

Visvesvaraya Technological University, Belagavi – 590018



PROJECT REPORT
ON

CLOUD-BASED SMART MONITORING SYSTEM FOR BABY HEALTH AND SAFETY

Submitted in partial fulfillment for the award of degree of

BACHELOR OF ENGINEERING
in
COMPUTER SCIENCE & ENGINEERING

Submitted by

| | |
|-------------------|------------|
| Aaron Tauro | 4SO21CS002 |
| Abhik L Salian | 4SO21CS004 |
| Akhil Shetty M | 4SO21CS013 |
| H Karthik P Nayak | 4SO21CS058 |

Under the Guidance of

Dr Sridevi Saralaya
Professor, Department of CSE



DEPT. OF COMPUTER SCIENCE AND ENGINEERING
ST JOSEPH ENGINEERING COLLEGE

An Autonomous Institution

(Affiliated to VTU Belagavi, Recognized by AICTE, Accredited by NBA)

Vamanjoor, Mangaluru - 575028, Karnataka

2024-25

ST JOSEPH ENGINEERING COLLEGE

An Autonomous Institution

(Affiliated to VTU Belagavi, Recognized by AICTE, Accredited by NBA)

Vamanjoor, Mangaluru - 575028, Karnataka

DEPT. OF COMPUTER SCIENCE AND ENGINEERING



CERTIFICATE

Certified that the project work entitled "Cloud-Based Smart Monitoring System for Baby Health and Safety" carried out by

| | |
|--------------------------|-------------------|
| Aaron Tauro | 4SO21CS002 |
| Abhik L Salian | 4SO21CS004 |
| Akhil Shetty M | 4SO21CS013 |
| H Karthik P Nayak | 4SO21CS058 |

the bonafide students of VIII semester Computer Science & Engineering in partial fulfillment for the award of Bachelor of Engineering in Computer Science and Engineering of the Visvesvaraya Technological University, Belagavi during the year 2024-2025. It is certified that all corrections/suggestions indicated during Internal Assessment have been incorporated in the report. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the said degree.

Dr Sridevi Saralaya
Project Guide

Dr Sridevi Saralaya
HOD-CSE

Dr Rio D'Souza
Principal

External Viva:

Examiner's Name

Signature with Date

1.

.....

2.

.....

ST JOSEPH ENGINEERING COLLEGE

An Autonomous Institution

(Affiliated to VTU Belagavi, Recognized by AICTE, Accredited by NBA)

Vamanjoor, Mangaluru - 575028, Karnataka

DEPT. OF COMPUTER SCIENCE AND ENGINEERING



DECLARATION

We hereby declare that the entire work embodied in this Project Report titled “**Cloud-Based Smart Monitoring System for Baby Health and Safety**” has been carried out by us at St Joseph Engineering College, Mangaluru under the supervision of **Dr Sridevi Saralaya**, for the award of **Bachelor of Engineering in Computer Science & Engineering**. This report has not been submitted to this or any other University for the award of any other degree.

Aaron Tauro USN:4SO21CS002

Abhik L Salian USN:4SO21CS004

Akhil Shetty M USN:4SO21CS013

H Karthik P Nayak USN:4SO21CS058

Acknowledgement

We dedicate this page to acknowledge and thank those responsible for the shaping of the project. Without their guidance and help, the experience while constructing the dissertation would not have been so smooth and efficient.

We sincerely thank our Project guide **Dr Sridevi Saralaya**, Professor, Computer Science and Engineering for his guidance and valuable suggestions which helped us to complete this project. We also thank our Project coordinators **Ms Supriya Salian** and **Dr Saumya Y M**, Dept of CSE, for their consistent encouragement.

We owe a profound gratitude to **Dr Sridevi Saralaya**, Head of the Department, Computer Science and Engineering, whose kind support and guidance helped us to complete this work successfully

We are extremely thankful to our Principal, **Dr Rio D'Souza**, Director, **Rev. Fr Wilfred Prakash D'Souza**, and Assistant Director, **Rev. Fr Kenneth Rayner Crasta** for their support and encouragement.

We would like to thank all faculty and staff of the Department of Computer Science and Engineering who have always been with us extending their support, precious suggestions, guidance, and encouragement through the project.

We also extend our gratitude to our friends and family members for their continuous support.

Abstract

The need for reliable infant monitoring systems has grown due to the high demands of modern parenting and the importance of ensuring infant safety. This project presents a “Cloud-Based Smart Monitoring System for Baby Health and Safety,” which monitors key health metrics such as body temperature, heart rate, room temperature, humidity, and posture. By providing real-time notifications and alerts, the system offers parents peace of mind and enhances infant safety.

Recent advancements in non-contact health monitoring utilize technologies like remote photoplethysmography and computer vision for detecting health parameters. However, existing systems often lack comprehensive capabilities or rely on contact-based sensors that may cause discomfort to infants. This project overcomes these challenges by integrating contactless sensors and machine learning techniques, creating a holistic and user-friendly monitoring solution.

The methodology involves developing a mobile application that interacts with a cloud-based system and sensors to analyze infant health data in real time. The system employs computer vision algorithms to monitor baby posture and detect unsafe positions, such as tummy sleeping, potentially preventing sudden infant death syndrome (SIDS). Experimental results confirm the system’s reliability and accuracy under various environmental conditions, providing immediate alerts during abnormalities.

This work significantly enhances infant safety by reducing the need for constant parental monitoring while offering peace of mind. The system demonstrates a valuable contribution to infant health care by combining advanced technology with practical usability.

Table of Contents

| | |
|--|------------|
| Acknowledgement | i |
| Abstract | ii |
| Table of Contents | iii |
| List of Figures | vi |
| List of Tables | vii |
| 1 Introduction | 1 |
| 1.1 Background | 1 |
| 1.2 Problem statement | 2 |
| 1.3 Objectives | 2 |
| 1.4 Scope | 2 |
| 2 Literature Survey | 4 |
| 2.1 MyVox-Device for the Communication Between People: Blind, Deaf, Deaf-Blind and Unimpaired | 4 |
| 2.2 Mobile Lorm Glove - Introducing a Communication Device for Deaf-Blind People | 4 |
| 2.3 Tactile Board: A Multimodal Augmentative and Alternative Communication Device for Individuals with Deafblindness | 5 |
| 2.4 An Efficient Communication System for Blind, Dumb and Deaf People | 6 |
| 2.5 Multimodal Communication System for People Who Are Deaf or Have Low Vision | 7 |
| 2.6 A Communication System for Deaf and Dumb People . . | 8 |
| 2.7 On Improving GlovePi: Towards a Many-to-Many Communication Among Deaf-blind Users | 8 |
| 2.8 HaptiComm: A Touch-Mediated Communication Device for Deafblind Individuals | 9 |

| | | |
|----------|--|-----------|
| 2.9 | HaptiComm: A Touch-Mediated Communication Device for Deafblind Individuals | 10 |
| 2.10 | Proposed system | 10 |
| 2.11 | Comparison of existing methods / Summary | 12 |
| 3 | Software Requirements Specification | 14 |
| 3.1 | Functional requirements | 14 |
| 3.2 | Non-Functional requirements | 14 |
| 3.3 | User Interface Designs | 15 |
| 3.4 | Hardware and Software requirements | 15 |
| 3.4.1 | Hardware Requirements: | 15 |
| 3.4.2 | Software Requirements: | 16 |
| 3.5 | Performance Requirements | 16 |
| 3.6 | Design Constraints | 17 |
| 3.7 | Other Requirements | 18 |
| 4 | System Design | 19 |
| 4.1 | Architecture Design | 19 |
| 4.2 | Decomposition Description | 19 |
| 4.3 | Data Flow Design | 20 |
| 4.4 | Use case Diagram | 21 |
| 5 | Implementation | 23 |
| 5.1 | Audio Extraction | 23 |
| 5.2 | Speech Separation | 23 |
| 5.2.1 | Sepformer | 24 |
| 5.3 | Speech Enhancement | 24 |
| 5.3.1 | Lite Audio Visual Speech Enhancement | 24 |
| 5.3.2 | Spectral Subtraction | 25 |
| 5.4 | Speaker Detection | 26 |
| 6 | System Testing | 27 |
| 6.1 | Testing Objectives | 27 |
| 6.2 | Types of Testing conducted | 27 |
| 7 | Results and Discussion | 29 |
| 7.1 | Face detection | 29 |
| 7.2 | Speaker recognition | 29 |
| 8 | Conclusion and Future work | 34 |

List of Figures

| | | |
|-----|--|----|
| 4.1 | System Architecture Diagram | 19 |
| 4.2 | Flow chart | 20 |
| 4.3 | Dataflow design | 21 |
| 4.4 | Use case diagram for customer | 22 |
| 5.1 | code snippet for audio extraction | 23 |
| 5.2 | code snippet for speech separation | 24 |
| 5.3 | code snippet for speech enhancement using LAVSE | 25 |
| 5.4 | code snippet for speech enhancement using spectral subtraction | 26 |
| 5.5 | code snippet for speaker detection | 26 |
| 6.1 | testcases | 28 |
| 7.1 | Face detection | 29 |
| 7.2 | Speaker recognition 1,person 1 | 30 |
| 7.3 | Speaker recognition 1,person 2 | 30 |
| 7.4 | Speaker recognition 2,person 1 | 31 |
| 7.5 | Speaker recognition 2,person 2 | 31 |
| 7.6 | Speaker recognition 3,person 1 | 32 |
| 7.7 | Speaker recognition 3,person 2 | 32 |

List of Tables

| | | |
|-----|---|----|
| 2.1 | Comparison of Existing Projects | 13 |
| 6.1 | Work Flow | 28 |

Chapter 1

Introduction

1.1 Background

In contemporary society, communication serves as the cornerstone of human interaction, fostering connections, disseminating information, and enabling the exchange of ideas. However, for individuals confronting sensory impairments, particularly those who are both deaf and blind, these essential channels of communication become profoundly challenging to navigate. The deaf-blind community, often overlooked in technological advancements, faces formidable barriers to accessing and participating in text-based communication independently, thus exacerbating their isolation and limiting their ability to engage meaningfully with the world. Despite remarkable strides in technology that have revolutionized communication for many, the unique challenges faced by the deaf-blind community persist. Traditional modes of communication, such as sign language and tactile sign language, while invaluable, often require face-to-face interaction or the presence of an interpreter. These limitations become increasingly pronounced in our digital age, where accessibility to efficient and portable communication solutions is paramount. As a consequence, the deaf-blind community remains largely excluded from the benefits of the digital world, hindering their capacity to connect with others, access information, and engage in everyday conversations.

Recognizing this significant gap in accessibility, our project proposes the development of a Multimodal Communication System specifically tailored to the needs of individuals who are both deaf and blind. This transformative initiative aims to empower the deaf-blind community by providing them with a means to engage effectively in text-based conversations, bridging the communication gap and enhancing their ability to connect with the world. The project's genesis lies in a commitment to inclusivity and the belief that technological innovation can serve as a powerful equal-

izer. By addressing the communication challenges faced by the deaf-blind community, we aspire to foster a more inclusive society where everyone, regardless of their abilities, can actively participate in the digital age. The subsequent sections outline the scope, methodology, and feasibility considerations of our project, providing a comprehensive roadmap for the development of the work.

1.2 Problem statement

Video digest revolves around developing algorithms and techniques to automatically create concise and coherent summaries of longer videos while retaining the essential content and context of the original footage. The primary goal is to provide an efficient representation of the video, making it easier for users to understand the video's content quickly without having to watch the entire duration. In this project we develop a query based dynamic summarization tool using deep learning techniques.

1.3 Objectives

The aim/goal should be explained here.

