B-FIT GLASSES – VISUAL ASSISTANCE GLASSES

FOR VISUALLY IMPAIRED

2023-108

Final Report

Kakulandara P.M

(IT20256814)

B.Sc. (Hons) Degree in Information Technology.

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

February 2023

B-FIT GLASSES – VISUAL ASSISTANCE GLASSES

FOR VISUALLY IMPAIRED

2023-108

Final Report

B.Sc. (Hons) Degree in Information Technology.

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

February 2023

DECLARATION

To the best of our knowledge and belief, this proposal does not contain any previously published or written by another person content, with the exception of where credit is given in the text. I hereby declare that this is my own work, and this proposal does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university or Institute of higher learning.

Name	Student ID	Signature
Kakulandara P.M	IT20256814	

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

……………………………………….. …………………………….

Signature of the supervisor Date

(Dr. Dharshana Kasthurirathna)

ACKNOWLEDGEMENT

I would like to take a moment to express my profound gratitude to my supervisor, Dr. Dharshana Kasthurirathna, and co-supervisor, Hansi De Silva, for their unwavering support and invaluable guidance throughout this transformative journey of research. Their continuous encouragement, mentorship, and direction have been the pillars on which this endeavor was built, especially during the challenging and trying moments when their expertise and wisdom truly shone.

I am also deeply indebted to Ms. Sanjeevi Chandrasiri, Ms. Narmada Gamage, and Ms. Karthiga Rajendran for their generous contributions in the form of recommendations and insightful comments. Their constructive feedback and meticulous attention to detail played a pivotal role in shaping the outcome of my research. Their belief in my work and their willingness to share their knowledge have been instrumental in my successful completion of this project.

Lastly, I would like to extend my heartfelt thanks to my family and the dedicated research team who have been by my side throughout this journey. Their unwavering support, encouragement, and inspiration have been the driving force behind my determination to push the boundaries of knowledge. Without their love and encouragement, this achievement would not have been possible.

In closing, I want to express my deepest appreciation to everyone who has played a role in this journey, from mentors to colleagues to loved ones. Your support has been the wind beneath my wings, propelling me toward the realization of my research goals, and for that, I am eternally grateful.

ABSTRACT

The creation of smart glasses equipped with machine learning algorithms and various hardware components such as sensors, cameras, microphones, Arduinos, and mobile phones as servers has the potential to significantly improve the lives of visually impaired people. These glasses use machine learning to aid in everyday activities such as obstacle detection, text recognition, and fire detection.

The sensors and cameras embedded in the glasses can identify obstacles and notify the user with a voice guide to provide information on the distance to the obstacle, allowing visually impaired people to traverse their surroundings in a safe and secure manner. Furthermore, the glasses can identify text on a variety of surfaces, such as books, letters, and labels, and convert it into speech to assist with reading.

Furthermore, the use of panel pc and laptop as servers provides for data processing and storage, increasing the total efficiency and accuracy of the glasses. The combination of these technologies offers a comprehensive solution for the visually impaired, with the ability to improve their quality of life considerably.

In conclusion, the use of smart glasses outfitted with machine learning algorithms and various hardware components has the potential to revolutionize the way visually impaired people engage with the world, providing a safer, more accessible, and more autonomous way of life.

Keywords - smart glasses, machine learning, sensors, cameras, microphones, Arduinos, servers, obstacle detection, text recognition, fire detection, visually impaired.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
SLIIT	Sri Lanka Institute of Information Technology
MFCCs	Mel-frequency cepstral coefficients
CNN	Convolutional Neural Networks
R-CNN	Region-based Convolutional Neural Network
VAD	Vocal Activity Detection
DNN	Deep Neural Network
LSTM	Long Short-term Memory Networks
VPS	Visual Positioning System
UTF-8	(Unicode Transformation Format 8-bit)
NLTK	Natural Language Tool Kit

Table 1 List of Abbreviations

1 INTRODUCTION

1.1 Background

Visual impairment is a worldwide disorder that affects millions of individuals. While there are many assistive technologies available for the visually impaired, smart glasses with machine learning and hardware components have recently gained popularity due to their versatility and potential to help with a wide range of everyday activities.

Smart glasses for the visually impaired usually include sensors, cameras, and microphones that use machine learning algorithms to help people navigate their surroundings, recognize text, and identify potential hazards. These glasses are frequently combined with panel pc and laptop to improve data processing and storage capabilities, thereby improving total performance and accuracy.

Machine learning algorithms are especially useful in smart glasses because they can be taught on large datasets of visual and auditory data, enabling the glasses to learn and identify various objects, faces, and text. These algorithms can also help with obstacle detection and distance estimation, enabling the glasses to warn users of possible hazards.

Smart glasses equipped with machine learning and hardware components can help with duties such as facial recognition, wayfinding, voice commands, voice detection and in addition to obstacle detection and text recognition. These glasses can learn to identify various faces, navigate complex environments, and spot potential hazards such as fires using deep learning algorithms.

Overall, smart glasses for the visually impaired that use machine learning and various hardware components have the potential to significantly improve the quality of life for people who have visual impairments by allowing them to navigate their environments more securely and independently.

1.2 Literature Survey

Image-based Visual Assistance for the Blind. Th [1]is study describes an image processing-based visual aid device for the blind. [2]

Google Vision API is used in the design and implementation of voice-activated smart glasses for people with visual impairments. The design and execution of a voice-assisted smart glasses system for the blind using the Google Vision API are discussed in this article. [3]

Using a Raspberry Pi, a visual assistant for blind people. This article describes an object recognition, obstruction avoidance, and text-to-speech conversion Raspberry Pi-based visual assistant system for the blind. [1]

A list of voice input commands for assistive smart glasses applications and users with visual impairments. This study examines the possible uses of assistive smart glasses and provides a list of vocal input commands for users with vision impairments. [4]

Smart Glasses for Visual Perception Assistance for the Blind. This article explains an object recognition, text-to-speech translation, and guidance support smart glasses-based visual perception aid system for the blind. [5]

In general, speech commands and voice detection tasks can be performed using machine learning algorithms. ACAM, LSTM, and CNN are just a few examples of deep learning models that can be used to carry out these duties. In order to precisely identify IoT devices linked to a network, machine learning algorithms can also be applied to network traffic data.

1.3 Research Gap

Feature	**Research 1 [1]**	**Research 2 [2]**	**Research 3 [3]**	**Research 4 [4]**	**Research 5 [5]**	**Proposed Solution**
Inbuilt Voice Assistance with a bot (Bi-directional)	✗	✗	✗	✗	✗	✓
Voice Input	✗	✓	✓	✓	✗	✓
Voice Output	✓	✓	✓	✓	✓	✓

Table 2 Summary of Research Gap

Integrating an AI bot that can comprehend commands from users in various situations is one possible study area that needs to be covered in order to create spectacles for people who are visually challenged. Although modern AI-powered spectacles have features like auditory explanations of the surroundings, they might not always be able to react to specific user instructions.

For instance, a blind individual donning AI-enhanced spectacles might ask the bot to interpret a particular document or sign even in a busy or noisy environment, the bot would need to be able to identify the user's vocal instructions and accurately understand the user's command meaning. The creation of low-cost smart glasses for the blind represents another potential research gap. There is a need to create smart glasses systems that are specially made for people with visual impairments because the majority of current smart glasses systems are made for people with normal vision. Additionally, methods that are accessible and affordable for people with low incomes need to be developed. (Himadri Nath Saha, October 2017)

1.4 Research Problem

These days, it is quite difficult for individuals to adopt a visually impaired person into society. While trying to connect with users without correct identification, blind people must endure a truly difficult time. Also, if they do not properly recognize obstacles when trying to go outside, they run the risk of falling or getting wounded. Also, because they cannot read books, paperwork, etc., blind students have a hard time learning. All of the aforementioned issues will be solved by a wearable system we have created. The system uses a wearable device with a panel pc, camera, microphone, and earphones to do a variety of tasks, including detecting text from a paper and extracting notes. Moreover, it can recognize nearby users and monitor the owner of the glasses. By the use of computer vision and audio processing algorithms, the device will analyze the collected data and extract relevant information about the user's surroundings.

When the data has been processed, the server will use machine learning algorithms to do more analysis and processing. Machine learning algorithms will use named entity recognition and convolutional neural networks (CNN) in conjunction with natural language processing (NLP) techniques to successfully identify characters in a document (NER). The technology will provide the user with real-time information via auditory signals so they can interact with others and better control their surroundings. The wearable gadget will be connected to a panel pc, which will serve as the system's command center.

