PRESENTLY - IN ADVANCED PREPARATION FOR THE PRESENTATIONS

2021-02

Project Proposal Report

Rupasinghe Arachchige Kethmini Wandana

B.Sc. (Hons) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

July 2021

PRESENTLY - IN ADVANCED PREPARATION FOR THE PRESENTATIONS

2021-02

Project Proposal Report

B.Sc. (Hons) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

July 2021

Declaration of the Candidate and Supervisor

I declare that this is my work, and this proposal does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university or institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Name	Student ID	Signature
R.A.K.Wandana	IT18227550	

The above candidate is carrying out research for the undergraduate Dissertation under my supervision.

Signature of the supervisor: Date

Acknowledgement

This paper and the research behind it would not have been possible without the exceptional support of my supervisors, my Research supervisor Dr. Shyam Reyal, Co-supervisor Ms. Samanthi Eranga Siriwardene, and External supervisor Ms. Veerandi Kulasekara. Their enthusiasm, knowledge, and exacting attention to detail have been an inspiration and kept my work on track. Besides my supervisors, I would like to thank the rest of my group members: Himanshi, Chathurdi, Naveen for their encouragement, insightful comments, and every other support. And finally, thanks to my parents who always offering support and love.

Abstract

Emotions and body language are important in human communication and public speaking. Therefore, understanding emotional expressions and hand gestures in presentations and improving presentation skills will be of great value for a presenter or a speaker. The speech will be more impactful if emotions are added to the speech delivery. There is a lack of tool support to help conduct an efficient and in-depth analysis to get feedback for the presentations in advance[1]. The proposed system “Presently”, a mobile responsive web application allows speakers to assess their own presentation skills. Emotion Recognition is the process of interpreting a person's thinking and current mental state by analyzing facial expressions, body posture, gestures, and voice. In this paper, the proposed solution is to analyze users’ emotions via input video based on Computer Vision and Machine Learning techniques and provide effective feedbacks and ratings to users to enhance their presentation skills. Recognizing emotions and body language from images or video is a simple operation for the human eye, but it's a difficult challenge for machines and requires numerous image processing techniques for feature extraction. The proposed solution will extract unique features of the face from the input video and analyze them using Machine Learning algorithms and provide effective feedback as the output.

Key words – Emotions, Video Analysis, Feature Extraction, Computer Vision

Table of Contents

Table of Figures

List of Tables

List of Abbreviations

Abbreviation	Description
SVM	Support Vector Machine
PCA	Principal Component Analysis
RBF	Radial Basis Function

1. INTRODUCTION

1.1. Background and Literature Survey

Emotion is how we feel, whether that is anger, fear, sadness, happiness, anxiety, guilt, shame, or nervousness. Most people would prefer to leave emotions out of their presentations and speeches. When delivering a presentation, a professional speaker will appear to be more confident. But Nonprofessional speakers like university students will not appear to be confident most of the time. Not only using emotions but also using body language in presentation the right way can help speakers to engage more with the audience and be confident during the presentation. As per Figure 1-1, many of the university students agree that making eye contact, positive head moments, maintaining a confident posture will be helping to more interact with the audience.

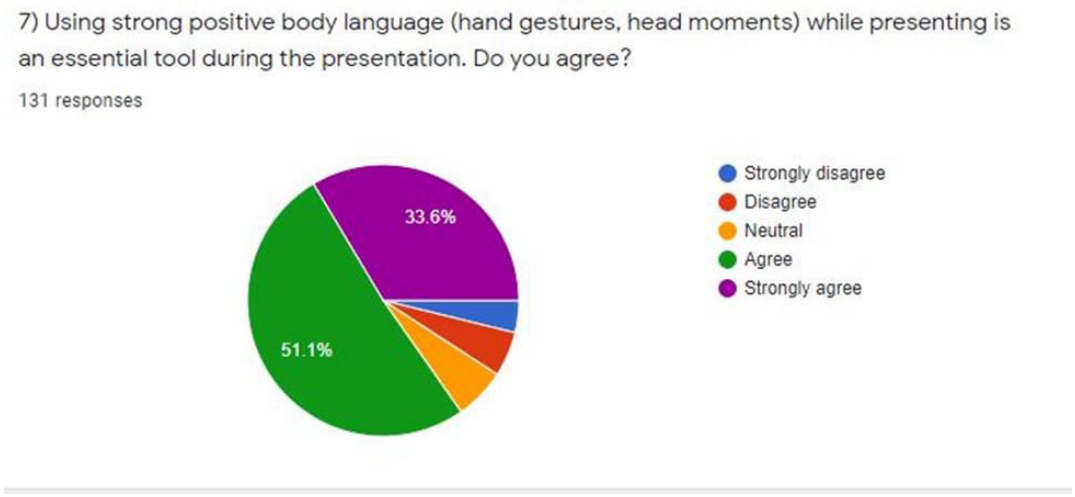

Figure 1-1 Summary of Responses to Using strong positive during a Presentation

Presentation is a way to conveys information, message, or a fact to an audience from a speaker. Preparation is the most important part of doing a successful presentation. Before preparing for the presentation, it is important to recognize what kind of audience will be there for the presentation. By structuring the presentation, learn how to control emotions while doing a presentation, learn how to attract and interact with the audience will help to do a presentation with ease.