Example: The aim of the proposed project work are:

1. To create a our own medicinal plant dataset
2. To develop a deep learning algorithm with high accuracy to identify the plant
3. To develop a web and mobile application to automatically identify the medicinal plant based on its image

1.4 Scope

The scope of query-based video synopsis is to create a condensed and representative version of a longer video that is specifically tailored to a user's query. This allows users to quickly and efficiently find relevant information within a large video dataset. Query-based video synopsis is important because it offers a solution to the problem of information overload in large video datasets. With the exponential growth of video data, it is becoming increasingly difficult for users to manually sift through large amounts of

video content to find the information they need. Query-based video synopsis provides an automated and efficient way to extract and present the most relevant parts of the video data, making it easier for users to access and use the information they need. Another important aspect of query-based video synopsis is its potential applications in various industries and domains. For example, in law enforcement, query-based video synopsis can help investigators quickly identify relevant video footage related to a specific crime or suspect. In education, it can be used to create condensed versions of lecture videos for students who need to review specific topics or concepts. In marketing, it can be used to identify and extract key moments in promotional videos that are likely to be of interest to customers. Overall, query-based video synopsis has significant scope and importance in enabling users to efficiently and effectively access relevant information from large video datasets, and it is likely to continue to play an increasingly important role in a wide range of industries and applications.

Chapter 2

Literature Survey

2.1 MyVox-Device for the Communication Between People: Blind, Deaf, Deaf-Blind and Unimpaired

The paper^{ref1} describes the development of the MyVox device, a communication system designed for individuals who are deaf-blind. The device is powered by an ARM-based computer, specifically the Raspberry Pi, and includes a USB keyboard for input, a speaker for speech synthesis, a braille display, a vibration motor for notifications, and a real-time clock for time perception.

The inputs of the device include a USB keyboard for text input, which can be customized to suit the user's needs, and the system also incorporates a real-time clock for time perception. The outputs of the device consist of an LCD display for text output, a speaker for speech synthesis, a braille refreshable display for tactile output, and a vibration motor for notifications. These components enable the device to provide communication capabilities for individuals who are deaf-blind, catering to their specific sensory needs and facilitating interaction with others.

The paper concludes by discussing future work, including the potential for internet access and custom applications, as well as plans to make the device available to more people in need. Overall, the MyVox device represents an important step forward in addressing the communication challenges faced by deaf-blind individuals and promoting social inclusion.

2.2 Mobile Lorm Glove - Introducing a Communication Device for Deaf-Blind People

The paper^{ref2} introduces the Mobile Lorm Glove, a communication device designed to support deaf-blind people's communication and enhance their independence. The hardware prototype of the Mobile Lorm Glove includes

fabric pressure sensors, vibrating motors, an ATmega328 microcontroller, shift registers, darlington transistor arrays, and a Bluetooth module for data transmission. The actuators on the glove are placed at varied distances to adjust stimulus duration, catering to individual tactile sensitivity and lorming speed. This customization allows for the adjustment of the maximal applied intensity and lorming speed to serve the user's needs.

The application scenarios for the Mobile Lorm Glove include mobile communication over distance and simultaneous translation. The glove enables the composition of text messages and their transmission to a receiver's handheld device, where the message can be read directly or translated into the Lorm alphabet using the Mobile Lorm Glove. Additionally, the glove functions as a simultaneous translator, allowing communication with individuals who are not familiar with Lorm, thus broadening the spectrum of people with whom deaf-blind individuals can engage. The device also enables parallel one-to-many communication, which can be particularly helpful in educational and learning contexts.

The Mobile Lorm Glove's input unit consists of fabric pressure sensors on the palm and a rectangular sensor on the wrist, which detect tactile input from the user's hand. These sensors transmit data to a microcontroller using a matrix design. The output unit comprises vibrating motors on the back of the glove, which provide haptic feedback based on the input received. The control unit, integrated with the microcontroller and a Bluetooth module, manages the data transmission between the glove and the user's handheld device. The user composes text messages by lorming onto their own left hand with the glove, and the data is transmitted to the handheld device for further processing or translation.

2.3 Tactile Board: A Multimodal Augmentative and Alternative Communication Device for Individuals with Deafblindness

The Tactile Board, a mobile Augmentative and Alternative Communication (AAC) device, is introduced as a groundbreaking solution to communication challenges faced by individuals with deafblindness^{ref3}. Characterized by dual sensory loss, deafblind individuals encounter difficulties compensating for impaired sight and hearing. The Tactile Board translates text and speech into real-time vibrotactile signs displayed through a haptic wearable, enabling deafblind individuals to communicate without direct assistance.

Based on interviews with 60 individuals with deafblindness across Europe, the Tactile Board features a 4-by-4 haptic matrix, customizable vocabulary database, and a haptic vest with coin-vibration motors. This facilitates two-way communication, addressing barriers to social interactions within the deafblind community. The paper underscores the limitations of existing technologies in meeting the unique challenges posed by dual sensory impairment, positioning the Tactile Board as a promising advancement in assistive technology.

The device's potential applications include communication with strangers, conveying environmental information, and supporting interactions in scenarios where direct touch is impractical, particularly relevant during the COVID-19 pandemic. The paper envisions future evaluations to assess the Tactile Board's impact on the confidence of individuals with deafblindness in initiating social interactions, contributing significantly to assistive technology for the deafblind community. The Tactile Board employs a Samsung Galaxy Tab S2 tablet, Android OS, and Google's NLP API for speech recognition. RealTime Framework facilitates communication, while a Raspberry Pi and Python script translate data to on-off commands for coin-vibration motors in a haptic vest. The system incorporates 3D-printed cases and tactile tablet covers for accessibility.

2.4 An Efficient Communication System for Blind, Dumb and Deaf People

The paper^{ref4} outlines a project aimed at enhancing the lives of the blind, deaf, and dumb population in India, which comprises around 70 million people. The proposed system focuses on facilitating communication through sign language recognition, utilizing hand gestures for interaction between humans and computers. The system involves recording a user's voice, converting it to text on a server, classifying the text, generating corresponding signs, and transmitting them to applications for deaf or dumb individuals. Additionally, a reverse system for sign-to-speech communication is envisioned for visually impaired people. The background study references related work in object modeling, sign language character recognition, and American Sign Language translation through image processing. Technologies to be used include Blob Detection, Template Matching, and Skin Color Detection. The proposed system's characteristics include video initiation, hand sign recognition, a graphical user interface (GUI), and the ability to operate windows based on recognized signs. The system archi-

ture involves a server for text conversion and sign generation, promoting efficient communication for the visually and hearing impaired. The conclusion emphasizes the system's potential to bridge communication gaps through image processing, enabling the recognition, storage, and use of sign language for various computer operations.

The proposed system for enhancing communication among the blind, deaf, and dumb in India utilizes image processing techniques. Technologies include Blob Detection, Template Matching, and Skin Color Detection. The system incorporates algorithms for recognizing hand signs, translating speech to text, and generating corresponding signs for improved interaction through sign language.

2.5 Multimodal Communication System for People Who Are Deaf or Have Low Vision

The paper^{ref5} addresses communication challenges confronting individuals with deafness or low vision, particularly concerning standard consumer products. Profoundly deaf individuals predominantly rely on text and graphics, lacking access to verbal communication channels like radio and television. The elderly, commonly affected by visual impairment, further complicates communication, and additional disabilities can exacerbate these challenges. The research aims to develop an inexpensive real-time transformation method for verbal messages to assist the profoundly deaf and visually impaired, emphasizing the necessity for a solution with minimal visual resource requirements.

The proposed system involves transforming textual information into visual color patterns, utilizing color tagging, Morse code, and the Phonetic Alphabet for temporal coding. The design incorporates light sources, such as LEDs, with brightness modulation for improved text visualization. The limitations of Morse code for real-time communication are discussed, prompting the suggestion of a single LED coupled with glasses to enhance direct viewing. This approach aims to improve text translation and perception for the deaf and visually impaired compared to traditional codes.

In conclusion, the novel light code variant, offering enhanced dynamic perception, holds potential benefits for the deaf, visually impaired, and as a peripheral display for wearable computer applications. The proposed approach demonstrates promise in advancing communication and visual perception for diverse user groups.

2.6 A Communication System for Deaf and Dumb People

The paper **ref6** addresses communication challenges for deaf and mute individuals, highlighting speech's significance. Sign language, their primary form of communication, limits interaction with the outer world. Technology, particularly hand gesture recognition within human-computer interaction (HCI), is identified as a potential solution. An "artificial speaking mouth" concept is introduced, converting captured hand gestures into speech. This innovative technique could empower mute individuals to convey thoughts naturally. The literature survey underscores the communication disadvantage faced by mute individuals compared to blind counterparts using traditional language. A proposed solution involves dumb communication interpreters, translating hand gestures into speech. Flex Sensors in a digital glove facilitate sign language translation into speech, fostering communication between mute communities and the public.

The paper introduces a real-time hand gesture recognition system for HCI, utilizing the Kinect sensor. It outlines three modules: real-time hand tracking, training gestures, and gesture recognition using hidden Markov models. Challenges in recognizing small, complex hand articulations are acknowledged, proposing Kinect for improved human-computer interaction. The proposed system aims to bridge communication gaps for speech and hearing-impaired individuals, focusing on offline recognition through four processes: Image Acquisition, Preprocessing, Feature Extraction, and Classification. Proper segmentation, feature extraction, and classification are highlighted for successful recognition, reducing data dimensionality. The procedure, implemented and tested with 26 hand gesture images, plays gesture audio files upon recognition, enabling two-way communication and enhancing speech-hearing impaired individuals' communication capabilities.

2.7 On Improving GlovePi: Towards a Many-to-Many Communication Among Deaf-blind Users

The paper **ref7** introduces an enhanced version of GlovePi, a low-cost wearable device designed to facilitate communication for deaf-blind individuals. The authors emphasize the significance of assistive technologies in promoting the inclusion, integration, and independence of people with disabilities, particularly those with deaf-blindness. The proposed extension

of GlovePi's architecture aims to enable many-to-many communication, addressing the need for enhanced social interaction and daily activities for deaf-blind users. By focusing on improving communication capabilities, the authors aim to enhance the overall quality of life for deaf-blind individuals. The proposed enhancements in GlovePi demonstrate a commitment to leveraging technology to address the unique communication challenges faced by individuals with deaf-blindness, ultimately contributing to their social inclusion and well-being.

2.8 HaptiComm: A Touch-Mediated Communication Device for Deafblind Individuals

The paper^{ref8} presents HaptiComm, a touch-mediated communication device designed to facilitate communication for Deafblind individuals. The device uses an array of electrodynamic actuators to reproduce tactile sensations of fingerspelling, allowing for communication through touch. The paper describes the design and implementation of the device, including the actuator guidance system, which was developed to cancel magnetic interference between close actuators and keep the intended skin contacts. The study evaluated the device's ability to reproduce tactile sensations of fingerspelling and the participants' ability to recognize the type and number of activated actuators. The results showed that the device was able to reproduce three of the five contact types of fingerspelling, and that participants were able to accurately recognize the type and number of activated actuators. The authors suggest that HaptiComm has the potential to improve communication for Deafblind individuals and that further investigations are needed to explore its full potential. They plan to refine the timing and speed parameters of actuation and estimate the device's individual and sequenced letter recognition rate compared to human fingerspelling. They also plan to quantify the learning curve of the device, which is an essential element in the adoption of assistive technology.

Overall, the paper presents an innovative device that has the potential to improve communication for Deafblind individuals. The study provides valuable insights into the development and evaluation of HaptiComm, highlighting its strengths and limitations and suggesting avenues for future research.