The main goal of this system is to create a reasonable solution for personal safety, social interaction, and ambient supported living. By combining a variety of hardware elements and machine learning approaches, the system will provide the user with a complete and efficient solution for navigating their surroundings and interacting with others.

2 OBJECTIVES

2.1 Main Objective

The primary goal of using voice commands and detection in glasses for visually impaired people is to provide users with an intuitive and hands-free method to navigate their environment and perform a variety of activities. The glasses can interpret speech commands and provide real-time feedback to the user by leveraging machine learning and sophisticated hardware components such as sensors, cameras, speakers, and microphones.

The glasses can detect the users commands and Using AI bot understand the user command based on several instructions, the glasses can also be programmed to perform a variety of tasks, such as adjusting the user's route, providing audio feedback on nearby landmarks or obstacles, and reading content in a document.

Overall, the goal of incorporating speech commands and detection into glasses for the visually impaired is to provide a powerful and versatile instrument for navigation and task management. The glasses can provide visually impaired individuals with a highly customizable and intuitive solution for navigating their environment and performing a wide range of daily tasks, without the need for traditional user interfaces or physical input devices, by leveraging the latest advances in machine learning and hardware technology.

2.2 Specific Objectives

Voice command

1. Accurate voice recognition: Even in noisy environments, the glasses should be able to recognize a broad variety of voice commands spoken by the user.
2. Voice commands: The glasses should detect the user’s commands.
3. Real-time audio feedback: The glasses should provide the user with real-time audio feedback based on their voice commands, such as adjusting their path or giving information on local landmarks.
4. Identifying the users’ command: According to the users command glasses should understand the user command in different context

Voice detection

1. Accurate sound recognition: The glasses should be able to identify the sound of the user, to provide appropriate feedback to the user.
2. Real-time audio feedback: The glasses should provide the user with real-time audio feedback based on the sounds detected.

3 METHODOLOGY

3.1 Project Overview

1. **Microphone**: captures the user's voice and converts it into an analog signal that the system can handle.
2. **Speaker**: Output the guidance that required for the user to navigate and understanding signs, documents, currency, and people.
3. **Pre-processing**: The analog data is pre-processed to remove noise and improve signal quality.
4. **Feature extraction**: The pre-processed signal is examined in order to derive features that can be used to recognize the user's voice and speech.
5. **Vocal recognition**: To identify the user and detect their speech, the extracted features are compared to a pre-defined database of vocal patterns.
6. **User tracking**: The system monitors the user's position and movements in order to provide context for voice commands and interactions.

In general, a voice command and voice detection system with user tracking allows users to engage with technology using their voice while also giving context and feedback to enhance the system's accuracy and usability.

The system gives the user feedback to confirm that their vocal commands were recognized and to guide them through the interaction.

Requirements Gathering and Analysis

Functional implementation

1. **Voice Recognition and Detection**: Implement a robust voice recognition and detection system that can accurately capture and interpret voice commands from users. This system should be able to handle multiple similar meaning keywords for a single command.
2. **Client-Server Architecture**: Set up a client-server architecture where a panel PC on the client side communicates with a laptop server for

processing. Ensure smooth data transfer between the two using sequential threads.

3. **Text-to-Speech and Speech-to-Text**: Utilize Google Text-to-Speech and Speech-to-Text APIs to convert text to speech and speech to text, facilitating communication between the user and the system.
4. **Natural Language Processing (NLP)**: Implement NLP techniques using libraries like NLTK for preprocessing voice commands. This step is crucial for understanding user instructions accurately.
5. **Object Detection**: Enable object detection using computer vision techniques to identify objects in the environment. This can help users locate and interact with objects around them.
6. **Known Person Detection**: Implement a feature to recognize known individuals by voice commands. The system should be able to provide the user with the name or identity of the recognized person.
7. **Sign Detection**: Enable the system to detect and interpret signs, providing auditory feedback to the user to help them navigate their surroundings safely.
8. **Document Reading**: Implement a document reading function that can read text from documents aloud to the user, making printed material accessible to them.
9. **Money Recognition**: Enable the system to recognize and provide information about different denominations of currency notes, assisting users with financial transactions.
10. **Voice-Guided Assistance**: Implement a voice-guided assistance feature that can provide real-time guidance to the user, helping them navigate their environment and perform various tasks.
11. **User Feedback and Learning**: Collect user feedback to continuously improve the system's performance and user experience. Implement

mechanisms for the system to learn from user interactions and adapt over time.

12. **Privacy and Security**: Ensure that user data is handled securely, and that the system complies with privacy and security standards. Consider user consent and data protection.
13. **Accessibility Standards**: Make sure the system adheres to accessibility standards like WCAG to ensure it is accessible to all users with disabilities.
14. **Documentation and Training**: Provide clear documentation for users and potentially offer training sessions to help visually impaired individuals effectively use the system.
15. **Integration with Assistive Technologies**: Explore integration with other assistive technologies to enhance the overall user experience and compatibility with existing tools.
16. **Regular Updates and Testing**: Keep the system and its components up to date and conduct rigorous testing and quality assurance to maintain functionality and performance.

By focusing on these functional aspects, we can create a comprehensive and effective system for assisting visually impaired individuals in various aspects of their daily lives.