Most of the time people need to practice before doing a presentation or a speech. Nowadays most students had to do online presentations due to the pandemic situation in the world. Hence, “Presently” will help those students to self-evaluate their presentation performance before doing an actual presentation or a speech. Most people agree that emotional control and using positive body language while doing a presentation is an essential tool during the presentation as shown in Figure 1-2.

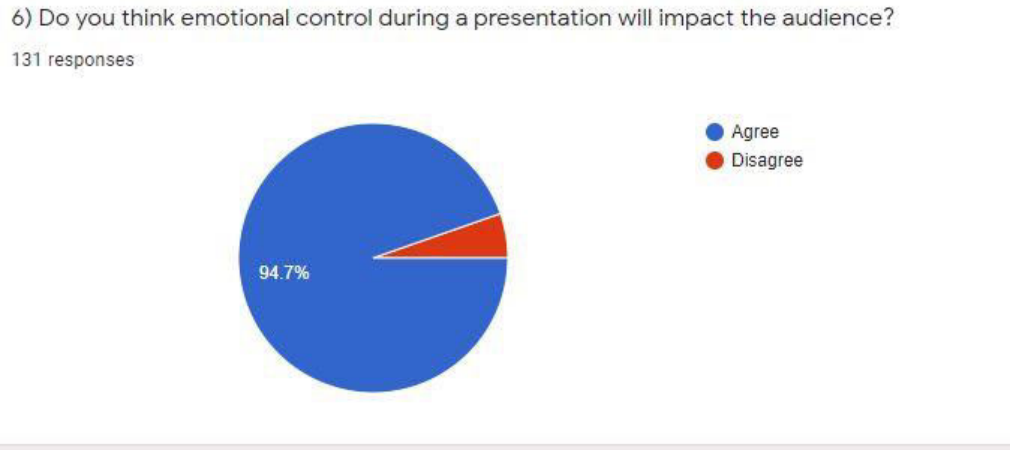

Figure 1-2 Summary of Responses to Emotional Control during a Presentation

Academic students are only focused on the completion of the presentation or a speech and get higher grades. Therefore, they should be practice beforehand for the presentation and get some idea about their presentation quality. There is a lack of tool support to self-evaluate the presentation skills in Sri Lanka among academic students as well as professional speakers. There are not any self-evaluation presentation skills systems or tools to analyze the emotions of the speaker within Sri Lanka especially. Most of the existing systems do not check other than the basic emotions like happy, sad, angry, surprise, not the feelings when speakers get while doing a presentation like nervousness, confidence, excitement. So most professional speakers or academic students need to continue with manual evaluating methods instead of using self-evaluating software. Most of the existing research is limited to basic emotions and emotion analysis is not used in the presentation quality evaluation field. Hence “Presently” will be a great tool to prepare the presentations in advance.

1.2. Literature Review

The main goal of this research work is to identify presenters' emotional expressions from the human face and body language. Therefore, a survey on various existing research works to analyze the emotions and body language using presentation videos are reviewed and discussed. EmoCo: Visual Analysis of Emotion Coherence in Presentation Videos [1] is a system studied by Haipeng Zeng et al[1] to analyze emotion coherence across different behavioral modalities in presentation videos. It analyses three modalities: face, text, and audio and eight emotions (happiness, sad, anger, disgust, contempt, fear surprise, and neutral) with the accuracy of the first video on the face, text and audio channels was 85.4%, 79.3%, and 89.6%, respectively, and the accuracy of the second video on the face, text and audio channels was 95.8%, 87.3% ,and 81.8%, respectively.

The type of human emotion represented by facial features is identified using a face emotion detection system. In 2015, Ibrahim A. Adeyanju et al. [2] has researched "Performance Evaluation of Different Support Vector Machine Kernels for Face Emotion Recognition". In that research, researchers analyze four different SVM kernels and seven emotions and SVM multiclass classification is used. The used dataset is manually cropped to retain only the face and remove other parts of the body. Then the colored images of the faces are converted into grayscale images and features extracted and dimensionality reduction of image matrix researchers use the PCA method. According to the image dimensions and SVM kernels, accuracy will change. The average highest accuracy of 99.33% performed in Quadratic Function kernel when the 200 x 200 reduced pixel images were used. Computation time is low for small dimensional images[2].

In 2017, Binh T. Nguyen et al. [3] studied Efficient real-time Emotion Detection using camera and facial Landmarks. One of the most important prerequisites for developing a real-time emotion detection system is the ability to recognize facial landmarks: eyes, eyebrows, eyelids, nose, lips, and jaw. In this work[3], an approach of Kazemi and

Sullivan is used to extract 68 key points in human faces effectively. The research considers Positive, Negative, and Blank emotion types. To build an appropriate emotion classification model, researchers use different supervised classification models: multi-class Support Vector Machine (multi-SVM) with two kernels (linear and RBF kernels), decision tree, and random forest. The results of this research showed that multi-class SVM with RBF kernel is chosen for the final system, and it detects human emotions naturally with 70.65% accuracy.

