur (D17 - C, 64)
Pravin Tripathi (D17 - C, 65)
Pratik P. Watwani (D17 - C, 70)

Project Mentor
Mrs. Sharmila Sengupta

(2017-18)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
Department of Computer Engineering

Certificate

This is to certify that _____ of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on ***"PANTOMATH 'BO - DISCOVER CROSS MODAL HUMAN BEHAVIOR ANALYSIS"*** as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor ***Prof. Sharmila Sengupta*** in the year 2017-2018.

This thesis/dissertation/project report entitled ***Pantomath 'Bo - Discover Cross Modal Human Behavior Analysis*** by ***Ankit Shaw, Swastika Thakur, Pravin Tripathi, Pratik P. Watwani*** is approved for the degree of _____.

Program Outcomes	Grade
PO1, PO2, PO3, PO4, PO5, PO6, PO7, PO8, PO9, PO10, PO11, PO12, PSO1, PSO2	

Date:
Project Guide:

_____

Acceptance Letter

CereLabs
Know • Connect • Invent

28th July, 2017

To,
Dr. (Mrs.) Nupur Giri,
The Head of Department,
Computer Engineering Department,
VESIT

Subject: Acceptance Letter for BE Project

Dear Madam,

With reference to the above mentioned subject I would like to bring to your notice that the under mentioned students of Computer Engineering Department, VESIT, would be working under Cere Lab's guidance for their BE project, **"Pantomath 'bo : Discover Cross Modal Human Behavior Analysis"**, under their internal guide Prof. Sharmila Sengupta from the period July 2017 to March 2018.

I hereby state and confirm that upon successful completion of the project, a completion certificate will be issued to all the students and their internal guide.

Thanking you.

Members:
Pratik P. Watwani
Pravin Tripathi
Ankit Shaw
Swastika Thakur

Yours sincerely,

Devesh Rajadhyax
Founder & CEO

Cere Labs Pvt. Ltd.
10 – Jamnadas Industrial Estate, Near Jawahar Cinema, Dr. R.P. Road, Mulund (West), Mumbai – 400 080, Maharashtra India. **Tel : +91-22-65653301**
CIN: U72900MH2015PTC263117

Letter of Completion of Project

CereLabs
Know • Connect • Invent

22nd April, 2018

To
Dr. (Mrs.) Nupur Giri
Head of the Department
Computer Engineering
Vivekanand Education Society's Institute of Technology
Chembur

Completion and acceptance of the Final Year (B.E) Project

Dear Madam,
I am writing this letter to certify that the Final Year students, namely, **Ankit Shaw, Swastika Thakur, Pravin Tripathi and Pratik P. Watwani**, of Computer Engineering, VESIT has successfully completed their final year (B.E) project, "***Pantomath 'Bo: Discover Cross-Modal Human Behavior Analysis***" subordinate to my guidance (Cere Labs Pvt. Ltd.). And their internal guide **Prof. Mrs Sharmila Sengupta.**

As agreed upon acceptance, this letter is intended for the purpose of acceptance of their project and ultimately a successful completion under my belief.

Thank you.

Members:

Ankit Shaw
Swastika Thakur
Pravin Tripathi
Pratik P. Watwani

Internal Guide

Prof.(Mrs.) Sharmila Sengupta

Yours Sincerely

Cere Labs Private Limited * Mumbai *

Devesh Rajadhyax
Founder and CEO

Cere Labs Pvt. Ltd.
10 – Jamnadas Industrial Estate, Near Jawahar Cinema, Dr. R.P. Road, Mulund (West), Mumbai – 400 080, Maharashtra India. **Tel : +91-22-65653301**

Project Report Approval
For
B. E (Computer Engineering)

This thesis/dissertation/project report entitled ***Pantomath 'BO - Discover Cross Modal Human Behavior Analysis*** by ***Ankit Shaw, Swastika Thakur, Pravin Tripathi, Pratik P. Watwani*** is approved for the degree of

_____.

Internal Examiner

_____

External Examiner

_____

Head of the Department

_____

Principal

_____

Date:

Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

---	---
Ankit Shaw (Roll No - 55)	Swastika Thakur (Roll No - 64)
---	---
Pravin Tripathi (Roll No - 65)	Pratik P. Watwani (Roll No - 70)

Date:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Associate Professor **Mrs. Sharmila Sengupta** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement at several times.

Computer Engineering Department
COURSE OUTCOMES FOR B.E PROJECT

Learners will be to:

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solution for the problem.
CO 4	Able to interpret the data and datasets to be utilized.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one's competency.
CO 11	Able to develop professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

Abstract

Job interviews are a predominant part of any hiring process to evaluate applicants. It is used to evaluate applicant's knowledge, skills, abilities, and behavior in order to select the most suited person for the job. Recruiters make their opinion, on the basis of both verbal and nonverbal communication of an interviewee. Our behavior and communication in daily life are cross modal in nature. Facial expression, hand gestures and body postures are closely linked to speech and hence enrich the vocal content. Nonverbal communication plays an important role in what we are saying and what we actually mean to say. It carries relevant information that can reveal social construct of a person as diverse as his personality, state of mind, or job interview outcome; they convey information in parallel to our speech. In this paper we present an automated, predictive expert system framework for computational analysis of HR Job interviews. The system includes analysis of facial expression, language and prosodic details of the interviewees and thereby quantifies their verbal and nonverbal behavior. The system predicts the rating on the overall performance of the interviewee and on each behavior traits and hence predict their personality and hireability.

We introduce a computerized, predictive expert system framework of HR Job interviews. The framework incorporates the examination of facial expression appearance, dialect and prosodic subtle elements of the interviewees and accordingly evaluates their verbal and nonverbal conduct. The framework predicts the rating on the general execution of the interviewee and on every conduct attributes and henceforth foresee their identity and hireability.

Keywords: cross modal, expert system, facial expression, HR Interviews, Behavior Analysis, Cross-modal analysis.

Table of Content

Index

Chapter No.	Title	Page No.
	Certificate	2
	Acceptance Letter of Industrial Project	3
	Letter of Completion of Project (By Industry)	4
	Project Report Approval	5
	Declaration	6
	Acknowledgement	7
	Course outcomes for B.E Project	8
	Abstract	9
1	**Introduction**	15
1.1	Introduction to the project	16
1.2	Motivation for the project	17
1.3	Problem Definition	17
1.4	Relevance of the Project	18
1.5	Methodology used	18
2.	**Literature Survey**	24
2.1	Literature survey and its various sources	25
2.1.a	Abstract of the research paper	25
2.1.b	Inference drawn from the paper	25
3.	**Requirement of Proposed System**	32
3.1	Functional Requirements	33
3.2	Non-Functional Requirements	33

3.3	Constraints	33
3.4	Hardware, Software, Technology and tools – Requirements	34
3.5	Selection of the Hardware, Software, Technology and tools	36
4.	**Proposed Design**	39
4.1	Block Diagram of the System	40
4.2	Modular Diagram Representation of the System	40
4.3	Detailed Design of the Proposed System	42
4.3.a	Data Flow Diagram (Level 0,1,2)	42
4.3.b	Flowchart for the Proposed System	44
4.3.c	State Transition Diagram/Activity Diagram	45
4.4	Project Scheduling and Tracking using Gantt Chart	46
5	**Implementation Details**	49
5.1	Algorithms for respective Modules	50
5.2	Comparative Analysis with the existing algorithms	55
5.3	Evaluation of the Developed System	55
6	**Testing**	61
6.1	Unit Testing	62
6.2	Integration Testing	62
7	**Result Analysis**	63
7.1	Parameters Control	64
7.2	Screenshots of the User Interface	64
7.3	Graphical Output	69
7.4	Reports Generated/Tables Obtained	69
8	**Conclusion**	70
8.1	Limitation	71

8.2	Conclusion	71
8.3	Future Scope	71
	References	72
	Review Sheet	73
	Appendix	76
	CD	

LIST OF FIGURES

Sr. No	Name	Page No.
1	Figure 1.1 Project Framework	16
2	Figure 1.5 System (Interview) Setup	19
3	Figure 4.1 System Diagram	40
4	Figure 4.2.1 Input Module of the System	40
5	Figure 4.2.2 Analysis Module of the System	41
6	Figure 4.3.1.1 Data Flow Diagram Level 0	42
7	Figure 4.3.1.2 Data Flow Diagram Level 1	42
8	Figure 4.3.1.3 Data Flow Diagram Level 2	43
9	Figure 4.3.2 Flowchart Diagram of the System	44
10	Figure 4.3.3 State Transition Diagram/ Activity Diagram	45
11	Figure 4.4.1 Project Schedule	46
12	Figure 4.4.2 Gantt Chart of the Project	47
13	Figure 5.1.1 PRATT Audio Analysis	50
14	Figure 5.1.2.1 Head Tracking for Head Nod Detection	52
15	Figure 5.1.2.2 Eye Detection Trackers	52
16	Figure 5.1.2.3 Facial landmarks Detection	53
17	Figure 5.3.1 Hidden Layer Output, CNN	55
18	Figure 5.3.2 Goodness of Fit for Lasso Model	57
19	Figure 5.3.3 Goodness of Fit for SVR Model	57
20	Figure 5.3.4 Goodness of Fit for Ridge Model	58
21	Figure 5.3.5 Correlation coefficient comparison of a different set of features.	59
22	Figure 5.3.6 MSE Results for SVR and Ridge Models between Ground Scores and Predicted Scores.	59

23	Figure 5.3.7 MSE Results for SVR and Lasso Models between Ground Scores and Predicted Scores.	60
24	Figure 7.2.1 Login Page UI	64
25	Figure 7.2.2 – 7.2.3 Registration Page UI	65
26	Figure 7.2.3 Interviewer's Screen UI	66
27	Figure 7.2.4 Interview Progress UI	66
28	Figure 7.2.5 System Loading and Initiating UI	67
29	Figure 7.2.6 Proceedings of Interview UI	67
30	Figure 7.2.7 Output and Analysis Generation UI	68
31	Figure 7.2.8 Candidates List and Comparison UI	68
32	Figure 7.3 Smile Score Variation Graph	69
33	Figure 7.4 Reports Generated on Interviewer's UI	69

LIST OF TABLES

Sr. No	Name	Page No.
1	Table 1 Interview Questions	20
2	Table 2 Annotation Form Questions	21
3	Table 3 Prosodic Features	51
4	Table 4 Facial Landmarks Features Extracted	53
5	Table 5 Statistical analysis of Audio Cues	56

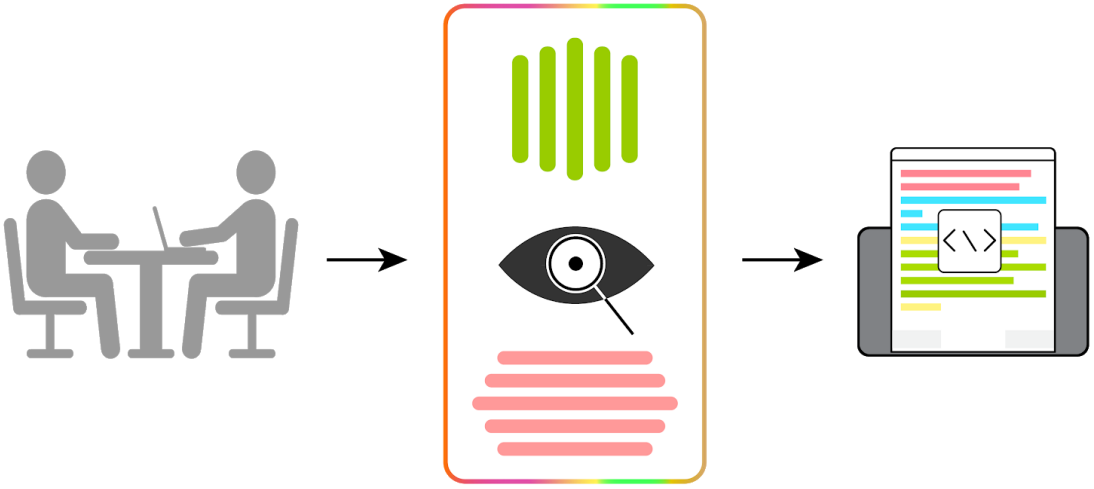
CHAPTER 1

INTRODUCTION

Chapter 1: Introduction

1.1 Introduction

Nonverbal communication is as significant as verbal communication while interviewing for a job. In up close and personal communication, our nonverbal conduct passes on data about our identity and characteristics notwithstanding our discourse. Nonverbal correspondence is an oblivious procedure and difficult to control, hence has an imperative impact in the result of the employment interview. The nonverbal correspondence incorporates visual cues, sound cues, and dialect or lexical cues. Visual cues or facial cues can have a capable impact on interview performance. Our physical appearance and how we dress are just piece of visual cues which are essential for an interview. Alongside this, visual cues likewise incorporate our facial expressions (grinning, nonpartisan, astound) amid discussion. Hand gestures, eye to eye connection, body introduction likewise adds to the visual signs amid correspondence.



Another nonverbal cue is the sound cues or paralinguistic cues which alludes to the vocal (prosodic) qualities and highlights. What the interviewee says as well as how (s)he says can make variations in the result of the interview. Voice adjustment, variety in pitch, the rate of discourse, intensity and so on are a portion of the imperative vocal signals that may break or make your odds to be employed. Monotone voice, limited pitch range are some portrays that are not favored in an interview. At last, the verbal signals like lexical highlights give us an incredible detail of data about the style of

talking about the interviewee and its substance. The decision of word we utilize while talking uncovers distinctive parts of our personality.

This project addresses the challenge of automated computational analysis of cross modal communication, including visual, language and prosody of job interview videos, to create a predictive expert system that helps predict hireability and rate different trait of the interviewees. The results are then displayed on the dashboard for the interviewer.

This predictive system can be used for screening, hiring candidates for a job interview and act as an expert support system for hiring managers to take appropriate decisions. This system can also be used for a fully automated screening of candidates in online interviews, this can help with initial shortlisting of suitable candidates for further domain specific evaluation.

1.2 Motivation

The motivation behind this project came from the placement going on in the college for our 2017-18 BE batch. In a placement process there are different rounds through with a candidate has to go through; from qualifying aptitude test to technical interview and finally the HR interview. HR interview is the last round but is as important as any other round. During such interview along with our verbal communication, the nonverbal behaviour play a very important role. Research says that application who use their nonverbal communication proper have high chances of getting hired. So this motivated us to take a project on automated human behaviour analysis in Job Interviews.

1.3 Problem Definition

To make an Expert System, a computational predictive framework that could analyse the candidate based on various communication cues and predict hireability and rate different traits.

This system provides solution to identify the personality type characteristics of an interviewee and classification according to the levels exhibited in respective personality characteristic through facial expression recognition techniques along with

vocal analysis to match with the personality characteristics of Passion, Cooperation, Confidence and Emotional Stability for job where high level of communication need to be done when performing the duties.

1.4 Relevance of the Project

Pantomath 'BO is being created with the point of helping the HR group of any association with prospective employee meetings. HR Interviews can be extreme and exceptionally unmistakable. Pantomath 'BO is a framework intended to assess and break down the unmistakable and various enthusiastic outward appearances of a candidate to help decide their qualities.

These qualities can go from confidence, stress, happiness, fear et al. The analytics of these characteristics will enable the HR team to decide the best-fit possibility for their job vacancies. Not just the system proposes to work with outward appearance yet in addition with the voice fierceness. Voice profundity, its variety, power and pitch can depict a considerable measure about the condition of the individual and their identity. As talked about, works with video as well as cumulatively with sound to give a remarkable examination of a hopeful execution in a prospective employee meeting. This is a framework which is intended to be extraordinary, helping an HR group of an organization as well as which can be used with preparing applicants. We can even convey this framework, for training institutes for training of a candidate for their dream employments. In any case, with certain change, this framework can even be reached out to simply adapt up to HR meets as well as any sort of interview be it Psychometric, Competency, Group Interviews and so on.

1.5 Methodology Used

Dataset Description

The first part of the project is creating an interview dataset.

a) Interview Collection- Data Gathering

The first part of the project research is collecting interview dataset. For this purpose, we conduct mock interviews for an open position. Around 50 video interviews are to be collected.

1) Interview Setup and Participants: Interview is taken under realistic setting. For the setup to conduct interview we need a camera to face the interviewee and two microphones one for the interviewee and another the interviewer. Interview is conducted in a room with 2 chairs and a desk. Camera is used to collect facial expressions and body cues while the microphone is to collect audio conversations during the course of interview. We need participants who are English speakers. Interview is conducted by professional counsellors.

Figure 2: Interview Setup

2) Process: Before the interview we conduct psychometric/personality test for each interviewee. This is done for all the interviewees for data gathering. Interview conducted was structured in nature. Total of seven questions were asked. Sequence of asking question remained same. We mapped these 7 question to different personality traits we were looking for through each questions. After the data collection, 6 independent judges assessed the interviews and rated them.

Sr. No.	Questions	Traits
1	Tell me about yourself?	General
2	Any past experience where stress was correctly managed by you?	Stress Resistance
3	What motivates you to do a good job?	Passion
4	Explain why we should hire you?	Confidence
5	Has anything ever irritated you about people and how do you deal with it?	Cooperation
6	Tell me about a time when you showed leadership skills.	Leadership
7	Tell me about your strengths and weaknesses.	General

b) Annotation- Data Labelling

Next step in data set building is labelling of the interviews, for which independent judges rate all the interviews manually. This is necessary to get the ground truth about the interview. Rating form is given to independent judges who will watch the whole interview and give scores for different assessment questions in the form. Since human ratings are subjective, we ask 2-3 different judges to rate the interview by answering assessment questions and then take average of their ratings.

The questions about "Overall Rating" and "Recommend Hiring" represent the overall performance. The remaining questions have been selected to evaluate several high-level behavioural dimensions such as warmth (e.g., "friendliness", "smiling"), presence (e.g., "engagement", "excitement", "focused"), competence (e.g. speaking rate), and content (e.g., "structured").

Sr.No	Annotation Form	
1	The overall performance rating!	Overall rating
2	How likely is the person to be hired?	Recommend Rating
3	Did he use engaging tone?	Engagement
4	Did he seem excited?	Excitement
5	Did he maintain proper eye contact?	Eye contact
6	Did he smile appropriately?	Smile
7	Did he seem Friendly?	Friendliness
8	Speaking rate was good?	Speaking Rate
9	Did he use too many filler words?	No fillers
10	Did he pause appropriately?	Paused
11	Did he seem authenticate?	Authentic
12	Was he appearing calm?	Calm
13	Was he seem focused?	Focused
14	Were his answers structured?	Structured
15	Was he stressed?	Not Stressed

Data Analysis

This is the second part of the project in which we perform data analysis on the interview recorded. This is done for feature extraction from the interview videos.

a) Audio Cues

The process of automatically detecting the person speaking during an interaction is known as speaker diarization, and consists of three main steps: the first is the classification of the audio stream into speech and non-speech segments; the second is the detection of speaker transitions and is referred to as speaker segmentation; the third is speaker clustering and its purpose is to group speech segments based on speaker characteristics. Prosodic features (voiced rate, pitch, and energy) are extracted from the audio. Several software packages to extract prosodic features are publicly available, such as Praat, Wavesurfer, or the MIT Media Lab speech feature extraction code.

For noise reduction and editing purposes Audacity – a free multitrack audio editor and recorder can be used. We will extract prosodic features of the interviewee's speech using the open-source speech analysis tool PRAAT.

The important prosodic features include the pitch information, vocal intensities, characteristics of the first three formants, and spectral energy, which have been reported to reflect our social traits.

b) Video Cues

Detecting faces robustly and accurately constitutes a strict requirement for the problem of automatically extracting kinesic cues related to facial expressions, gaze, and head nods. We extract facial features of the interviewees from each video frame. Head gestures such as nods and shakes are extracted from each video frame, and treat their average values as features.

c) Language Cues

The words we use while speaking or writing reveal many aspects of our identity. Several representations of verbal content have been successfully used for the prediction of social variables. Widely used in natural language processing, n-grams are a data-driven representation of verbal content which encode the occurrence of n subsequent words, and have been successfully used for the inference of Big-Five personality traits.

Incorporating word counts often results in sparse high-dimensional feature vectors, and suffers from the "curse of dimensionality" problem, especially for a limited sized corpus. Instead of using raw unigram counts, we can employ counts of various psycholinguistic word categories defined by the tool "Linguistic Inquiry Word Count" (LIWC). The LIWC categories include negative emotion terms (e.g., sad, angry), positive emotion terms (e.g., happy, kind) and various content word categories (e.g., anxiety, insight).

Predictive Framework / Classification

Regression models will be trained to predict the interview scores and other interview-specific traits such as excitement, friendliness, engagement, awkwardness, etc. For the prediction framework, we automatically extracted various features from the videos of the interviews. Then we trained two regression algorithms – SVR, LASSO and Ridge.

CHAPTER 2

LITERATURE SURVEY

Chapter 2: Literature Survey

2.1 Research Paper

T. DeGroot and J. Gooty. "Can nonverbal cues be used to make meaningful personality attributions in employment interviews?" Journal of Business and Psychology, 24(2):179– 192, 2009.

2.1.a Abstract

a. This paper studies the relationships between nonverbal cues and interview performance ratings and also examines the role of personality attributions.
b. Conscientiousness attributions explain the relationship between visual cues and interview ratings, extraversion attributions mediate the relationship between vocal cues and interview ratings. Neuroticism attributions had a suppressing effect for both visual and vocal cues.
c. The structured behavioural interview is rich in behavioural and non-behavioural information available to the interviewer. While the non behavioural information might consist of right or wrong answers related to the content of questions, the behavioural information is less structured and therefore more open to interpretation.
d. In this paper three types of nonverbal cues are studied - dynamic, static, and paralinguistic cues. Dynamic cues are easily changed such as eye contact, body orientation, smiling, gesturing, and head movement. Static cues are demographic variables and physical attractiveness. Paralinguistic cues refer to vocal characteristics such as speech rate, volume, tone, and pausing.
e. The Brunswik lens model shows the potential impact of personality attributions when both channels of information are used.