3.2 System Overview Diagram

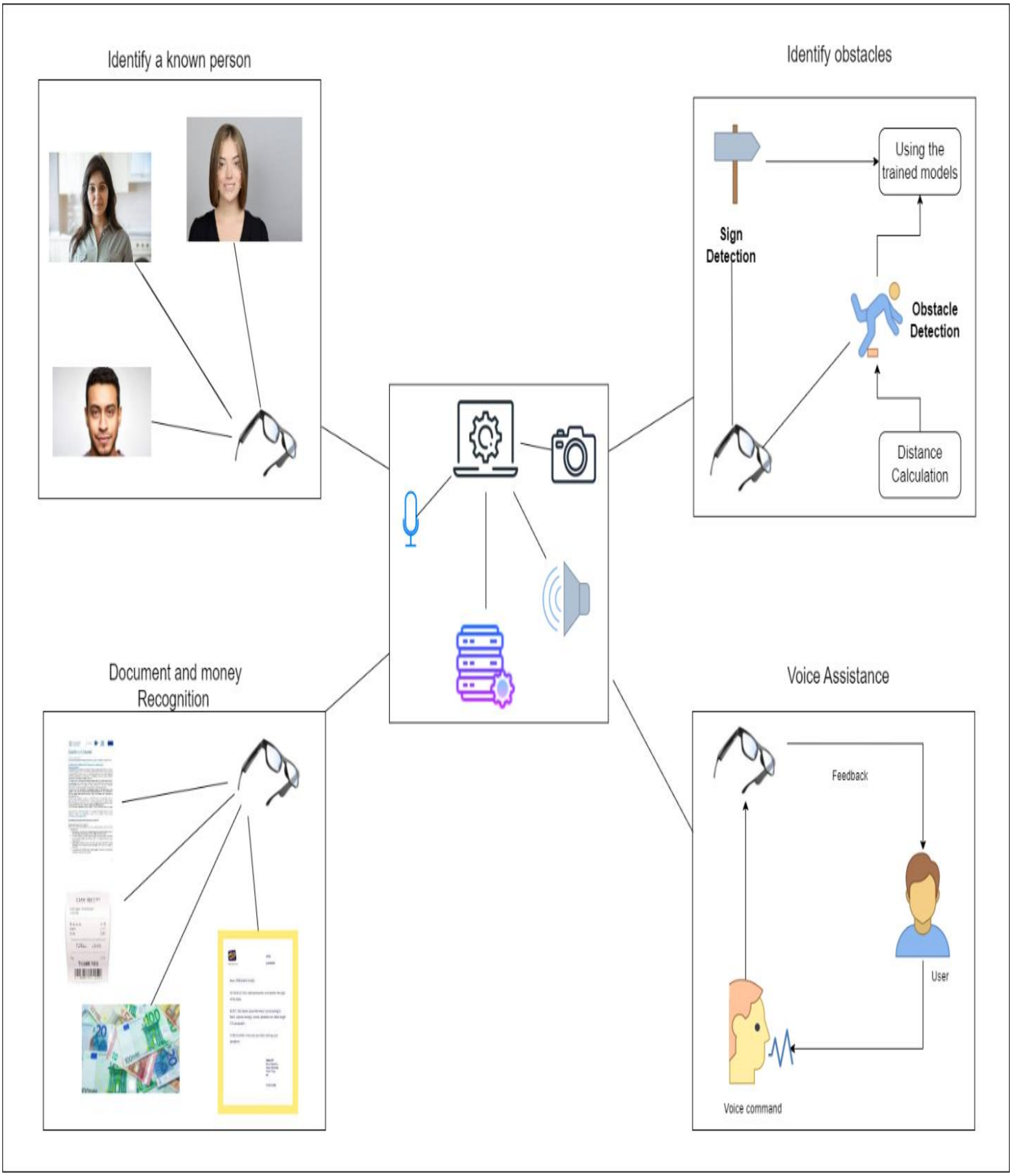

Figure 1 System overview diagram

3.3 Design Phase - Individual Component

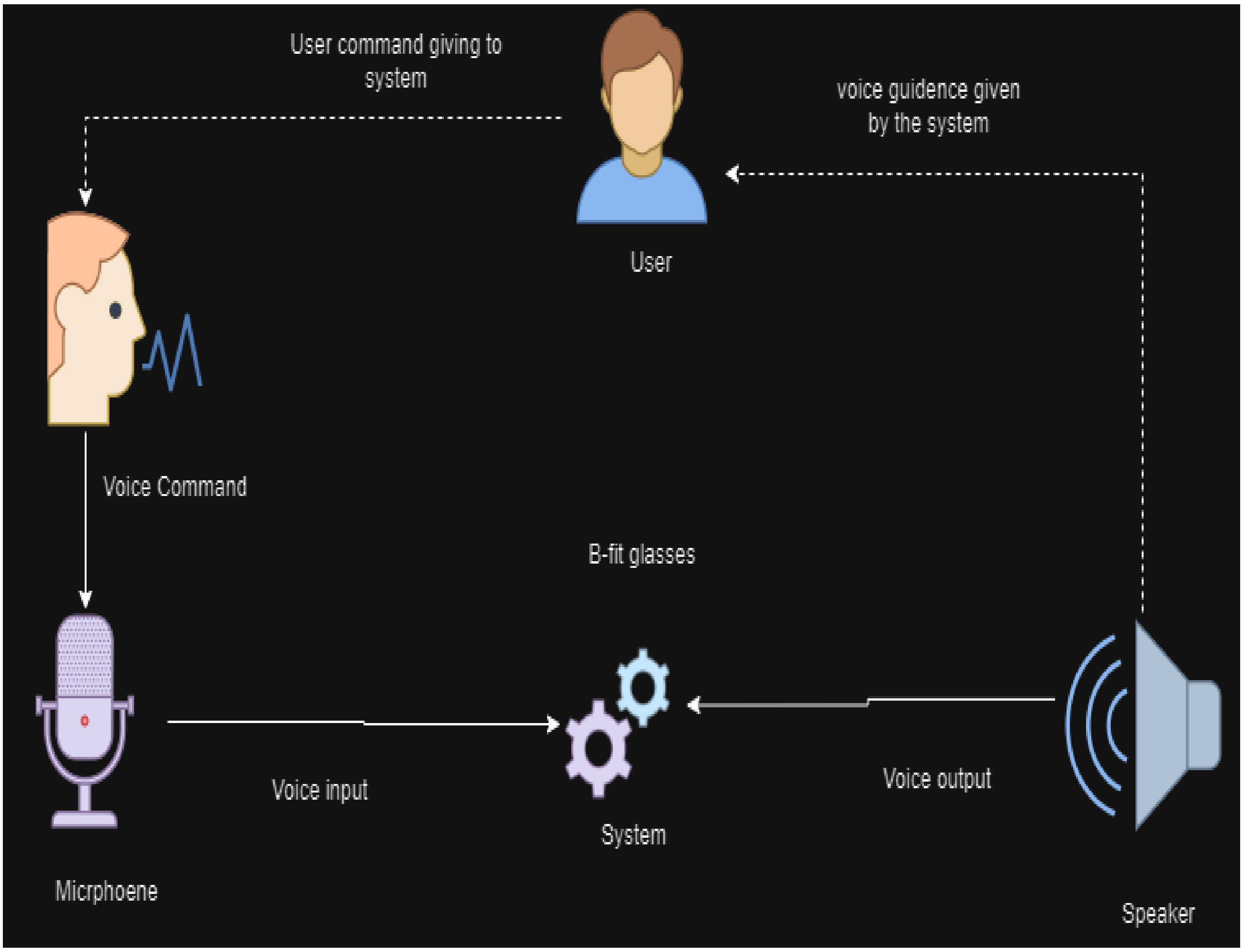

Figure 2 Individual component

3.4 Project Requirements

3.4.1 Functional Requirements

- Voice Recognition and Command Processing:
  - The system must accurately recognize and process voice commands from users.
  - It should be capable of handling multiple similar meaning keywords for a single command.
- Client-Server Communication:
  - Implement a reliable communication protocol between the client-side panel PC and the server-side laptop for data transfer.
- Text-to-Speech Conversion:
  - The system should use Google Text-to-Speech to convert text-based information into audible speech for the user.
- Speech-to-Text Conversion:
  - Utilize Google Speech-to-Text to convert spoken words from the user into text for processing.
- Natural Language Processing (NLP):
  - Implement NLP techniques for preprocessing voice commands to enhance understanding and accuracy.
- Voice-Guided Assistance:
  - The system should provide real-time voice guidance to assist the user in navigation and tasks.

3.4.2 Non-Functional Requirements

- Performance:
  - The system should respond to user commands with minimal latency, ensuring a smooth and real-time user experience.
- Scalability:
  - The architecture should be scalable to handle an increasing number of users and data without significant performance degradation.

- Reliability:
    - The system should operate reliably without frequent failures or downtime.
    - It should be able to recover gracefully from unexpected errors or interruptions.
- Availability:
    - The system should be available for use 24/7 or during specified operational hours, with high uptime.
- Security:
    - Implement robust security measures to protect user data and privacy.
    - Ensure secure communication between the client and server components.
- Usability:
    - The system should have an intuitive and user-friendly interface to cater to visually impaired users.
    - It should be easy for users to understand and navigate the voice commands.
- Accessibility:
    - Ensure that the system complies with accessibility standards, making it usable for people with various degrees of visual impairment.
    - Support screen readers and Braille displays.
- Compatibility:
    - Ensure compatibility with a range of devices and operating systems, including different screen readers and assistive technologies.
- Maintainability:
    - The system should be designed in a modular and maintainable way to allow for updates and improvements.
- Documentation:
    - Provide comprehensive documentation for system administrators and end-users to facilitate installation, configuration, and use.
- Performance Monitoring:
    - Implement performance monitoring and logging to track system usage, identify bottlenecks, and address performance issues.
- Data Backup and Recovery:
    - Regularly back up user data and ensure that there are mechanisms in place for data recovery in case of data loss.
- Compliance:
    - Ensure that the system complies with relevant laws and regulations, including data protection and accessibility standards.
- Interoperability:
    - If applicable, ensure that the system can integrate with other assistive technologies and software used by visually impaired individuals.
- Response Time:

    - Define acceptable response time thresholds for various system functions to maintain user satisfaction.
- Training and Support:
    - Provide training and support resources for system administrators and users to troubleshoot issues and maximize the system's benefits.
- Environmental Considerations:
    - Consider the environments in which the system will be used and ensure that it functions reliably under various conditions (e.g., noise levels).

3.5 System Design

- Voice Detection:
    - Implement a voice detection module on the client-side that can continuously listen for voice input.
    - Utilize voice activity detection algorithms to determine when the user is speaking.
- Voice Command Recognition:
    - Develop a voice command recognition module that can process the user's speech.
    - Train a natural language processing model (using Keras and NLTK) to understand and interpret voice commands.
- Keyword Expansion:
    - Enable the system to recognize multiple similar keywords for a single command. This can be achieved through natural language understanding and synonym matching.
- Client-Server Communication:
    - Establish communication protocols between the panel PC (client) and the laptop (server) to pass voice data and commands.
    - Use network sockets or other suitable methods for data transfer.
- Voice Feedback:

- Implement text-to-speech (TTS) capabilities on both the client and server sides. I mentioned using Google's TTS service, which can be integrated into the system.
- Provide voice feedback to confirm user commands, acknowledge successful actions, and relay information.

- Machine Learning Models:

  - Train machine learning models for identify user commands.
  - Continuously update and fine-tune these models to improve accuracy.
- Concurrent Processing:

  - Use multi-threading on both the client and server sides to ensure concurrent processing of voice detection and command execution.
  - This allows the system to remain responsive and handle real-time interactions.