Emotion Recognition is the process of analyzing a person's facial expressions, body posture, gestures, and voice to understand their thinking and current mental state[4]. In this "Audience Feedback Analysis using Emotion Recognition" system [4], Mrs. Madhura Prakash et al. are used DNN for detecting the faces from the input provided because DNN has the highest accuracy in detecting the correct potion of the face and the number of faces detected is high compared to the other techniques. Other than that, the same process as previously mentioned researches[2-3] is applied for this research also. As the result of this research, researchers found the average emotional state of the data gathering and graphical representation of the times each emotion has appeared[4].

In 2018, Vengal Rao Guttha et al. [5] proposed a system for generating automated feedback using Facial Emotion Recognition. The system consists of two parts: "Emotion Recognition" and "Feedback Generation". A pre-trained model based on Histogram of Oriented Gradients and SVM linear classifier is used for face detection and SVM with linear kernel classifier is used for train the dataset [5]. After recognizing the emotions there are three methods to generate the feedback. In recognizing the six emotions (happy, sad, anger, disgust, surprise and neutral), an accuracy of 90.60 ±1.65% was reached. The Automated Feedback Generation System [5] was able to process images, identify emotions with reasonable accuracy, and generate feedback in a form that could be utilized to evaluate customer satisfaction of advertisements.

Hand gestures are one of the most often utilized gestures in everyday human communication[6]. Multimodal systems allow Human Computer Interaction applications

to incorporate several sorts of modalities that occur concurrently in a more efficient and reliable manner. This paper covers the emotion of the face such as happy, anger, disgust, surprise, neutral and fear. and hand motions, as well as information on the multimodal analyzer based on facial emotion identification[6].

1.3. Research Gap

According to the research papers I have read in the past few days, there are not many research papers that have been conducted in the Presentation analysis area. There is no proper research paper to evaluate presenter's emotions and body language using video analysis at the same time. Many research articles and systems have been developed to detect users' emotions, but an expert system to detect and recognize emotions and body language while presenting a presentation has yet to be developed. Most of the research papers [1] [4] [5] [7] measure the emotions of Happy, Sad, Fear, Surprise, Anger, Neutral. There are only few research papers[8] focused on emotions like confidence and disappointment. But there is not any research paper that analyze the Nervousness. Nervousness is one of the main emotions that used by the nonprofessional users when doing a presentation. When reading more research papers, I found that most of the systems are worked as desktop applications not as mobile responsive web applications. Also, there are lack of research papers which provides an efficient feedback about emotions.

Features	EmoCo [1]	Research [9]	Research [10]	Research [8]	PRESENTLY
Happy	√	√	X	√	√
Sad	√	√	X	X	√
Anger	√	√	X	X	√
Surprise	√	√	X	√	√
Fear	√	√	X	X	√
Confidence	X	X	X	√	√
Neutral	√	X	X	√	√
Nervousness	X	X	X	X	√

Body Language	X	X	√	X	√
Providing Feedbacks	X	X	X	X	√
Mobile Responsive Application	X	X	X	X	√

Table 1 Research Comparison Table

"Presently" personal trainer application, which is proposed, is designed with many more functionalities than other researches which are currently prevailing. By the proposed solution, the "Presently" would be able to analyze the users' emotions and body language which used in uploaded video, to suggest suitable emotions and body language, to provide efficient feedbacks and rating to the user that he/she needs to use for a presentation.

1.4. Research Problem

Non-Professional speakers like academic students face many difficulties while preparing for a speech or a presentation. Not practicing enough, become nervous, avoiding eye contact, failing to engage emotionally are some of the main difficulties people had to face when doing a presentation. Therefore, practice in advance to the presentation helps to stable the emotions while doing the presentation. Getting feedback before doing the presentation will help the speaker to improve their presentation skills. Most speakers use manual methods to get feedback. There is no proper method to self-evaluate presentation performance. As for Figure 1-3, many university students have not used any existing systems to self-evaluate their presentation skills before the actual presentation.

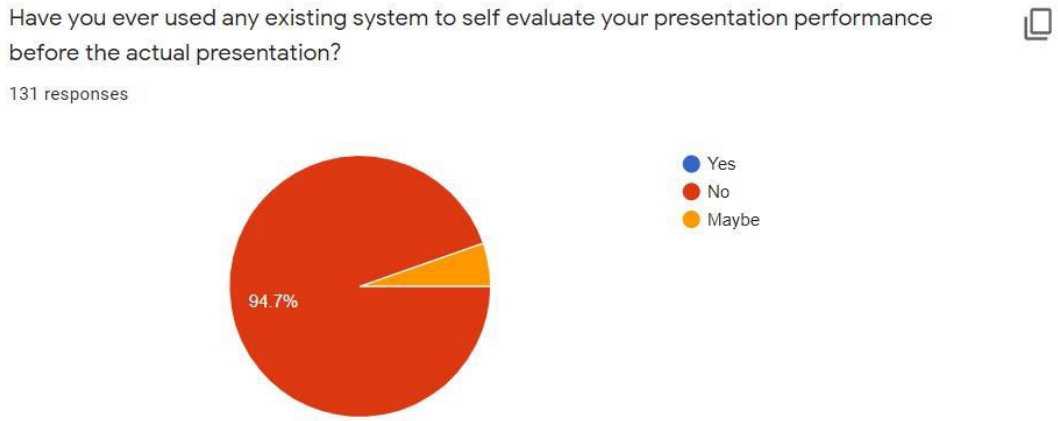

Figure 1-3 Summary of Responses about Existing system

For the problem that has not the proper method to pre-evaluate the presentation skills, this proposed solution "Presently" is expecting to provide feedback in advance to the users to improve their presentation skills.

