Mobile Base Sinhala Book Reader for Visually Impaired Individuals

TMP-23-198

Status Document 2

Semini J.P.D.L. – IT20241346

Bachelor of Science (Hons) Degree in Information Technology

Specializing in Information Technology

Department of Information Technology

Faculty of Computing

Sri Lanka Institute of Information Technology

Sri Lanka

September 2023

1. Microsoft Teams Details

1.1 Teams group channels

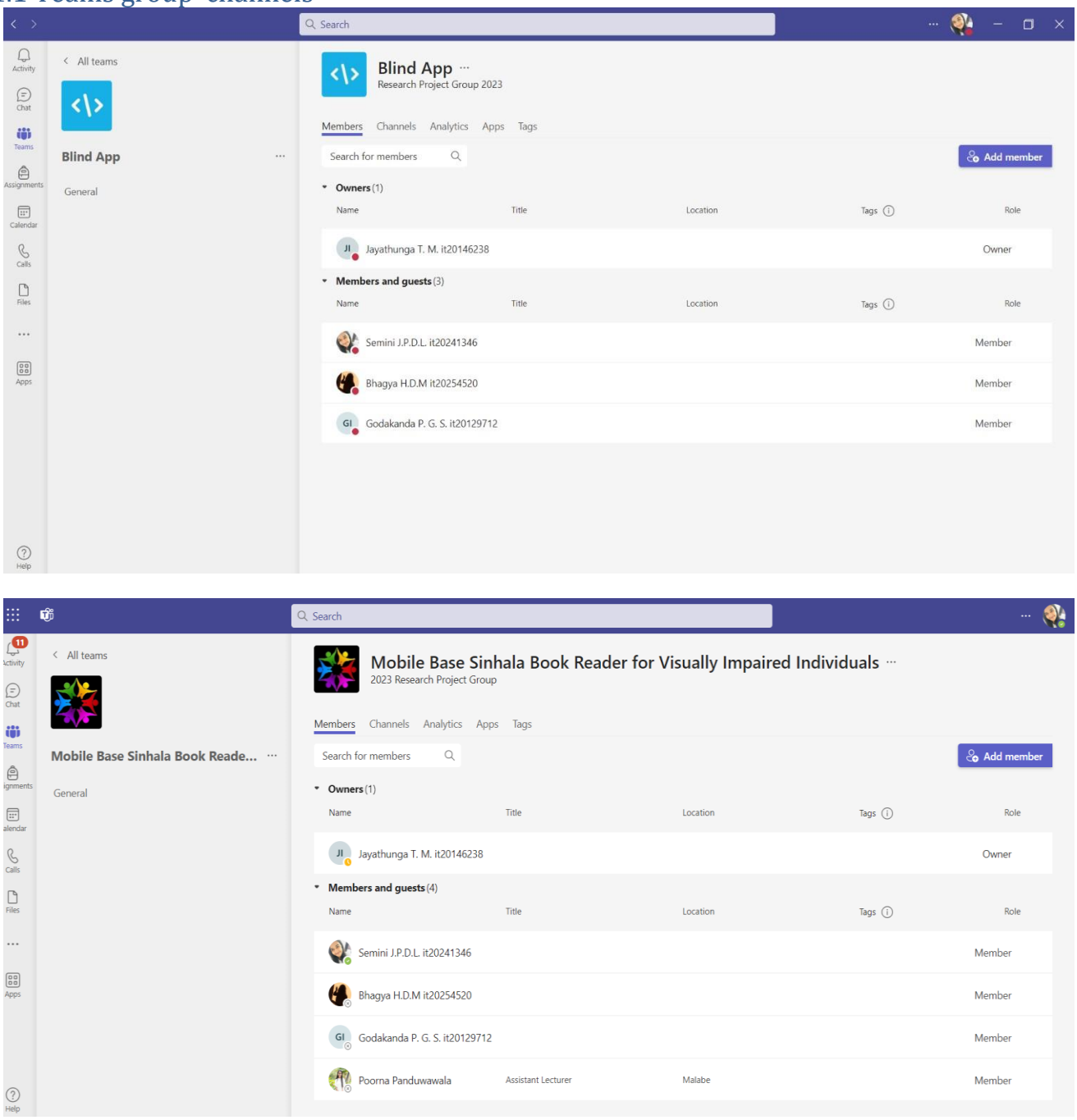