- User Guidance:

  - Implement voice-guided prompts and instructions to assist users in using the system effectively.
  - Provide clear and concise guidance on available commands and how to navigate through the system.

- Output Presentation:

  - Design the system to present output in an easily comprehensible manner.
  - When the system identifies known persons, signs, money, or reads documents, ensure that this information is communicated clearly through voice feedback.

3.6 Work Breakdown Structure and Gantt Chart

3.6.1 Work Breakdown Structure

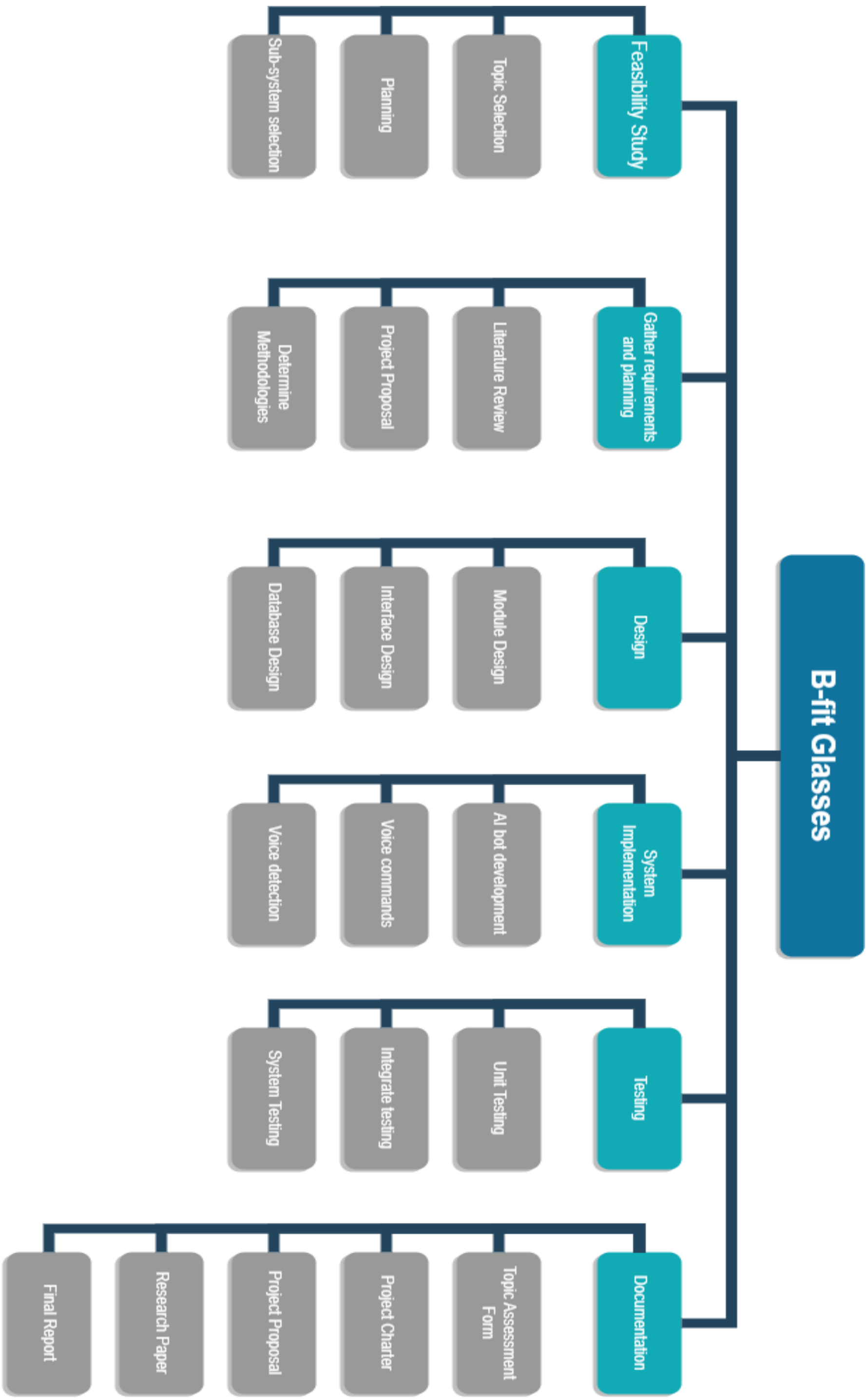

Figure 3 Work breakdown structure

3.6.2 Gantt Chart

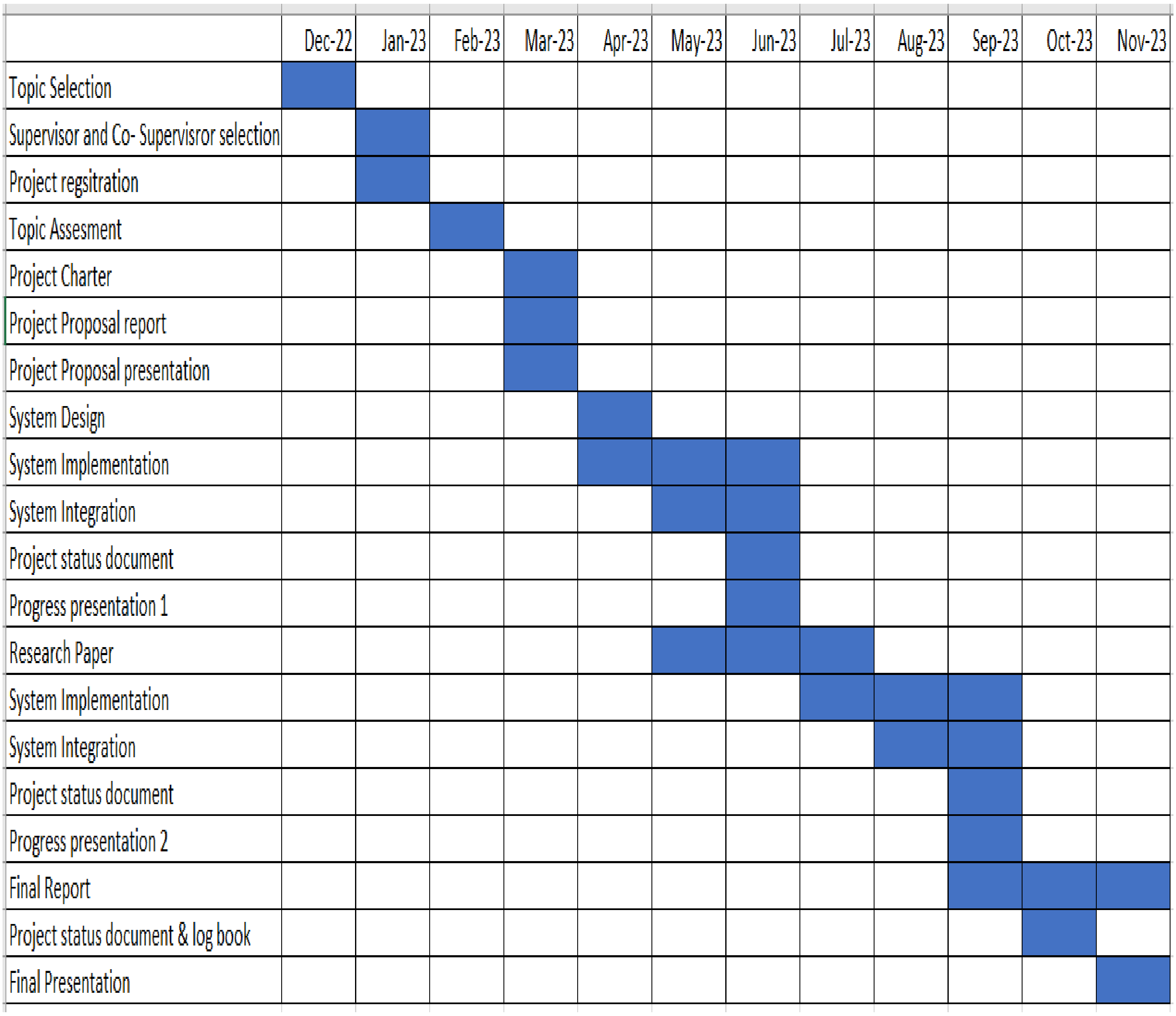

Figure 4 Gantt chart

3.7 Deployment, Marketability and Commercialization

1. Deployment
    a. Documentation:
        i. Develop comprehensive documentation that includes user manuals, installation guides, and troubleshooting resources.
        ii. Ensure the documentation is user-friendly and accessible to both technical and non-technical users.
        iii. Make the documentation available in digital formats and offer customer support channels for assistance.

    b. Pilot Version:

        i. Launch a pilot version of our system to a select group of users or organizations.
        ii. Use this phase to gather feedback, identify potential issues, and make necessary improvements.
        iii. Ensure that the pilot version is stable and user-friendly.