2. OBJECTIVES

2.1 Main Objective

The main objective of developing “Presently”, a mobile responsive web application is to ensure that speakers can prepare in advance for the presentations to deliver a successful presentation to the audience. Presently will be able to monitor speakers’ presentation skills and provide feedback based on their emotion control, body language, pronunciation, tonality, and presentation slides quality.

2.2 Specific Objectives

To achieve the main objective, the specific objectives that need to be attained are in the proposed system as follows,

Specific Objective 01: To correctly extract the related emotions and body language postures from the uploaded video.

Specific Objective 02: To create a dataset that includes non-basic feelings like, nervousness, confidence, anxiety.

Specific Objective 03: To suggest user, what are the incorrect emotions, hand gestures that she/he used in the presentation.

Specific Objective 04: Identify the presenters’ emotions and body language using video analysis and suggest to the user what emotions and enhancements are used to present the speech.

3. METHODOLOGY

3.1 Project Overview

Targeting professional and non-professional speakers, the proposed system should be simpler and easier to use, and also it should not occupy the unnecessary amount of time and cost to learn use which is compatible with the Sri Lankan domain with the online situation. Therefore, when creating the system, user friendly interfaces must be added. Human-Computer Interaction, Machine Learning, Image and Text Classification, Natural Language processing are the most important applied areas of the proposed system. By using the proposed system "Presently" will help users to get self-evaluate their presentation skills and get feedback about their presentation slides separately. From the "Presently" users can analyze their presentation skills in a variety of areas including, incorrect pronunciation and vocabulary mistakes, match or mismatch between topic tone and prosody, emotions and body language used to present the story, quality of the presentation slides.

The proposed system "Presently" will guide any user to login to the system and create an account in the system. Then he/she can upload video or audio clip or presentation slides to the system. If user wants to upload either video only or audio only or presentation slides only, system will facilitate user to upload them separately. Using uploaded video clip system will recognize emotions and body language of the presenter and provide a feedback. From uploaded video, system will extract the audio file or from the uploaded audio file, system will check for the pronunciation mistakes, vocabulary errors, tonality, prosody of the presenter and provide a feedback for the presentation. System will facilitate to upload the presentation slides separately and it will check the quality of the presentation slides by image classification. Finally, the overall system will provide feedback and rating for the uploaded file. The system overview diagram, system overview, and resources and technologies described below will help in understanding the systematic, theoretical analysis of the methodologies used during "Presently" research field of study.