2.2.b Inference drawn from the paper

a. Coefficients for the model are computed by regressing the combined personality attributions on both dependent variables, job performance on one side of the lens model and expert interview ratings on the other side of the lens model.

b. This suggests that interview raters and job performance raters almost equally use personality attributions when making their ratings.
c. Understanding the impact of effects of personality in interviews has great organizational relevance. Extraversion has been shown to be a much-desired managerial trait.
d. One conclusion which can be made is that no matter how much an interview is structured; nonverbal cues cause interviewers to make attributions about candidates.

Andrew S. Imada and Milton D. Hakel .“Influence of Nonverbal communication and Rater Proximity on Impressions and Decisions in Simulated Employment Interviews”Journal of Applied Psychology 1977, Vol. 62, No. 3, 295-300.

2.1.a Abstract

a. This paper studies various ways to assess job applicants and examines the role of nonverbal communication in hiring process.
b. Nonverbal behaviors enable the applicant to convey an "immediacy" dimension in interviews. Immediacy in an interaction between two individuals involves greater perceptual availability of two persons.
c. It investigates the effects of the applicant's nonverbal behaviors and rater proximity on interview impressions and decisions.

2.2.b Inference drawn from the paper

a. According to the paper, through the use of eye contact, gestures, smiling, smaller interpersonal distances, an attentive posture, and a more direct body orientation, the applicant will be perceived and described as a warmer and more enthusiastic person.
b. When nonverbally immediate, the applicant will be perceived and rated as being more likely to be accepted, more successful, more qualified, better liked, as having more desirable characteristics, being more motivated, more competent, and more satisfied if given the position, and as a result will be recommended for the position.
c. Interviewers feel very comfortable and satisfied when interacting with a nonverbally immediate applicant and uncomfortable and dissatisfied when interacting with a non-immediate applicant. This, however, is not the case for observers.

Vikrant Soman and Anmol Madan. "SOCIAL SIGNALING: PREDICTING THE OUTCOME OF JOB INTERVIEWS FROM VOCAL TONE AND PROSODY" IEEE In'tl Conference on Acoustics, Speech and Signal Processing, Dallas TX March 2009.

2.1.a Abstract

a. Recruiters and psychologists believe that speaking style, confidence and behavior plays a very important role in the face-to-face interview process.
b. Non-linguistic verbal cues play an important role in the outcomes of interviews.
c. Social signalling measures are quite effective in predicting behavioral outcomes in different social interactions.
d. Using features extracted from student's vocal tone and prosody, non-linguistic speaking style can be quantified.
e. Successful candidates have a characteristic speaking style and these vocal features can be used to build a predictive model of the interview outcomes, with over 85% accuracy.

2.2.b Inference drawn from the paper

a. Highly-rated candidates have, on average, higher activity levels. These candidates also show lower levels of vocal emphasis during the interview conversation.
b. Ongoing social signalling between an interviewee and interviewer during a job interview can be modelled using vocal tone and prosody features.
c. The extracted features have a high correlation and predictive value with regard to the overall rating and engaged in discussion ratings.
d. Non-Linguistic communication channel can be accurately quantified using automated methods, and even used as a basis to predict future outcomes.

Naim Iftekhar, Tanveer Md, Gildea Daniel & Hoque Ehsan. (2015). Automated Prediction and Analysis of Job Interview Performance: The Role of What You Say and How You Say It.. 10.1109/FG.2015.7163127.

2.1.a Abstract

a. A common perception surrounding job interviews is that the content of the interviewee's answers is the most important determinant for success.
b. However, empirical studies show that nonverbal behaviour is as important as the verbal response in job interviews. Nonverbal behaviours are subtle, fleeting, subjective, and sometimes even contradictory, posing a significant challenge for any prediction framework.
c. In this paper, an automated prediction framework is designed and implemented for quantifying the outcome of job interviews, given the audio-visual recordings.
d. The proposed prediction framework automatically extracts a diverse set of multimodal features (lexical, facial, and prosodic), and quantifies the overall interview performance, the likelihood of getting hired, and 14 other social traits relevant to the interview process.

2.2. b Inference drawn from the paper

a. The outcome of job interviews often depends on a subtle understanding of the interviewee's response. A momentary mistake can ruin the interview outcome.
b. Therefore, anomaly detection techniques are better.
c. Facial features, particularly the smile, are found to be significant for predicting friendliness. Besides smiling, people who speak more words associated with "We" than "I" are perceived as being friendlier.
d. Therefore, it is recommended to speak more fluently, use less filler words, speak as "we" (vs. "I"), use more unique words, and smile more.

L. Nguyen, A. Marcos, M. Marron, and D. Gatica. Multimodal analysis of body communication cues in employment interviews. In ACM International Conference on Multimodal Interaction (ICMI), 2013.

2.1.a Abstract

a. Gestures are an essential component of body communication as they are used to enrich the vocal content and aid listener comprehension by augmenting the attention, activating images or representations in the listener's mind, and increasing the recall of what is being said.
b. Job hireability impressions and self-rated personality can be predicted using body communication cues.
c. Speaking status can be used to improve the prediction performance of personality and hireability
d. In order to analyse the predictive validity of body posture with respect to self-rated personality traits and hireability impressions, a regression problem was defined which aims at predicting the exact hireability and personality scores, where each social variable is considered as an independent regression task.

2.2.b Inference drawn from the paper

a. Except communication and conscience, all hireability measures are significantly correlated with each other.
b. Extraversion is significantly and positively correlated with three hirability scores: hiring decision, conscience, and stress resistance.
c. The automatic features based on hand speed and hand acceleration are conditioned on the speaking status.
d. The best prediction results for the hiring decision and stress resistance are achieved using automatic cues based on activity histograms.

A. T. Rupasinghe, N. L. Gunawardena, S. Shujan and D. A. S. Atukorale, "Scaling personality traits of interviewees in an online job interview by vocal spectrum and facial cue analysis," 2016 Sixteenth International Conference on Advances in ICT for Emerging Regions (ICTer), Negombo, 2016, pp. 288-295.

2.1.a Abstract

a. Online interviews have become a trend in the process of selecting an appropriate candidate. Synchronous interviews directed in ongoing over the Web can be utilized in the place of traditional face-to-face interviews.
b. Online interviews can be used as an initial study to the formation of first impressions for initial screening and shortlisting purposes.
c. An ICT based solution is provided in this paper to identify the personality type characteristics of an interviewee.
d. It is then classified according to the levels exhibited in respective personality characteristic through facial expression recognition techniques along with vocal analysis to match with the personality characteristics of Passion, Cooperation, Confidence and Emotional Stability for IT job positions where high level of communication need to be done when performing the duties.

2.2.b Inference drawn from the paper

a. The accuracy levels can be increased by applying the model only for the male(M) or female(F) participants as the models which are developed by merging both male and female participants result in reducing the accuracy levels.
b. Same classifier cannot be applied to all the selected personality traits in order to achieve higher accuracy levels.
c. When less number of classification classes are used and if some classes have low number of datasets, there‘s a high possibility of the results achieving extreme ends.
d. To improve the accuracy of the outcome, sample size should be larger while maintaining a balanced gender distribution.

CHAPTER 3

REQUIREMENTS OF PROPOSED SYSTEM

Chapter 3: Requirements of Proposed System

3.1 Functional Requirement

I. With priority, to find out how the candidate is acknowledging with facial expressions and their tone of speech and thereby conduct the required analysis.
II. Utilizing the voice of the candidate to know more about the competency of the candidate (Physical behavior of nervousness, confidence, shyness el at.)
III. It should give an overall candidate score to make better decision.
IV. It should be able to handle multiple candidate simultaneously.
V. It should analyze the interview videos and automatically extract features from them.

3.2 Non- Functional Requirements

I. It should use Open Source Softwares to minimize the cost of development.
II. It should use Raspberry Pi for deployment on client side.
III. High data storage capacities in servers.
IV. System should be portable if required to move to a different platform, requiring not more than 20% of module adjustment
V. Should consume the least amount of memory and computational resources
II. The operating constraint includes the presence of at least one operator, required for initiating, questioning and finishing the session, and one interviewee who would be sole participant of the session
III. Legality should be maintained, as open source software is used, legality issues should be taken care in prior to deployment of the product.

3.3 Constraints

I. Storage of raw data.
II. High storage demand.
III. High Computational Requirement. (GPUs)

3.4 Hardware, Software, Technology and tools- Requirements

I. Hardware Specification:

a. Pi Camera -

The basic purpose of camera is to capture the video of the interviewee.

b. Microphone -

It will be used for capturing audio of the interviewee.

c. Raspberry Pi kit 3 -

It will be used for physical deployment of the project. It will act as a front end of the project.

II. Software Specification:

a. Python
b. Praat / Audacity
c. Apache Server
d. MySQL Database

III. Technology

a. Image Processing -

Digital image processing is the use of computer algorithms to perform image processing on digital images. As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing.

b. Prosodic Analysis -

Analysis of a language based on its patterns of stress and intonation in different contexts. In systemic grammar, prosodic analysis is regarded as an essential foundation for the analysis of syntax and meaning.

c. Machine Learning -

Machine learning is a field of computer science that gives computer systems the ability to "learn" with data, without being explicitly programmed.

d. Natural Language Processing -

Natural-language processing is an area of computer science and artificial intelligence concerned with the interactions between computers and human languages.

IV. Tools

Tools and ML libraries:

a. Tensorflow

TensorFlow is an open-source software library for machine learning across a range of tasks. It is a symbolic math library, and also used as a system for building and training neural networks to detect and decipher patterns and correlations, analogous to human learning and reasoning.

b. Keras

Keras is an open source neural network library written in Python. It is capable of running on top of MXNet, Deeplearning4j, Tensorflow, CNTK or Theano. Designed to enable fast experimentation with deep neural networks, it focuses on being minimal, modular and extensible. In 2017, Google's TensorFlow team decided to support Keras in TensorFlow's core library. Chollet explained that Keras was conceived to be an interface rather than an end-to-end machine-learning framework. It presents a higher-level, more intuitive set of abstractions that make it easy to configure neural networks regardless of the backend scientific computing library.

c. CUDA Toolkit

CUDA is a parallel computing platform and application programming interface (API) model created by Nvidia. It allows software developers and software engineers to use a CUDA-enabled graphics processing unit (GPU) for general purpose processing – an approach termed

GPGPU (General-Purpose computing on Graphics Processing Units). The CUDA platform is a software layer that gives direct access to the GPU's virtual instruction set and parallel computational elements, for the execution of compute kernel.

d. PRAAT

Praat is a free computer software package for the scientific analysis of speech in phonetics. The program supports speech synthesis, including articulatory synthesis.

e. Opencv

OpenCV (*Open Source Computer Vision*) is a library of programming functions mainly aimed at real-time computer vision. It was originally developed by Intel. OpenCV supports the Deep Learning frameworks TensorFlow, Torch/PyTorch and Caffe.

3.5 Selection of the Hardware, Software, Technology and tools

I. Hardware:

a. Pi Camera

The basic purpose of camera is to capture the video of the interviewee.

b. Microphone -

It will be used for capturing audio of the interviewee.

c. Raspberry Pi kit 3 -

It will be used for physical deployment of the project. It will act as a front end of the project.

II. Software:

a. Python -

The project is implemented in python since its easy to use and is simple and minimalistic language. It contains large repository of packages needed for machine learning.

b. Praat/Audacity -

Audacity was used to clean the audio files of the interviews taken to get rid of noise present in it.

Praat is for the scientific analysis of speech in phonetics. We used it to extract audio features from the audio files.

c. Apache Server –

The project is implemented as a standalone project so for handling request from client machine we are using open source software. This will open a port on the Raspberry pi and later client can talk to it via not network protocols

III. Technology

a. Image Processing -

Image processing is used to access webcam. We use it to get record interviews in real time. It is used to detect faces in video frames before video cues extraction. It is used to detect head nod of the candidates.

b. Prosodic Analysis -

Prosodic analysis is needed to extract audio cues from the audio files of the interview.

c. Machine Learning -

Machine Learning is first needed in video cue extraction; in which we used it to detect emotions shown by the candidate. We also use it to plot facial key points. ML is also used to train our predictive model to predict ratings for different traits of interviewee.

IV. Tools

a. Tensorflow -

It was used to train the FER dataset for emotion detection. It was used as backend in Keras to train and then predict the emotion in real time.

b. Keras -

Keras is the high level API we used to access tensorflow functions.

c. CUDA Toolkit -

CUDA Toolkit was to to enable the gpu for parallel processing for our project.

d. Praat -

Praat was used to extract prosodic/audio features from the audio files of the interview.

e. OpenCV -

OpenCV was used for real time processing for the interview being recorded. It was used for head nod detection using optical flow technique, it was also used to detect faces in the frame using haar cascade.

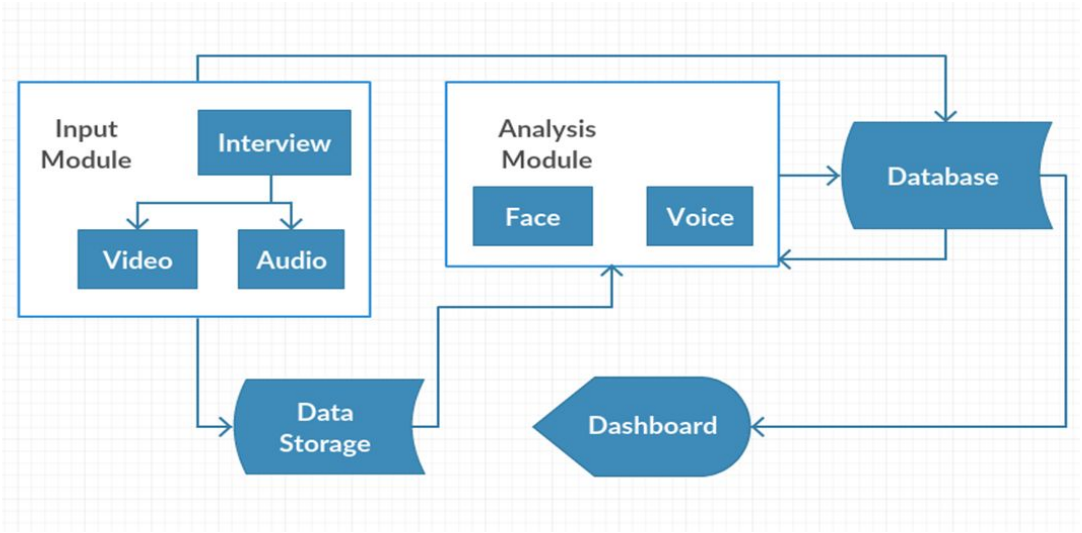
CHAPTER 4

PROPOSED DESIGN

Chapter 4: Proposed Design

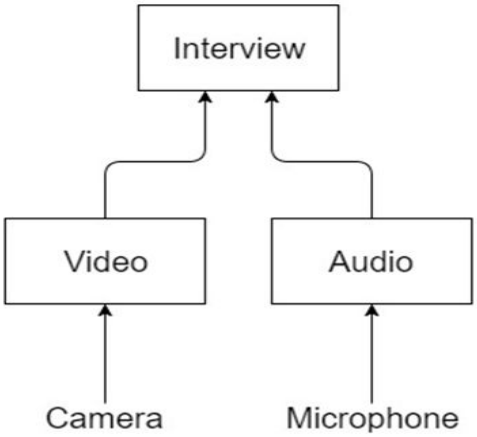
4.1 Block Diagram

The system has two modules, Input, and Analysis respectively. The input module is used to automatically collect video and audio from the input devices and store the data into a suitable data storage. The analysis module takes the data (recorded interviews) singularly to perform the analysis on them and thereby extract features from audio, video and language cues.



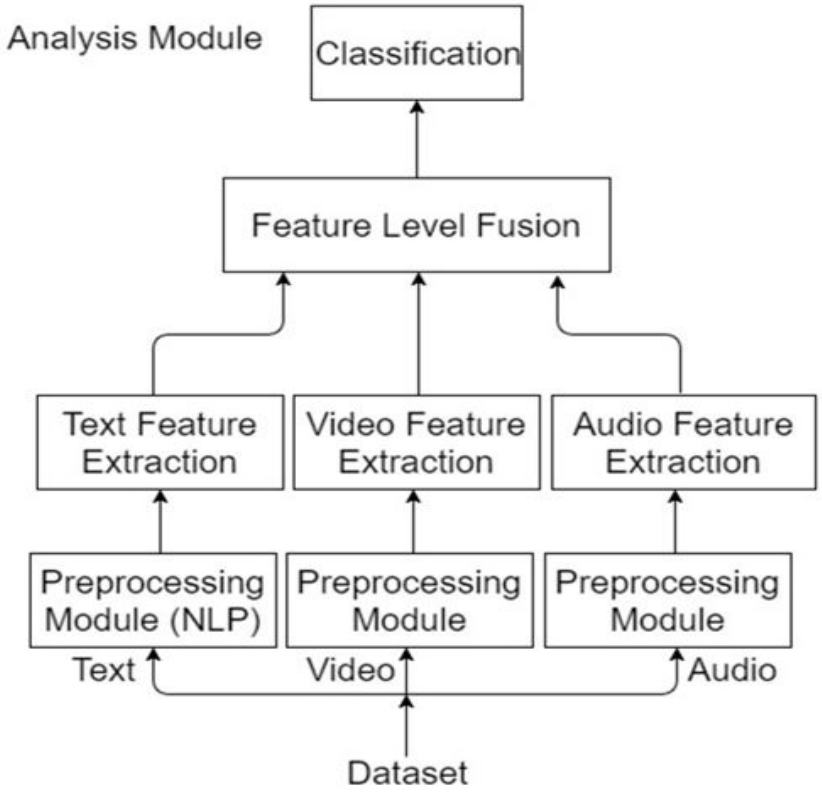
4.2 Modular Design

4.2.1 Input Module



The input module is used to collect the interview. We collect the video from the camera and audio from the microphone. For real time recording, we used threading to collect video and audio simultaneously.

4.2.2 Analysis Module

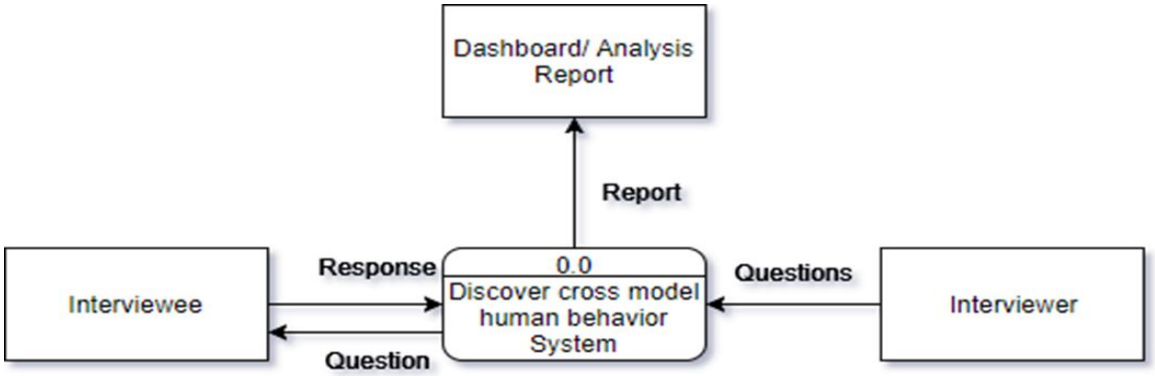


The analysis module is used for analysis of interview, starting from pre-processing till classification. First we pre-process the text, video and audio then we extract their features. After features of the text, video and audio are extracted, we merge the features into single feature vector. This feature vector is then put into the classifier to predict the ratings.