2. Marketability and Commercialization:

    a. Content Marketing - Video Infographics:

        i. Create engaging video infographics that showcase the system's features and benefits.
        ii. Highlight real-life scenarios where the system proves valuable for visually impaired individuals.
        iii. Share these videos on social media, our website, and video-sharing platforms to raise awareness.

    b. Flyer:

        i. Design visually appealing flyers that provide a concise overview of our system's capabilities.
        ii. Include compelling visuals and a clear call to action.
        iii. Distribute physical copies at relevant events and digitally through email campaigns.

    c. Email Marketing Strategy:

        i. Develop a targeted email marketing strategy:

ii. Segment our email list to tailor messages to different user groups.
iii. Send regular newsletters, product updates, and success stories.
iv. Use email automation to nurture leads and engage with subscribers.

d. User Testimonials and Case Studies:

i. Collect user testimonials and create case studies that illustrate how our system has positively impacted visually impaired individuals.
ii. Share these success stories on our website, in marketing materials, and during sales pitches.

e. Partnerships and Collaborations:

i. Seek partnerships with organizations, schools, and institutions serving the visually impaired community.
ii. Collaborate on joint marketing efforts and distribution channels to reach a broader audience.

f. Accessibility and Inclusivity Promotion:

i. Emphasize the accessibility and inclusivity features of our system in all marketing materials.
ii. Highlight how the system enhances independence and improves the quality of life for visually impaired users.

g. Feedback Mechanism:

i. Establish a feedback mechanism that allows users to provide input and suggestions.
ii. Demonstrate that we value user feedback by actively incorporating it into product improvements.

h. Pricing and Subscription Models:

i. Define competitive pricing strategies and subscription models that cater to different user segments.
ii. Consider offering trial periods or discounts to encourage adoption and loyalty.

i. Social Media Presence:

i. Maintain an active and engaging presence on relevant social media platforms.
ii. Share regular updates, educational content, and user-generated content to foster a strong online community.

j. Analytics and Performance Monitoring:

i. Implement analytics tools to track the performance of marketing efforts.
ii. Monitor metrics such as website traffic, conversion rates, email open rates, and social media engagement.
iii. Use these insights to refine our marketing strategy.

k. Regulatory Compliance and Certifications:

i. Ensure our system complies with accessibility and data privacy regulations.
ii. Obtain any necessary certifications or endorsements that can instill trust and confidence in potential users.

l. Continuous Improvement:

i. Continuously gather market feedback, analyze competitor activities, and adapt our marketing and commercialization strategies to evolving market conditions.

4 IMPLEMENTATION AND TESTING

4.1 Implementation

1. System Architecture Design:

- Begin by defining the overall architecture of our system. Decide how the various components will interact and communicate with each other.
- Choose the appropriate technologies and frameworks for each component, keeping in mind scalability and compatibility.

2. Development Environment Setup:

- Set up development environments on both the client-side (panel PC) and server-side (laptop) systems.
- Install the necessary programming languages, libraries, and development tools.

3. Device model Design:
    - Design a wearable device which is used to set up the camera, sensors, and Arduino board.
    - Consider the weight and usability of the device model.

4.1.1 Data Collection

In the voice detection component of our system, the goal is to accurately capture and understand voice commands from users. To achieve this, we utilize a customized dataset that contains predefined commands. This dataset serves as a reference for the system to recognize specific commands and trigger corresponding actions. The dataset typically consists of a variety of commands that visually impaired users might use to interact with the system effectively.


```
{"intents": [


  {"tag": "active_doc_read",
   "patterns": ["active document read", "document read active","document read" ],
   "responses": ["ok document reading activated"]
  },
  {"tag": "active_person_recog",
   "patterns": ["active person recognition system", "person recognition system active", "person recognition", "active person recognition"],
   "responses": ["ok person recognition system activated"]
  },
  {"tag": "active_sign_recog",
     "patterns": ["active sign recognition system", "sign recognition system active", "sign recognition", "active sign recognition"],
     "responses": ["ok sign recognition system activated"]
  },
  {"tag": "bot_name",
     "patterns": ["what is your name", "your name", "who are you"],
     "responses": ["my name is b-fit", "hellow i am b-fit", "You can call me b-fit"]
  }
]
}
```


Figure 5 Dataset

4.1.1.1 Voice detection

Voice detection is a core functionality that captures and interprets user's spoken commands. It includes audio capture, noise reduction, speech-to-text conversion, parsing, and matching with a customized dataset of predefined commands. When a match is found, the system triggers the associated action, providing users with voice-guided assistance. Regular updates, including similar meaning keywords, enhance accuracy and adaptability.

4.1.1.2 Voice command

Voice commanding is the user interaction process where the system recognizes and executes specific actions or functions based on voice commands spoken by the user. It involves capturing and interpreting voice input, matching it with predefined commands, and triggering corresponding actions, such as object detection, known person recognition, sign detection, document reading, or money note recognition. Voice commanding enables visually impaired users to control and receive information from the system through spoken instructions.

4.1.2 Data Pre-Processing

Data preprocessing for voice commands is the initial step in making sense of user-spoken instructions. This process involves cleaning and structuring the captured audio input to improve the accuracy of speech recognition. It may include noise reduction, audio segmentation, and speech-to-text conversion, transforming the raw audio into a textual representation that the system can understand and act upon. Data preprocessing ensures that voice commands are clear, noise-free, and ready for accurate interpretation and execution by the system.

4.1.2.1 Convert speech to text

In our system, we utilize Google Speech-to-Text technology to convert spoken words or audio input into written or textual form. This conversion process is essential for interpreting user voice commands. Once the spoken content is accurately converted into text, it is further encoded into the UTF-8 format. UTF-8 (Unicode Transformation

Format 8-bit) is a character encoding standard that allows the representation of a broad range of characters and symbol sets from different languages and scripts. Encoding the text into UTF-8 ensures compatibility with various character sets, making it suitable for handling diverse user inputs and commands effectively within the system.

4.1.2.2 Removing unwanted characters and tokenization

In our system's data preprocessing phase, we

employ the Natural Language Toolkit (NLTK) to enhance the quality and structure of textual data, which is obtained from spoken voice commands converted into text. This process involves two key steps:

Removing Unwanted Characters: Initially, the system identifies and removes unwanted or irrelevant characters from the text. These may include special symbols, punctuation marks, numbers, or any other non-essential characters that do not contribute to the understanding of the command. The purpose is to clean the text and make it more suitable for analysis.

Tokenization: Following the character removal step, the NLTK library is utilized for tokenization. Tokenization is the process of breaking down the cleaned text into individual units or tokens, typically words or phrases. NLTK's tokenization tools segment the text into these meaningful units, which are then used for further analysis, understanding, or processing of the voice command.

4.1.3 Model Training

4.1.3.1 Voice detection and commanding

Model comparison for voice command and detection is a critical phase in developing a system tailored for visually impaired users. It begins by selecting a set of candidate machine learning models, each with its unique architecture and attributes. These models are then subjected to a rigorous evaluation process. The dataset is meticulously split into training, validation, and testing subsets to facilitate model training and assessment. During training, the models adjust their internal parameters based on the training data, and hyperparameter tuning further refines their performance. The validation dataset helps monitor the models for accuracy and any potential overfitting or underfitting issues. Subsequently, the models undergo final testing using a separate dataset to evaluate their real-world effectiveness. Performance metrics, including accuracy and precision are computed for each model. The model that consistently demonstrates the highest accuracy and meets other relevant criteria is selected as the most suitable choice for the system. Once chosen, this model can be deployed to provide accurate voice command recognition and detection, enhancing the user experience for visually impaired individuals.

4.1.4 Results and Discussion

4.1.4.1 Voice commanding and detection

The results of the voice commanding and detection model training are promising, with a commendable accuracy rate of 90%. This high level of accuracy signifies a significant milestone in the development of a system designed to assist visually impaired individuals. Achieving such precision in recognizing voice commands suggests that the model has a strong grasp of the intended user instructions, leading to more effective interactions and responses. While this accuracy rate is a substantial achievement, it is important to note that further fine-tuning and testing will be essential to ensure consistent performance across various real-world scenarios and user voices. Additionally, continuous data collection and model refinement are crucial for accommodating the diverse range of commands and user preferences that may arise. This accomplishment paves the way for enhancing the system's functionality and usability, ultimately providing visually impaired users with a more seamless and empowering experience.