2.9 HaptiComm: A Touch-Mediated Communication Device for Deafblind Individuals

The paper^{ref8} presents HaptiComm, a touch-mediated communication device designed to facilitate communication for Deafblind individuals. The device uses an array of electrodynamic actuators to reproduce tactile sensations of fingerspelling, allowing for communication through touch. The paper describes the design and implementation of the device, including the actuator guidance system, which was developed to cancel magnetic interference between close actuators and keep the intended skin contacts. The study evaluated the device's ability to reproduce tactile sensations of fingerspelling and the participants' ability to recognize the type and number of activated actuators. The results showed that the device was able to reproduce three of the five contact types of fingerspelling, and that participants were able to accurately recognize the type and number of activated actuators. The authors suggest that HaptiComm has the potential to improve communication for Deafblind individuals and that further investigations are needed to explore its full potential. They plan to refine the timing and speed parameters of actuation and estimate the device's individual and sequenced letter recognition rate compared to human fingerspelling. They also plan to quantify the learning curve of the device, which is an essential element in the adoption of assistive technology.

Overall, the paper presents an innovative device that has the potential to improve communication for Deafblind individuals. The study provides valuable insights into the development and evaluation of HaptiComm, highlighting its strengths and limitations and suggesting avenues for future research.

2.10 Proposed system

PLEASE NOTE: REMOVE THE QUESTION AND WRITE ONLY THE ANSWERS

Why the Chosen Problem/Project is Important:

The proposed Bidirectional Communication System for Deaf-Blind Individuals is of paramount importance as it addresses a critical and often overlooked issue, aiming to break down communication barriers and foster inclusivity. By providing a tailored solution for individuals who are both deaf and blind, the project empowers this community to actively engage in text-based communication, educational activities, and personal inter-

actions. The system's integration in educational settings ensures equal opportunities for deaf-blind students, enabling their participation in classroom discussions and access to digital educational materials. Beyond education, the project promotes social connection by facilitating text-based communication through various channels. Moreover, it contributes to technological advancements for accessibility, reflecting a human-centric approach that recognizes and responds to the specific needs of the deaf-blind community. In essence, the project embodies the ethical responsibility of the technological community to create inclusive solutions, ultimately improving the quality of life for individuals facing unique challenges.

What is Novel in Proposed Project Work:

The proposed project distinguishes itself through a novel integration of speech-to-text conversion and braille representation, providing a multi-modal communication system for deaf-blind individuals. Notably, it addresses the educational context, empowering deaf-blind students to actively engage in classroom activities and digital learning. The inclusion of real-time braille hardware interaction, an intuitive user interface designed for the unique needs of deaf-blind users, and human-centric usability testing further set this project apart. By combining these elements, the project contributes to inclusive technological innovation, emphasizing accessibility and usability to enhance the communication experience for the deaf-blind community.

How it Would Advance the State-of-the-Art:

The proposed project would advance the state-of-the-art by pioneering a multimodal communication system for deaf-blind individuals, seamlessly integrating speech-to-text conversion with real-time braille hardware interaction. Its focus on educational integration and an intuitive user interface tailored to the unique needs of deaf-blind users represents a groundbreaking contribution to inclusive education and user-centered design. The project's emphasis on human-centric usability testing ensures that technological advancements align with the practical needs and preferences of the deaf-blind community, setting a new standard in assistive technology for those with combined auditory and visual impairments. Overall, the project's comprehensive approach positions it at the forefront of innovative and inclusive communication solutions, advancing the current state-of-the-art in assistive technology.

How it Differs from Existing Works:

Unlike existing solutions that often focus on singular aspects of communication for the deaf-blind, the proposed project innovates through its multimodal approach, seamlessly combining speech-to-text conversion with real-time interaction with dedicated braille hardware. Its unique emphasis on educational integration, user interface tailored for the deaf-blind, and human-centric usability testing further distinguish it. The project's comprehensive communication solution, covering various facets of accessibility, sets it apart from current works, marking a pioneering advancement in assistive technology for individuals with combined auditory and visual impairments..

2.11 Comparison of existing methods / Summary

In Table 2.1, we've laid out a comparison of different research studies. We're taking a close look at how each study has approached its work, what methods they've used to put their ideas into action, the conclusions they've come to, and the results they've achieved. We're also noting down the challenges each study has faced.

Table 2.1: Comparison of Existing Projects

| Project Title | Problem Addressed | Methodology | Implementation and Results | Inference and Results | Limitation/Future Scope |
|--|--|--|---|--|---|
| Mobile Lorm Glove-Introducing a Communication Device for Deaf-Blind People (February 2012) | Communication challenges for deaf-blind individuals | Uses fabric pressure sensors, vibrating motors, and a Bluetooth module for communication | Enables mobile communication, simultaneous translation, and one-to-many communication | Enhances independence and communication for deaf-blind individuals | Thickness of the glove |
| Tactile Board: A Multimodal Augmentative and Alternative Communication Device for Individuals with Deafblindness (November 2020) | Communication challenges for individuals with deafblindness using a mobile AAC device | Utilizes a 4-by-4 haptic matrix, customizable vocabulary database, and a haptic vest | Employs Samsung Galaxy Tab S2, Android OS, Google's NLP API, Raspberry Pi, and Python script | Potential applications include communication with strangers and conveying environmental information. | Future evaluations are envisioned, especially during the COVID-19 pandemic |
| Multimodal Communication System for People Who Are Deaf or Have Low Vision (January 2002) | Communication challenges for individuals with deafness or low vision | Involves real-time transformation of verbal messages into visual color patterns | Uses LEDs with brightness modulation for improved text visualization. | Shows promise for real-time communication for individuals with hearing and vision impairments | Acknowledges limitations of Morse code and proposes a novel light code variant. |
| On Improving GlovePi: Towards a Many-to-Many Communication Among Deaf-blind Users (January 2018) | Communication challenges for deaf-blind individuals, emphasizing many-to-many communication | Enhanced version of GlovePi with sensors, Raspberry Pi, mobile devices, and a tuple center | Focuses on improving communication capabilities for enhanced social interaction | Aims to contribute to the social inclusion and well-being of deaf-blind individuals | Future work involves integrating output sensors for tactile feedback. |
| MyVox-Device for the Communication Between People: Blind, Deaf, Deaf-Blind and Unimpaired (October 2014) | Developed for individuals who are deaf-blind, addressing their communication challenges | Powered by Raspberry Pi, includes USB keyboard, speaker, braille display, vibration motor, and real-time clock | Provides customized inputs and outputs for text, speech, and tactile communication | Represents an important step in addressing the communication challenges faced by deaf-blind individuals | Future work involves internet access, custom applications, and broader availability |
| HaptiComm: A Touch-Mediated Communication Device for Deafblind Individuals (April 2023) | Communication challenges for Deafblind individuals through touch-mediated communication using electrodynamic actuators | Utilizes an array of electrodynamic actuators to reproduce tactile sensations of fingerspelling, with a focus on canceling magnetic interference and addressing shaking and vibrations | Successfully reproduces three of the five contact types of fingerspelling, participants accurately recognize the type and number of activated actuators | Further investigations are needed to explore its full potential, including refining timing and speed parameters and estimating letter recognition rates compared to human fingerspelling | Acknowledges susceptibility to shaking and vibrations, plans to refine actuation parameters, estimate letter recognition rates, and quantify the learning curve |

Chapter 3

Software Requirements Specification

3.1 Functional requirements

Speech-to-Text Conversion:

Integrate robust speech recognition tools or APIs, such as Google Cloud Speech-to-Text or Python's SpeechRecognition library. Capture and transcribe spoken words into written text. Ensure high accuracy in speech-to-text conversion to facilitate precise communication.

Text-to-Braille Conversion:

Develop a sophisticated algorithm capable of translating transcribed text into Braille characters. Support different Braille standards and languages. Efficiency to minimize processing time for text-to-Braille conversion.

Braille Hardware Integration:

Integrate with Braille hardware systems i.e. device containing the sensor and actuators. Enable real-time sensory updates for seamless interaction.

User Interface Development:

Design an intuitive user interface for easy interaction. Facilitate smooth communication between the application and Braille hardware.

Hardware Interaction:

Develop a system that interfaces seamlessly with the chosen Braille hardware. Ensure the application can send Braille characters to the hardware for physical representation.

3.2 Non-Functional requirements

Security Measures:

Implement robust security protocols to ensure user data privacy and secure communication.

Usability Testing:

Conduct extensive usability testing with deaf-blind users to evaluate sys-

tem functionality. Gather feedback for continual improvement.

Accessibility Standards Compliance:

Ensure compliance with accessibility standards to cater to the specific needs of the deaf-blind community. Test and enhance the application's compatibility with different screen reader software.

Language and Braille Standards Support:

Support Braille standards to enhance versatility and stay updated with the latest Braille standards and ensure compatibility

3.3 User Interface Designs

Intuitive Design:

Simple Navigation: Design a straightforward navigation system that is easy for deaf-blind users to comprehend. Utilize clear and concise menu structures to facilitate intuitive interaction.

Accessibility Features:

Tactile Feedback Options: Integrate tactile feedback options within the user interface to enhance the user experience for deaf-blind individuals. Provide customizable settings for feedback intensity and type.

Design specifications:

Maintain a consistent design language across the website and application to provide a unified user experience. Design the website to be responsive across different devices, ensuring accessibility on desktops, tablets, and smartphones.

3.4 Hardware and Software requirements

3.4.1 Hardware Requirements:

Sensors for Braille Input:

Deploy sensors capable of detecting Braille characters either through touch or proximity sensors. Ensure the sensors are responsive to user input for a seamless interaction experience.

Actuators for Tactile Feedback:

Integrate actuators to provide tactile feedback corresponding to the Braille characters displayed. Design the actuators to deliver precise and distinguishable tactile sensations for each Braille character.

Braille Symbol Actuator/Sensor:

Employ a hardware system as the primary hardware interface. Ensure the

device can dynamically represent different Braille characters based on user inputs and can sense.

3.4.2 Software Requirements:

Speech-to-Text Conversion Software:

Utilize reliable speech recognition tools such as APIs, PyAudio, SpeechRecognition, and librosa by Python library for accurate conversion of spoken words into written text. Select a technology stack that supports real-time speech-to-text conversion.

Text-to-Braille Conversion Algorithm:

Develop a robust algorithm for translating the transcribed text into Braille characters using Python. Ensure the algorithm supports various Braille standards and languages.

Braille to Hardware Translation Software:

Implement software to translate the Braille characters into signals that can be understood by the hardware. Developing a communication protocol using a hardware device that can be used by sensing for seamless interaction between the software and hardware components.

Website or Application:

Create a user-friendly website or application interface for text-based communication. Include features for speech-to-text conversion, text-to-Braille conversion, and seamless interaction with the hardware.

Operating System Compatibility:

Ensure compatibility with major operating systems, such as Android and iOS, for mobile applications. For websites, ensure compatibility across different web browsers.

Integration with ROS (Robot Operating System):

Implement the necessary software components to integrate with ROS. Ensure smooth communication between different software modules.

3.5 Performance Requirements

Real-time Speech-to-Text Conversion:

Achieve near-instantaneous speech-to-text conversion. Evaluate system response time for spoken words to text. Ensure real-time transcription for effective communication.

Efficient Text-to-Braille Translation:

Swift translation of text to Braille characters. Assess the speed of the text-

to-Braille conversion algorithm. Minimize delays in Braille representation.

Seamless Hardware Interaction:

Establish real-time communication with Braille hardware. Monitor time for Braille characters to be transmitted and displayed. Achieve responsive updates on the Braille hardware.

Scalability:

Ensure optimal performance with increased user interactions. Evaluate system performance under varying loads. Maintain optimal performance with a growing user base and data load.

Resource Utilization:

Optimize resource usage for efficient operation. Assess CPU, memory, and network utilization. Ensure resource-efficient operation on diverse devices.