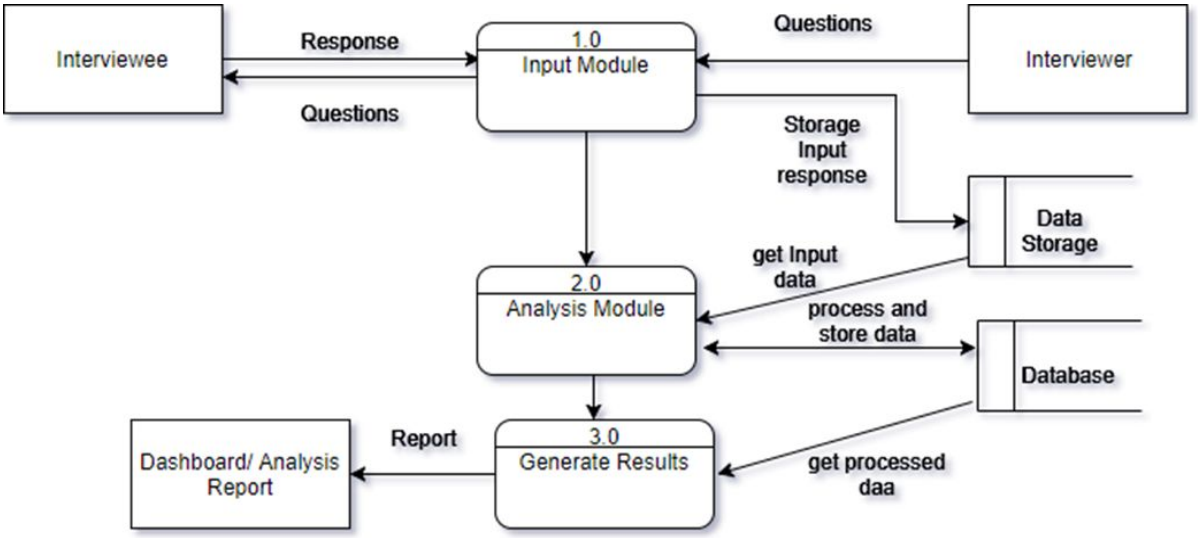
4.3 Detailed Design

4.3.1 Data Flow Diagram

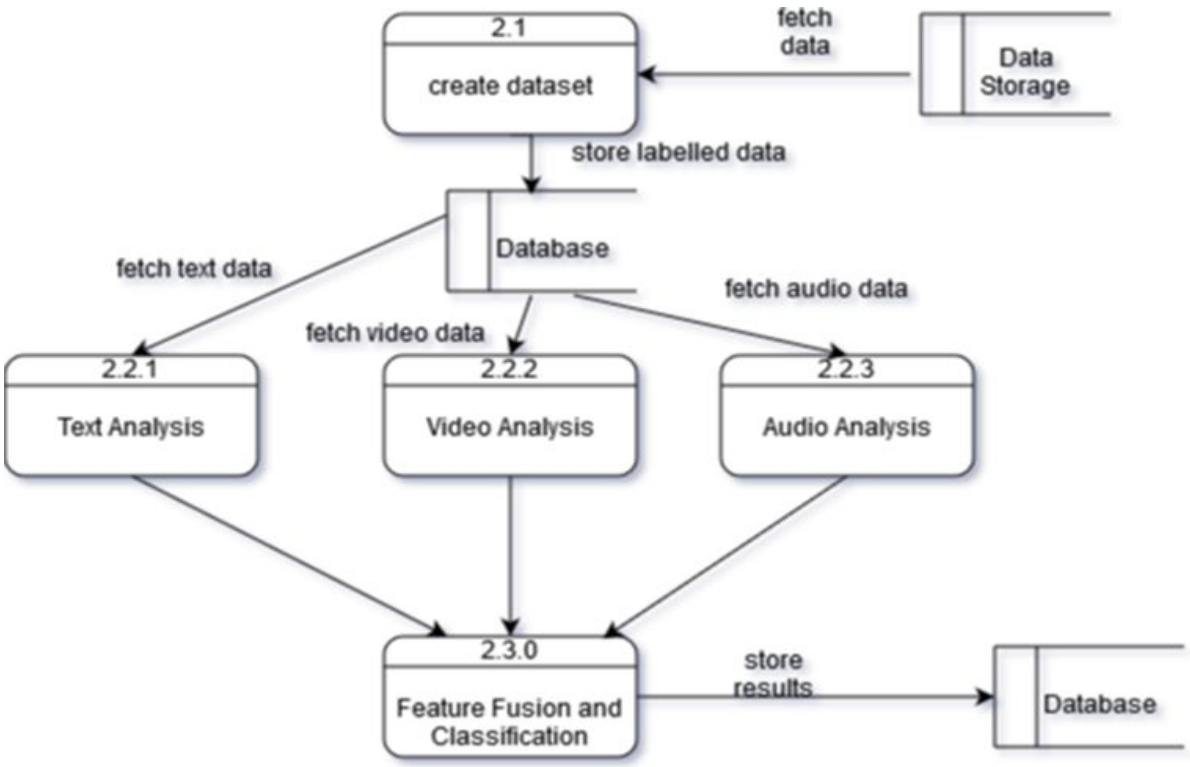
Level 0:



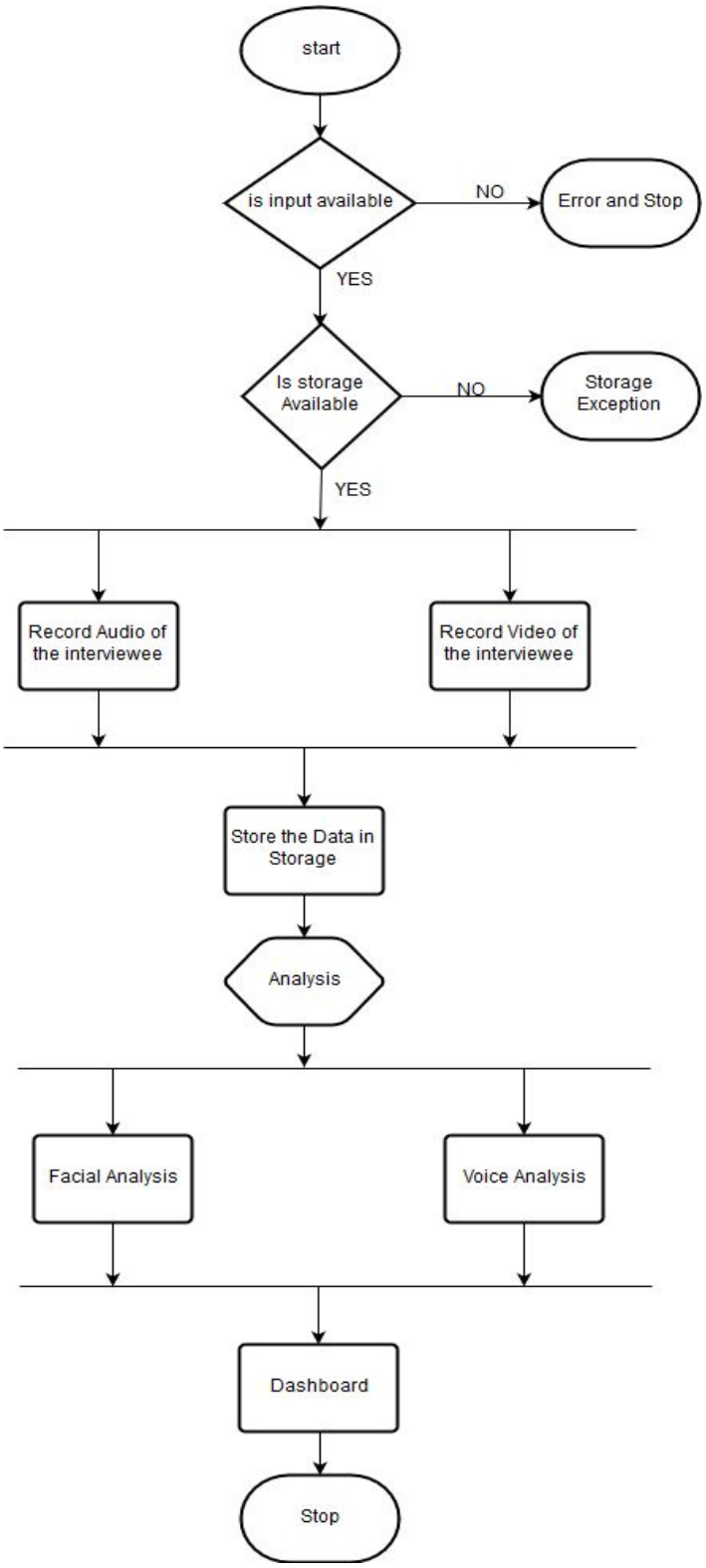
Level 1:



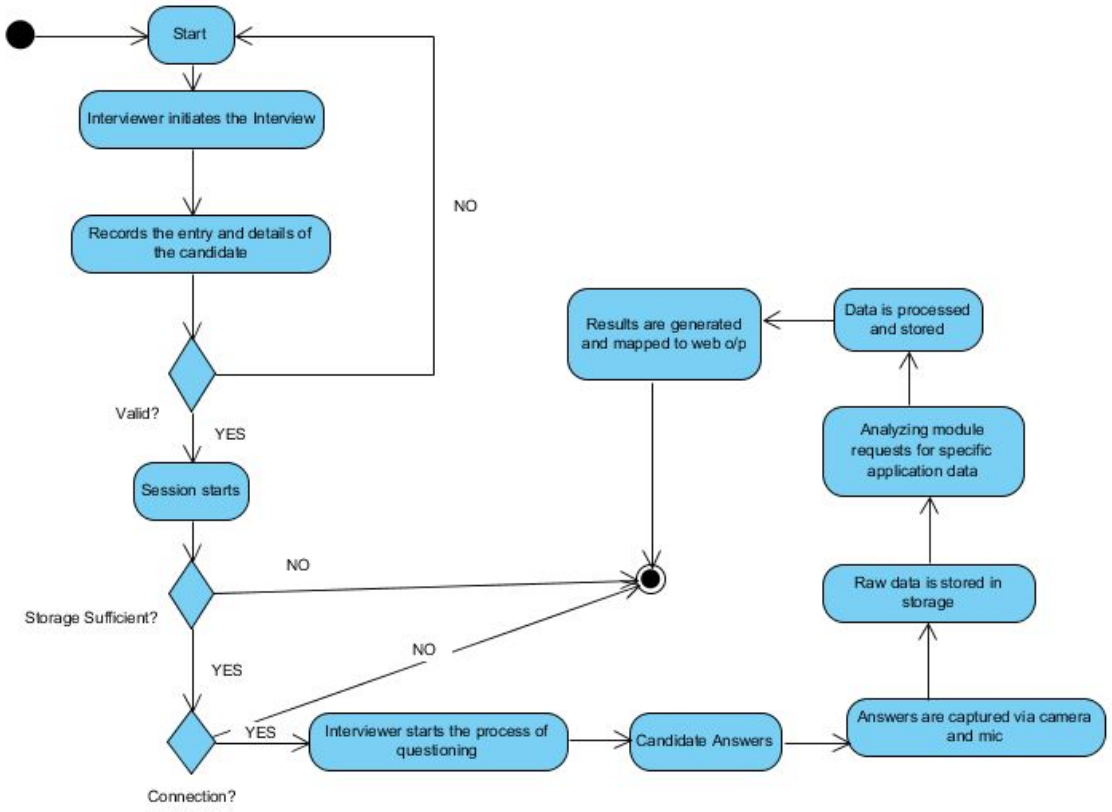
Level 2:


2.1
create dataset
fetch
data
Data
Storage
store labelled data
Database
fetch text data
fetch video data
fetch audio data
2.2.1
Text Analysis
2.2.2
Video Analysis
2.2.3
Audio Analysis
2.3.0
Feature Fusion and
Classification
store
results
Database

4.3.2 Flowchart of the System



4.3.3 State Transition Diagram/ Activity Diagram

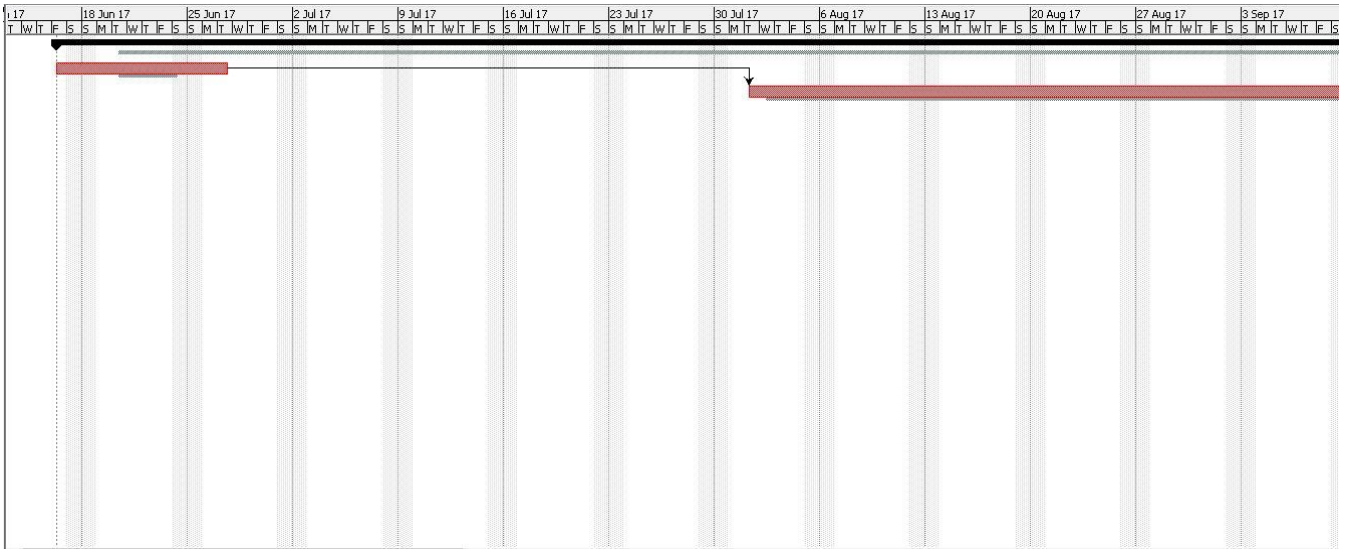


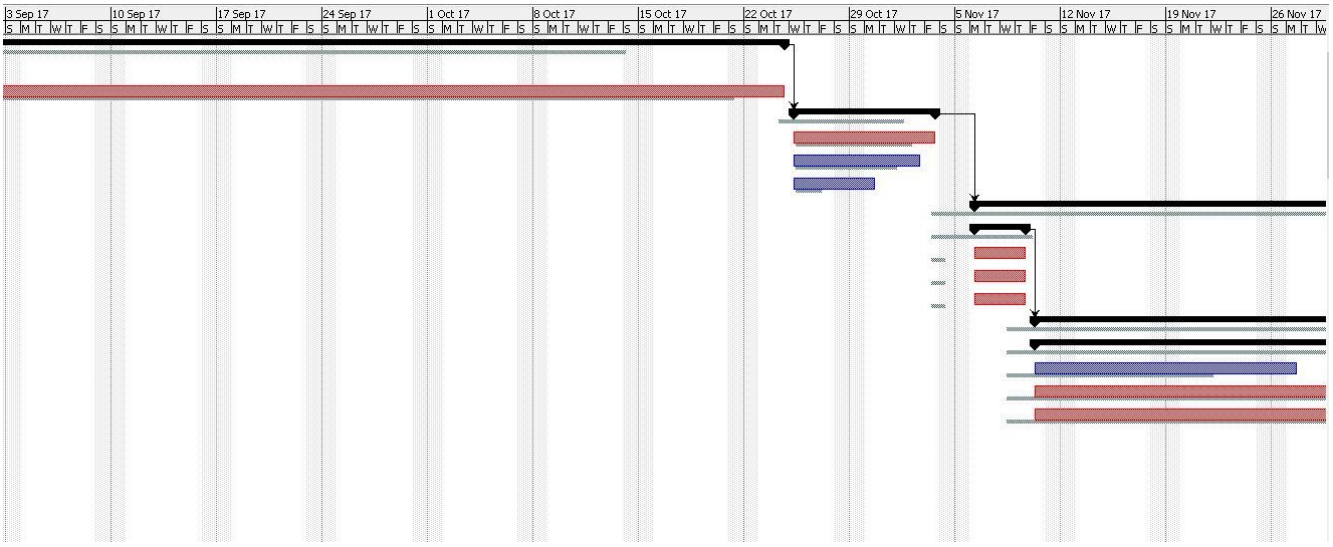
4.4 Project Scheduling & Tracking using Timeline / Gantt Chart

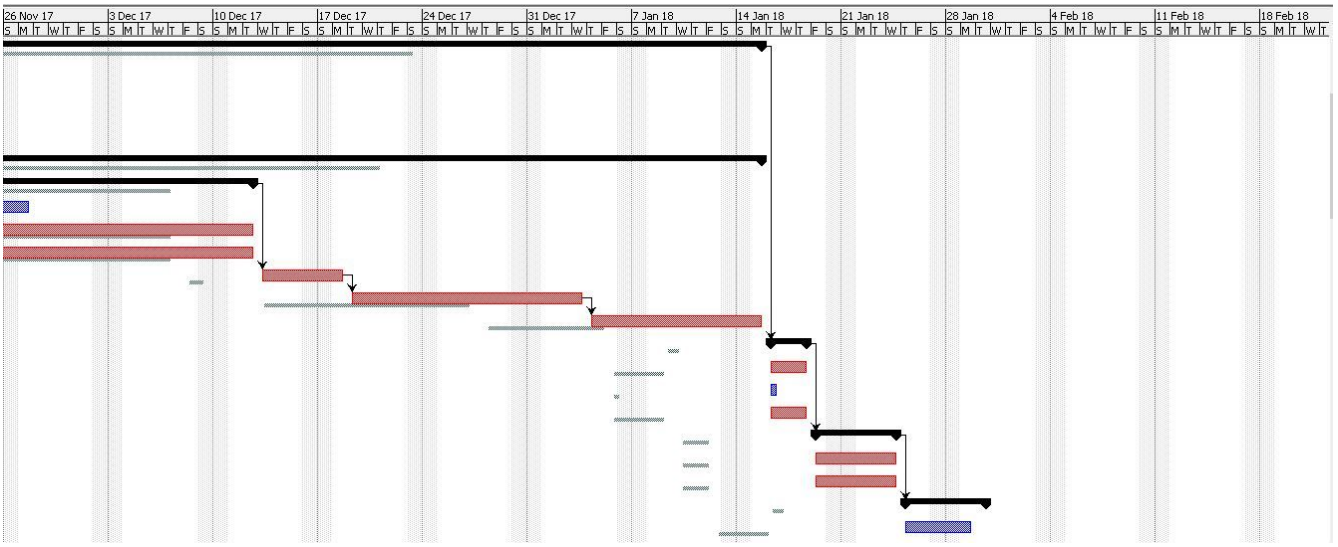
4.4.1 Project Schedule

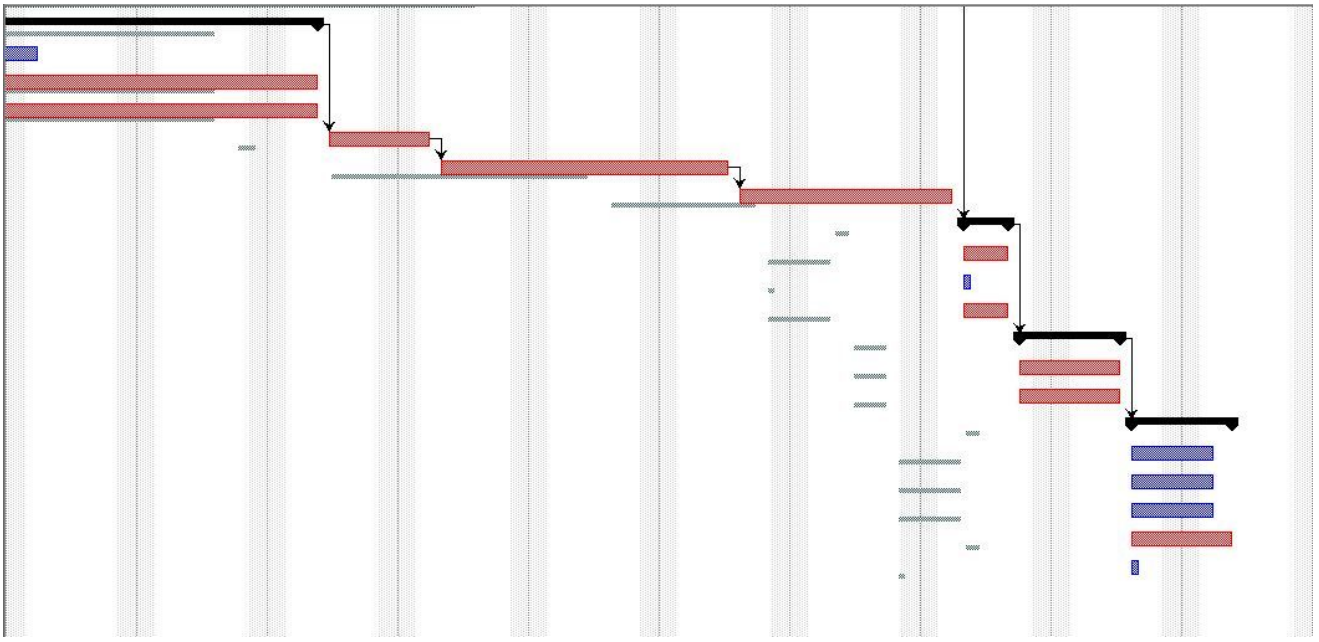
	Name	Duration	Start	Finish	Predecessors
	Requirement and Analys	**93 days**	**6/16/17 8:00 AM**	**10/24/17 5:00 PM**	
	Interview Question Prepara	8 days	6/16/17 8:00 AM	6/27/17 5:00 PM	
	Dataset Preparation	61 days	8/1/17 8:00 AM	10/24/17 5:00 PM	2
	Design	**8 days**	**10/25/17 8:00 AM**	**11/3/17 5:00 PM**	**1**
	Architecture Design	8 days	10/25/17 8:00 AM	11/3/17 5:00 PM	
	Dashboard UI design	7 days	10/25/17 8:00 AM	11/2/17 5:00 PM	
	Database Design	4 days	10/25/17 8:00 AM	10/30/17 5:00 PM	
	Implementation	**51 days**	**11/6/17 8:00 AM**	**1/15/18 5:00 PM**	**4**
	Input Module	**4 days**	**11/6/17 8:00 AM**	**11/9/17 5:00 PM**	
	Audio Module developmer	4 days	11/6/17 8:00 AM	11/9/17 5:00 PM	
	Video Module developmer	4 days	11/6/17 8:00 AM	11/9/17 5:00 PM	
	Database Handler Module	4 days	11/6/17 8:00 AM	11/9/17 5:00 PM	
	Analysis Module	**47 days**	**11/10/17 8:00 AM**	**1/15/18 5:00 PM**	**9**
	Feature Extraction	**23 days**	**11/10/17 8:00 AM**	**12/12/17 5:00 PM**	
	Audio Feature Extractic	12 days	11/10/17 8:00 AM	11/27/17 5:00 PM	
	Text Feature Extractior	23 days	11/10/17 8:00 AM	12/12/17 5:00 PM	
	Video Feature Extractio	23 days	11/10/17 8:00 AM	12/12/17 5:00 PM	
	Feature level Fusion	4 days	12/13/17 8:00 AM	12/18/17 5:00 PM	14
	Training Model	12 days	12/19/17 8:00 AM	1/3/18 5:00 PM	18
	classification	8 days	1/4/18 8:00 AM	1/15/18 5:00 PM	19
	Integeration	**3 days?**	**1/16/18 8:00 AM**	**1/18/18 5:00 PM**	**8**
	Audio Module Integeration	3 days?	1/16/18 8:00 AM	1/18/18 5:00 PM	
	Video Module Integeration	1 day?	1/16/18 8:00 AM	1/16/18 5:00 PM	
	Analysis Module Integeratic	3 days?	1/16/18 8:00 AM	1/18/18 5:00 PM	
	Testing	**4 days**	**1/19/18 8:00 AM**	**1/24/18 5:00 PM**	**21**
	Input Module Testing	4 days	1/19/18 8:00 AM	1/24/18 5:00 PM	
	Analysis Module Testing	4 days	1/19/18 8:00 AM	1/24/18 5:00 PM	
	Deployment	**4 days**	**1/25/18 8:00 AM**	**1/30/18 5:00 PM**	**25**
	Hardware Installation	3 days	1/25/18 8:00 AM	1/29/18 5:00 PM	
	Server Installation	3 days	1/25/18 8:00 AM	1/29/18 5:00 PM	
	Software Deployment	3 days	1/25/18 8:00 AM	1/29/18 5:00 PM	
	Final Testing	4 days	1/25/18 8:00 AM	1/30/18 5:00 PM	
	Handover	1 day	1/25/18 8:00 AM	1/25/18 5:00 PM	