3.2 System Overview Diagram

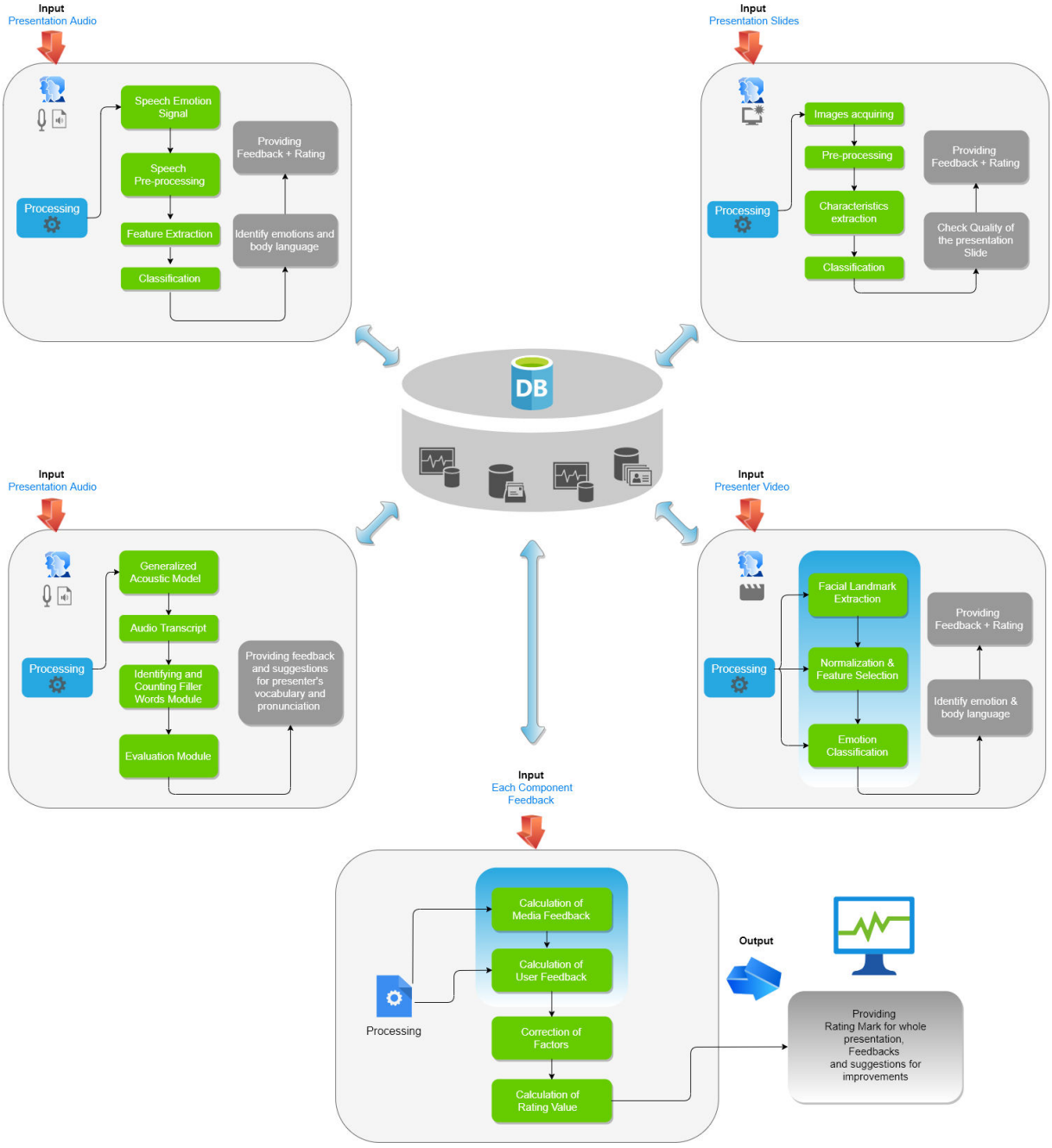

Figure 3-1 System Overview Diagram

3.3 Individual Component - Suggest the user what emotions and enhancements used to present the story using video analysis.

3.3.1 System Architecture Diagram

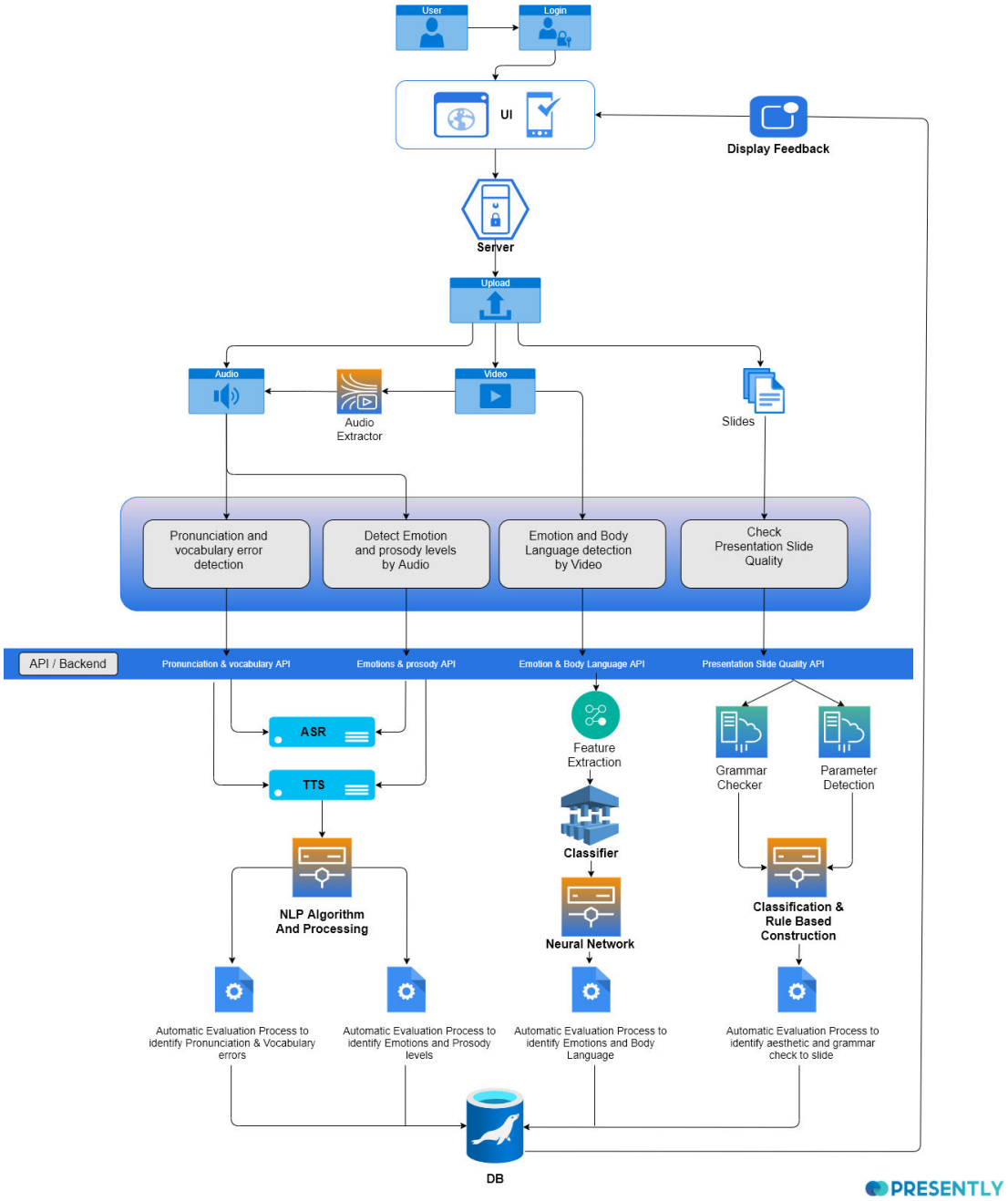

Figure 3-2 System Architectural Diagram

3.3.2 The flow of the System – Individual Component

Many people had to face some difficulties when doing a presentation or a speech in front of an audience. Not practicing enough, avoiding eye contact, failing to engage emotionally are some of the main difficulties that people had to face when doing a presentation. Therefore, practicing beforehand to the presentation helps to stable the emotions while doing the presentation.