Error Handling:

Implement effective error-handling mechanisms. Evaluate the system's ability to manage errors. Gracefully handle errors to minimize disruption.

Usability Testing:

Conduct usability testing based on user feedback. Gather feedback on system responsiveness and ease of use. Regular testing and iterative improvements for user satisfaction.

3.6 Design Constraints

Portability:

The system must be designed for portability, considering use across different devices. Optimize the user interface and functionalities for seamless operation on various platforms, including mobile devices and desktop computers.

Device Compatibility:

Ensure compatibility with a variety of devices commonly used by the deaf-blind community. Design the system to adapt to different screen sizes, resolutions, and hardware configurations for widespread accessibility.

Real-time Communication:

Address the need for real-time communication between the application and hardware. Optimize data transmission and processing to minimize latency, providing users with immediate updates on the Braille hardware.

Usability for Deaf-Blind Users:

Prioritize usability for individuals with dual sensory impairments. Conduct usability testing with the deaf-blind community, incorporating their feedback to optimize the system's accessibility and ease of use.

3.7 Other Requirements

Long-term Support Plans:

Develop strategies for long-term system support and updates. Establish a framework for ongoing maintenance, addressing evolving technological standards and user needs.

Training Programs:

Provide comprehensive training programs for users, educators, and support staff. Design training materials and sessions to ensure effective usage and support, promoting accessibility and user empowerment.

Chapter 4

System Design

paragraph contents...

4.1 Architecture Design



Figure 4.1: System Architecture Diagram

This Figure 4.1 illustrates a high-level overview of the audio visual speech separation system. It is important to note that the specific techniques, algorithms, and models used in each component can vary depending on the implementation approach and the requirements of the system.

4.2 Decomposition Description



Figure 4.2: Flow chart

Figure 4.2 represent the flow chart of the proposed system. In audio visual speech separation, the goal is to decompose an audio signal containing multiple overlapping speakers into individual speech signals corresponding to each speaker. The decomposition process involves separating the desired speech signals from the background noise and other interfering sounds.

4.3 Data Flow Design

The audio input undergoes pre-processing, while the visual input is processed to extract relevant cues. The pre-processed audio and processed visual data are then integrated. From the integrated representation, features are extracted. These features are utilized in the speech separation stage, where individual speech signals are separated from the mixture. Post-processing techniques are applied to enhance the quality of the separated speech signals. Finally, the individual speech signals are outputted as the result of the system. The data flow design ensures a sequential flow of operations, starting from capturing and processing the inputs, integrating the audio-visual information, extracting features, performing speech separation, applying post-processing, and generating the output. This design allows for effective processing and separation of audio visual data to obtain distinct speech signals from overlapping speakers.

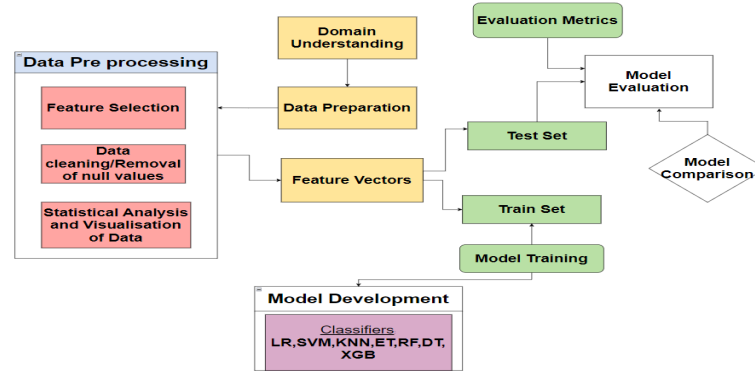


Figure 4.3: Dataflow design

4.4 Use case Diagram

use cases represent the main functionalities and tasks involved in the audio visual speech separation system. Each use case contributes to the overall process of capturing, processing, integrating, separating, and post-processing the audio and visual data to achieve the desired outcome of individual speech signal separation.

Pre-process Audio: This use case involves pre-processing the captured audio data. It may include operations like filtering, noise reduction, and echo cancellation to improve the quality of the audio signals.

Process Visual: This use case involves processing the captured visual data. It includes tasks such as face detection, facial landmark tracking, or lip motion analysis to extract relevant visual cues associated with speech production.

Integrate Audio-Visual: This use case represents the integration of the pre-processed audio data and processed visual data to create a synchronized audio-visual representation, aligning the audio and visual streams.

Extract Features: This use case involves extracting relevant features from the integrated audio-visual representation. It may include computing spectrograms, MFCCs, facial landmarks, or other visual and audio features.

Perform Speech Separation: This use case focuses on the actual speech separation process. It utilizes the extracted audio and visual features to separate the individual speech signals from the mixture, using techniques such as blind source separation or deep learning-based models.

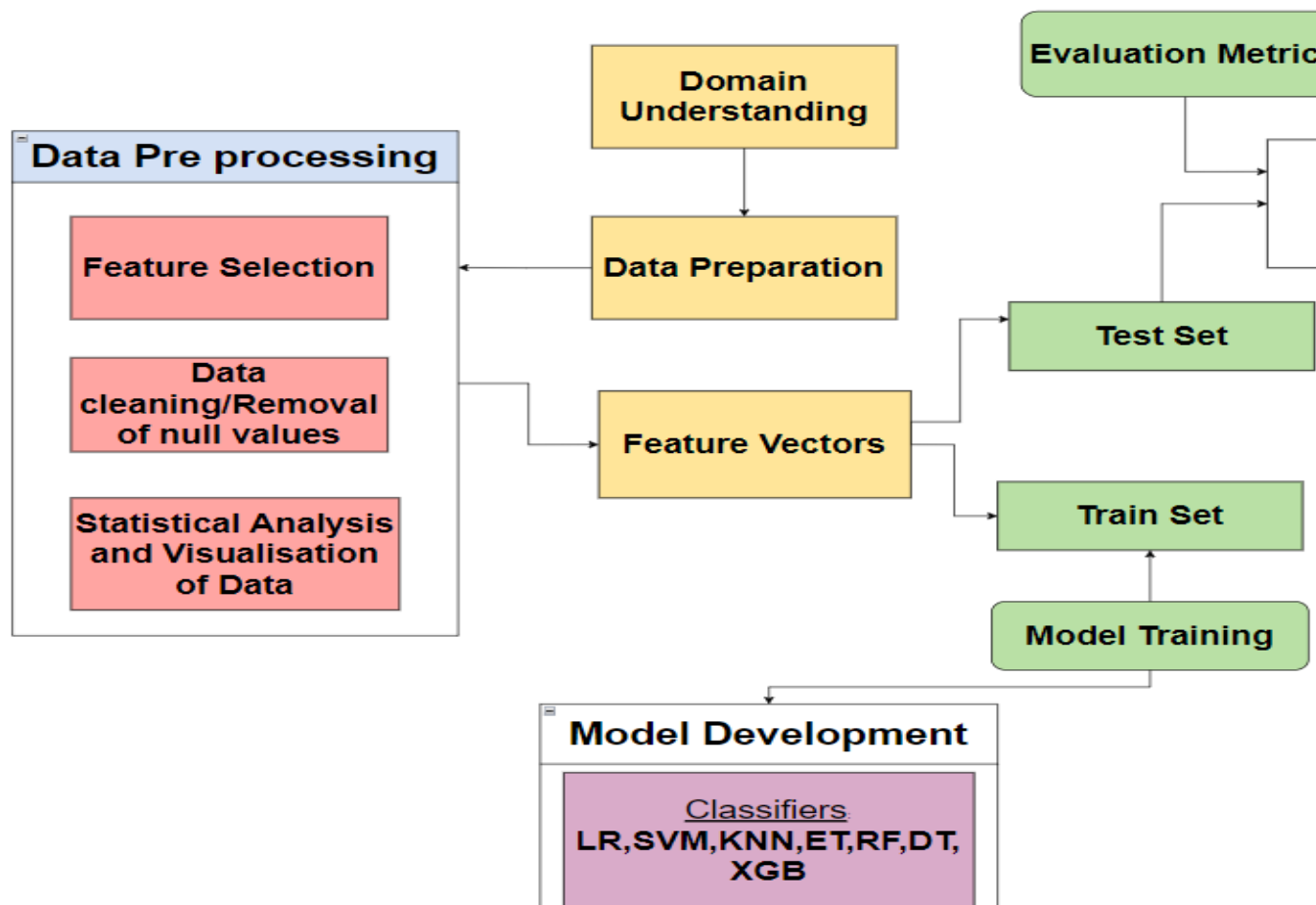


Figure 4.4: Use case diagram for customer

Chapter 5

Implementation

5.1 Audio Extraction

Audio extraction is the process of isolating and extracting the audio content from a multimedia source, such as a video file. It involves separating the audio track from the accompanying video or other elements to obtain a standalone audio file representing the sound present in the source material.



Figure 5.1: code snippet for audio extraction

5.2 Speech Separation

SpeechBrain is an open-source framework

5.2.1 Sepformer

SepFormer is an algorithm for speech separation that utilizes self-attention mechanisms. It employs a transformer-based architecture to capture long-range dependencies and model the relationships between time-frequency points in the audio mixture, enabling the separation of multiple speech sources from the mixture.



Figure 5.2: code snippet for speech separation

5.3 Speech Enhancement

5.3.1 Lite Audio Visual Speech Enhancement

Lite AVSE algorithm is used for the separation and enhancement of the speech. The system includes two visual data compression techniques and removes the visual feature extraction network from the training model, yielding better online computation efficiency. As for the audio features, short-time Fourier transform (STFT) is calculated of 3-second audio segments. Each time-frequency (TF) bin contains the real and imaginary parts of a complex number, both of which used as input. Power-law compression used to prevent loud audio from overwhelming soft audio. The same processing is applied to both the noisy signal and the clean reference signal.

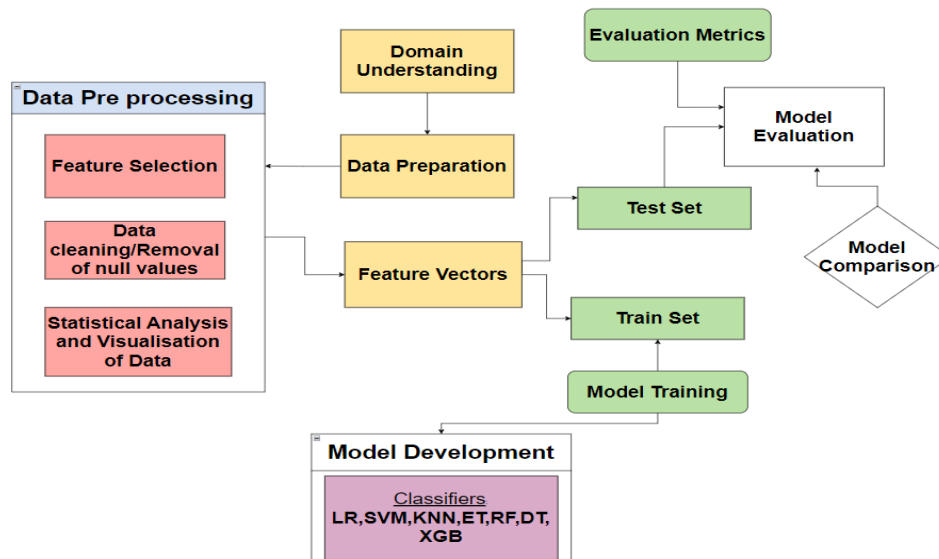


Figure 5.3: code snippet for speech enhancement using LAVSE

5.3.2 Spectral Subtraction

Spectral subtraction is a technique used in audio signal processing to reduce background noise from an audio signal. It involves estimating the noise spectrum from a noisy signal and subtracting it from the noisy spectrum to enhance the desired signal. The resulting spectrum is then transformed back into the time domain to obtain a cleaner audio signal as in Figure 5.4



Figure 5.4: code snippet for speech enhancement using spectral subtraction

5.4 Speaker Detection

The cv2 functions provide methods to load the pre-trained models, apply them to images or video frames, and draw bounding boxes around the detected faces. By leveraging cv2's face detection capabilities, you can automate tasks such as facial recognition, emotion analysis, or face tracking in various applications like surveillance, biometrics, or augmented reality.