4.4.2 Gantt Chart









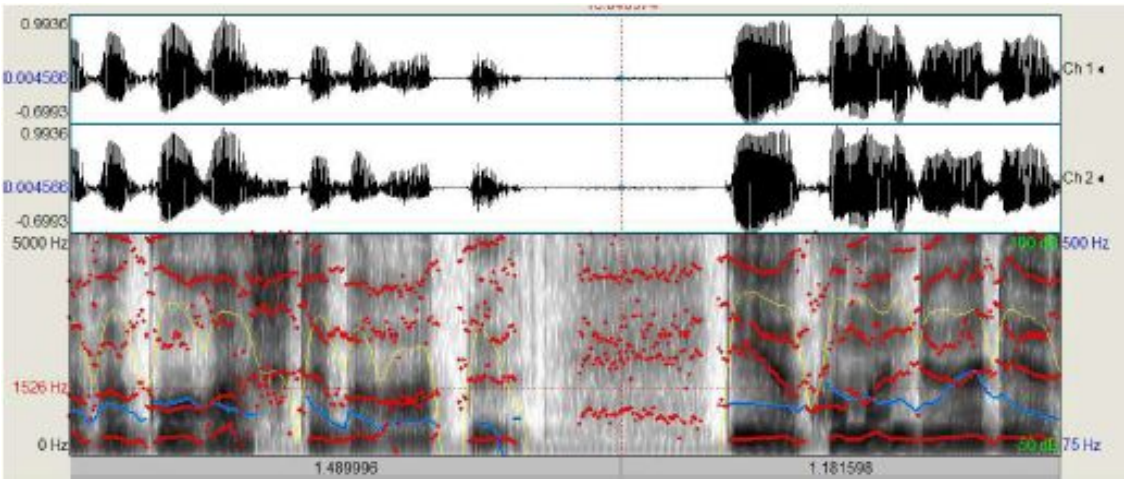
CHAPTER 5
IMPLEMENTATION DETAILS

Chapter 5: Implementation Details

5.1 Algorithms for the respective modules developed

5.1.1 Audio Cues

The preliminary step is to detect the individual who is speaking during the interview, the interviewer or the interviewee. Whilst during this process, we also need to segment the speech and non-speech chunks. Secondly, we need to make sure, that, the existence of any noise should be eliminated from the audio, if present. For audio cues, we extract the prosodic features such as Pitch, Band energy, and the Speech Rate respectively.



The tool utilized for this purpose is PRAAT, an open-source audio analysis tool for extracting the prosodic features. We first split the audio files into different chunks, wherein these chunks consist the answers from the interviewee and construct the collection of prosodic features of these splitted chunks. We used python script to access praat library (Parselmouth-praat) and extracted features from all the audio files. We averaged the value of each of the features for each of the questions for a particular interview.

Important features for audio cues are vocal intensities, pitch information, formant characteristics and spectral energy. These features reflect our social traits while communicating.

Algorithm Used:

Mean, Standard Deviation, Ratio were calculated for pitch, intensity Spectral energy, formants and were used as features in the dataset for training the model. Different features extracted from the audio clips of the interviews are –

Prosodic Features Extracted
Length
Average Band Energy
Average Intensity
Max Intensity
Mean Intensity
Range Intensity
SD Intensity
Average Pitch
Max Pitch
Mean Pitch
Range of Pitch (Max – Min)
SD Pitch
Mean F1
Mean F2
Mean F3
Mean B1
Mean B2
Mean B3
SD F1
SD F2
SD F3
Mean F2/F1
Mean F3/F1
SD F2/F1
SD F3/F1

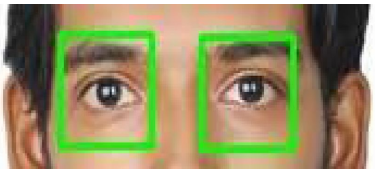
5.1.2 Visual Cues

Visual cues are the facial features of the interviewee during the interview. Different facial cues extracted are facial expression, gaze, and head nods. For every video frame, facial features of the interviewee are extracted. The first step in this is to detect the face itself in the frames. The Haar Cascade classifier for face detection is used to perform this task of face detection. Various facial expressions such as smiling, surprise or neutral are detected using convolution neural network. The model is trained on facial expression recognition dataset.

Second visual cue extracted is head nods. This is used to detect Yes and No when signaled during interview. Studies have proposed the use of Hidden Markov Model to detect head nodding.

The third cue in facial feature extraction is the eye gaze. Eye contact is an important nonverbal behavior, it is used to monitor feedback, reflect cognitive activity. We extract the amount of time the interviewee looks at the interviewer while answering a question or during the course of the interview.

Apart from these three cues we also extracted some more cues using facial action units. We used the dlib library to detect facial landmarks. Facial landmarks help us localize different regions of the face like eyes, eyebrows, nose, mouth, and jawline.



The dlib uses a pre-trained facial landmark detector that estimates the location of 68 coordinates that map to the face. Figure 5.3 represents these 68 points that maps the facial structure on the face. These points are then used to calculate more facial

features like eyebrow raise distance from the eye, eye-opening distance, lip corner distance etc.

Algorithm Used:

Face detection – Haar Cascade/ Viola Jones method

Smile Detection – Convolutional Neural Network

Head Tracking – Optical Flow motion tracking.

Facial Landmark Features – Euclidean Distance to get distance of our interest. Different features extracted from the facial Landmarks are:

Outer Eyebrow Height	Inner Eyebrow Height
Outer Lip Height	Inner Lip Height
Inner Eyebrow Distance	Lip Corner Distance
Eye Opening	

5.1.3 Lexical Cues

While speaking, what we say and how we say tells a lot about us. We can use "VADER" standing for Valence Aware Dictionary and sentiment Reasoner for verbal cues/ lexical cues. The VADER categories include negative emotion terms (e.g., sad, angry), positive emotion terms (e.g., happy, kind) and various content word categories (e.g., anxiety, insight). It is one of the most used lexicons for analyzing lexical cues, and especially when it comes to sentimental words. Each and every word in the dictionary is assigned to one or more categories. On top of this, VADER also has the capabilities of understanding sentiment-laden slang words.

Algorithms Used:

VADER Sentiment Analysis

5.1.2 Predictive Model:

After extracting features from the facial cues, audio cues, and verbal cues, we create a single feature vector joining the features from the facial, audio and verbal cues for each interview. This finally generates the dataset. This dataset has ratings done by the judges as well. This dataset is used against model prediction to find the similarity between the predicted score and actual judges score. We have used a regression model to predict the hireability and overall score of the interviewed candidate. Mean Square Error (MSE) helps us to find the best model since it gives the error made by the actual judge's score. Among Support Vector Regression (SVR), Lasso Regression and Ridge Regression, MSE of Ridge and SVR were lesser and predicted the results closer to judges score.

Algorithms Used:

SVR, LASSO, RIDGE

5.2 Comparative Analysis with the existing algorithms

While most existing studies rely on audio nonverbal cues (prosody, speaking turn-based features) or visual cues to predict personality, few studies have investigated the use of body communication. Most systems have not used multimodal nature of communication to analyze and have been limited to only audio cues or video cues. Our System is an end to end standalone application. In our system we use cross modal nature of the communication to analyze the interview. We use video cues, prosodic/audio cues as well as lexical/ language cues to analyze the interview. And finally create a predictive model to rate the candidates and determine their hireability.

5.3 Evaluation of the developed system

The proposed system produced a variety of results, from a single personality trait to the overall hireability recommendation of the candidate. We produced graph plots for every aspect, from internal structures, models to the final output, so as to represent the accuracy and working of the system constructed.

Smile Detection CNN's one of the hidden layer output :

Figure 13 is a representation of the input of an image to the CNN architecture. This image exists on the second hidden convolutional layer of the CNN architecture; used

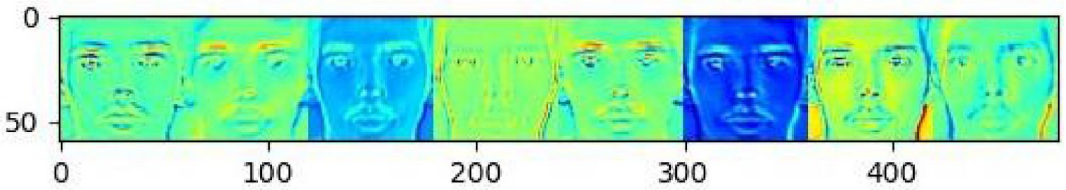


Figure 13: Hidden Layer output, CNN

for emotion detection on the face of the interviewee. This layer produces 8 different images due to the fact that the second hidden convolution layer in the CNN architecture has the 8 different filters, where each filter has the capability to identify a different feature. Since these filters are exhibited by the second layer, they provide the opportunity to represent higher level features on the basis of the low-level features existing on the first hidden layer of the architecture. This thus results in an output image having a size of 60*60 pixels.

Statistical Results:

There are two categories of statistical results furnished from the output produced by the predictive system.

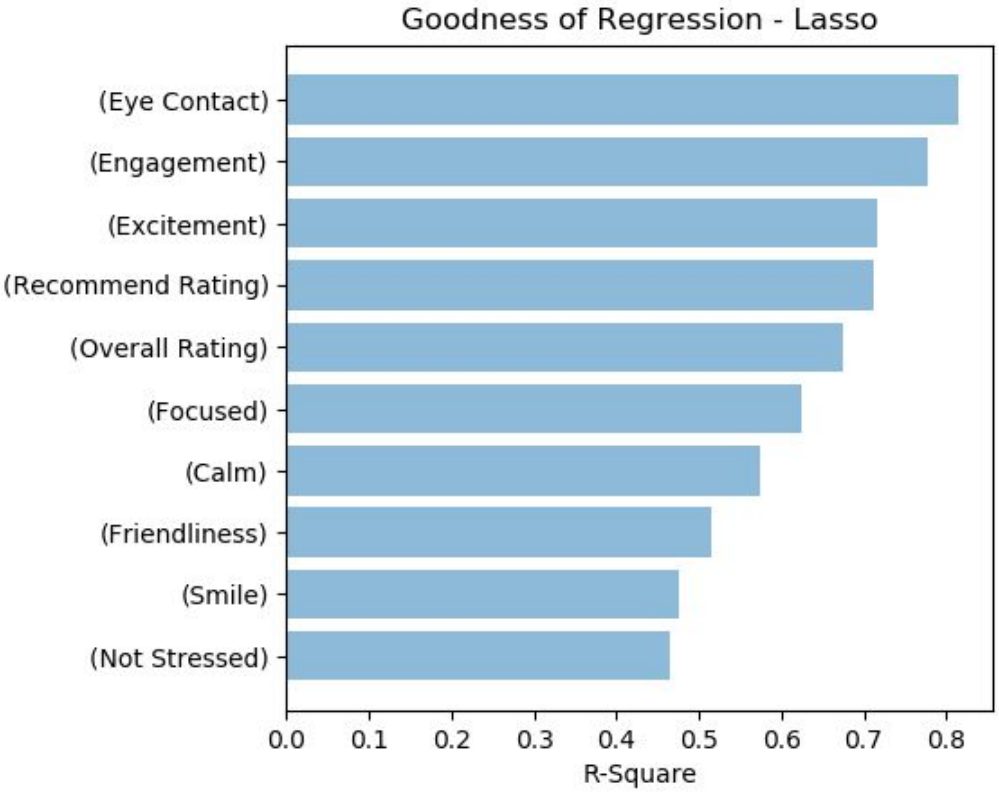
The first section is the statistics just about the audio cues. Referring to the Table I. we have the different statistical values produced; Average, mean, standard deviation and maximum on the four most important features which were extracted from audio; Formants, intensity, pitch and band energy.

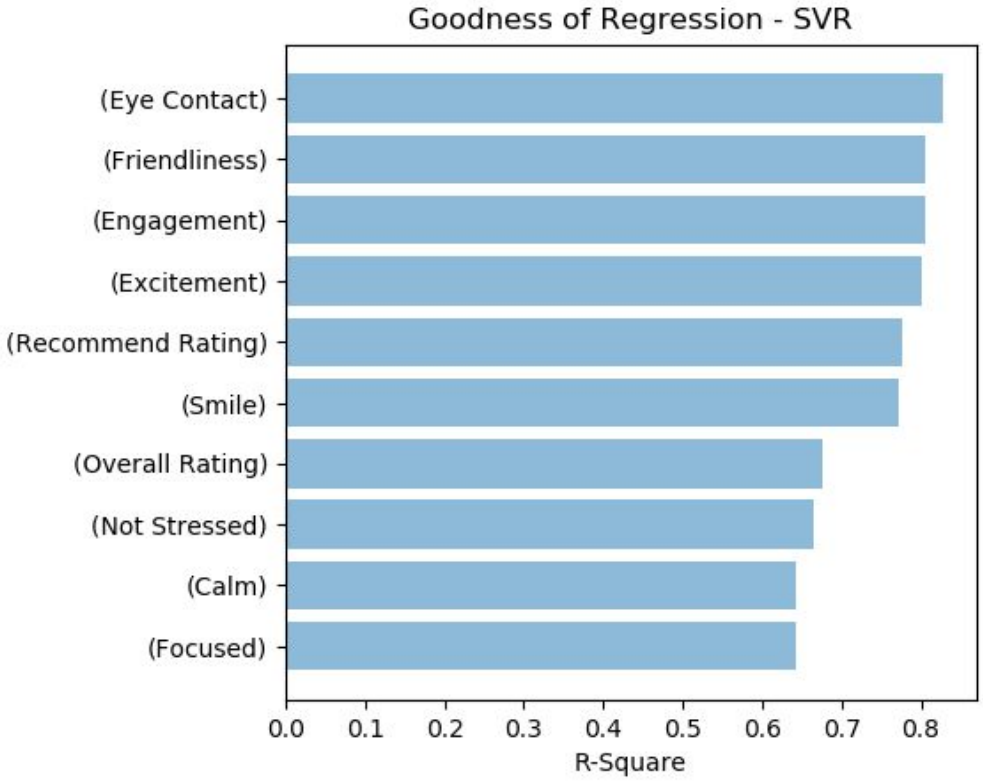
	Features			
	Formant	Intensity	Pitch	Band Energy
Average	-	63.76	113.58	0.0783
Mean	855.96	2.93	99.32	-
Standard Deviation	289.15	14.98	23.62	-
Max	-	78.56	569.24	-

When the system produces outputs on the UI with the overall hireability scores of all the candidates, it beforehand makes a comparison of all the candidates, this comparison considers both verbal and nonverbal cues. So, in order to carry out this comparison, the system has to have some concrete or comparison values, with which it has to determine which candidate is better than which of the other candidate. The Table 5 represents all the audio statistics of a single interview, when these values are compared to, say, a minimum criteria values, it becomes easier to then determine a better comparison activity.

The second section talks about various statistical analysis of different models used with the system and their fitness for the purpose.

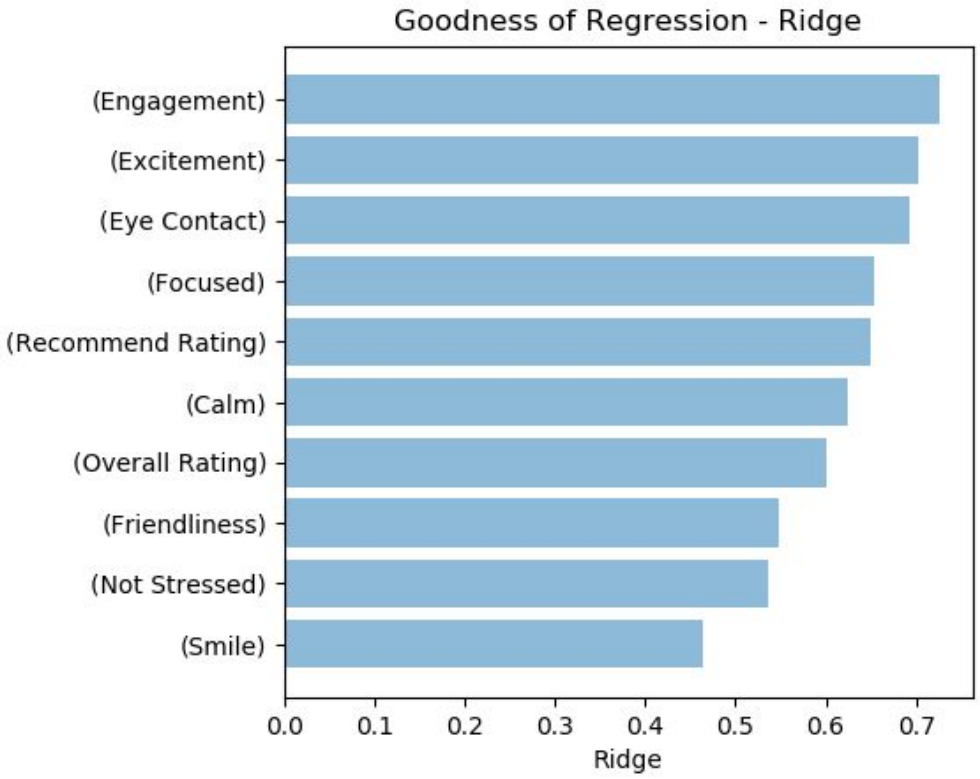
Below figure is graph plot of Lasso Model, portraying the goodness capacity, it represents the models predicting accuracies across different personality traits considered by us at earlier stages





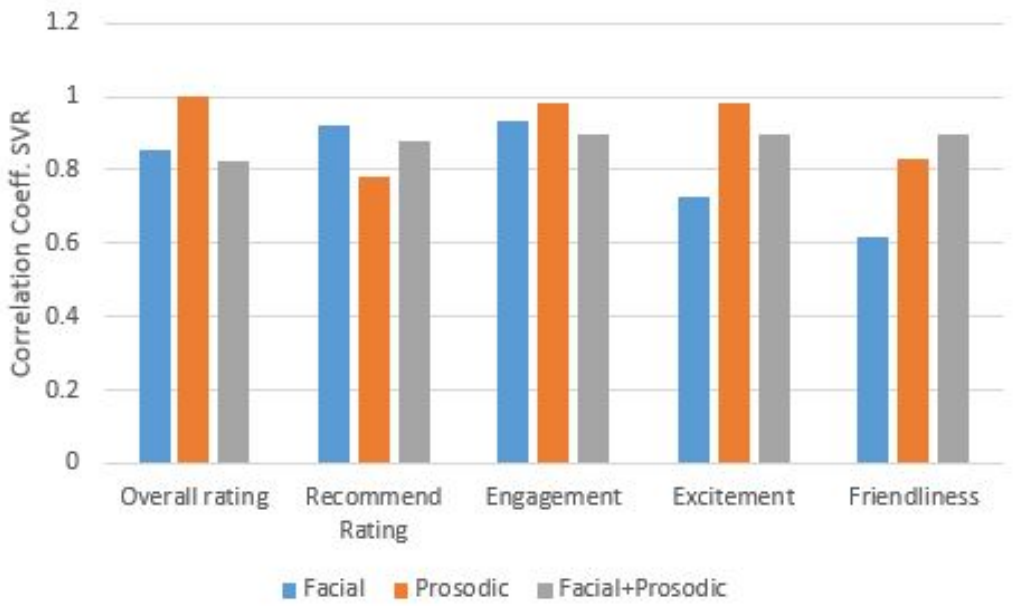
The above figure portrays the goodness capacity of the Support Vector Regression Model for the different traits.

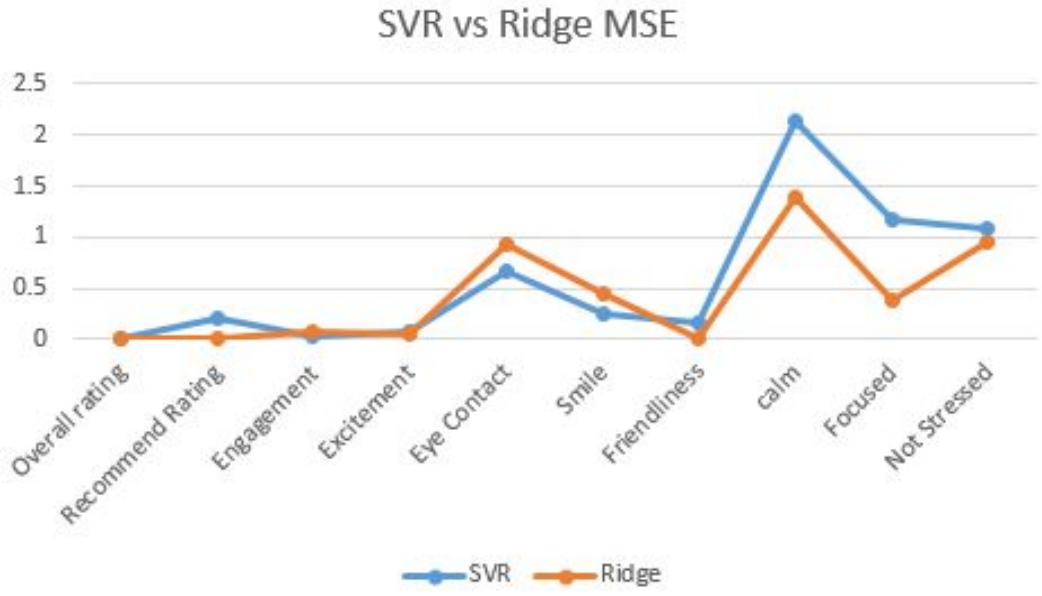
The below figure represents the goodness capacity of the Ridge model for the different traits.