The user can upload the video to the proposed mobile responsive web application. From the system, only the video clip will be extracted from the uploaded video. Then the system will convert the video clip into image frames and detect users' faces using Face Recognition System. Then the proposed system will identify presenters' emotions, the body language from that image frames, and analyze those emotions separately using video analysis and emotion analysis using Computer Vision. At the end of the process, the proposed system will give a rating and feedback to the user using Machine Learning Algorithms. This grade and feedback might assist speakers in gaining an understanding of their presentation abilities. Receiving comments on their speech can help them improve their ability to control their emotions while giving a presentation.

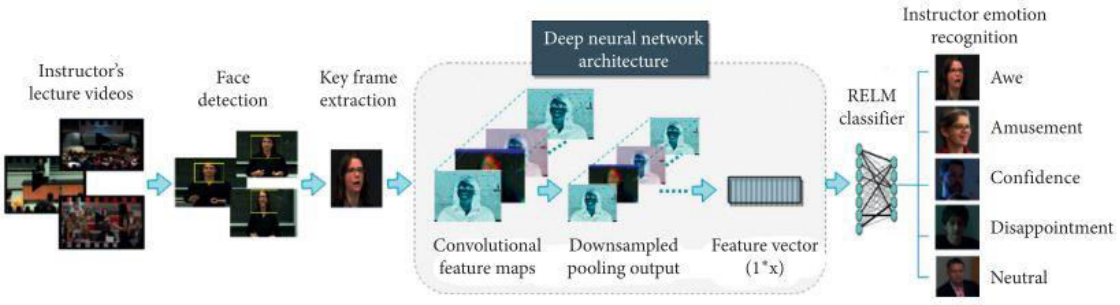

Figure 3-3 Process of Emotion Recognition System [4]

3.3.3 The flow of the system diagram – Individual Component

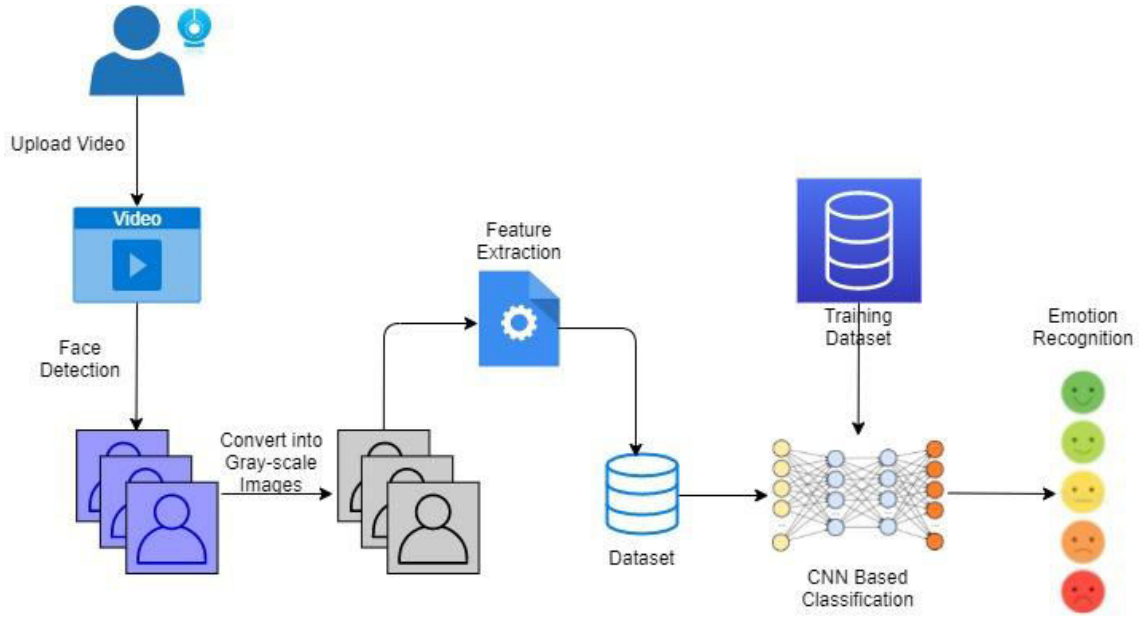

Figure 3-4 System Architectural Diagram - Individual Component

3.4 Tools and Technologies

 *Figure 3-5 Image of Python*	Python is an interpreted, high-level, general programming language. [10] Python is used to train the data set through machine learning. Machine learning is a type of significant intelligence that provides computers with the ability to learn without being explicitly programmed. Machine Learning focuses on the development of computer programs that can change when exposed new data. [11]
 *Figure 3-6 Image of Git*	Used for Version controlling. This will help developers to track and manage changes to a software project's code. GitHub is a website and cloud-based service that helps developers to store and manage their code. [12]
 *Figure 3-7 Image of Amazon S3 Bucket*	Amazon S3 bucket is a Storage service offered by Amazon Web Services. Amazon S3 offers object storage through web interface. Files, Images, Videos can store in a S3 bucket. [13]
 *Figure 3-8 Image of FastAPI*	FastAPI is a web framework for developing APIs with Python 3.6+ based on standard Python type hints that is modern and fast (high-performance). [14]
 *Figure 3-9 Image of OpenCV*	OpenCV supports a wide variety of programming languages such as C++, Python, Java, etc. It is used to detect faces in systems. [14]