4.2 Technology Stack

Library	Keras
Language	Python

Editor	PyCharm
Packages	tensorflow-cpu, gTTS==2.3.1, keras==2.12.0, nltk==3.8.1, numpy==1.23.5
Version Controlling	Git Lab

Table 3 Technology Stack

4.3 Testing

Testing the voice command and detection functionality is a pivotal phase in the development of our system tailored for visually impaired users. With an acute focus on accuracy and reliability, we meticulously evaluate the system's performance across diverse test scenarios. These scenarios encompass a broad spectrum of voice commands and user interactions, including typical requests and challenging edge cases. Our testing environment is deliberately designed to mimic real-world conditions, introducing various acoustic challenges such as background noise, differing accents, and varying microphone qualities. We employ rigorous evaluation metrics, including accuracy and precision to gauge the system's proficiency in recognizing and correctly responding to voice commands. Furthermore, user feedback is an integral part of the testing process, allowing us to assess usability and cater to user-specific requirements. As we navigate this testing phase, the ultimate goal is to refine and fine-tune the system, ensuring it consistently delivers an accurate and empowering experience to visually impaired individuals. Continuous testing, coupled with model adaptation and scalability assessments, forms the cornerstone of our commitment to providing a dependable and user-centric voice command and detection system.

Test Case 01

Test Case ID	01
Test Priority	High
Associated User	User
Pre-conditions	-

Test Case Title	Testing user commands in different environment
Testing Process	Activate functions according to the user command
Expected Result	Should identify user command and trigger the function relevant to the command
Actual Result	Expected results are achieved.
Pass/Fail	Pass

Table 4 Test case 1

Test Case 02

Test Case ID	02
Test Priority	High
Associated User	User
Pre-conditions	-
Test Case Title	Testing user commands with different accent
Testing Process	Activate functions according to the user command
Expected Result	Should identify user command and trigger the function relevant to the command.
Actual Result	Expected results are achieved.

Pass/Fail	Pass

Table 5 Test case 2

Test Case 03

Test Case ID	03
Test Priority	High
Associated User	User
Pre-conditions	-
Test Case Title	Testing user commands with different microphones
Testing Process	Activate functions according to the user command
Expected Result	Should identify user command and trigger the function relevant to the command.
Actual Result	Expected results are achieved.
Pass/Fail	Pass

Table 6 Test case 3

5 SOCIAL AND ETHICAL CONCERNS

Integrating voice command and detection technology into systems for visually impaired individuals is a transformative step towards enhancing accessibility and usability. However, this innovation comes with significant social and ethical considerations. Privacy and data security are paramount, demanding robust measures to safeguard user voice data and guarantee its protection. Transparency and informed consent are equally crucial, ensuring that users understand how their data will be used. Accessibility and inclusivity must be at the forefront, ensuring that the technology benefits all users regardless of their abilities. Bias and fairness concerns require continuous efforts to minimize discrimination in voice recognition algorithms. Ownership of voice data and its ethical use necessitate clear guidelines. User empowerment, adherence to accessibility standards, and ongoing education are key to responsible implementation. Providing redundancy and alternative interaction methods is vital for situations where voice recognition may fall short. Lastly, a commitment to continuous monitoring and improvement underscores our dedication to addressing evolving ethical concerns and meeting the unique needs of visually impaired users. In navigating these social and ethical dimensions, we aim to create a voice command and detection system that is not only technologically advanced but also ethically sound and user centric.

6 DESCRIPTION OF PERSONAL AND FACILITIES

Member	**Component**	**Task**
Kakulandara P.M	Voice command and detection	Architect the system's overall architecture for integrating smart glasses with a server for voice detection and commanding.
		Develop and refine the voice detection model, including dataset curation, preprocessing, model design, and iterative training and fine-tuning.
		Ensure data privacy and compliance with relevant regulations in data management.
		Integrate the trained model with the smart glasses and server, including designing efficient communication

		and optimizing model performance for the smart glasses' computational resources.
		Design and execute comprehensive test scenarios, real-world experiments, and performance metrics to validate the system's reliability and precision.
		Utilize the high-performance computing cluster within the research for model development, exploring deep learning frameworks, adjusting hyperparameters.
		Collaborate with a multidisciplinary team to ensure user-centric design, providing a seamless, intuitive, and user-friendly experience for visually impaired individuals using smart glasses.

Table 7 Personal

Facilities:

Facility	Description
Internet Connection	This is crucial to enable real-time communication between the server and the smart glasses. It must be able with Google Text to Speech.
Panel PC	acts as a dedicated computing unit in the client side. It provides the necessary processing power to run the converting text to speech, speech to text, encoding decoding voice efficiently.
3D printed device model	a useful tool for producing and developing smart glasses parts. It enables the rapid production of specialized components while also allowing for customization, ensuring consumers have an ergonomic fit. The smart glasses are also made more user-friendly by supporting the iterative design and development process.

Table 8 Facility

7 RESULTS AND DISCUSSION

7.1 Results

7.1.1 Voice detection and commanding

```
Recognized activa document print
1/1 [==============================] - 0s 169ms/step
Recognized document read
1/1 [==============================] - 0s 37ms/step
Recognized dog breed like you
1/1 [==============================] - 0s 36ms/step
Recognized document read active
1/1 [==============================] - 0s 50ms/step
i cant hear you properly tell again plz
1/1 [==============================] - 0s 33ms/step
Recognized your name
1/1 [==============================] - 0s 51ms/step
Recognized what's your name
1/1 [==============================] - 0s 32ms/step
i cant hear you properly tell again plz
1/1 [==============================] - 0s 30ms/step
i cant hear you properly tell again plz
1/1 [==============================] - 0s 41ms/step
time out internet ping to high ..
```

Figure 6 User commands and activate functions

```
Server response: Server received:  you'r friend nipun is here
Server response: Server received:  you'r friend hirosha is here
Server response: Server received:  you'r friend pamuditha is here
```

Figure 7 Known person detection response

After identifying the known person system output voice saying “your friend (friends name) is here”.

```
Server response: Server received:  show this signs 1 stop
Server response: Server received:  show this signs (no detections)
Server response: Server received:  show this signs 1 crossing
Server response: Server received:  show this signs (no detections)
Server response: Server received:  show this signs 1 no_entry, 1 stop
Server response: Server received:  show this signs 1 bus_stop
```

Figure 8 Sign detection response

After identifying the sign system output voice saying "1 (sign name) detected”.

```
20
Server response: Server received:  this is 20 sri lankan rupee bill
100
Server response: Server received:  this is 100 sri lankan rupee bill
100
Server response: Server received:  this is 100 sri lankan rupee bill
500
Server response: Server received:  this is 500 sri lankan rupee bill
500
Server response: Server received:  this is 500 sri lankan rupee bill
500
Server response: Server received:  this is 500 sri lankan rupee bill
1000
Server response: Server received:  this is 1000 sri lankan rupee bill
1000
Server response: Server received:  this is 1000 sri lankan rupee bill
```

Figure 9 Money note detection response

After identifying the money note system output voice saying, “this is (value of the money note) Sri Lankan rupee bill”.

```
GOOD IDEA
SECRET SEVEN ADVENTURE had moustaches, this one fitted che bllhat ruled them bill out he was but had no moustache! dark-haired and Could he Jeap high? Would he show that he cou
and   really   seemed to enjoy bears came On. themselves boxing with each other   and with their trainer. One little bear was s0 fond of its trainer that it kept hugging his le
```

Figure 10 Document read response

After identifying the document system output the content of the document.