Figure 1-Team members

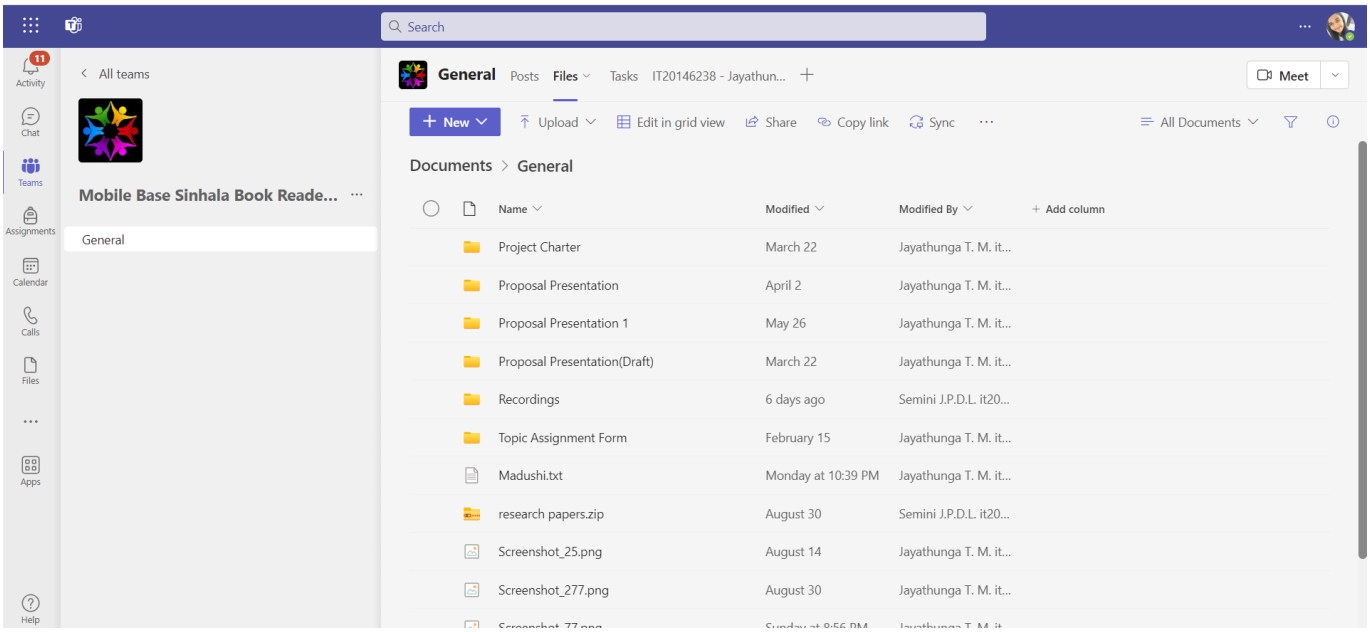


Figure 2-Team documents

1.2 Team Chat Meetings

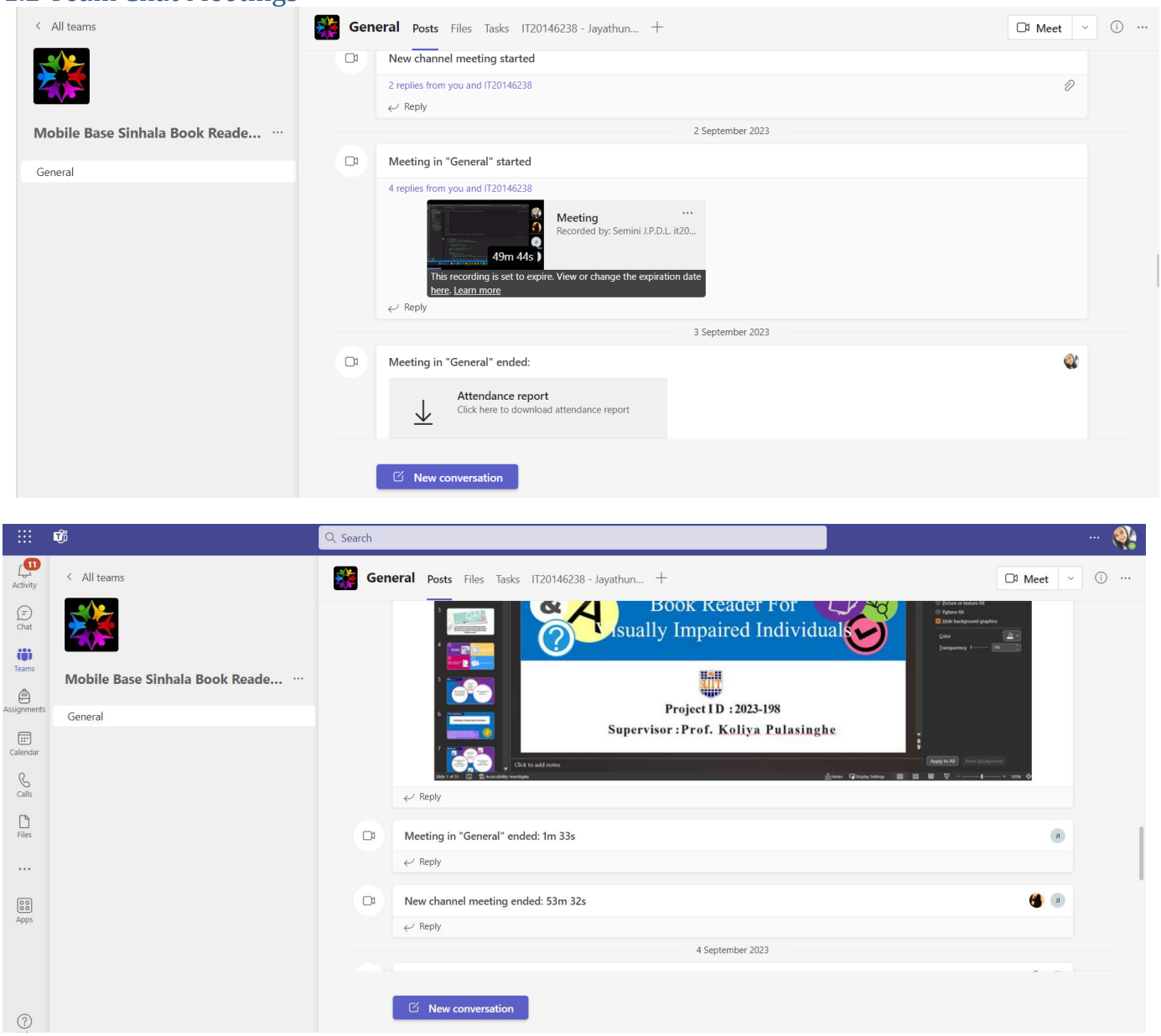


Figure 3-Team chat meetings

1.3 Teams chat group with the supervisor