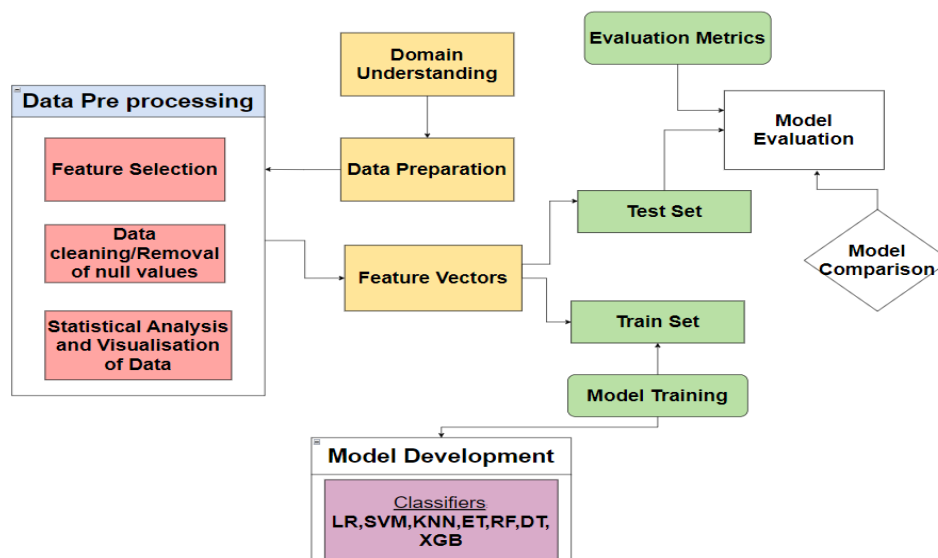


Figure 5.5: code snippet for speaker detection

Chapter 6

System Testing

Testing is a procedure of executing the program with unequivocal intension of **ref4** discovering mistakes, assuming any, which makes the program, fall flat. This stage is an essential piece of improvement.

It plays out an exceptionally basic part for quality affirmation and for guaranteeing unwavering quality of programming. It is the way toward finding the mistakes and missing operation and furthermore an entire confirmation to decide if the targets are met the client prerequisites are fulfilled.

The objective of testing is to reveal prerequisites, outline or coding blunders in the projects. Therefore, unique levels of testing are utilized in programming frameworks. The testing results are utilized amid upkeep. The testcases are shown in Figure 6.1

6.1 Testing Objectives

This area manages the points of interest in the various classes of the test which should be directed to approve capacities, imperatives and execution. This can be accomplished fundamentally by using the methods for testing, which assumes a crucial part in the improvement of a product.

6.2 Types of Testing conducted

The structure of the program is not being considered in useful testing. Test cases are exclusively chosen on the premise of the prerequisites or particulars of a program or module of program but the internals of the module or the program are not considered for determination of experiments**ref1**.

The program to be tried is executed with an arrangement of experiments and the yield of the program for the experiments is assessed to



Figure 6.1: testcases

decide whether the program is executing not surprisingly. The accomplishment of testing in uncovering mistakes in projects depends basically on the experiments. There are two fundamental ways to deal with testing Black Box or functional Testing and White Box or structural testing. Table 6.1 shows the workflow.

Table 6.1: Work Flow

| Sl No | Work | Duration(in Weeks) |
|-------|--|--------------------|
| 1 | Audio Extraction | 1 |
| 2 | Audio Enhancement using LAVSE | 4 |
| 3 | Audio Separation using Speechbrain | 3 |
| 4 | Noice Reduction using spectral subtraction | 2 |
| 5 | Image segmentation | 3 |
| 6 | Speaker Identification | 5 |

Chapter 7

Results and Discussion

7.1 Face detection



Figure 7.1: Face detection

Above Figure 7.1 shows initial face detection process using opencv and dlib. It convert the image to grayscale, apply the model using cv2. detectMultiScale(), and draw bounding boxes around the detected faces using cv2.rectangle(). Display or save the result using cv2.imshow() or cv2.imwrite().