7.2 Integrated Smart Glasses

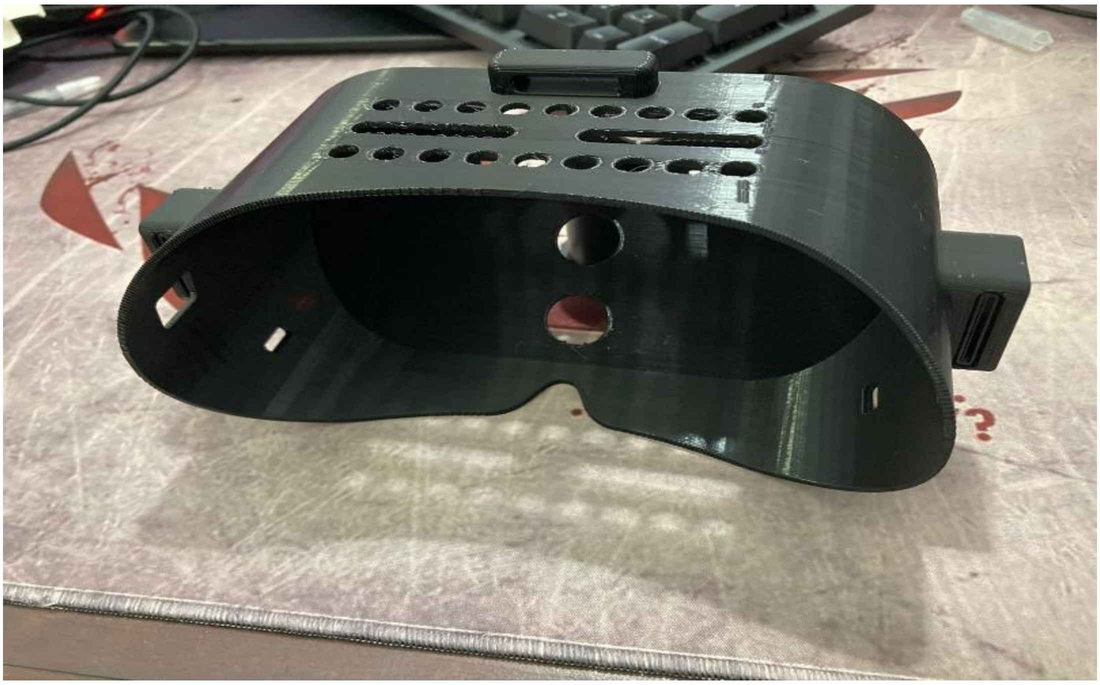

Figure 11 3D printed model (a)

Figure 12 3D printed model (b)

Figure 13 Setuped device (a)

Figure 14 Setuped device (b)

Figure 15 Device front view

Figure 16 Device side view

7.3 Machine Learning Models

7.3.1 Voice detection and command model

7.3.1.1.1 Model architecture:

The model architecture utilized to follows a neural network structure designed for intent recognition from voice commands. It begins with the preparation of a labeled dataset, where each voice command is associated with specific word patterns representing its intent. The input layer consists of neurons equal to the length of the word patterns, while the output layer matches the number of intent classes. Between these layers are two hidden layers of 128 and 64 neurons, respectively, with 'relu' activation functions for non-linearity. Dropout layers with a 50% dropout rate are incorporated to prevent overfitting and improve generalization. The model employs the stochastic gradient descent (SGD) optimizer with learning rate adjustments, decay, momentum, and Nesterov momentum for efficient training. The loss function used is categorical cross-entropy, suitable for multi-class classification tasks. Over a span of 200 epochs and with a batch size of 5, the model iteratively adjusts its parameters using the training data. This architecture aims to capture the intricate relationships between word patterns and intent labels, ultimately enabling accurate intent recognition for voice commands.

7.3.1.1.2 Learning process:

The learning process of the model is involves a supervised training approach, where the model learns to recognize intents from voice commands through an iterative process. The training begins with the preparation of a labeled dataset, where each voice command is associated with specific word patterns denoting its intent. During each training epoch, the model receives batches of voice command data and corresponding intent labels. The stochastic gradient descent (SGD) optimizer is employed, with a learning rate of 0.01, decay, momentum, and Nesterov momentum, to efficiently adjust the model's internal parameters. The model aims to minimize the categorical cross-entropy loss, which measures the dissimilarity between predicted and actual intent labels. Over the course of 200 epochs, the model adapts its parameters, fine-tuning its ability to accurately classify voice commands into predefined intent categories. This iterative learning process allows the model to progressively improve its performance, ultimately achieving a high level of accuracy in intent recognition, which is crucial for responsive and effective interactions with users.

7.3.1.1.3 Loss function

The model employs the categorical cross-entropy loss function as a pivotal component of the model's training process. Categorical cross-entropy is a well-suited choice for multi-class classification tasks, such as intent recognition from voice commands. It quantifies the dissimilarity between the model's predicted intent probabilities and the actual intent labels present in the training data. During training, the model continually adjusts its internal parameters to minimize this loss. By doing so, the model effectively learns to assign higher probabilities to the correct intent class while reducing the likelihood of misclassification. In essence, the categorical cross-entropy loss serves as a guiding metric, driving the model toward improved accuracy in intent recognition. It plays a critical role in training the neural network, facilitating the convergence of the model's predictions with the actual intent labels, and ultimately contributing to the system's ability to accurately respond to voice commands.

8 CONCLUSION

In the development and implementation of this visionary system tailored for visually impaired individuals, we have embarked on a transformative journey at the intersection of technology and human empowerment. From the inception of voice detection to the nuanced intricacies of voice command interpretation, this system represents more than a technical marvel; it embodies the spirit of accessibility and inclusivity. By adeptly capturing user voice input, processing it with precision, and offering context-aware responses, it transcends the conventional boundaries that often limit the experiences of visually impaired individuals. In essence, this system becomes a bridge that fosters independence, interaction, and engagement, ensuring that technology becomes a powerful ally for those who depend on it.

The impact of this system extends beyond mere convenience; it is a testament to the human capacity for innovation and empathy. By seamlessly integrating voice recognition and understanding, we have empowered visually impaired users to navigate the digital world with ease. Its role in providing real-time assistance, recognizing voice commands, and delivering spoken feedback is a reflection of our commitment to harnessing technology for the betterment of society. It embodies the idea that technology should adapt to human needs and be accessible to all, irrespective of physical abilities. This system, while remarkable in its current state, also serves as a foundation upon which future advancements in accessibility technology will be built.

As we look to the horizon, the potential of this system is boundless. With ongoing research and development, we anticipate even greater accuracy, expanded functionalities, and broader integration into everyday life. From guiding users through their environments to assisting in document reading, recognizing individuals, and identifying signs and currencies, the versatility of this system opens doors to countless possibilities. It reaffirms the idea that technology can serve as a powerful equalizer, ensuring that visually impaired individuals enjoy the same opportunities and experiences as their sighted counterparts. In doing so, it embodies the core values of inclusion and diversity.

In conclusion, this system represents more than a technological achievement; it is a testament to our commitment to creating a more accessible and inclusive world. By facilitating seamless voice interactions, it empowers visually impaired individuals to engage with technology and the digital realm in ways that were once unimaginable. It underlines our belief in the transformative power of innovation and its ability to break down barriers. As we move forward, we remain dedicated to pushing the boundaries of what technology can achieve, continually striving to enhance the lives of visually impaired individuals and all those who can benefit from the fruits of our labor.

9 BUDGET AND BUDGET JUSTIFICATION

Description	Amount (LKR)
Panel pc	45000.00
Night vision and day vision camera	21000.00
Ultra sonic sensor	1000.00
Microphone	500.00
Speaker	500.00
Battery Pack	3000.00
Database	Free

Arduino board	4000.00
Total	75000.00

Table 9 Budget and budget justification

10 REFERENCES

[1] D. K. B. N. W. Tejal Adep, Visual Assistant for Blind People using Raspberry Pi, 2021.

[2] P. S. Rajendran, P. Krishnan and D. J. Aravindhar, Design and Implementation of Voice Assisted Smart Glasses for Visually Impaired People Using Google Vision API, 28 December 2020.

[3] J.-Y. Lin, C.-L. Chiang, M.-J. Wu, C.-C. Yao and M.-C. Chen, Smart Glasses Application System for Visually Impaired People Based on Deep Learning, 07-15 February 2020.

[4] P. K. ,. S. R. John Aravindhar, Design and Implementation of Voice Assisted Smart Glasses for Visually Impaired People Using Google Vision API, 2020.

[5] S. L. Z. L. K. W. D. L. Jinqiang Bai, Smart Guiding Glasses for Visually Impaired People in Indoor Environment, 2017.

[6] S. L. Z. L. K. W. D. L. Jinqiang Bai, Smart guiding glasses for visually impaired people in indoor environment, 27 September 2017.

[7] R. D. H. N. S. Himadri Nath Saha, Low cost ultrasonic smart glasses for blind, October 2017.

[8] S. M. T. ,. V. S. Deepthi Jain B, Visual Assistance for Blind using Image Processing, 2018.

[9] R. S. &. E. S. Bineeth Kuriakose, Tools and Technologies for Blind and Visually Impaired Navigation Support: A Review, 27 Sep 2020.

[10] T. Asgher, Voice Recognization using Machine Learning Approach, November 2021.

[11] A. L. M. F. ,. F. O. M. ,. F. E. S. Aline Darc Piculo Dos Santos, Aesthetics and the perceived stigma of assistive technology for visual impairment, 2022 Feb;17.