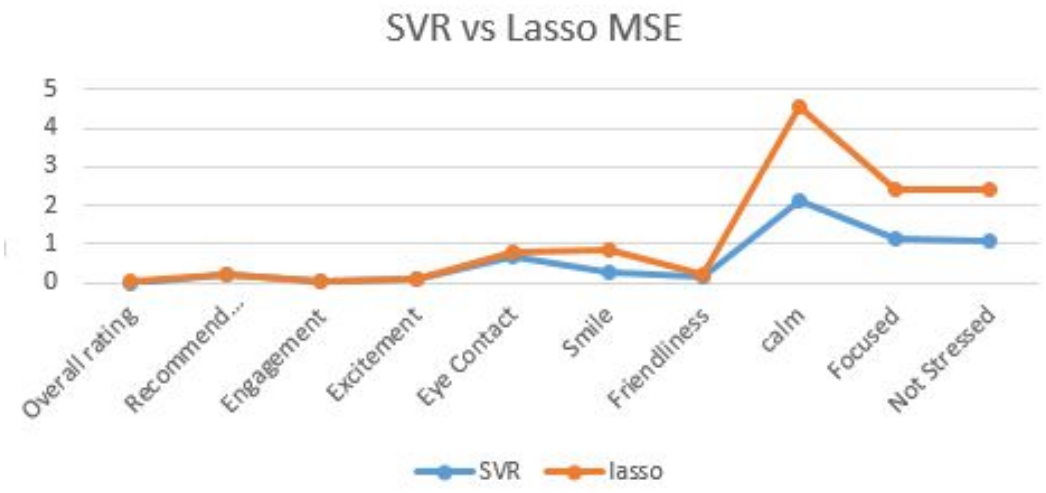


Looking at the goodness of fit for the different algorithms we can observe that the SVR model fits the best with R-Square value for all the different traits more than 0.6, while Lasso and Ridge's model falls short in fitting some traits when compared to SVR.

For the sake of understanding the outcomes, let us consider the top five traits judged represented in Figure. We take Overall rating, Recommend Rating, Excitement, Engagement, and Friendliness and compare their correlation coefficient of SVR model for a different set of features like Facial only, Prosodic only, and Facial and Prosodic together.


1.2
1
0.8
0.6
0.4
0.2
0
Correlation Coeff. SVR
Overall rating
Recommend Rating
Engagement
Excitement
Friendliness
Facial
Prosodic
Facial+Prosodic


SVR vs Ridge MSE
2.5
2
1.5
1
0.5
0
Overall rating
Recommend Rating
Engagement
Excitement
Eye Contact
Smile
Friendliness
calm
Focused
Not Stressed
SVR
Ridge



The above figure is of MSE results for SVR and Lasso models between Ground Scores and Predicted Scores.

The above all figure represents the comparison between the mean square errors of SVR model with the Lasso and Ridge Model. In Figure 6.7, the mean square error for Ridge model is lower for most of the traits compared to SVR model. While in Figure 6.8 the mean square error for the Lasso model is larger than the SVR model for all the traits.

CHAPTER 6

TESTING

Chapter 6: Testing

The framework being developed is destined to produce the following results with respect to the inputs given:

1. Detailed statistics of analysis of the interview
2. Interviewees scores w.r.t the emotional and voice analysis
3. Session details
4. General statistics about the interview
5. Competency report of the entire interview process of all the candidates

Following Testings are performed:

- Unit Testing: Unit testing is performed on Audio and video capture. Audio capture module work better when mic is placed closer to the user.Video Capture Module perform better when some expected traits are encountered. If expected traits are not encountered after few minutes it stops the processing.

- Integration Testing: Integration testing is performed between Analysis module and Audio and video capture module. Parallel execution of the module create some load on the streaming module. This causes real time smile score graph little delayed.If the system is working under stress condition then streaming feature is delayed and this can be noticed on the dashboard while interview is going.

Performance:

Modules	Normal Condition	Stressed Condition	Overall Performance
Audio Capture	Excellent	Good	Good
Video Capture	Excellent	Laggy	Good
Analysis	Excellent	Excellent	Excellent
Streaming	Average	Laggy	Good

CHAPTER 7
RESULTS

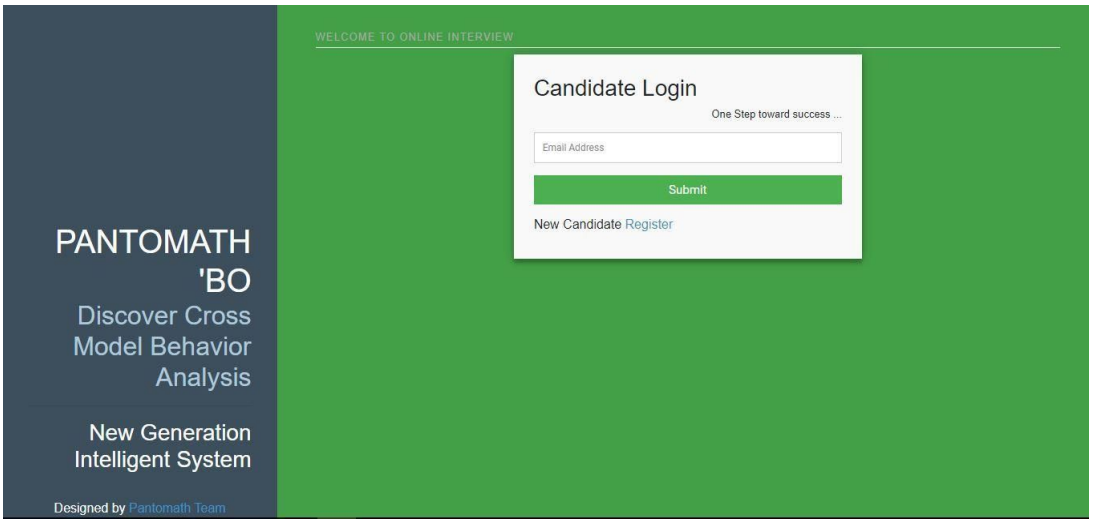
Chapter 7: Results

7.1 Parameters Considered

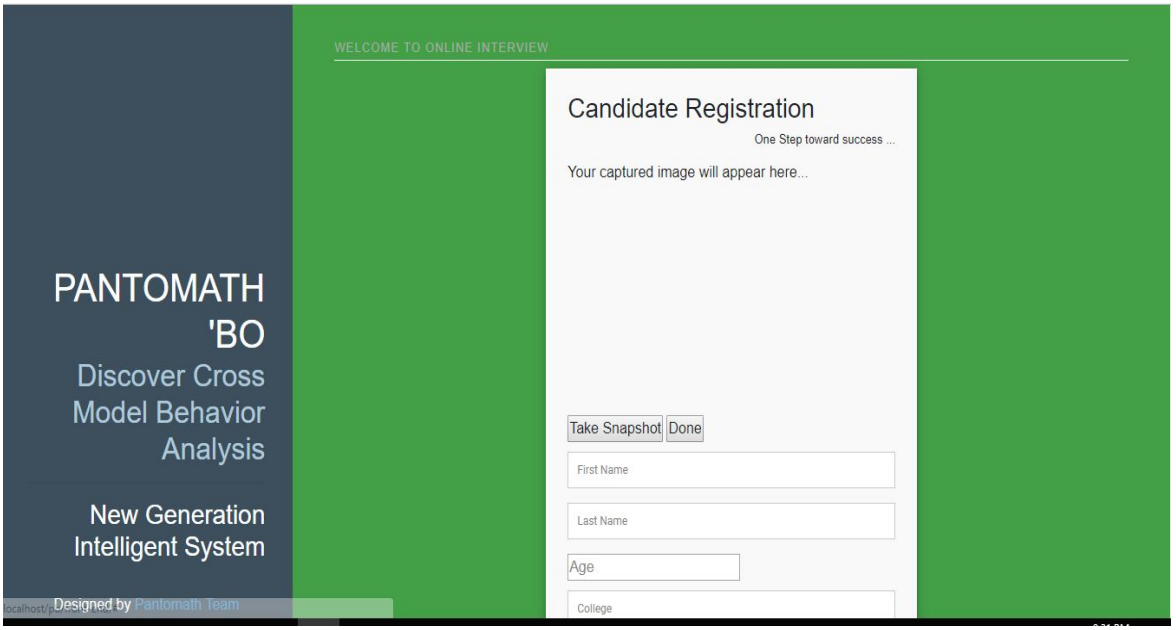
Various parameters considered are goodness of fit, correlation coefficient and mean square error. Looking at the goodness of fit for the different algorithms we observed that the SVR model fits the best with R-Square value for all the different traits more than 0.6, while Lasso and Ridge model falls short in fitting some traits when compared to SVR. We found that the mean square error for Ridge model is lower for most of the traits as compared to SVR model and the mean square error for the Lasso model is larger than the SVR model for all the traits.