Table 2 Tools and Technologies for the system

3.5 Gnatt Chart and Work Breakdown chart

Gnatt Chart

No	Assessment / Milestone	Start Date	End Date	2021-2022													
				April	May	June	July	August	September	October	November	December	January	February	March	April	May
1	**Project discussion workshop**	23-Apr-21	23-Apr-21														
2	**Topic evaluation**	15-May-21	30-Jul-21														
2a	Select a topic	15-May-21	20-May-20														
2b	Select a supervisor	20-May-21	23-May-21														
2c	Topic Evaluation form submission	23-May-21	25-Jun-21														
2d	Project charter submission	20-Jun-21	30-Jul-21														
3	**Project proposal report**	15-Jun-21	10-Aug-21														
3a	Create Project Proposal - individual	15-Jun-21	15-Jul-21														
3b	Create Project Proposal - group	15-Jul-21	06-Aug-21														
3c	Project proposal presentation	01-Aug-21	10-Aug-21														
4	**Develop the system**	06-Aug-21	20-Feb-22														
4a	Identifying functions	06-Aug-21	20-Aug-21														
4b	Database designing	20-Aug-21	12-Sep-21														
4c	Implementation	12-Sep-21	30-Dec-21														
4d	Unit testing	01-Jan-22	30-Jan-22														
4e	Integration testing	30-Jan-22	20-Feb-22														
5	**Progress Presentation - I**	01-Jan-22	06-Jan-22														
5a	Project Status document	01-Jan-22	06-Jan-22														
5b	Create presentation document	01-Jan-22	06-Jan-22														
5c	Progress Presentation – I (50%)	06-Jan-22	06-Jan-22														
6	**Research Paper**	18-Oct-21	18-Mar-22														
6a	Create the Research Paper	18-Oct-21	18-Mar-22														
7	**Progress Presentation - II**	22-Mar-22	29-Apr-22														
7a	Create presentation document	22-Mar-22	29-Apr-22														
7b	Progress presentation – II (90%)	29-Apr-22	29-Apr-22														
8	**Final Report Submission**	14-Apr-22	14-May-22														
8a	Final Report Submission	14-Apr-22	14-May-22														
8b	Application assessment	01-May-22	14-May-22														
8c	Project status document	14-May-22	14-May-22														
8d	Student logbook	14-May-22	14-May-22														
9	**Final Presentation & Viva**	14-Apr-22	25-May-22														
9a	Create final presentation	01-May-22	25-May-22														
9b	Final report submission	25-May-22	25-May-22														

Table 3 Gnatt Chart

Work Breakdown Chart

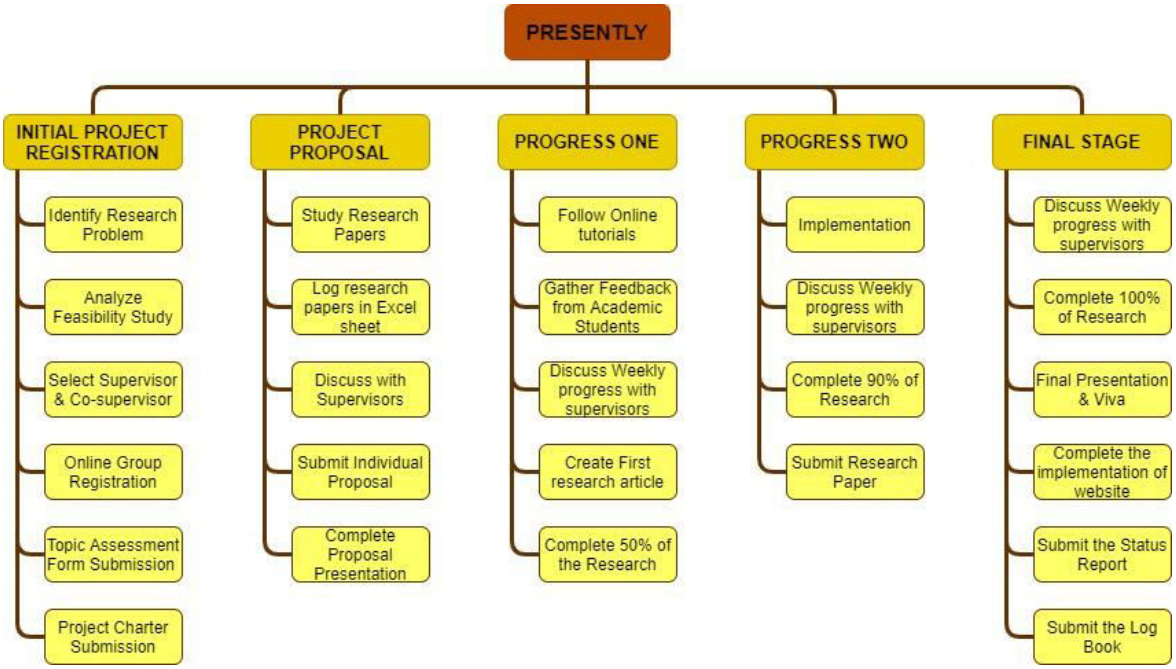

Figure 3-10 Work Breakdown Chart

3.6 Requirement Analysis

Functional Requirements

- User can create an account and should be able to login to the “Presently” using the account
- User should be able to upload the video to the system.
- System should be able to detect and recognize the presenters faces and extract the features from images
- System should be able to provide a feedback according to presenter’s video