[12] "voice-activity-detection," [Online]. Available: https://github.com/topics/voice-activity-detection.

[13] "Train Speech Command Recognition Model Using Deep Learning," [Online]. Available: https://www.mathworks.com/help/deeplearning/ug/deep-learning-speech-recognition.html.

[14] An Inventory of Voice Input Commands for Users with Visual Impairments and Assistive Smartglasses Applications, 2020.

[15] "An Inventory of Voice Input Commands for Users with Visual Impairments and Assistive Smartglasses Applications," 2020. [Online].

Appendices

Appendix A: Supervisor Project Endorsement

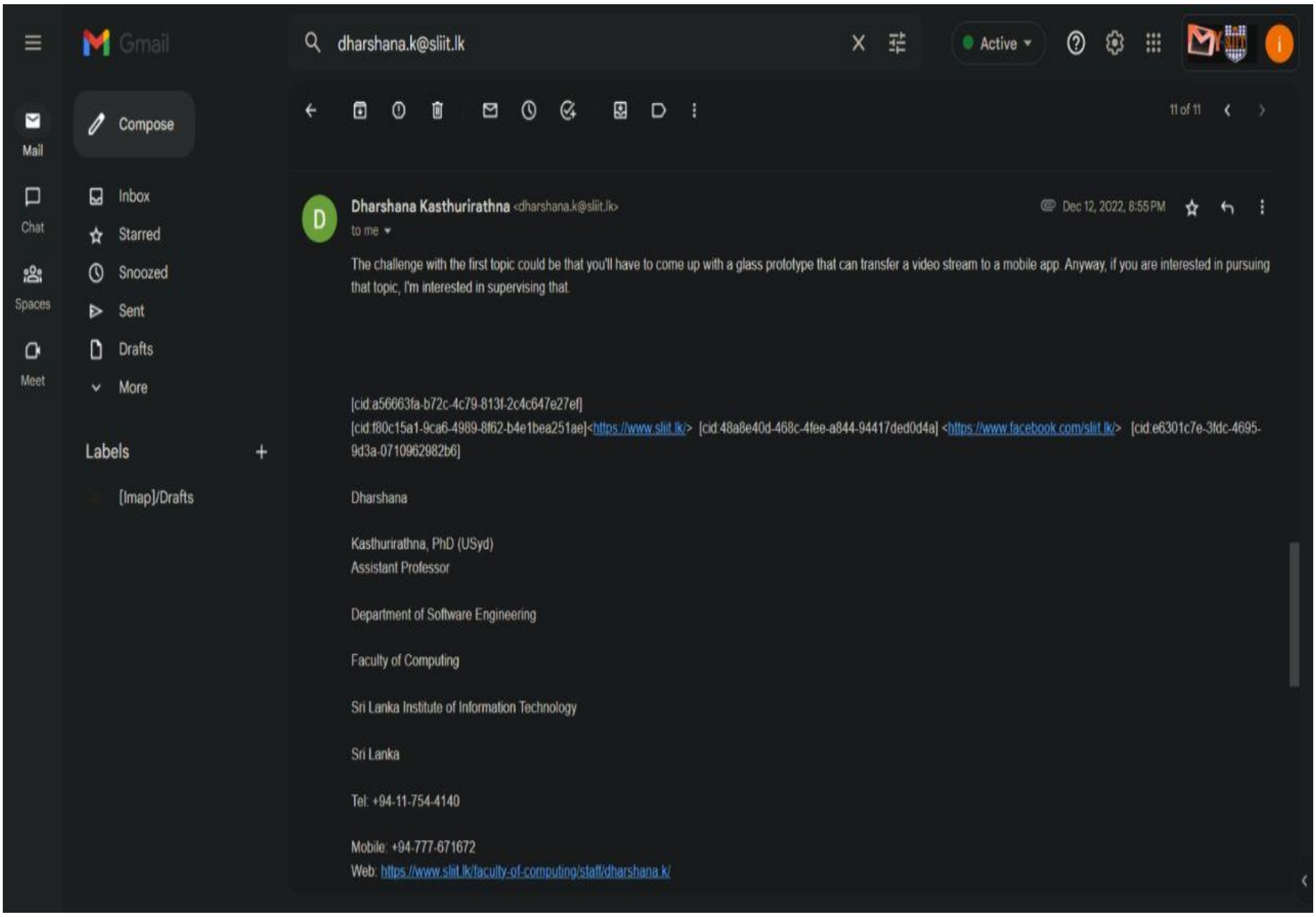

Figure 17 Supervisor project endorsement

Appendix B: Co-Supervisor Project Endorsement

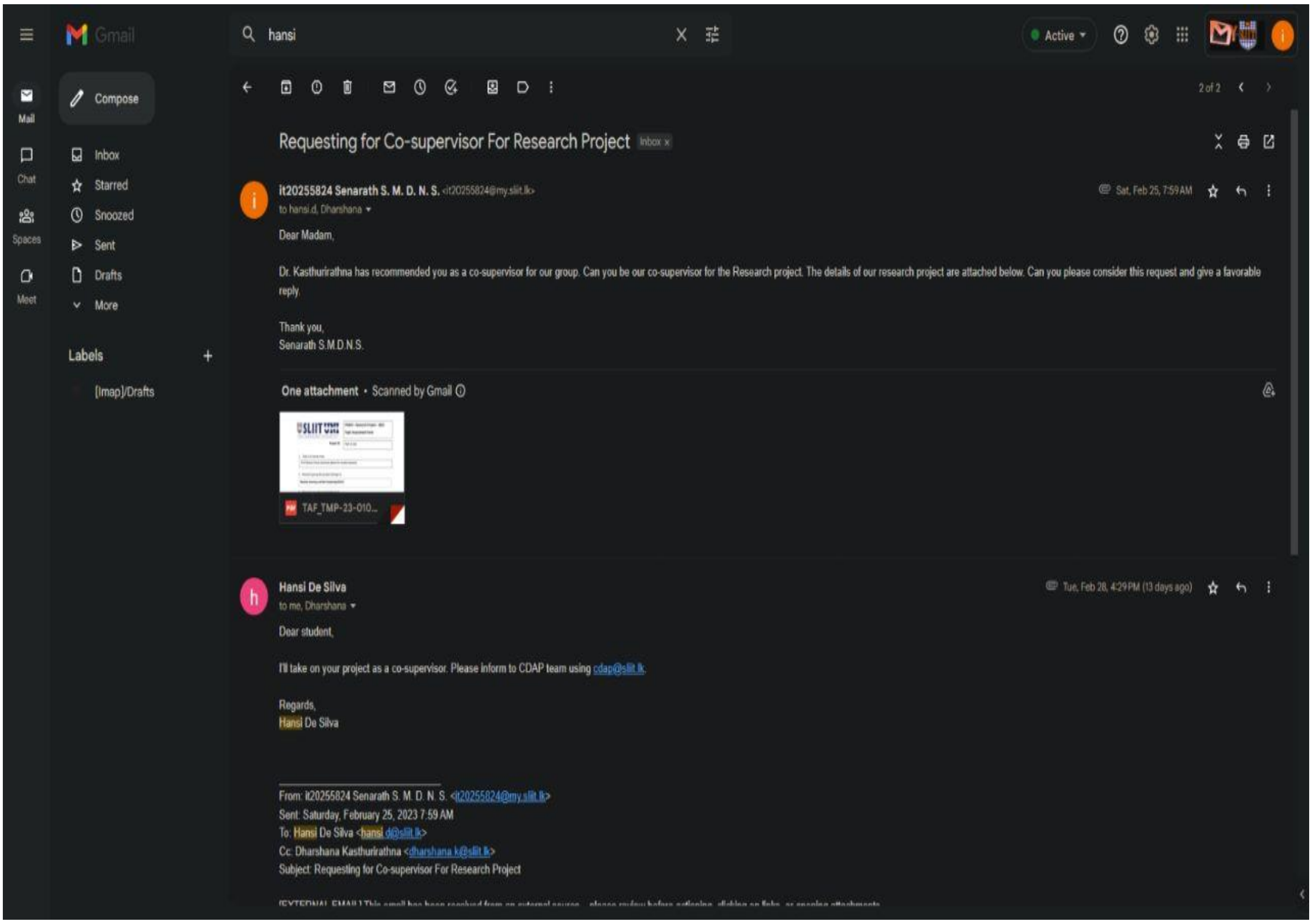

Figure 18 Co-Supervisor project endorsement