7.2 Speaker recognition

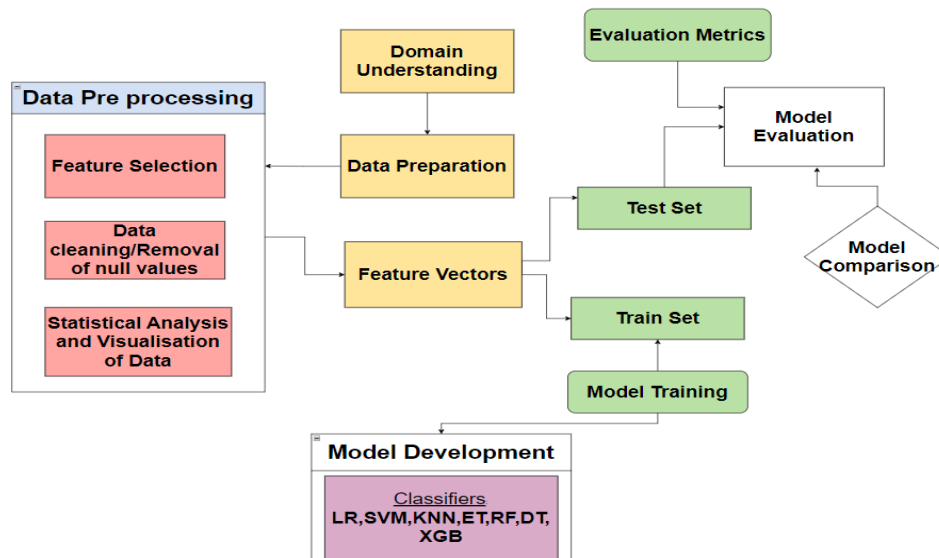


Figure 7.2: Speaker recognition 1, person 1

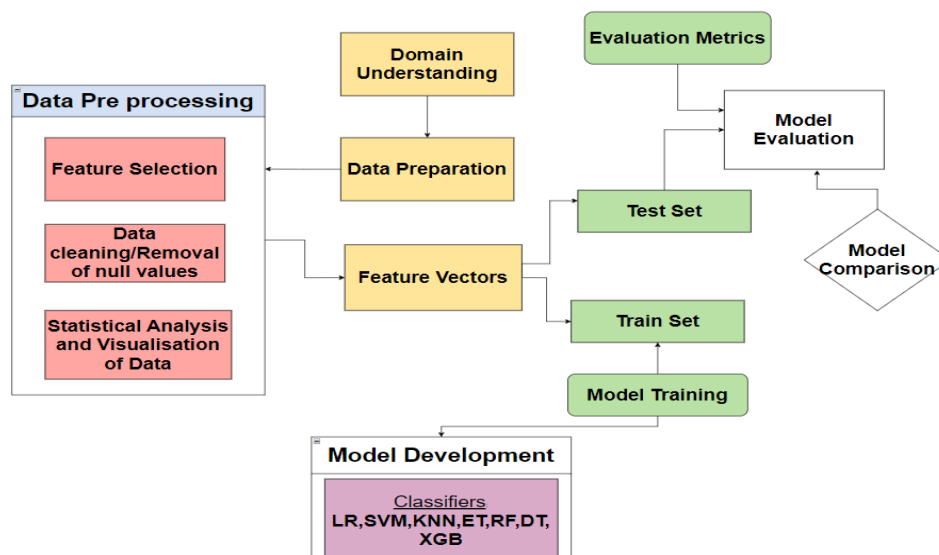


Figure 7.3: Speaker recognition 1, person 2

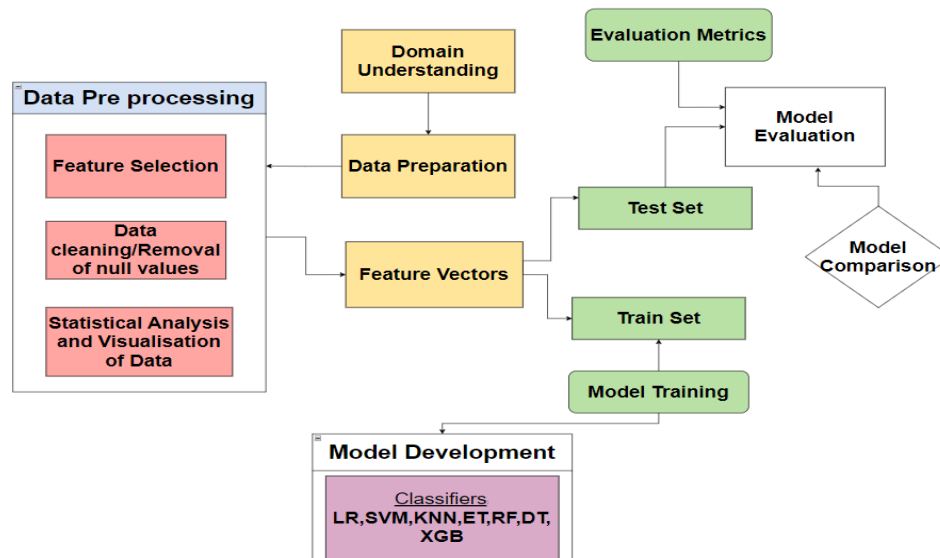


Figure 7.4: Speaker recognition 2, person 1

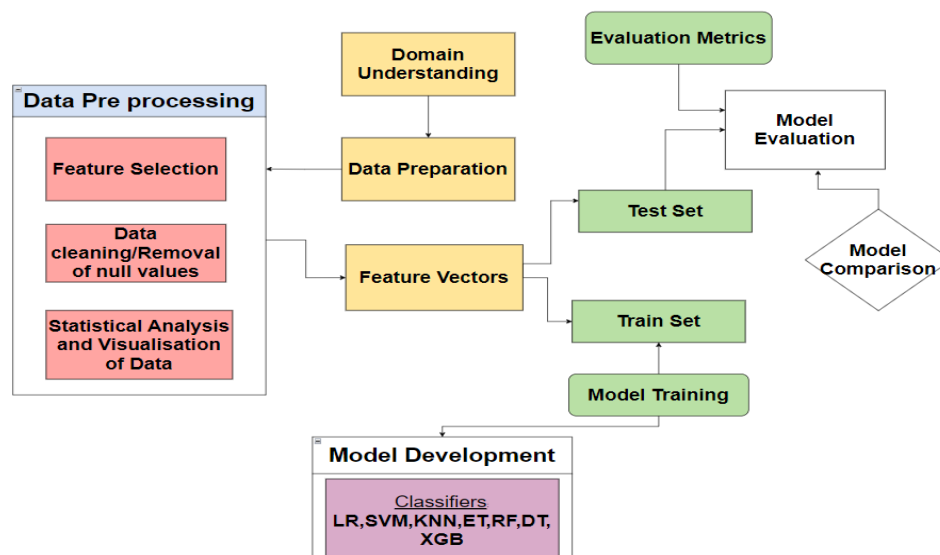


Figure 7.5: Speaker recognition 2, person 2

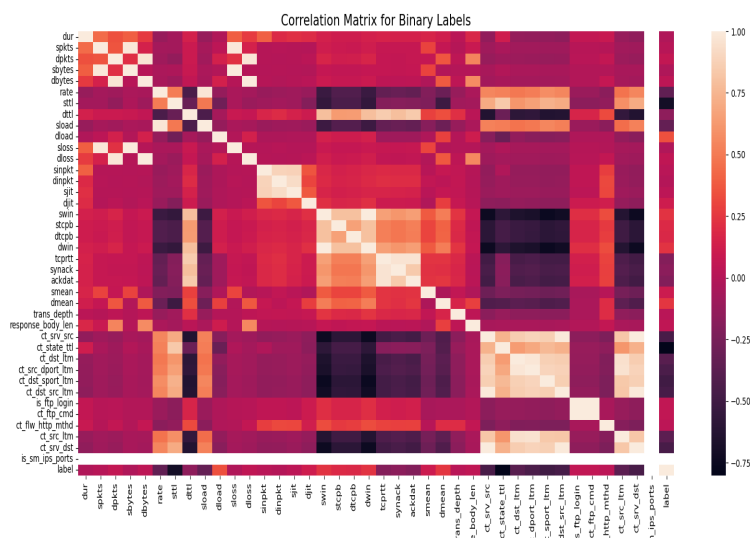


Figure 7.6: Speaker recognition 3, person 1

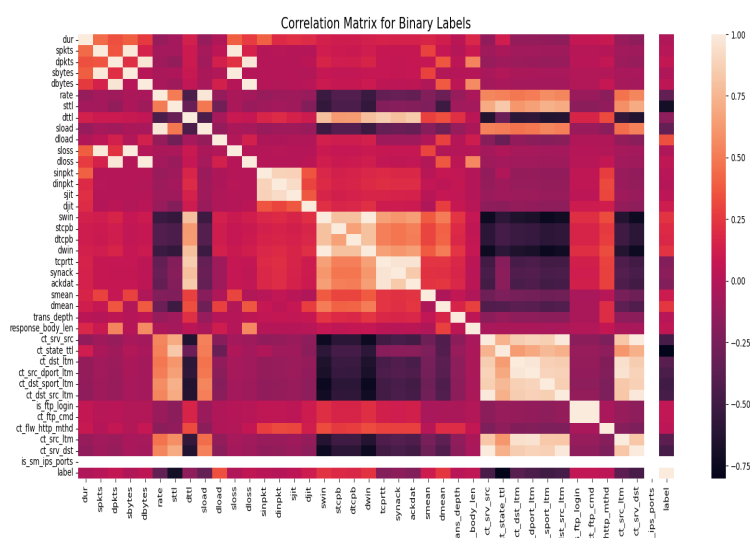


Figure 7.7: Speaker recognition 3, person 2

Above figures from 7.2 to 7.7 shows speaker recognition process using opencv and dlib. Speaker detection using cv2 and dlib involves utilizing dlib's pre-trained models along with cv2 functions to detect and locate human faces. By combining face detection with additional techniques such as audio analysis or lip movement tracking, speaker detection can be achieved in various applications like video conferencing or surveillance.

Chapter 8

Conclusion and Future work

The Project will help in narrowing the imprecise communication problem in real-time data using speech separation and speaker identification technique by Deep Learning and Image Processing algorithms. This will impact the communication and security sectors in a greater extent. Overall, this project aims to develop an application or method that can help to separate the audio-visual speech and enhance it based on speaker identification.

This project can be further developed as: • By incorporating more real-world testing and gathering feedback from individual units. • The system can be connected with communication devices or services to enable the users to communicate with others with ease.

This project has a great potential to make a positive impact on communication and security situations. Its continuous improvement will be important to make this impact even greater

References

- [1] Hu, G., Yang, Y., Yi, D., Kittler, J., Christmas, W.J., Li, S., & Hospedales, T.M. "When Face Recognition Meets with Deep Learning: An Evaluation of Convolutional Neural Networks for Face Recognition," 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), pp. 384–392, Dec. 2015, doi: 10.1109/iccvw.2015.58.
- [2] Parkhi, Omkar, Andrea Vedaldi, and Andrew Zisserman. "Deep face recognition." In BMVC 2015-Proceedings of the British Machine Vision Conference 2015. British Machine Vision Association, 2015.
- [3] Levitin, Anany. "Introduction to design and analysis of algorithms", 2/E. Pearson Education India, 2008.
- [4] Prabhu, "Understanding of Convolutional Neural Network (CNN) — Deep Learning", URL: <https://medium.com/RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>. Accessed on 23/07/2023
- [5] L. Blanger and A. R. Panisson, "A Face Recognition Library using Convolutional Neural Networks," International Journal of Engineering Research and Science, vol. 3, no. 8, pp. 84–92, Aug. 2017, doi: 10.25125/engineering-journal-ijoer-aug-2017-25.
- [6] R. Khedgaonkar, K. Singh, and M. Raghuwanshi, "Local plastic surgery-based face recognition using convolutional neural networks," Demystifying Big Data, Machine Learning, and Deep Learning for Healthcare Analytics, pp. 215–246, 2021, doi: 10.1016/b978-0-12-821633-0.00001-5.
- [7] P. J. Phillips, "A Cross Benchmark Assessment of a Deep Convolutional Neural Network for Face Recognition," 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 705–710, May 2017, doi: 10.1109/fg.2017.89.

- [8] Z. Huang, J. Zhang, and H. Shan, “When Age-Invariant Face Recognition Meets Face Age Synthesis: A Multi-Task Learning Framework,” 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7278–7287, Jun. 2021, doi: 10.1109/cvpr46437.2021.00720.
- [9] Z. Huang, J. Zhang, and H. Shan, “When Age-Invariant Face Recognition Meets Face Age Synthesis: A Multi-Task Learning Framework,” 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7278–7287, Jun. 2021, doi: 10.1109/cvpr46437.2021.00720.



The Report is Generated by DrillBit Plagiarism Detection Software

Submission Information

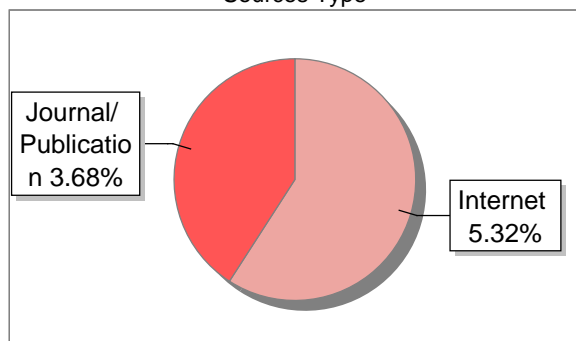
| | |
|--------------------------|---|
| Author Name | Saumya |
| Title | Gesture-Enhanced Presentation Control for Education |
| Paper/Submission ID | 3504295 |
| Submitted by | saumyam@sjec.ac.in |
| Submission Date | 2025-04-15 14:59:07 |
| Total Pages, Total Words | 6, 4458 |
| Document type | Article |

Result Information

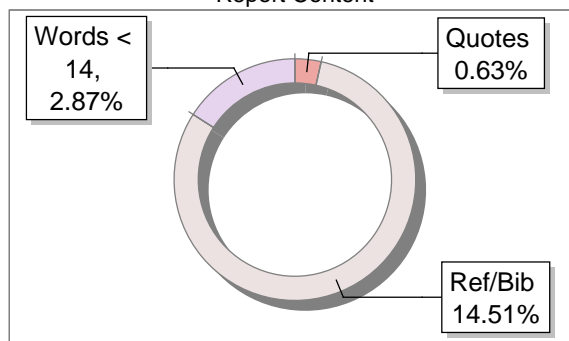
Similarity **9 %**



Sources Type



Report Content



Exclude Information

| | |
|-----------------------------|--------------|
| Quotes | Not Excluded |
| References/Bibliography | Excluded |
| Source: Excluded < 14 Words | Not Excluded |
| Excluded Source | 0 % |
| Excluded Phrases | Not Excluded |

Database Selection

| | |
|------------------------|---------|
| Language | English |
| Student Papers | Yes |
| Journals & publishers | Yes |
| Internet or Web | Yes |
| Institution Repository | Yes |

A Unique QR Code use to View/Download/Share Pdf File



Gesture-Enhanced Presentation Control for Education

Saumya Y M

Dept. of CSE

St Joseph Engineering College

Vamanjoor, India

saumyam@sjec.ac.in

JaishmaKumari B

Dept. of CSE

St Joseph Engineering College

Vamanjoor, India

jaishmab@sjec.ac.in

Colin Christon DCruz

Dept. of CSE

St Joseph Engineering College

Vamanjoor, India

colinchriston@gmail.com

Austin Dsouza

Dept. of CSE

St Joseph Engineering College

Vamanjoor, India

austindsz21@gmail.com

Daniel Loy Braggs

Dept. of CSE

St Joseph Engineering College

Vamanjoor, India

danielloy675@gmail.com

Elwin Jason Pereira

Dept. of CSE

St Joseph Engineering College

Vamanjoor, India

elwinjpereira02@gmail.com

Abstract—Presentation skills are vital in many areas of life. Giving presentations is probably a common experience for anyone, whether they are a worker, student, business owner, or employee of an organisation. The requirement to manage and manipulate the slides with a keyboard or other specialised device might make presentations seem tedious at times. Enabling users to control the slideshow with hand gestures is the aim of this work. Gestures have become increasingly common in human-computer interaction in recent years. Several PowerPoint functionalities have been attempted to be controlled by hand movements by the system. This system maps motions using multiple Python modules and uses machine learning to identify motions with minute variances. Creating the perfect presentation is becoming increasingly difficult due to a number of aspects, including the slides, the keys to switching the slides, and the audience's composure. An intelligent presentation system that is based on hand gestures makes it simple to update or modify the slides. Allowing viewers to explore and manipulate the slideshow with hand movements is the technology's main objective. The technique recognises various hand motions for a variety of tasks using machine learning. A means of recognition opens up a line of communication between people and machines.

Index Terms—*Gesture, Gesture Recognition, Human Computer Interaction, Presentation, Annotation, Slide change.*

I. INTRODUCTION

In today's ever-evolving education landscape, traditional classroom presentations are under- going a digital transformation. As digital learning gains prominence, there is a pressing need for a more dynamic and engaging means of controlling presentations. The existing tools not only limit the interactive potential of educators but also create accessibility challenges, particularly for those with physical disabilities. These issues hinder the effectiveness of teaching and can disrupt the flow of lessons [1]. Educators seek innovative ways to engage students using technology, making the "Gesture-Enhanced Presentation Control for Education" a highly relevant task.

In AR/VR, hand tracking is essential for facilitating natural engagement and communication [10], and it has been a subject of intense discussion in the field of study. For many years, research has been conducted on vision-based hand pose estimation [2]. The slides are editable by users. The interactive presentation system creates a more useful and approachable user interface for manipulating presentation displays by utilising state-of-the-art human-computer interaction techniques. When these hand gesture choices are used in place of a traditional mouse and keyboard control, the presentation experience is substantially improved. Nonverbal communication refers to the use of body language and gestures to convey a certain message. The Python framework was primarily employed in the construction of the system, together with NumPy, MediaPipe, openCV, and CV zone technologies. The goal of this approach is to improve presentations' usefulness and efficiency [5].

II. LITERATURE REVIEW

The paper [6] offers a thorough examination of computer vision based hand gesture recognition system. From mathematical algorithms like the row vector to machine learning approaches, it critically analyzes strengths and limitations. By scrutinizing methods such as edged image analysis and vector passing, the survey identifies research gaps and showcases deficiencies. This foundation justifies the chosen techniques, providing vital background and directionality. It aids in positioning the paper's contributions by benchmarking current challenges and showcasing field deficiencies, delivering crucial insights for hand gesture recognition understanding and system improvement. They employed the Row Vector Algorithm, the Diagonal Sum Algorithm, the Mean and Standard Deviation of the Edged Image, and the Edging and Row Vector Passing Algorithm.