Non-Functional Requirements

1) Security
    - Users should accept the Rules and Regulations provided by System.
    - System should allow the user to login to the system with valid Credentials.
    - System should not allow the access to view other users’ feedbacks.
2) Reliability
    - System should be mobile responsive.
3) Performance
    - Should be able to save all feedbacks of the user in user account.
    - System should be analyze the upload files in less time.
    - System loading time should be less.
4) Usability
    - System should be easily navigate.
    - System should be user friendly.

4. BUDGET AND BUDGET JUSTIFICATION

Component	Amount (LKR)
Internet	4000.00
Stationary	2000.00
Documentation and Printing Cost	3000.00
Server Cost	4000.00
Educational Survey Cost (Online Payments)	1000.00
Electricity	2000.00
Transport	1500.00
Total	**17500.00**

Table 4 Budget Analysis

5. REFERENCES

[1] H. Zeng *et al.*, "EmoCo: Visual analysis of emotion coherence in presentation videos," *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 1, pp. 927–937, 2020, doi: 10.1109/TVCG.2019.2934656.

[2] I. A. Adeyanju, E. O. Omidiora, and O. F. Oyedokun, "Performance evaluation of different support vector machine kernels for face emotion recognition," *IntelliSys 2015 - Proc. 2015 SAI Intell. Syst. Conf.*, pp. 804–806, 2015, doi: 10.1109/IntelliSys.2015.7361233.

[3] B. T. Nguyen, M. H. Trinh, T. V. Phan, and H. D. Nguyen, "An efficient real-Time emotion detection using camera and facial landmarks," *7th Int. Conf. Inf. Sci. Technol. ICIST 2017 - Proc.*, pp. 251–255, 2017, doi: 10.1109/ICIST.2017.7926765.

[4] Madhura Prakash, Aman Kumar, Anmol Saxena, and Somil Agrawal, Twinkle Singh, "Audience Feedback Analysis using Emotion Recognition," *Int. J. Eng. Res.*, vol. V9, no. 05, pp. 1090–1094, 2020, doi: 10.17577/ijertv9is050717.

[5] V. R. Guttha, H. K. Kondakindi, and V. Bhatti, "Automated feedback generation system using facial emotion recognition," *2018 3rd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. RTEICT 2018 - Proc.*, pp. 975–981, 2018, doi: 10.1109/RTEICT42901.2018.9012630.

[6] S. J. Goyal, A. K. Upadhyay, and R. S. Jadon, "Facial emotion recognition through hand gesture and its position surrounding the face," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 3, pp. 441–444, 2019.

[7] S. L. Happy, A. Dasgupta, P. Patnaik, and A. Routray, "Automated alertness and emotion detection for empathic feedback during E-learning," *Proc. - 2013 IEEE 5th Int. Conf. Technol. Educ. T4E 2013*, pp. 47–50, 2013, doi: 10.1109/T4E.2013.19.

[8] Y. K. Bhatti, A. Jamil, N. Nida, M. H. Yousaf, S. Viriri, and S. A. Velastin, "Facial Expression Recognition of Instructor Using Deep Features and Extreme Learning Machine," *Comput. Intell. Neurosci.*, vol. 2021, 2021, doi: 10.1155/2021/5570870.

[9] R. Colomo-Palacios, T. Samuelsen, and C. Casado-Lumbreras, "Emotions in software practice: Presentation vs. coding," *Proc. - 2019 IEEE/ACM 4th Int. Work. Emot. Aware. Softw. Eng. SEmotion 2019*, no. Section II, pp. 23–28, 2019, doi: 10.1109/SEmotion.2019.00012.

[10] G. Wei, L. Jian, and S. Mo, "Multimodal (Audio , Facial and Gesture) based Emotion Recognition challenge," pp. 57–60.

[10] "Python Introduction", [Online]. Available: https://www.w3schools.com/python/python_intro.asp

[11] GeeksForGeeks, "Introduction to Machine Learning using Python", [Online]. Available: https://www.geeksforgeeks.org/introduction-machine-learning-using-python/

[12] "A Beginner's Introduction to GitHub", [Online]. Available: https://www.atlassian.com/git

[13] Amazon S3 bucket Tutorial, "Python code samples for Amazon S3", [Online]. Available:

https://docs.aws.amazon.com/code-samples/latest/catalog/code-catalog-python-example_code-s3.html

[14] FastAPI Tutorial, "FastAPI", [Online]. Available: https://fastapi.tiangolo.com/

[15] OpenCV Tutorial, "FastAPI", [Online]. Available: https://docs.opencv.org/master/d0/de3/tutorial_py_intro.html