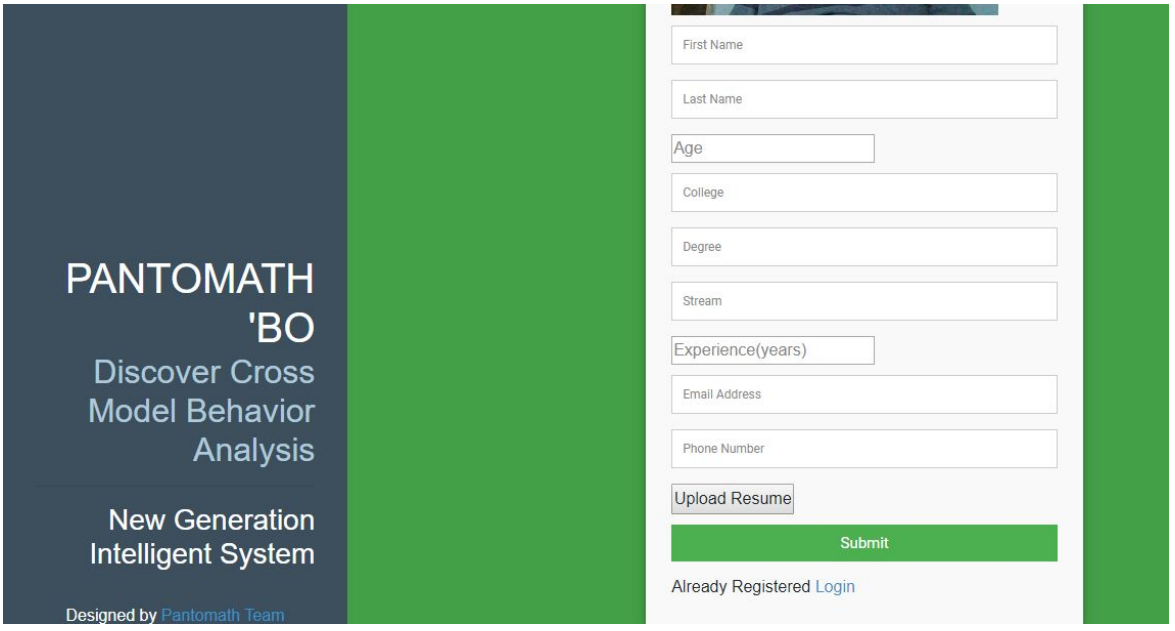
7.2 Screenshots of User Interface

7.2.1 Login

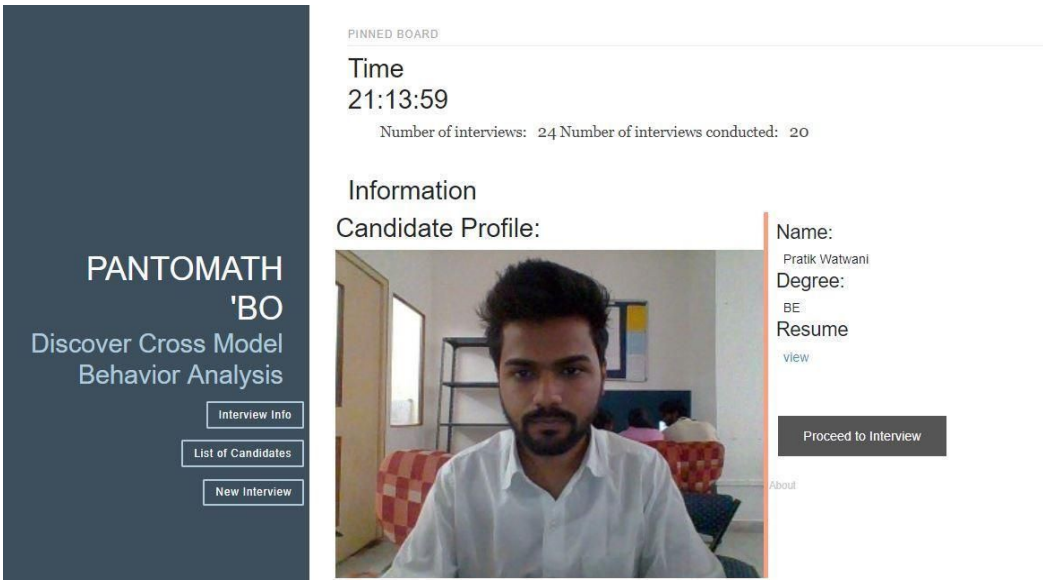


7.2.2 Registration

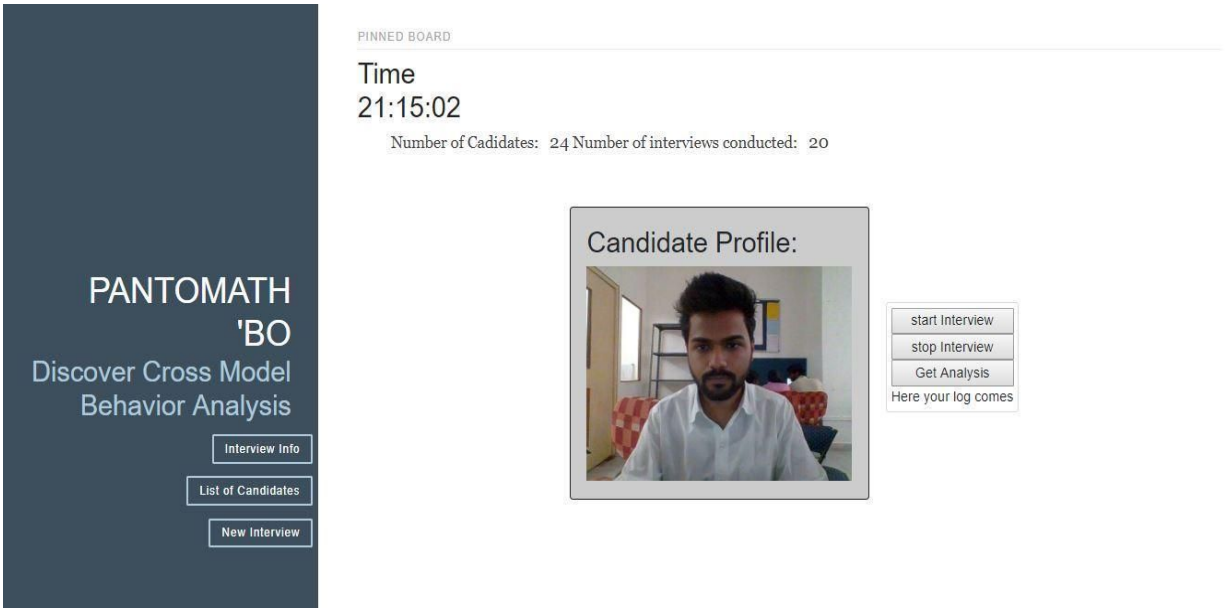




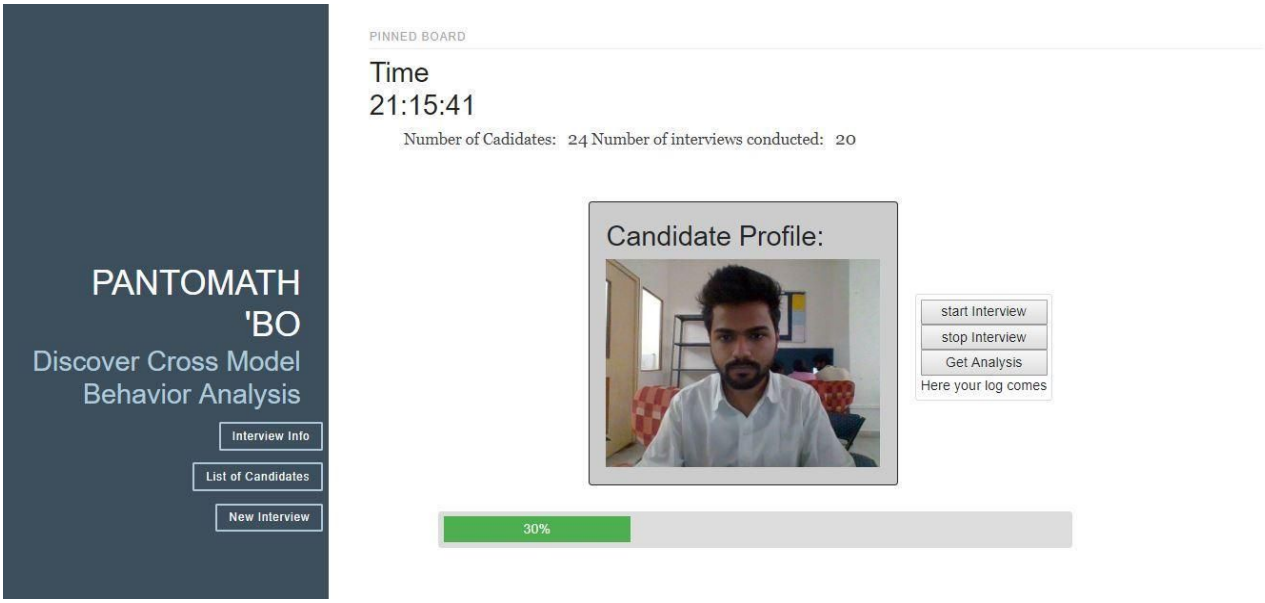
7.2.3 Interviewer's Screen



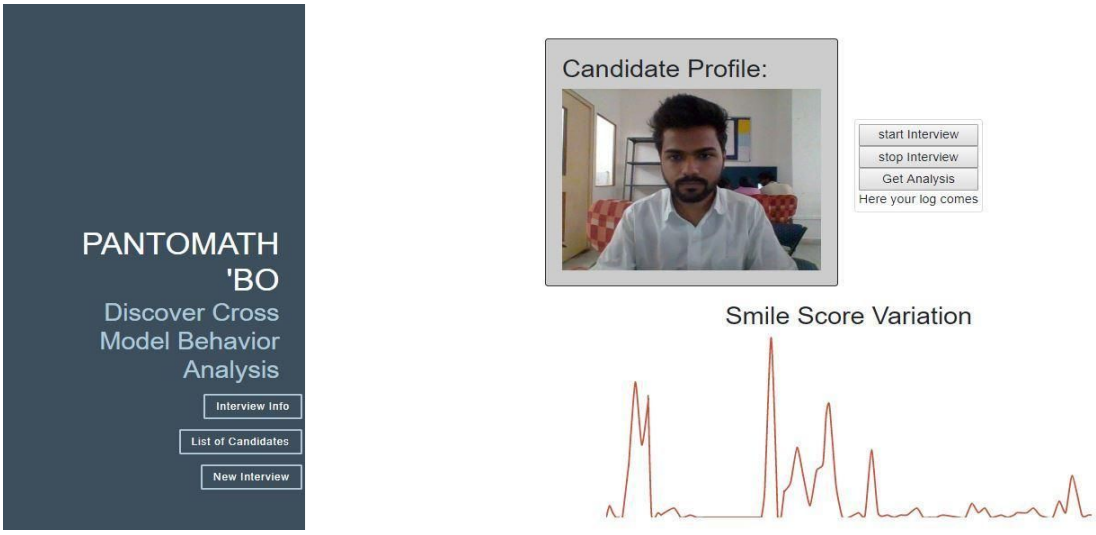
7.2.4 Take Interview



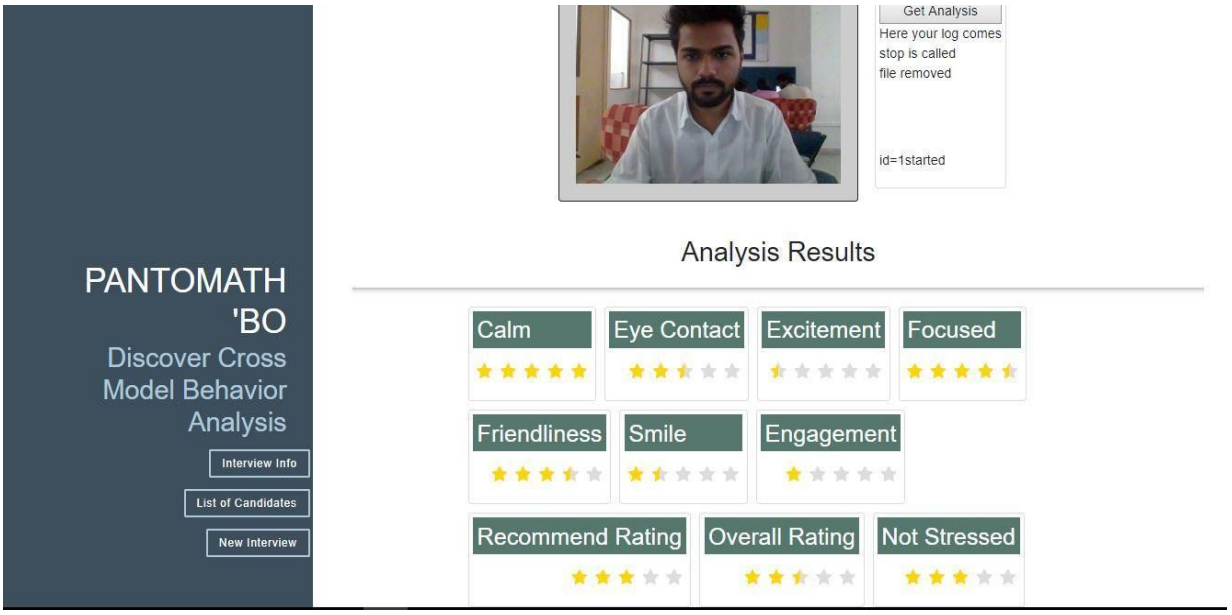
7.2.5 System Loading



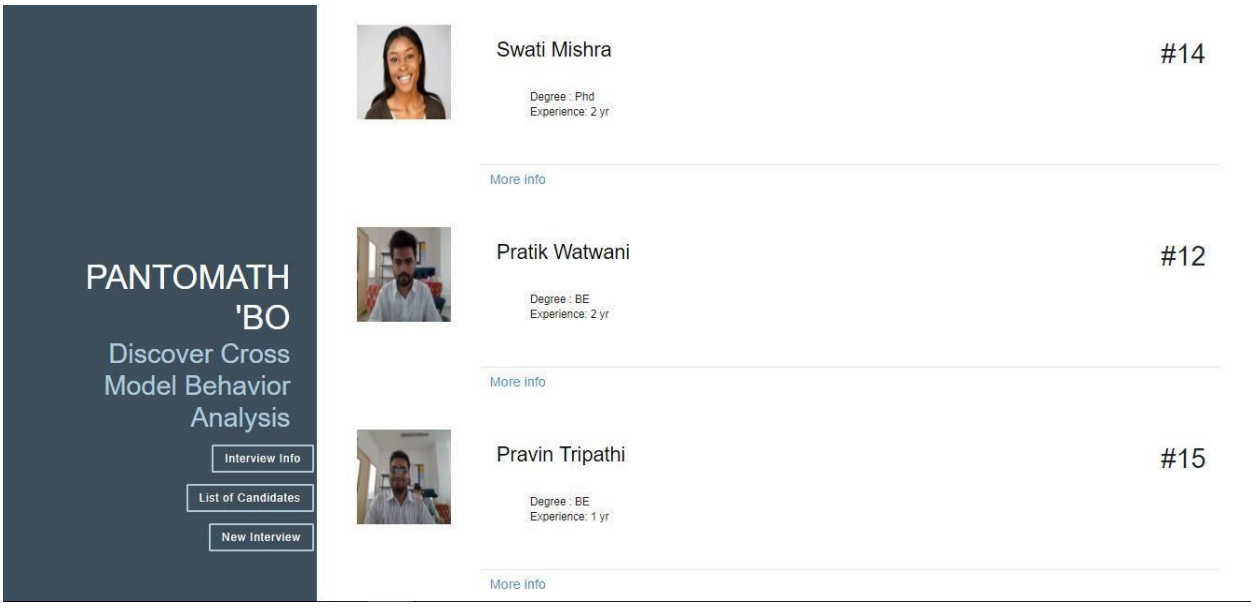
7.2.6 Interview Proceeding



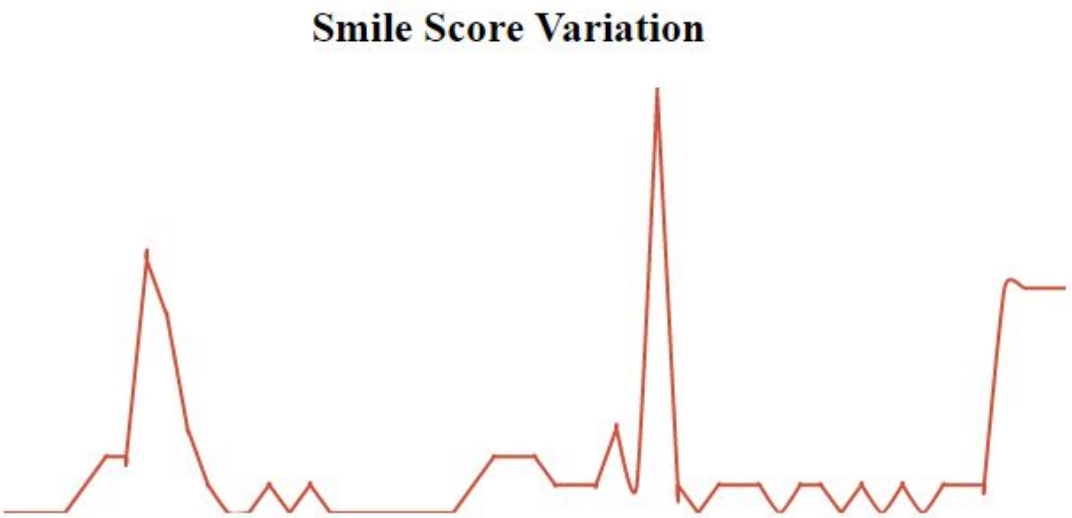
7.2.7 Generate Analysis



7.2.8 Candidates Comparison

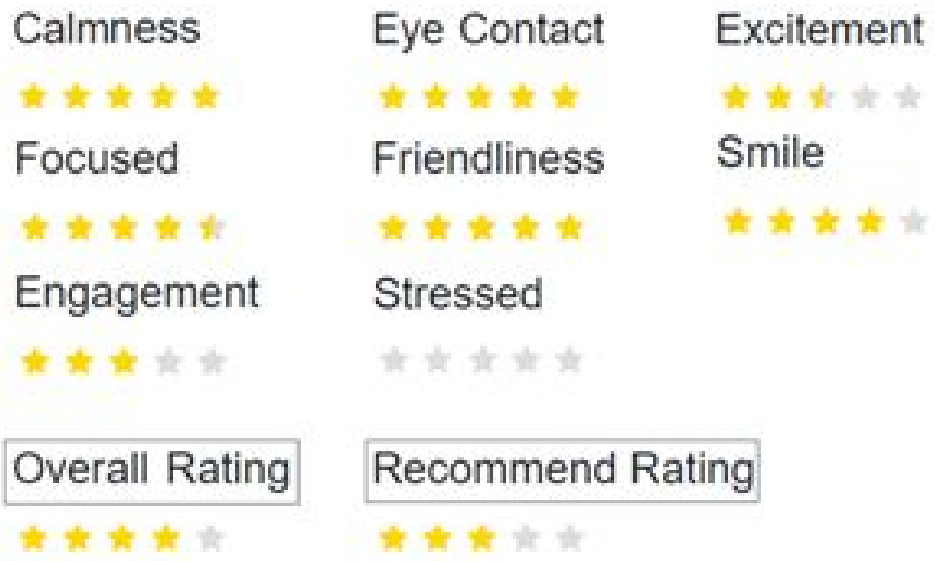


7.3 Graphical Output



Smile Score Variation, a real time graph that is displayed real time on the UI of Web App.

7.4 Reports Generated



An interviewee's final scores displayed to the interviewer on their UI dashboard at the end of the interview.

CHAPTER 8

CONCLUSION

Chapter 8: Conclusion

8.1 Limitation

- Only one Interview at a time can be conducted.
- Poor Performance when multiple Interviewer environment is considered
- Silent environment is needed for accurate result.
- In depth Audio analysis cannot be performed since single audio channel is used.

8.2 Conclusion

- This system conducts a systematic study on applicant's body communication in job interviews with respect to hireability impressions and self-rated personality.
- Our system is trained to predict ratings based on real time interview videos, and real time feedback can be received at the same time.
- Real-time feedback for different individual traits can be received on the interviewer dashboard.
- This prediction framework is independent and can be deployed as a standalone system on cloud Platform, IoT device or as a web App.

8.3 Future Scope

- Multiple Interviewer Environment capability can be added.
- Cloud Based support can be added which enable interviewee to give interview on the go at anytime from anywhere.
- Accuracy of the model need to be improved
- Support for adaptability of the environment can be added which let user to adjust the system according to their workspace.

CHAPTER 9
REFERENCES

Chapter 9: Reference

[1] T. DeGroot and J. Gooty. “Can nonverbal cues be used to make meaningful personality attributions in employment interviews?” Journal of Business and Psychology, 24(2):179–192, 2009.

[2] V. Soman and A. Madan, “Social signaling: Predicting the outcomeof job interviews from vocal tone and prosody,” in ICASSP. Dallas,Texas, USA: IEEE, 2010.

[3] M. A. Gotcher, “Nonverbal communication in the job Interview: first impressions and Nonverbal immediacy,” Senior Thesis, Texas Tech University, 1990.

[4] Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. Journal of Personality and Social Psychology, 77(6), 1296-1312.

[5] M. Knapp, J. Hall and T. Horgan, Nonverbal communication in human interaction. Boston, MA: Wadsworth Cengage Learning, 2014.

[6] A. S. ImadaandM. D. Hakel. Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. Journal of Applied Psychology, 62(3):295–300, 1977.

[7] R. W. Frick, “Communicating emotion: The role of prosodic features.” Psychological Bulletin, vol. 97, no. 3, p. 412, 1985.

[8] A. Kapoor and R. Picard, “A Real-Time Head Nod and Shake Detector,” Proc. Workshop Perspective User Interfaces, Nov. 2001.

[9] W. Tan and G. Rong. A real-time head nod and shake detector using HMMs. Expert Systems with Applications, 25(3):461–466, 2003.

[10] D. Sanchez-Cortes, P.Motlicek, and D. Gatica-Perez. Assessing the Impact of Language Style on Emergent Leadership Perception from Ubiquitous Audio. In Proceedings of the ACM International Conference on Mobile and Ubiquitous Multimedia (MUM), 2012.

REVIEW SHEETS

Cerelabs pvt. Ltd.
Devesh Rajadhyax. - Prateek Watwani - Absent

Class: D17 A/B/C
Group No.: 45

Project Evaluation Sheet 2017 - 18

Title of Project: Pantomath'BO Discover Cross Modal Human Behavior Analysis

Group Members: Ankit Shaw, Swastika Thakur, Pravin Tripathi, Pratik Watwani

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	~~3~~ 2	~~4~~ 5	2	2	1	1	2	2	2	4	4	~~39~~ 40

Comments: Very good topic, memory required for storage should be considered.

Indu Dokare
Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	2	5	2	2	2	1	2	2	2	4	4	40

Comments: Interesting Topic

Date: 26th September 2017

S. Sengupta

Apurva Chaudhari
Name & Signature Reviewer2

Project Evaluation Sheet 2017 - 18

GROUP NO.: 45

Title of Project: Discover Cross Modal Human Behavior Analysis

Group Members: Pratik Watwani, Pravin Tripathi, Swastika Thakur, Ankit Shaw

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg & Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	2	4	2	2	2	2	2	2	3	4	5	42

Comments: Good work.

Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg & Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	2	3	2	2	2	2	3	2	2	4	5	41

Comments:

Date: 26/02/18

Apurva Chaudhari

Name & Signature Reviewer2

Inhouse/ Industry:

Class: D17 A/B/C ✓

Group No.: 45

Project Evaluation Sheet 2017 - 18

Title of Project: Pantomath 'Bo : Discover Cross Modal Human Behavior Analysis

Group Members: Pratik P. Watwani, Praviin Tripathi, Ankit Shaw, Swastika Thakur

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Inter retation o. Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Profess ional Skills (5)	Innov ative Appr oach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	2	4	1	2	1	2	2	3	3	4	4	40
Comments: Almost done. 95%.															

Indu Dokare
Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Profess ional Skills (5)	Innov ative Appr oach (5)	Total Marks (50)
Review of Project Stage 1	4	4	5	2	4	1	2	1	2	2	3	3	4	4	41
Comments: Excellent															

Date: 15th March,2018

Apurva Chaudhari
Name & Signature Reviewer2

APPENDIX

Paper Publication in IEEE Xplore, Conference in 2nd International Conference in Electronics, Communication and Aerospace Technology (ICECA 2018).

Acceptance Letter

ICECA 2018

2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018)

29,30,31 March, 2018 | http://icoeca.org/index.html | smys375@gmail.com

2 March, 2018

To

Pravin Tripathi, PratikP. Watwani, Swastika Thakur , Ankit Shaw , Sharmila Sengupta
Department of Computer Engineering, Vivekanand Education Society's Institute of Technology
Chembur, Mumbai

Paper ID: ICECA 2018–616

Subject: Acceptance Letter –**2nd IEEE International conference on Electronics, Communication and Aerospace Technology (ICECA 2018)** – Reg.

Dear Author,

This is the notification to inform you that your Oral presentation proposal entitled Discover Cross Modal Human Behavior Analysis to the 2nd IEEE International conference on Electronics, Communication and Aerospace Technology (ICECA 2018) organized by RVS Technical Campus, Coimbatore, India on 29, 30, 31 March, 2018 has been accepted as a result of blind reviews.

All registered papers will be published in IEEE Xplore
On behalf of the organization committee I would like to congratulate you.

Yours sincerely,

Dr. S. Smys,
Conference chair – ICECA 2018,
RVS Technical Campus,
Coimbatore, India.
smys375@gmail.com

TECHNICAL SPONSORS

Plagiarism Report for Submitted Paper

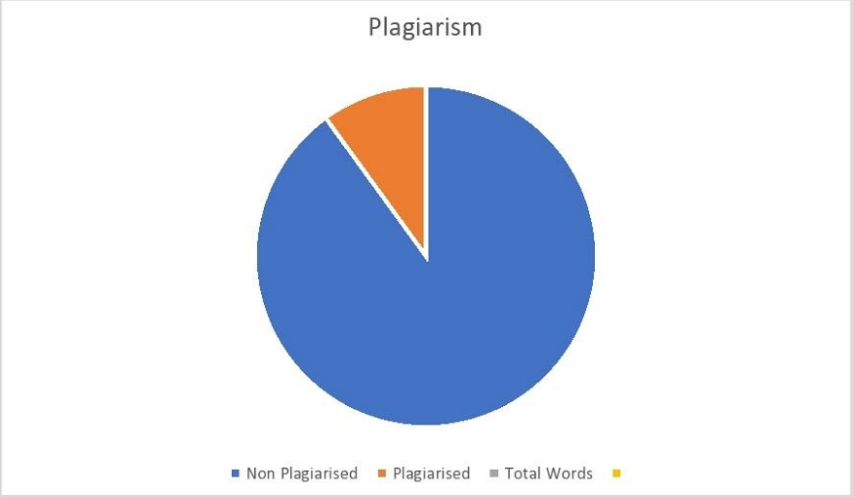
Plagiarism Check Originality Report

Paper Name: **Discover Cross-Modal Human Behavior Analysis**

Authors : **Pravin Tripathi, Pratik P. Watwani, Swastika Thakur, Ankit Shaw, Sharmila Sengupta**

Similarity Found: 4%

Date: 2nd March 2018.



Statistics: 431 words Plagiarized / 4315 Total words

Remarks: Low Plagiarism Detected - Your Document needs Optional Improvement.

Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018)
IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1-5386-0965-1

Discover Cross-Modal Human Behavior Analysis

Pravin Tripathi*, Pratik P. Watwani†, Swastika Thakur‡, Ankit Shaw§, Sharmila Sengupta ||
Department of Computer Engineering, Vivekanand Education Society's Institute of Technology
Chembur, Mumbai
Email: *tripathipravin95@gmail.com, †pratikpwatwani@gmail.com, ‡swastika.thakur1704@gmail.com,
§ankitshaw14@outlook.com, || sharmila.sengupta@ves.ac.in

***Abstract*— Job interviews are a predominant part of any hiring process to evaluate applicants. It is used to evaluate applicant's knowledge, skills, abilities, and behavior in order to select the most suited person for the job. Recruiters make their opinion, on the basis of both verbal and nonverbal communication of an interviewee. Our behavior and communication in daily life are cross-modal in nature. Facial expression, hand gestures and body postures are closely linked to speech and hence enrich the vocal content. Nonverbal communication plays an important role in what we are saying and what we actually mean to say. It carries relevant information that can reveal social construct of a person as diverse as his personality, state of mind, or job interview outcome; they convey information in parallel to our speech. In this paper, we present an automated, predictive expert system framework for the computational analysis of HR Job interviews. The system includes analysis of facial expression, language and prosodic details of the interviewees and thereby quantifies their verbal and nonverbal behavior. The system predicts the rating on the overall performance of the interviewee and on each behavior traits and hence predict their personality and hireability.**

***Keywords*— HR Interviews, Behavior Analysis, Cross-modal analysis.**

I. Introduction

Nonverbal communication is as crucial as verbal communication when interviewing for a job. In face-to-face communication, our nonverbal behavior conveys information about our personality and traits in addition to our speech. Nonverbal communication is an unconscious process and hard to manipulate, therefore plays an important part in the outcome of the employment interview. The nonverbal communication includes visual cues, audio cues, and language or lexical cues. Visual cues or facial cues can have a powerful effect on interview performance [1]. Our physical appearance and how we dress are only part of visual cues which are important for an interview. Along with this, visual cues also include our facial expressions (smiling, neutral, surprise) during the conversation. Hand gestures, eye contact, body orientation also contributes to the visual cues during communication.

Another nonverbal cue is the audio cues or paralinguistic cues which refers to the vocal (prosodic) characteristics and features. Not only what the interviewee says but how he/she says can make a lot of difference in the outcome of the interview [2]. Voice modulation, variation in pitch, the rate of speech, intensity etc. are some of the important vocal cues that may break or make your chances to be hired [3]. Monotone voice, narrow pitch range are some characteristics that are not preferred in an interview [3]. Finally, the verbal cues like lexical features give us great detail of information about the style of speaking of the interviewee and its content. The choice of the word we use while speaking reveals different aspects of our identity [4].

This paper addresses the challenge of automated computational analysis of cross-modal communication, including visual, language and prosody of job interview videos, to create a predictive expert system that helps predict hireability and rate different trait of the interviewees. The results are then displayed on the dashboard for the interviewer.

This predictive system can be used for screening, hiring candidates for a job interview and act as an expert support system for hiring managers to take appropriate decisions. This system can also be used for a fully automated screening of candidates in online interviews as shown in Figure. 2, this can help with initial shortlisting of suitable candidates for further domain specific evaluation.

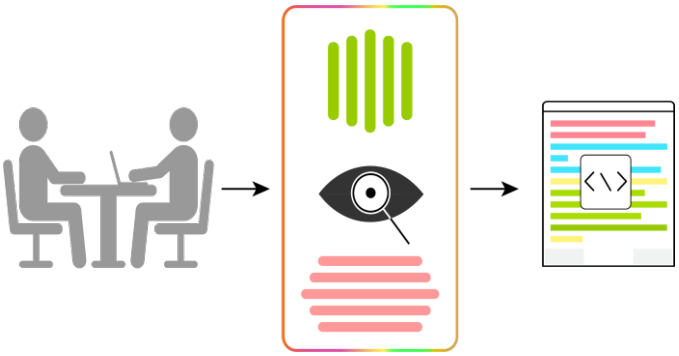


Figure. 1.1. Job Interview analysis model. From left to right 1. Interview Recording, 2. Feature Extraction, 3. Prediction Model for classification and rating of different traits of individuals.

The Figure. 1.1 shows the first part of the system that is the creation of a predictive framework, a classification model that will be able to rate an interview and different traits of the interviewee. The first step toward it is data gathering, for which we conducted interviews of college students. Next step was to extract verbal (word count, choice of word, sentiments), visual (facial emotions, eye gaze, head nods) and audio (speech rate, pitch, intensity, spectrogram, formants) cues from the interviews taken and physically rating the interviews to get the ground truth. The third was to test different classification algorithm to create a predictive model for the system, that would classify and rate different traits of individuals.

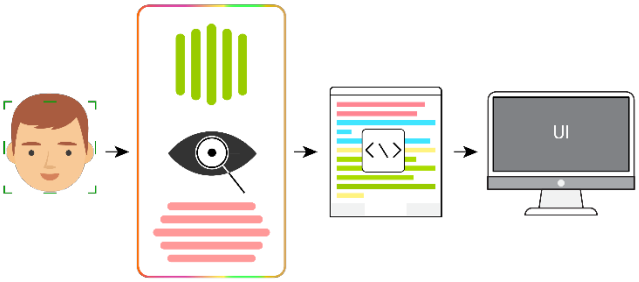


Figure. 1.2 Automated online interview. From left to right 1. Interview taken by candidate online, 2. Audio, Visual and Language Feature Extraction, 3. To predict hireability. 4. Results displayed.