This paper [2] introduces a system for controlling PowerPoint slides through hand gestures using a combination of a thermal camera and a webcam for robust hand tracking.

The methodology covers illumination invariant hand region extraction, gesture recognition through skin segmentation and SVM classification, and slide control mappings. Experiments demonstrate high accuracy in classifying gestures like swipe left and right to switch slides. The literature review analyzes existing research in gesture recognition and hand tracking, identifying challenges in accuracy and processing lag. The conclusion sums up key innovation, potential applications in interactive presentations, and limitations like small gesture vocabulary, suggesting enhancements through multidimensional dynamic time warping. The methodology included OpenCV, Haar Cascade Classifier, Skin Color Segmentation, and Gesture Recognition.

This paper [3] introduces an innovative vision based hand gesture recognition system designed for PowerPoint presentation control, encompassing both static and dynamic gestures. The literature review scrutinizes existing methodologies, highlighting limitations in accessibility and vocabulary across various approaches. Leveraging a sophisticated 7-layer convolutional neural network (CNN) built upon the 20BN-Jester baseline, the system extracts spatial and temporal features from dynamic hand gesture video frames, significantly improving accuracy. The training process, conducted on the 20BN-Jester dataset using PyTorch on a GPU system, results in a highly accurate model capable of real-time classification. The methodology involves multi-phase processing, from opening a PowerPoint presentation to capturing live webcam video of hand gestures, transformed into a 20-frame image array for classification. Python is the chosen programming language, with Tkinter for GUI, PyTorch for deep learning, OpenCV for computer vision, and PyAutoGUI for simulating virtual keyboard keypresses, ensuring a robust and versatile integration of functionalities. The system recognizes diverse gestures contributing to enhanced accessibility and user-friendliness in PowerPoint presentations. Their methodology used Convolutional Neural Network Architecture, 20BN-Jester dataset, multi-phase approach, PyTorch and PyAutoGU.

A hand gesture-controlled virtual mouse system for seamless human-computer interaction is presented in this paper [4]. Using a webcam to record the user's hand movements, the system uses computer vision and machine learning models to detect and identify pointing and clicking actions in real-time. These predicted hand poses are seamlessly translated into virtual cursor operations, allowing touchless spatial control. The literature review traces the evolution of gesture recognition techniques from initial glove-based tools to modern solutions, analyzing pros and cons of past approaches. The proposed methodology addresses limitations like hardware restrictions and system lag by blending MediaPipe, speech recognition, and natural language processing for an efficient and responsive interface. The Algorithms and Tools used here include Google's MediaPipe, Single Shot Detector model, Hand Landmark model.

This paper [5] presents a system for controlling presentations using hand gestures, built using OpenCV and Google's MediaPipe framework. A webcam captures video input of the

user's hand gestures, which are recognized by MediaPipe. Specific gestures like raising different numbers of fingers are then mapped to control commands for the presentation - changing between slides, accessing a pointer to draw on slides, and erasing drawings. The main technical challenge discussed is accurately recognizing gestures with background noise and variations in lighting. The system is designed to provide an intuitive hands free way of controlling presentations that could be used in real-world scenarios with basic hardware. Key libraries utilized include OpenCV for image processing and frame detection, MediaPipe for gesture recognition, and NumPy for numeric computing to transform the inputs into outputs. Overall, it demonstrates a practical application of computer vision and gesture recognition to facilitate more natural human-computer interaction. The Algorithms and Tools used here are BlazePalm, Hand Landmark Model, Hidden Markov Models (HMM), K-means clustering, Fast Fourier Transform (FFT), Non-maximum suppression and Encoder-decoder models.

The so-called Virtual Whiteboard, which is based on electronic pens and sensors, is given in the paper [8] and may offer an alternative to contemporary electronic whiteboards. With the tool in hand, the user can write, draw, and manipulate the contents of the whiteboard with just his or her hands. It is not necessary to have extra equipment like infrared diodes, infrared cameras, or cyber gloves. Dynamic hand gesture recognition is the foundation for user interaction with the Virtual Whiteboard computer application. When examining a video feed from a webcam connected to a multimedia projector that displays content from a whiteboard, gestures are identified. Kalman filtering helps to track the positions of hands in the image. In the paper the hardware and software of the Virtual Whiteboard is discussed with a special focus on applying Kalman filters for prediction of successive hand locations. The effectiveness of Kalman filter-supported recognition was evaluated for the motions used to manage the contents of the whiteboard, and the efficiency without filtering is provided.

The problem of estimating the entire 3D hand shape and pose from a single RGB image is a new and difficult one that is tackled in this study [7]. The majority of existing techniques for 3D hand analysis from monocular RGB images are limited to guessing the 3D positions of hand keypoints; they are unable to accurately convey the 3D shape of the hand. On the other hand, the research describes an approach based on Graph Convolutional Neural Networks (Graph CNNs) that can reconstruct a complete 3D mesh of the hand surface, which includes more detailed information on the 3D shape and attitude of the hand. They provide a large-scale synthetic dataset comprising both 3D postures and ground truth 3D meshes in order to train networks under complete supervision. Using the depth map as a weak supervision in training, the researcher presented a weakly supervised method for fine-tuning the networks using real-world datasets without 3D ground truth. Through rigorous evaluations on their suggested new datasets and two public datasets, proposed research indicate that proposed technique can build accurate and reasonable 3D

hand mesh and can accomplish superior 3D hand pose estimate accuracy when compared with state-of-the-art methods. The difficulties faced by patients receiving physical therapy are discussed in the paper [9], with a focus on the boredom of repeating exercises that may cause patients to lose enthusiasm. It offers a remedy in the shape of hand rehabilitation software, which makes use of hand gesture detection and recognition technologies to enhance patient engagement and enjoyment during rehabilitation. The MediaPipe Hands algorithm is used by the system to recognise gestures and detect hands.

The study [11] uses morphological processing and YCbCr thresholding to accomplish efficient gesture recognition for PowerPoint presentation control. The Hidden Markov Model is used to classify the gestures that have been identified. HMM is a statistical model that works well for tasks involving the recognition of patterns over extended periods of time.

The purpose of the paper [12] is to enable gesture-based control of PowerPoint presentations, and it does so by using multiple techniques. Machine learning algorithms are used in the study to identify and categorise hand gestures. By training the model to distinguish minor changes in movements, the system can accurately map these motions to specific actions, such as advancing or reversing slides. The Python programming language is used to implement the system, making use of Mediapipe and OpenCV packages.

The paper [13] provides a new way for controlling PowerPoint presentations using static hand gestures. This technique uses a webcam to record hand motions, making it a useful and user-friendly solution. The thinning method, a method for processing and analysing hand forms, is introduced in this study. The number of elevated fingers is determined by using the hand form parameters that are extracted using this procedure. This novel method improves gesture recognition precision. The fact that the suggested approach doesn't need any extra gear, like gloves, markers, or other gadgets, is one of its best qualities. This improves the system's accessibility and usability by enabling users to interact with their presentations using just their hands.

III. SYSTEM DESIGN

A. Architectural Diagram

The architectural design for gesture-enhanced presentation control as displayed in the above Fig. 1 begins with the webcam capturing the user's hand gestures, serving as the primary input method. OpenCV processes the video feed, extracting critical details such as hand position and shape. These details are then analyzed by MediaPipe, which employs sophisticated algorithms to recognize specific gestures based on predefined patterns. Following recognition, the identified gestures are relayed back to the presentation software, where they are interpreted into actions such as navigating slides or activating multimedia elements. This process involves several intermediary steps, including video capturing, framing, and hand detection as well as frames filtering to enhance accuracy. Feature extraction distills relevant information from the recognized gestures, which are then classified into predefined ac-

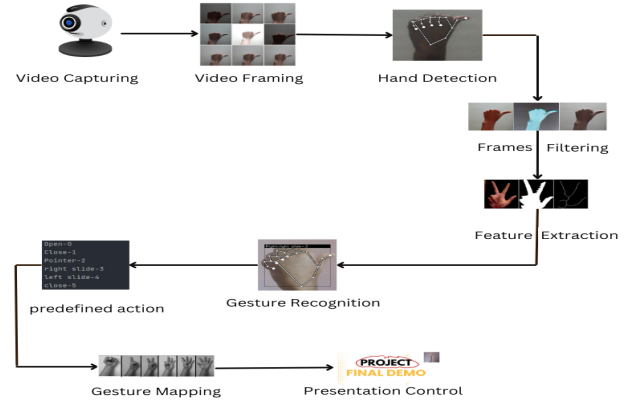


Fig. 1. Architecture Diagram

tions. The architecture further encompasses gesture mapping, where these classified gestures are matched with corresponding presentation control functions. Ultimately, the presentation control component interfaces seamlessly with the software, executing the mapped functions based on the recognized gestures. Throughout this interaction, the user plays a central role, activating the webcam input device and performing hand gestures within its view to control the presentation. Feedback mechanisms such as audible or visual signals confirm gesture recognition and execution ensuring a smooth and intuitive user experience.

B. Phases of Gesture Recognition

- **Phase 1: Video Acquisition (Webcam):** This phase involves capturing the video stream from a webcam or any other camera input device. The quality and resolution of the captured video are crucial for accurate hand gesture recognition. The video stream serves as the input for subsequent phases in the gesture recognition system.
- **Phase 2: Video Pre-processing:** Video pre-processing is essential for preparing the captured video stream for hand gesture recognition. This phase typically includes several tasks such as:
 - 1) **Frame Extraction:** The continuous video stream is divided into individual frames for analysis.
 - 2) **Background Removal:** Removing the background from each frame helps isolate the hands from the rest of the scene, reducing interference and improving accuracy.
 - 3) **Hand Region Detection:** Identifying and delineating the regions of interest containing the hands within each frame. This can involve techniques like skin tone detection or background subtraction to locate the hands within the frame accurately.
- **Phase 3: Feature Extraction:** In this phase, relevant features are filtered and extracted from the detected hand region. These features provide the basis for identifying and interpreting different hand gestures. Frame filtering tasks may include:

- 1) **Frame Rate Reduction:** The system can sample the video at a reduced frame rate, such as every second or third frame, to focus on key points in time where meaningful gestures occur. This eliminates redundant data and allows the model to concentrate on frames that contain significant hand movements.
- 2) **Background Subtraction:** Background subtraction techniques are applied to isolate the hand region from the background. This helps to filter out irrelevant objects and noise, ensuring that only the hand gesture is processed. Common methods include Gaussian Mixture Models (GMM) or simple thresholding techniques that detect motion in the foreground.
- 3) **Blurring and Smoothing:** Applying blurring or smoothing filters (such as Gaussian blur) to the frames can help remove minor noise or irregularities in the video feed. This step enhances the quality of the input image, making the subsequent feature extraction process more robust.
- 4) **Skin Detection and Masking:** Skin detection algorithms can be applied to identify regions of the frame corresponding to human skin tones, focusing specifically on hand regions. This creates a mask that highlights the hand while ignoring non-skin areas, leading to more precise hand detection and feature extraction.

Feature extraction tasks may include:

- 1) **Hand Pose Estimation:** Determining the orientation and configuration of the hand(s) in the frame, including finger positions and hand shape. These landmarks provide a detailed map of the hand's orientation, configuration, and shape. Algorithms like MediaPipe Hands can accurately detect these landmarks in real time.
 - 2) **Finger Tracking:** Tracking the movement and position of individual fingers within the hand region. By tracking the movement of each finger over time, the system can recognize dynamic gestures, such as a finger swipe or a specific finger motion sequence.
 - 3) **Motion Trajectory Analysis:** Analyzing the trajectory of hand movements over time to detect gestures involving motion, such as swipes or gestures with directional components. Recognizing when a gesture starts and ends is essential for temporal gestures. This involves detecting the initial movement and when the hand comes to rest or returns to a neutral position.
- **Phase 4: Gesture Recognition:** Gesture recognition involves two main steps:
 - 1) **Gesture Classification:** Classifying the extracted features into specific gesture classes or commands. This can be achieved using machine learning models such as convolutional neural networks (CNNs) or rule-based algorithms.

- 2) **Gesture Mapping:** Once gestures are classified, they are mapped to corresponding presentation control functions. For example, a specific hand pose or motion trajectory might correspond to commands like next slide, previous slide, or activate pointer mode.

- **Phase 5: Presentation Control:** In this final phase, the recognized gestures are used to control presentation software such as PowerPoint or Google Slides. This includes executing the mapped presentation control functions based on the recognized gestures, enabling seamless interaction with the presentation content. Presentation control functions may include slide navigation, pointer control, annotation or drawing tools activation, and other interactive features. The system interfaces with the presentation software through appropriate APIs or communication protocols to facilitate these actions.

IV. IMPLEMENTATION

The implementation of Gesture-Enhanced Presentation system commenced with the pivotal task of data collection and preprocessing. This is followed by selection and training the model for gesture recognition.

A. Data Collection and Preprocessing:

- **Data Collection:** A diverse set of hand gesture images representing various presentation commands like "Next Slide", "Previous Slide", "Start Presentation" and "Stop Presentation" is collected. Ensuring diversity, the dataset covers a wide range of hand shapes, positions, and lighting conditions. Quality assurance was maintained throughout the process to minimize noise and ensure clarity for effective model training.
- **Data Labeling:** Each image underwent a labeling process, where the images were manually annotated with the corresponding gesture it represents, including gestures like "Next Slide", "Previous Slide" and others. Key points within the hand gestures, such as the position of the index finger, were also labeled to provide crucial information for model training. Consistency in labeling was paramount to avoid confusion during model training and evaluation.
- **Data Preprocessing:** Before training the model, the collected dataset was preprocessed to enhance image quality and remove noise. This involved resizing images to a standard size, normalizing pixel values for consistent brightness and contrast, and applying various data augmentation techniques to increase dataset diversity. Relevant features, such as identifying landmarks or keypoints like the coordinates of the index finger, were extracted for effective gesture recognition.
- **Model Selection:** For the classification task TensorFlow's Keras API was opted. This choice was driven by the availability of pre-built deep learning models tailored for gesture recognition. By utilizing this framework, the model was able to efficiently

utilize computational resources and simplify the development process.

- **Training Data:** After preprocessing, the dataset was divided into training and validation sets. The goal here was to ensure a fair distribution of samples across different gesture classes, which is crucial for the model to generalize well to new, unseen data.
- **Model Training:** With TensorFlow as the training platform, a structured approach was followed. Techniques like transfer learning and fine-tuning of pre-trained models was applied to optimize the model's performance, especially given constraints such as limited training data or computational resources.
- **KeyPoint Classifier:** To facilitate the gesture recognition, the KeyPointClassifier class was implemented. This supported the deployment of a TensorFlow Lite interpreter, specifically configured with the chosen model file and thread specifications. With this setup, the landmark coordinates could be taken as input resulting in the accurate prediction of the class index, enabling precise classification of hand gestures based on these key points.

B. Real-time Gesture Recognition:

- **Camera Integration:** Integrate a camera module (e.g., webcam) with the system to capture real-time video input.
- **Gesture Detection:** Utilized libraries like OpenCV and MediaPipe to detect and track hand gestures in real-time video streams. Apply the trained gesture recognition model to classify detected gestures.
- **Feedback Mechanism:** Provide visual feedback to the user in real-time, indicating the recognized gesture and corresponding action.

C. GUI Development:

- **Graphical User Interface:** A user-friendly GUI was implemented using Tkinter framework, featuring an intuitive interface that enables users to interact with the presentation software using hand gestures.
- **Control Elements:** Implemented control elements such as buttons or sliders for common presentation functions (e.g., next slide, previous slide, start/stop presentation).
- **Integration with Gesture Recognition:** Integrated the gesture recognition module with the GUI, ensuring seamless interaction between gesture input and presentation control.

D. System Integration and Testing:

- **Component Integration and Function Testing:** All the components of the system, including gesture recognition, GUI, and presentation control logic are integrated. This is followed by a thorough testing to ensure the system functions as expected in different



Fig. 2. GUI without input pptx file

scenarios and environments. Test for accuracy, responsiveness, and robustness to variations in lighting conditions and hand gestures are performed.

V. RESULTS AND DISCUSSION

The results depict the Graphical User Interface (GUI) of the proposed system as displayed in the Fig. 2, illustrating the initial state with no file selected. Additionally, it showcases an array of recognized hand gestures and their corresponding actions seamlessly integrated into the interface.

Users can seamlessly navigate through slides, move left or right, annotate slides with a red-colored line, and erase annotations as needed, providing a dynamic and engaging presentation experience.

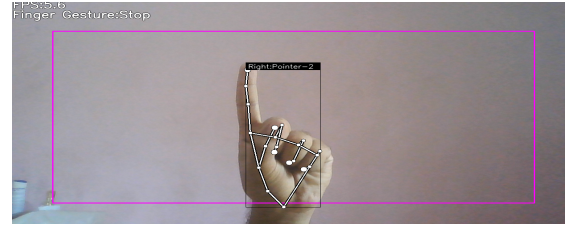


Fig. 3. Pointer to move cursor







In Fig. 3, the background image detection with landmarks is showcased, highlighting the system's ability to accurately detect gestures and provide real-time feedback to the user. This functionality empowers programmers with a flexible approach to working with gestures, ensuring precise recognition and seamless integration into presentation control. TABLE I provides a comprehensive overview of the different gestures supported by the proposed system. The table outlines the specific actions associated with each gesture, ensuring users understand how to utilize them. This detailed description facilitates a deeper comprehension of each gesture's functionality and its intended role within the proposed system.

VI. FUTURE WORK

Despite the successful implementation of the Gesture-Enhanced Presentation system, there are several avenues for future work and enhancements:

- 1) **Integration with Voice Commands:** Expanding the system to support voice commands alongside hand

TABLE I
GESTURES SUPPORTED BY THE PROPOSED SYSTEM

| Gesture | Action Performed | Description |
|---|---------------------------------------|---|
|  | Switch Between Annotation and Pointer | This gesture allows users to smoothly transition between annotation and pointer modes, as well as back and forth from pointer to annotation mode. |
|  | Clear Annotation | This gesture serves the purpose of clearing annotations made by the user previously. |
|  | Mouse Pointer | This gesture serves a dual purpose: first, it enables users to navigate the cursor during presentations, facilitating seamless control and highlighting of key areas. Additionally, it empowers users to make annotations, jot down notes, and mark significant sections within the presentation, enhancing engagement and interaction. |
|  | Next Slide | This gesture enables users to seamlessly transition to the next slide during presentations. |
|  | Previous Slide | This gesture enables users to seamlessly transition to the previous slide during presentations. |
|  | Exit Presentation | This gesture provides a way to the user to exit the presentation. |

gestures would provide users with additional control options and further enhance the user experience. Integrating voice recognition technology would enable presenters to navigate slides and execute commands using natural language.

- 2) **Enhanced Gesture Recognition:** Continuously improving the accuracy and robustness of gesture recognition algorithms is crucial for ensuring reliable performance across different environments and hand poses. Further research and development in this area could involve exploring advanced machine learning techniques and leveraging larger datasets for training.
- 3) **Multi-Modal Interaction:** Exploring the integration of multiple modalities such as hand gestures, voice commands and facial expressions could lead to more immersive and interactive presentation experiences. By combining different input modalities one can create a more versatile and adaptable system that caters to a wider range of user preferences and abilities.

VII. CONCLUSION

In conclusion, the development of the Gesture-Enhanced Presentation system for education has been a significant

endeavor aimed at revolutionizing the way educators and students interact with presentation materials. By leveraging hand gesture recognition technology, a user-friendly interface that allows presenters to control presentation slides seamlessly using intuitive gestures is proposed. This system offers an innovative and engaging approach to delivering educational content, enhancing the learning experience for both presenters and audiences. Through rigorous testing and iterative design improvements, it is ensured that the system meets the requirements of educational settings and delivers reliable performance.

REFERENCES

- [1] Bobo Zeng, Guijin Wang, Xinggang Lin. "A Hand Gesture Based Interactive Presentation System Utilizing Heterogeneous Cameras". TSINGHUA SCIENCE AND TECHNOLOGY ISSN11007-0214/15/181pp329-336 Volume 17, Number 3, June 2012.
- [2] Rida Zahra , Afifa Shehzadi , Muhammad Imran Sharif , Asif Karim , Sami Azam , Friso De Boer , Mirjam Jonkman , Mehwish Mehmood."Camera-based interactive wall display using hand gesture recognition",Computational Intelligence and Neuroscience, 2022.
- [3] Muhammad Idrees , Ashfaq Ahmad , Muhammad Arif Butt , and Hafiz Muhammad Danish. "Controlling Power Point using Hand Gestures in Python". PWebology (ISSN: 1735-188X) Volume 18, Number 6, 2021.
- [4] M. Kasar, P. Kavimandan, T. Suryawanshi, and S. Abbad, "AI-based real-time hand gesture-controlled virtual mouse," Australian Journal of Electrical and Electronics Engineering, pp. 1–10, Feb. 2024, doi: 10.1080/1448837x.2024.2313818.
- [5] Hajeera Khanum, Dr. Pramod H B." Smart Presentation Control by Hand Gestures Using Computer Vision and Google's MediaPipe", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 09 Issue: 07 — July 2022 .
- [6] Munir Oudah , Ali Al-Naji and Javaan Chahl "Hand Gesture Recognition Based on Computer Vision: A Review of Techniques". IEEE conference on computer vision - 23 July 2020
- [7] Lihao Ge, Zhou Ren, Yuncheng Li, Zehao Xue, Yingying Wang, Jianfei Cai, and Junsong Yuan. "3D Hand shape and Pose Estimation from a Single RGB image". In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 10833–10842, 2019.
- [8] M. Lech, B. Kostek, and A. Czyzewski, "Virtual Whiteboard: A gesture-controlled pen-free tool emulating school whiteboard," Intelligent Decision Technologies, vol. 6, no. 2, pp. 161–169, Feb. 2012, doi: 10.3233/idt-2012-0132.
- [9] Yeng, Angelina Chow Mei, et al. "Hand Gesture Controlled Game for Hand Rehabilitation." International Conference on Computer, Information Technology and Intelligent Computing (CITIC 2022). Atlantis Press, 2022.
- [10] H.S. Shrisha, V. Anupama, "NVS-GAN: Benefit of generative adversarial network on novel view synthesis", International Journal of Intelligent Networks, Volume 5, 2024, 184-195,doi.org/10.1016/j.ijin.2024.04.002.
- [11] Cahya, Rahmad., Arief, Prasetyo., Riza, Awwalul, Baqy. "PowerPoint slideshow navigation control with hand gestures using Hidden Markov Model method." 12 (2022).:7-18. doi: 10.31940/matrix.v12i1.7-18
- [12] K., P., Kumari., Bandaram, Bharath, Goud., Kalvakuntla, Sumana., Bathula, Naresh., Bellamkonda, Harish. "Automated Gesture Controlled Presentation Using Machine Learning." International Journal For Science Technology And Engineering, 10 (2022).:1248-1251. doi: 10.22214/ijras.2022.47517
- [13] Savitha, M. "Static Hand Gesture Recognition for PowerPoint Presentation Navigation using Thinning Method." International Journal on Recent and Innovation Trends in Computing and Communication, 6 (2018):187-189.