II. Literary Survey

Personality characteristics can be comprehended by finding the connection between the nonverbal and verbal cues. Utilizing just video recorded interviews will not provide with sufficient information about the personality traits and thus the results would be partial. However, on utilizing both, visual and vocal cues, they provide a way to estimate top five personality traits and better relationship between verbal and nonverbal cues. There may be a situation where the interviewer might miss out the nonverbal cues of a candidate, thus this method would allow an interviewer to focus on the job profile. [5]

In a face to face interview, besides the conversational content, behavioral characteristics like confidence, speaking rate, demeanor, etc. play important role in succeeding in a job interview. Combination some of these characteristics help in predicting behavioral results in different social interactions. Pentland proposed a function of prosody and speaking style, which attempts to capture the social signaling between individuals. Signaling measures can be obtained by identifying voicing/ non-voicing and speaking /non-speaking rate from raw audio. Based on non-linguistics speaking style rating, candidates can be grouped into two classes (highly-rated, poorly-rated). Highly rated candidate shows more engagement with interviewer than the poorly rated candidate. A key limitation was a limited amount of labeling of interview data. [6]

In many Interview nonverbal behavior is given more importance than verbal contents by employers. More than 80 percent of the interviewee who maintained steady eye contact, showed high energy level by smiling, hand gestures, and general body movement, varied voice intensity and pitch to express appropriate affect, and answered majority question with little hesitation and fluidity were called for the second interview round. [3]

Vocal and Gestural Cues:

Vocal cues are the major source of information which is used by the interviewer to make attributions about an interviewee's personality traits. The way interviewee speaks relay emotions and increase or decrease immediacy. [3]

With the help of gestures, we can communicate a variety of messages. Gestures such as stop, get lost, shaking a fist and stick are interpreted as non-immediate gestures and such people are avoided. [3]

Face, Eye Behavior, and Immediacy:

Facial expression may express some emotions like happiness, sadness, interest, smiling, etc. facial expression complements the meaning of spoken messages and sometimes replaces them. Eye also serves many functions like interaction, emotions, attentiveness, attitude, and persuasion. Those who maintain increased gaze are seen as competent, sociable, and composed therefore increasing immediacy. [3]

It is found that 38 percent of the emotional information transmitted by a given message may be attributed to vocal cues, 55 percent to facial expressions, and only one percent to words. [3]

III. Dataset Description

A. *Interview Collection- Data Gathering*

The initial stage of this research is collecting interview dataset. Hence, accordingly, we conducted mock interviews for an open position. A data set consisting of 50 videos have been created.

1) Interview Setup and Participants: Interview is taken in a realistic setting. The setup is fairly simple, the hardware utilized is a camera, which would face the interviewee and a single unidirectional microphone, for the interviewee. Two chairs for the seating and a table for the positioning and placement of the camera. The camera is used to collect facial expressions and body cues while the microphone is used to collect audio conversations during the course of the interview. The primary medium of interaction between the interviewer and the interviewee is considered to be with the English language.

Figure. 3.1 Visual representation of the Interview and the system Setup

The interviews conducted were undertaken by professional and trained experts having industrial experience in a professional environment.

2) *Process:* Before these interviews are conducted, a questionnaire is prepared consisting of 7 questions which are

classified into different personality characteristics. Seven basic traits, each of them is being supported by a question from the questionnaire. This mapping is essential; thus, each question helps to determine that specific trait of the candidate.

Refer to the following table, which portrays the mapping of each trait with the question constructed:

TABLE I.

QUESTIONNAIRE QUESTIONS MAPPED WITH RESPECT TO THEIR MATCHING TRAITS

Question	Trait
Tell me about yourself	General
Any past experience where stress was correctly managed by you?	Stress Resistance
What motivates you to do a good job?	Passion
Explain why we should hire you?	Confidence
Has anything ever irritated you about people? How do you deal with it?	Cooperation
Tell me about a time when you demonstrated leadership skills.	Leadership
Tell me about your strengths and weaknesses.	General

Along with these, one of the most contributing trait, i.e. Eye contact has also been taking into consideration.

B. *Annotation- Data Labelling*

Next step in dataset building is labeling of the interviews, for which a chair of independent judges rate all the interviews manually. This process is essential so as to get the ground truth about the interview statistics. A rating form was circulated to the chair who watched the entire interview of every candidate and accordingly assigned scores for different assessment questions in the form. Since human ratings are subjective in nature, the chair consisted of 6 judges, for them to rate the interview and finally taking the average of the final ratings.

IV. PROPOSED SYSTEM

The system has two modules, Input, and Analysis respectively. The input module is used to automatically collect video and audio from the input devices and store the data into a suitable data storage. The analysis module takes the data (recorded interviews) singularly to perform the analysis on them and thereby extract features from audio, video and language cues. The proposed system diagram for the interview analysis is represented by Figure. 3

After the features have been extracted from each source, they are then used in the prediction model for predicting the ratings for each specific trait for the interviewee, ultimately producing the prediction of hireability of the interviewee.

The results from the analysis and predictions are then displayed on the dashboard for the hiring manager to see.

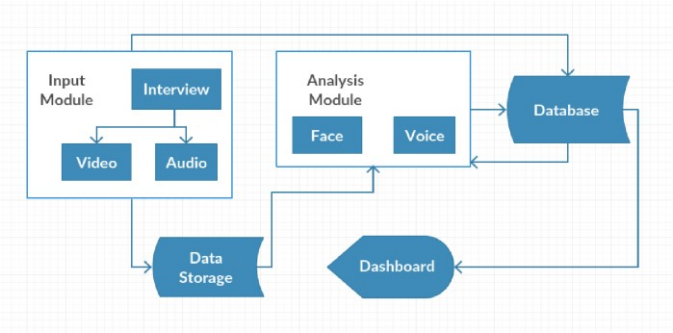


Figure. 4.1 Interview analysis system architecture diagram.

V. METHODOLOGY

This section details down the process of analysis on the recorded interviews and accordingly extracted features. Subsequently, we present the methods required to extract the audio, visual and language cues from the recorded data.

1) Audio cues: The major concern while extracting audio cue features is on the prosodic features. The preliminary step is to detect the individual who is speaking during the interview, the interviewer or the interviewee. Whilst during this process, we also need to segment the speech and non-speech chunks. Secondly, we need to make sure, that, the existence of any noise should be eliminated from the audio, if present. For audio cues, we extract the prosodic features of Pitch, Band energy, and the Speech Rate respectively.

The tool utilized for this purpose is PRAAT, an open-source audio analysis tool for extracting but not limited to the above-mentioned prosodic features. We first split the audio files into different chunks, wherein these chunks consist the answers from the interviewee and construct the collection of prosodic features of these splitted chunks.

Important features for audio cues are vocal intensities, pitch information, formant characteristics, spectral energy. These features reflect our social traits while communicating [7]. All these features can be extracted using the PRAAT tool.

2) Visual cues: Visual cues are the facial features of the interviewee during the interview. For every video frame, facial features of the interviewee are extracted. The first step in this is to detect the face itself in the frames. The Haar Cascade classifier for face detection can be used to do perform this task of face detection. Different facial cues extracted are facial expression, gaze, and head nods. There are different methods to detect facial expressions. Machine Learning is one of the approaches to detect facial expressions. Since it's a classification problem we can classify facial images using ML into different expressions like *neutral, joy, sadness, contempt, anger, disgust, fear and surprise.*

Second visual cue extracted is head nods. This is used to detect Yes and No during when signaled using head nods. Studies have proposed the use of Hidden Markov Model to detect head nodding [8, 9].

Figure. 5.1 The red dot, representing the center of the head, for keeping the track of head nods.

The third cue in facial feature extraction is the eye gaze. We know the importance of eye contact is an important nonverbal behavior, it is used to monitor feedback, reflect cognitive activity. We extract the amount of time the interviewee looks at the interviewer while answering a question or during the course of the interview.

Apart from these three cues we also extracted some more cues using facial action units. We used the **dlib library** [10] to detect facial landmarks. Facial landmarks help us localize different regions of the face like eyes, eyebrows, nose, mouth, and jawline.

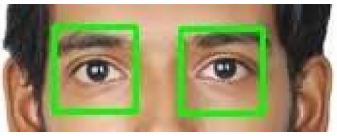


Figure. 5.2 Eye movement tracking indicators

The **dlib** uses a pre-trained facial landmark detector that estimates the location of 68 coordinates that maps to the face. Figure 5.3 represents these 68 points that maps the facial structure on the face. These points are then used to calculate more facial features like eyebrow raise distance from the eye, eye-opening distance, lip corner distance etc.

Figure. 5.3. Dots plotted on a facial Figures to help understand the production of various expression according to the movements of these points.

3) Language/Lexical Cues: While speaking, what we say and how we say tells a lot about us [11]. We can use "VADER" standing for Valence Aware Dictionary and sentiment Reasoner for verbal cues/ lexical cues. The VADER categories include negative emotion terms (e.g., sad, angry), positive emotion terms (e.g., happy, kind) and various content word categories (e.g., anxiety, insight). It is one of the most used lexicons for analyzing lexical cues, and especially when it comes to sentimental words. Each and every word in the dictionary is assigned to one or more categories. On top of this, VADER also has the capabilities of understanding sentiment-laden slang words [12].

4) Predictive Model: After extracting features from the facial cues, audio cues, and verbal cues, we create a single feature vector joining the features from the facial, audio and verbal cues for each interview. This finally generates the dataset. This dataset has ratings for the judges as well. This dataset is used against model prediction to find the similarity between the prediction and actual judges score. We have used a regression model to predict the hireability and overall score of the interviewed candidate. Mean Square Error (MSE) helps us to find the best model since it gives the error made by the actual judge's score. Among Support Vector Regression (SVR), Lasso Regression and Ridge Regression, MSE of Ridge and SVR were lesser and predicted the results closer to judges score.

VI. PROTOTYPE

The following image, figure 6.1, is a demonstration of the prototype created for the proposed system. This prototype comprises of the following hardware:

1. Raspberry Pi 3 Model B
2. Raspberry Pi Camera
3. Generic USB Microphone

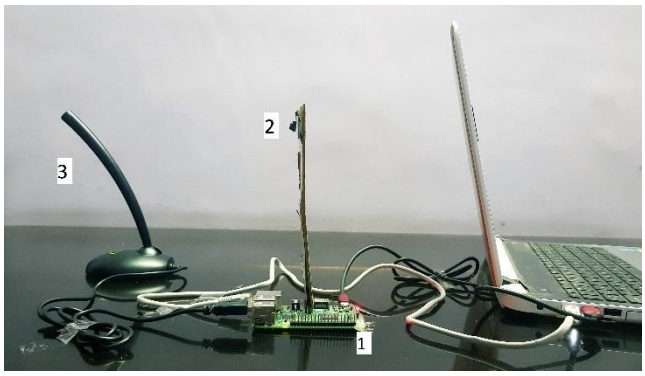


Figure. 6.1. Prototype setup of the proposed system.

The above-mentioned hardware is collectively referred to as kit in this paper. The system is powered by the official Raspbian OS designed specifically for the Raspberry Pi.

The prototype is implemented considering that we have developed a stand-alone system, where the kit (The above-mentioned components) will perform the analysis tasks of visual and vocal processing, in real time. This processing will generate a visual and vocal feature value and that value is provided to machine learning algorithm to generate the user, an "Overall Recommended Score" along with rest of the scores of different features like Excitement, Friendliness etc.

To access the kit, the user has to connect it with the web interface or via an HDMI display and once accessed, the user can control the interview as well as view the real-time statistical graphs of the candidate.

VII. RESULTS

The proposed system produced a variety of results, from a single personality trait to the overall hireability recommendation of the candidate. We produced graph plots for every aspect, from internal structures, models to the final output, so as to represent the accuracy and working of the system constructed.

1) Internal Architectural Representations: Figure 7.1 is a representation of the input of an image to the CNN architecture. This image exists on the second hidden convolutional layer of the CNN architecture; used for emotion detection on the face of the interviewee. This layer produces 8 different images due to the fact that the second hidden convolution layer in the CNN architecture has the 8 different filters, where each filter has the capability to identify a different feature. Since these filters are exhibited by the second layer, they provide the opportunity to represent higher level features on the basis of the low-level features existing on the first hidden layer of the architecture. This thus results in an output image having a size of 60*60 pixels.

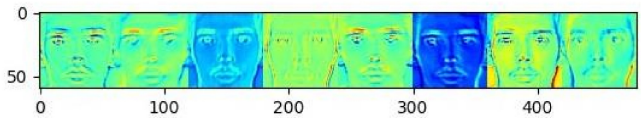


Figure 7.1. Visual representation of an image in the second hidden layer of CNN architecture.

2) Statistical Results: There are two categories of statistical results furnished from the output produced by the predictive system.

The first section is the statistics just about the audio cues. Referring to the Table II. we have the different statistical values produced; Average, mean, standard deviation and maximum on the four most important features which were extracted from audio; Formants, intensity, pitch and band energy.

TABLE II.
STATISTICAL ANALYSIS OF AUDIO CUES

	FEATURES			
	FORMANT	INTENSITY	PITCH	BAND ENERGY
AVERAGE	-	63.76382	133.5861	0.078372131
MEAN	855.9622	2.92734	99.31801	-
STANDARD DEVIATION	289.1454	14.9847	23.6255	-
MAX	-	78.56844	569.2424	-

When the system produces outputs on the UI with the overall hireability scores of all the candidates, it beforehand makes a comparison of all the candidates, this comparison considers both verbal and nonverbal cues. So, in order to carry out this comparison, the system has to have some concrete or comparison values, with which it has to determine which candidate is better than which of the other candidate. The Table II. represents all the audio statistics of a single interview, when these values are compared to, say, a minimum criteria values, it becomes easier to then determine a better comparison activity.

The second section talks about various statistical analysis of different models used with the system and their fitness for the purpose.

Figure. 7.2 is graph plot of Lasso Model, portraying the goodness capacity, it represents the models predicting accuracies across different personality traits considered by us at earlier stages. Followed by Figure. 7.3 which portrays the goodness capacity of the Support Vector Regression Model for the different traits.

Figure 7.2. The goodness of regression of Lasso model on the system.

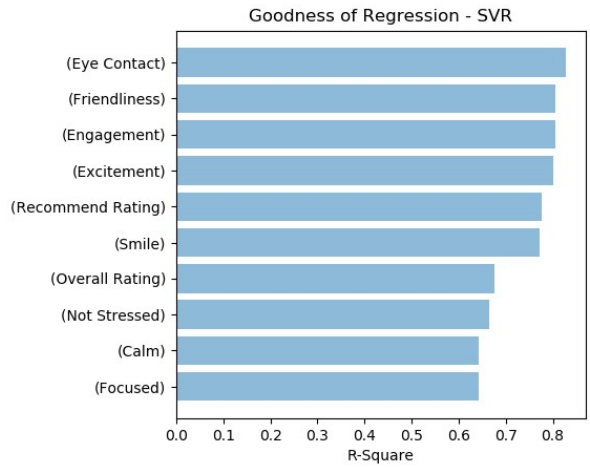


Figure 7.3. The goodness of regression of SVR model on the system.

Figure 7.4 represents the goodness capacity of the Ridge model for the different traits. Looking at the goodness of fit for the different algorithms we can observe that the SVR model fits the best with R-Square value for all the different traits more than 0.6, while Lasso and Ridge's model falls short in fitting some traits when compared to SVR.

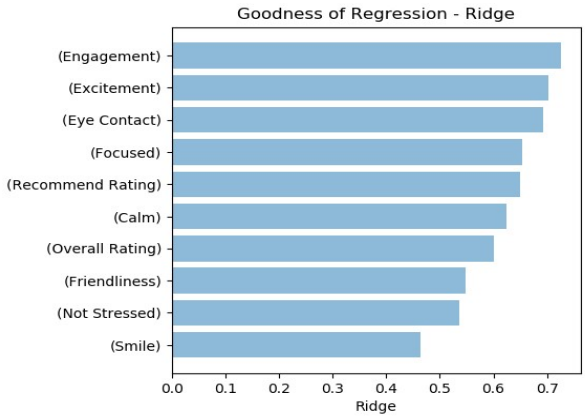


Figure 7.4. The goodness of regression of Ridge model on the system.

For the sake of understanding the outcomes, let us consider the top five traits judged represented in Figure 7.5. We take Overall rating, Recommend Rating, Excitement, Engagement, and Friendliness and compare their correlation coefficient of SVR model for a different set of features like Facial only, Prosodic only, and Facial and Prosodic together.

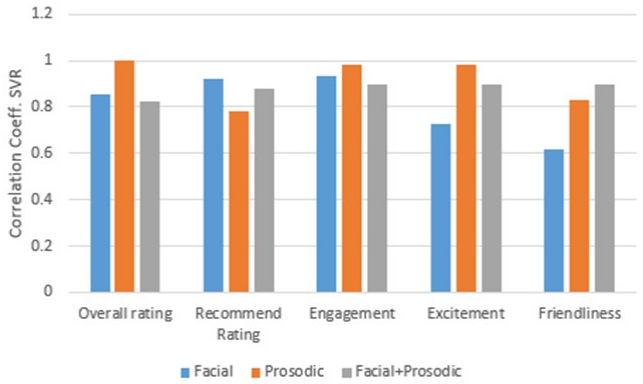


Figure 7.5. Correlation coefficient comparison of a different set of features.

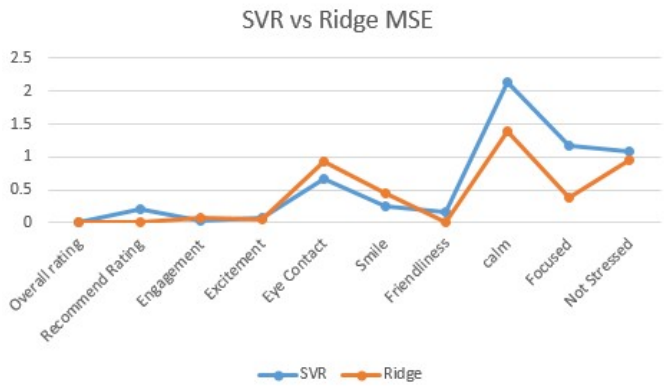


Figure 7.6. MSE results for SVR and Ridge models between Ground Scores and Predicted Scores.

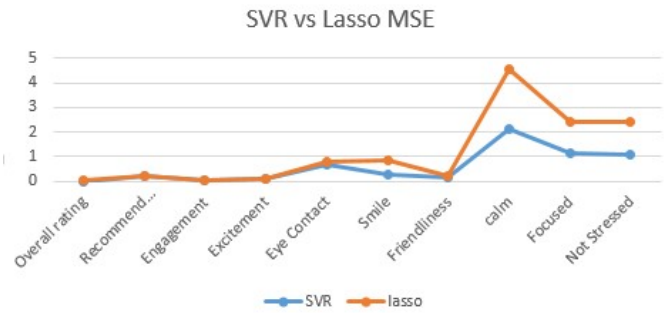


Figure 7.7. MSE results for SVR and Lasso models between Ground Scores and Predicted Scores.

Figure 7.6 and Figure 7.7 represents the comparison between the mean square errors of SVR model with the Lasso and Ridge Model. In Figure 7.7, the mean square error for Ridge model is lower for most of the traits compared to SVR model. While in Figure 7.8 the mean square error for the Lasso model is larger than the SVR model for all the traits.

3) User Interface Results: User Interface outputs are purely choice based, depending upon the requirements of the system user.

The UI that we designed for this project allowed an interviewer to observe in real time; a smile score variation graph.

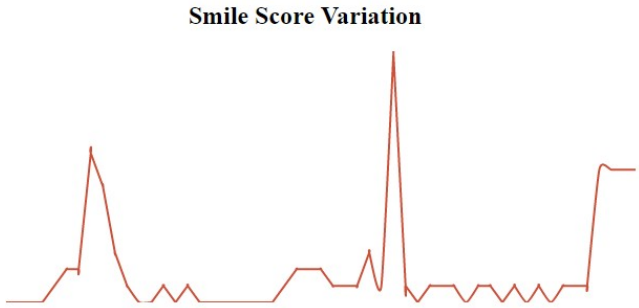


Figure 7.8. Smile Score Variation Graph as visible to the interviewer on their User Interface.

Figure 7.8 represents the real-time smile score variation shown to the HR manager in the UI. As the interview proceeds the graph changes depending on the user's smile intensity.

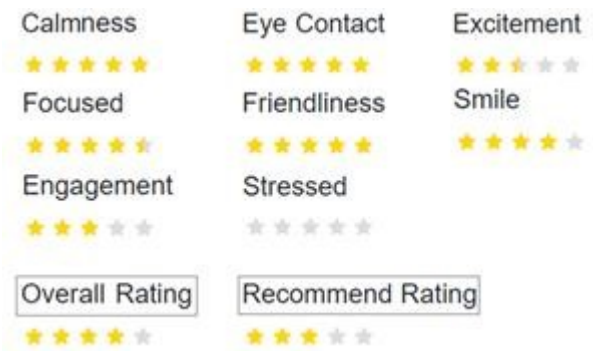


Figure 7.9. An interviewee's final scores displayed to the interviewer on their UI dashboard at the end of the interview.

Finally, when an interview ends, the system accordingly makes its predictions, and thusly produces a series of ratings to the interviewer. These ratings belong to various parameters; as visible, Calmness, eye contact, excitement, focused, friendliness, et al. Not only the individual parameters but the system also predicts an overall and recommended ratings of the interviewee. These parameters are predicted with different inputs provided by the interviewee in a combination.

VIII. RESULTS

The system developed here uses a multimodal approach, we have used Audio, Video as well as Lexical cues for preparing the prediction model. This modal simplifies the task of an organization for conducting HR based interviews, by allowing the interview process to be automated, carrying out the tasks of evaluating the verbal and nonverbal signatures of the interview candidate.

We cannot prioritize only one from both verbal and nonverbal signatures. Both are equally important when it comes to HR interviews. One of the most important case to consider is that the existing systems do not propose a multimodal nature existence. They are either dependent on audio or on video cues for analysis.

The dataset was generated with the help of video recording sessions of interviews of a number of candidates, they were then manually rated by a chair of judges for the sake of ground truth, so a comparison could be made when the system predicts the scores and ratings for the candidates. While generating data sets it became a necessity to ensure that the video quality was maintained, the candidate's facial points were visible in the recordings, not only that, but, we had to make sure that any kind of noise recorded, whilst during the interview was supposed to be minimized if not eliminated.

To choose the predictive model we tested our dataset on three different algorithms, Support Vector Regression, Lasso Regression and Ridge Regression. We judged the three trained models of SVR, Lasso, and Ridge on both their R-Square values and mean square error (MSE). From their R-Square and MSE values, we found that SVR fits our data the best and is the most accurate compared to Lasso and Ridge Models.

ACKNOWLEDGMENT

We would like to express our special thanks of gratitude to the Principal, HOD and Deputy HOD of the Computer Department, Vivekanand Education Society's Institute of Technology, for giving us an opportunity to showcase our skills and capabilities. Our mentor, Prof. (Mrs.) Sharmila Sengupta has been a guiding light in our project who has helped us through every stage of this project, mentoring us with this research, by that providing us with exposure to new learning and possibilities. Also, we would like to extend a grateful thanks to Mr. Devesh Rajadhyax, CEO, Cere Labs Pvt. Ltd who has encouraged and motivated us in providing an industrial exposure to our research. He has helped us through some of the tough scenarios. To all the individuals involved in this research, we are extremely thankful to them for their utmost support and care to us.

REFERENCES

1. T. DeGroot and J. Gooty. "Can nonverbal cues be used to make meaningful personality attributions in employment interviews?" Journal of Business and Psychology, 24(2):179– 192, 2009.
2. V. Soman and A. Madan, "Social signaling: Predicting the outcome of job interviews from vocal tone and prosody," in ICASSP. Dallas, Texas, USA: IEEE, 2010.
3. M. A. Gotcher, "Nonverbal communication in the job Interview: first impressions and Nonverbal immediacy," Senior Thesis, Texas Tech University, 1990.
4. Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. Journal of Personality and Social Psychology, 77(6), 1296-1312.
5. DeGroot, T. & Gooty, Can Nonverbal Cues be Used to Make Meaningful Personality Attributions in Employment Interviews? J. J Bus Psychol (2009)24:179. https://doi.org/10.1007/s10869-009-9098-0
6. A. Madan, R. Caneel, S. Pentland, "VibeFones: Socially aware Mobile Phones:, International Symposium of Wearable Computing (ISWC), Switzerland, 2006.
7. R. W. Frick, "Communicating emotion: The role of prosodic features." Psychological Bulletin, vol. 97, no. 3, p. 412, 1985.
8. A. Kapoor and R. Picard, "A Real-Time Head Nod and Shake Detector," Proc. Workshop Perspective User Interfaces, Nov. 2001.
9. W. Tan and G. Rong. A real-time head nod and shake detector using HMMs. Expert Systems with Applications, 25(3):461–466, 2003.
10. Davis E. King. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research* 10, pp. 1755-1758, 2009
11. D. Sanchez-Cortes, P.Motlicek, and D. Gatica-Perez. Assessing the Impact of Language Style on Emergent Leadership Perception from Ubiquitous Audio. In Proceedings of the ACM International Conference on Mobile and Ubiquitous Multimedia (MUM), 2012.
12. Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

Certificate of Participation in the Conference

IEEE

Certificate of Participation

This is to certify that

Ankit Shaw

has successfully presented a paper entitled

Discover cross - model Human Behavior Analysis.

in the Second International Conference on Electronics, Communication and Aerospace Technology (ICECA 2018), held during 29-31, March 2018, at the RVS Technical Campus, Coimbatore, Tamilnadu, India.

Session Chair Conference Chair Director

Technical Sponsor

ELECTRON DEVICES SOCIETY

IEEE

Certificate of Participation

This is to certify that

Swastika Thakur,

has successfully presented a paper entitled

Discover cross-model Human Behavior Analysis.

in the Second International Conference on Electronics, Communication and Aerospace Technology (ICECA 2018), held during 29-31, March 2018, at the RVS Technical Campus, Coimbatore, Tamilnadu, India.

Session Chair **Conference Chair** **Director**

Technical Sponsor

ELECTRON DEVICES SOCIETY

IEEE

Certificate of Participation

This is to certify that

Pravin Tripathi.

has successfully presented a paper entitled

Discover cross-modal Human Behavior Analysis.

in the Second International Conference on Electronics, Communication and Aerospace Technology (ICECA 2018), held during 29-31, March 2018, at the RVS Technical Campus, Coimbatore, Tamilnadu, India.

Session Chair **Conference Chair** **Director**

Technical Sponsor

ELECTRON DEVICES SOCIETY

IEEE

Certificate of Participation

This is to certify that

Pratik P. Watwani,

has successfully presented a paper entitled

Discover cross-modal Human Behavior analysis.

in the Second International Conference on Electronics, Communication and Aerospace Technology (ICECA 2018), held during 29-31, March 2018, at the RVS Technical Campus, Coimbatore, Tamilnadu, India.

Session Chair **Conference Chair** **Director**

Technical Sponsor

ELECTRON DEVICES SOCIETY

Certificate of Participation in Lokmanya Tilak College National Level Project Competition, "TANTRAGYAN 2K18" organized on 27th March 2018.

L.T.J.S.S.'s

LOKMANYA TILAK COLLEGE OF ENGINEERING

TANTRAGYAN 2K18

IEEE

CERTIFICATE

OF PARTICIPATION

THIS CERTIFICATE IS PROUDLY PRESENTED TO

Mr./Ms. ANKIT SHAW

OF VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

FOR THE PROJECT TITLED

DISCOVER CROSS-MODAL HUMAN BEHAVIOR ANALYSIS.

IN THE NATIONAL LEVEL PROJECT COMPETITION

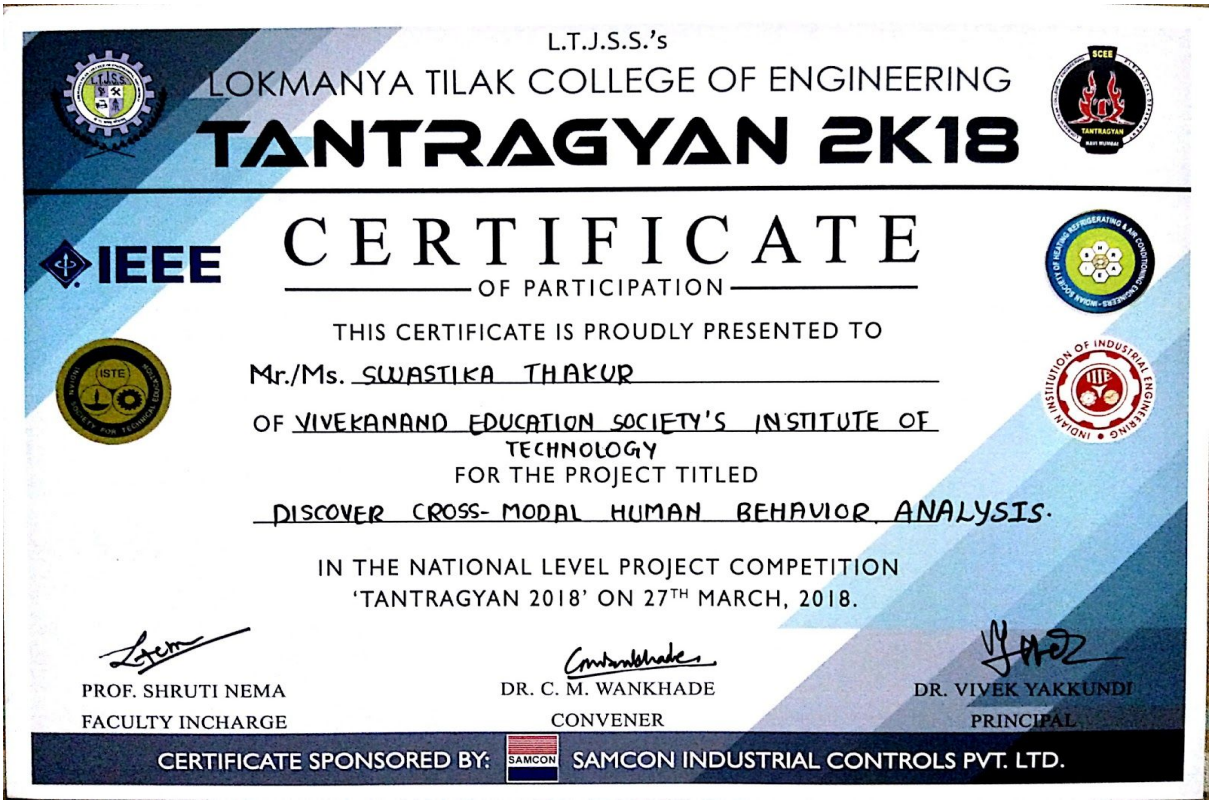
'TANTRAGYAN 2018' ON 27TH MARCH, 2018.

PROF. SHRUTI NEMA, FACULTY INCHARGE

DR. C. M. WANKHADE, CONVENER

DR. VIVEK YAKKUNDI, PRINCIPAL

CERTIFICATE SPONSORED BY: SAMCON INDUSTRIAL CONTROLS PVT. LTD.



L.T.J.S.S.'s

LOKMANYA TILAK COLLEGE OF ENGINEERING

TANTRAGYAN 2K18

IEEE

CERTIFICATE

OF PARTICIPATION

THIS CERTIFICATE IS PROUDLY PRESENTED TO

Mr./Ms. SWASTIKA THAKUR

OF VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

FOR THE PROJECT TITLED

DISCOVER CROSS-MODAL HUMAN BEHAVIOR ANALYSIS.

IN THE NATIONAL LEVEL PROJECT COMPETITION

'TANTRAGYAN 2018' ON 27TH MARCH, 2018.

PROF. SHRUTI NEMA, FACULTY INCHARGE

DR. C. M. WANKHADE, CONVENER

DR. VIVEK YAKKUNDI, PRINCIPAL

CERTIFICATE SPONSORED BY: SAMCON INDUSTRIAL CONTROLS PVT. LTD.

L.T.J.S.S.'s

LOKMANYA TILAK COLLEGE OF ENGINEERING

TANTRAGYAN 2K18

IEEE

CERTIFICATE

OF PARTICIPATION

THIS CERTIFICATE IS PROUDLY PRESENTED TO

Mr./Ms. PRAVIN TRIPATHI

OF VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

FOR THE PROJECT TITLED

DISCOVER CROSS - MODAL HUMAN BEHAVIOR ANALYSIS.

IN THE NATIONAL LEVEL PROJECT COMPETITION 'TANTRAGYAN 2018' ON 27TH MARCH, 2018.

PROF. SHRUTI NEMA
FACULTY INCHARGE

DR. C. M. WANKHADE
CONVENER

DR. VIVEK YAKKUNDI
PRINCIPAL

CERTIFICATE SPONSORED BY: SAMCON INDUSTRIAL CONTROLS PVT. LTD.

L.T.J.S.S.'s

LOKMANYA TILAK COLLEGE OF ENGINEERING

TANTRAGYAN 2K18

IEEE

CERTIFICATE

OF PARTICIPATION

THIS CERTIFICATE IS PROUDLY PRESENTED TO

Mr./Ms. PRATIK P. WATWANI

OF VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

FOR THE PROJECT TITLED

DISCOVER CROSS-MODAL HUMAN BEHAVIOR ANALYSIS.

IN THE NATIONAL LEVEL PROJECT COMPETITION 'TANTRAGYAN 2018' ON 27TH MARCH, 2018.

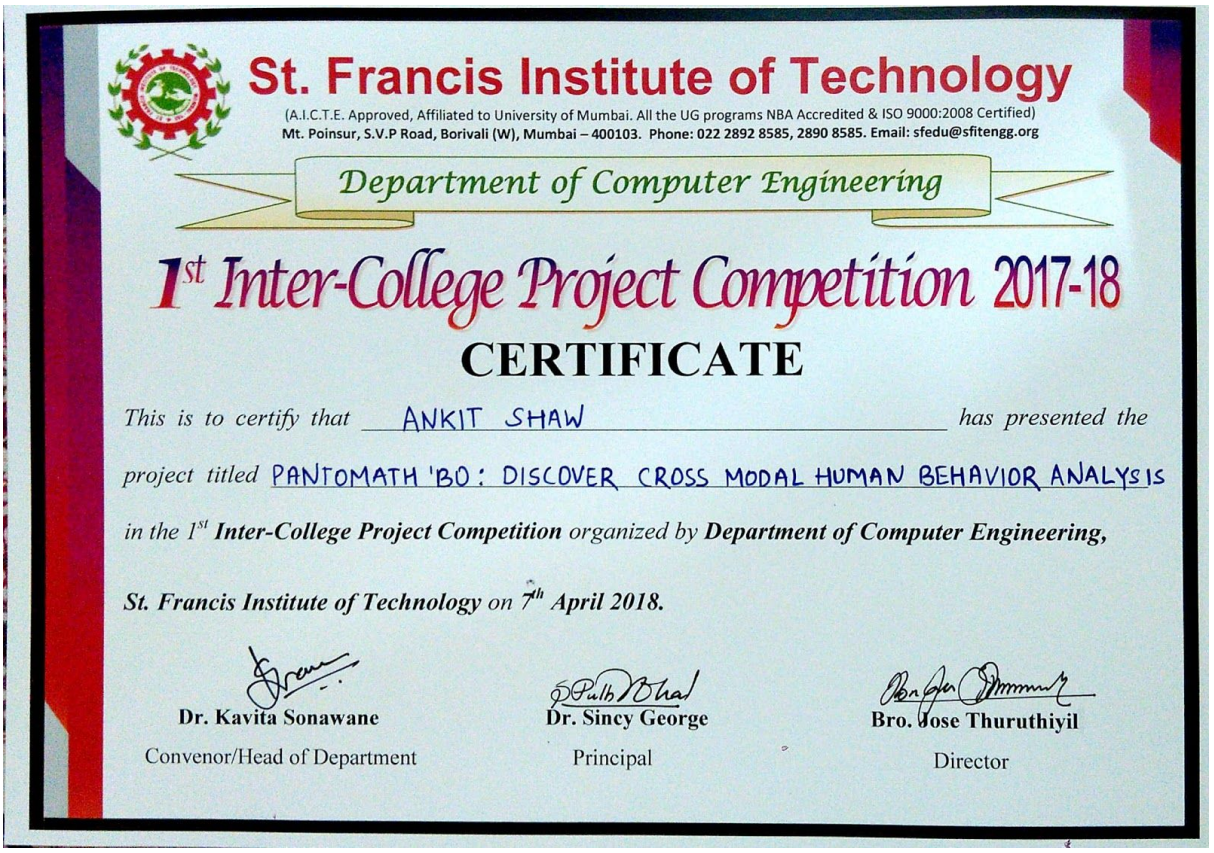
PROF. SHRUTI NEMA
FACULTY INCHARGE

DR. C. M. WANKHADE
CONVENER

DR. VIVEK YAKKUNDI
PRINCIPAL

CERTIFICATE SPONSORED BY: SAMCON INDUSTRIAL CONTROLS PVT. LTD.

Certificate of Participation in St. Francis Institute of Technology's Inter College Project Competition on 7th April 2018.



St. Francis Institute of Technology

(A.I.C.T.E. Approved, Affiliated to University of Mumbai. All the UG programs NBA Accredited & ISO 9000:2008 Certified)
Mt. Poinsur, S.V.P Road, Borivali (W), Mumbai – 400103. Phone: 022 2892 8585, 2890 8585. Email: sfedu@sfitengg.org

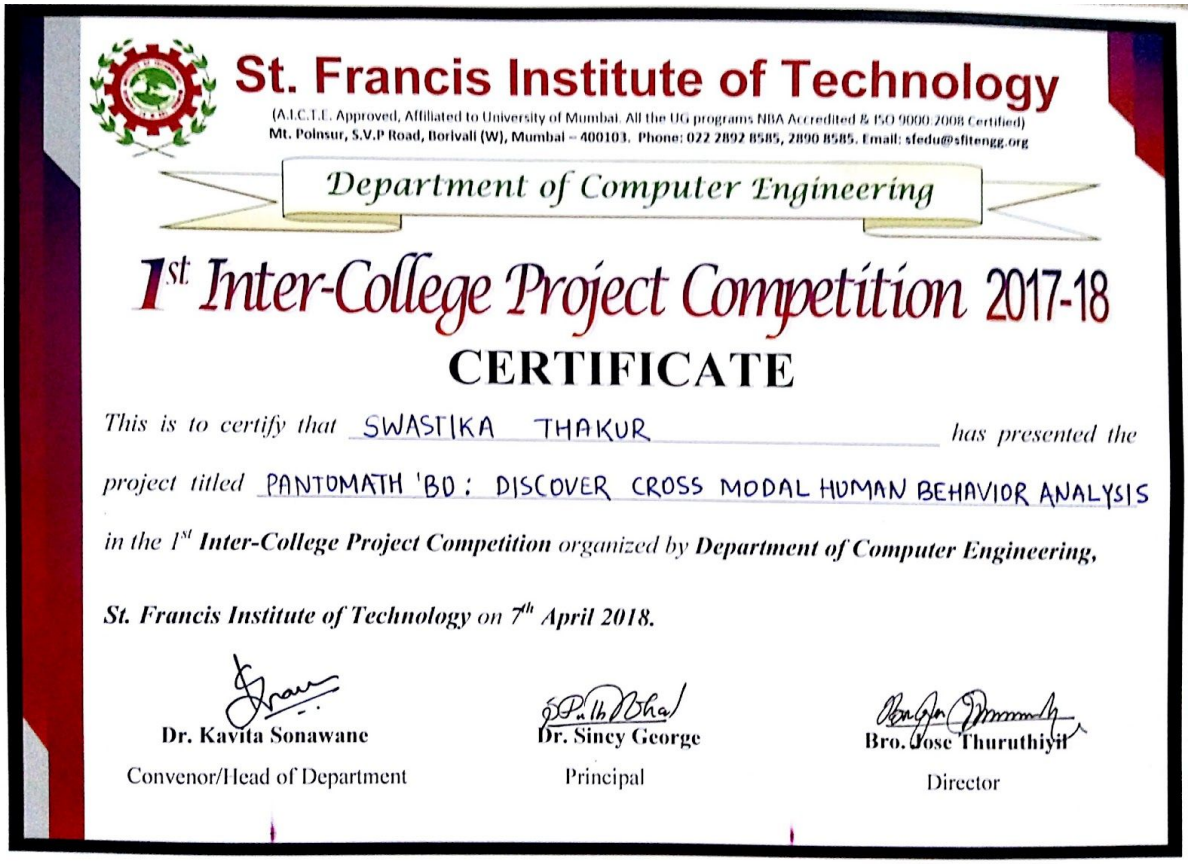
Department of Computer Engineering

1st Inter-College Project Competition 2017-18

CERTIFICATE

This is to certify that ANKIT SHAW *has presented the project titled* PANTOMATH 'BO : DISCOVER CROSS MODAL HUMAN BEHAVIOR ANALYSIS *in the 1st* ***Inter-College Project Competition*** *organized by* ***Department of Computer Engineering, St. Francis Institute of Technology*** *on* ***7th April 2018.***

Dr. Kavita Sonawane	Dr. Sincy George	Bro. Jose Thuruthiyil
Convenor/Head of Department	Principal	Director



St. Francis Institute of Technology

(A.I.C.T.E. Approved, Affiliated to University of Mumbai. All the UG programs NBA Accredited & ISO 9000:2008 Certified)
Mt. Poinsur, S.V.P Road, Borivali (W), Mumbai – 400103. Phone: 022 2892 8585, 2890 8585. Email: sfedu@sfitengg.org

Department of Computer Engineering

1st Inter-College Project Competition 2017-18

CERTIFICATE

This is to certify that SWASTIKA THAKUR *has presented the project titled* PANTOMATH 'BO : DISCOVER CROSS MODAL HUMAN BEHAVIOR ANALYSIS *in the 1st* ***Inter-College Project Competition*** *organized by* ***Department of Computer Engineering, St. Francis Institute of Technology*** *on* ***7th April 2018.***

Dr. Kavita Sonawane	Dr. Sincy George	Bro. Jose Thuruthiyil
Convenor/Head of Department	Principal	Director

St. Francis Institute of Technology

(A.I.C.T.E. Approved, Affiliated to University of Mumbai. All the UG programs NBA Accredited & ISO 9000:2008 Certified)
Mt. Poinsur, S.V.P Road, Borivali (W), Mumbai – 400103. Phone: 022 2892 8585, 2890 8585. Email: sfedu@sfitengg.org

Department of Computer Engineering

1st Inter-College Project Competition 2017-18

CERTIFICATE

This is to certify that PRAVIN TRIPATHI has presented the project titled PANTOMATH 'BO: DISCOVER CROSS MODAL HUMAN BEHAVIOR ANALYSIS in the 1st **Inter-College Project Competition** organized by **Department of Computer Engineering, St. Francis Institute of Technology** on 7th April 2018.

Dr. Kavita Sonawane	**Dr. Sincy George**	**Bro. Jose Thuruthiyil**
Convenor/Head of Department	Principal	Director

St. Francis Institute of Technology

(A.I.C.T.E. Approved, Affiliated to University of Mumbai. All the UG programs NBA Accredited & ISO 9000:2008 Certified)
Mt. Poinsur, S.V.P Road, Borivali (W), Mumbai – 400103. Phone: 022 2892 8585, 2890 8585. Email: sfedu@sfitengg.org

Department of Computer Engineering

1st Inter-College Project Competition 2017-18

CERTIFICATE

This is to certify that PRATIK P. WATWANI has presented the project titled PANTOMATH 'BO: DISCOVER CROSS MODAL HUMAN BEHAVIOR ANALYSIS in the 1st **Inter-College Project Competition** organized by **Department of Computer Engineering, St. Francis Institute of Technology** on 7th April 2018.

Dr. Kavita Sonawane	**Dr. Sincy George**	**Bro. Jose Thuruthiyil**
Convenor/Head of Department	Principal	Director