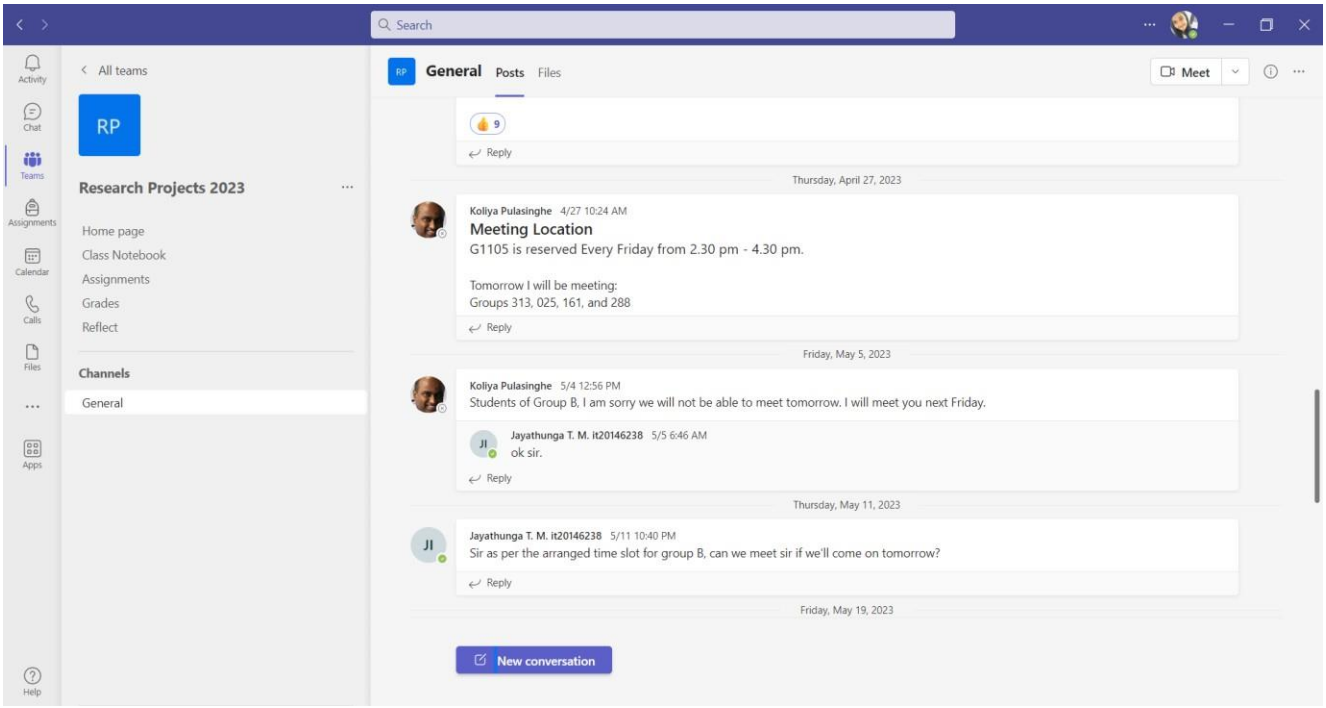
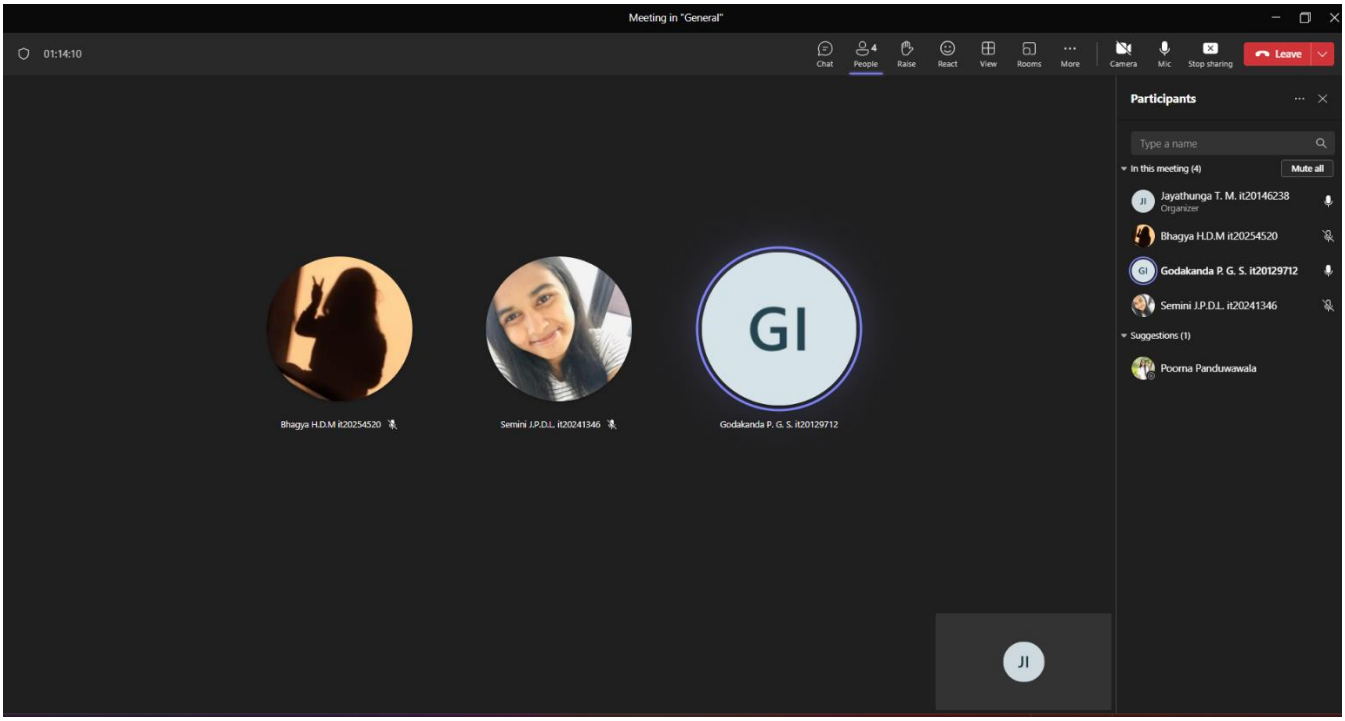
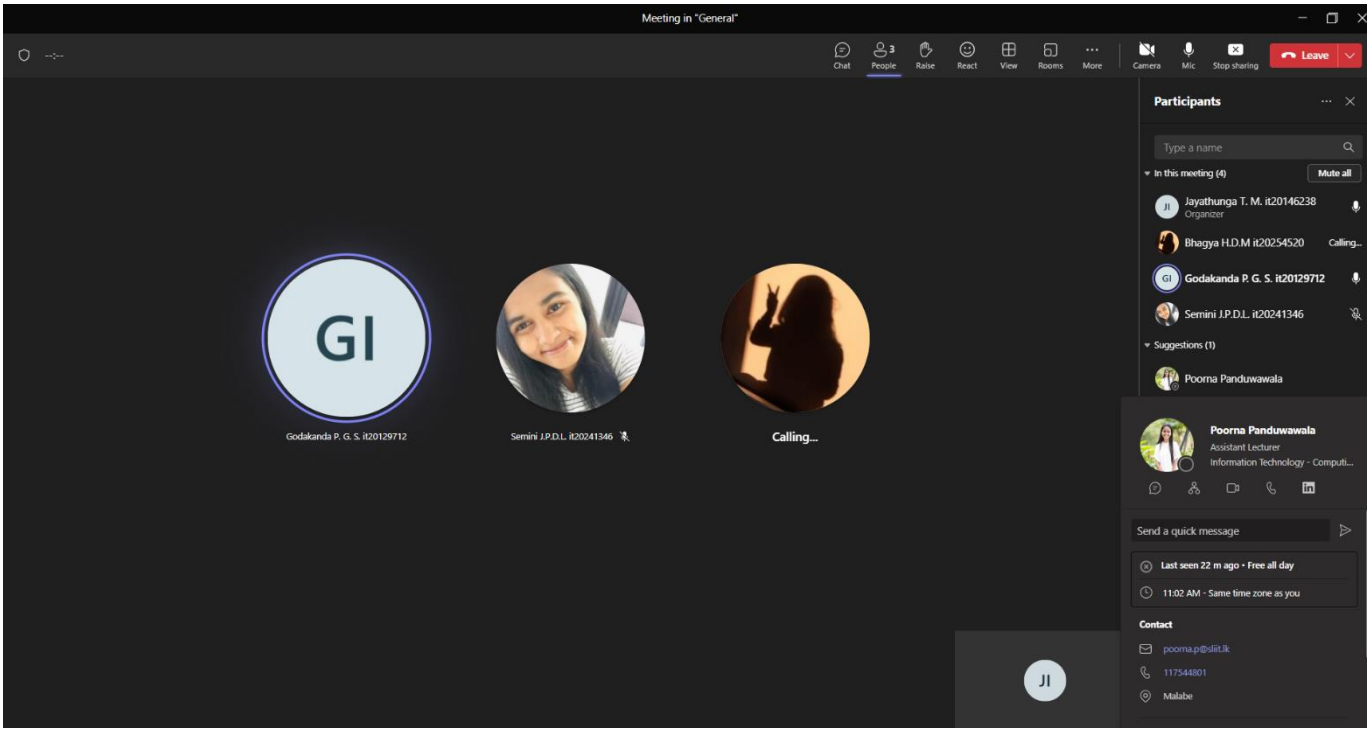
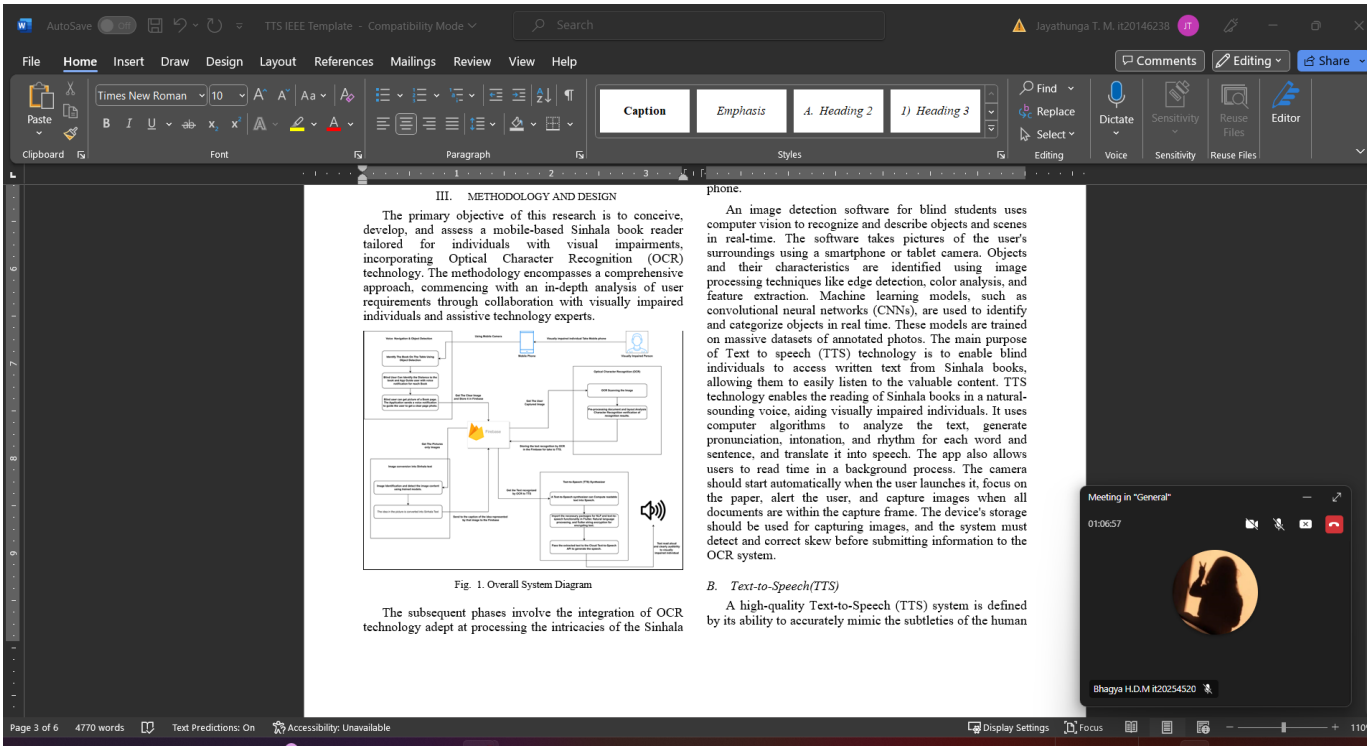


Figure 4-Team group with the supervisor

1.4 Teams calls with research team







III. METHODOLOGY AND DESIGN

The primary objective of this research is to conceive, develop, and assess a mobile-based Sinhala book reader tailored for individuals with visual impairments, incorporating Optical Character Recognition (OCR) technology. The methodology encompasses a comprehensive approach, commencing with an in-depth analysis of user requirements through collaboration with visually impaired individuals and assistive technology experts.

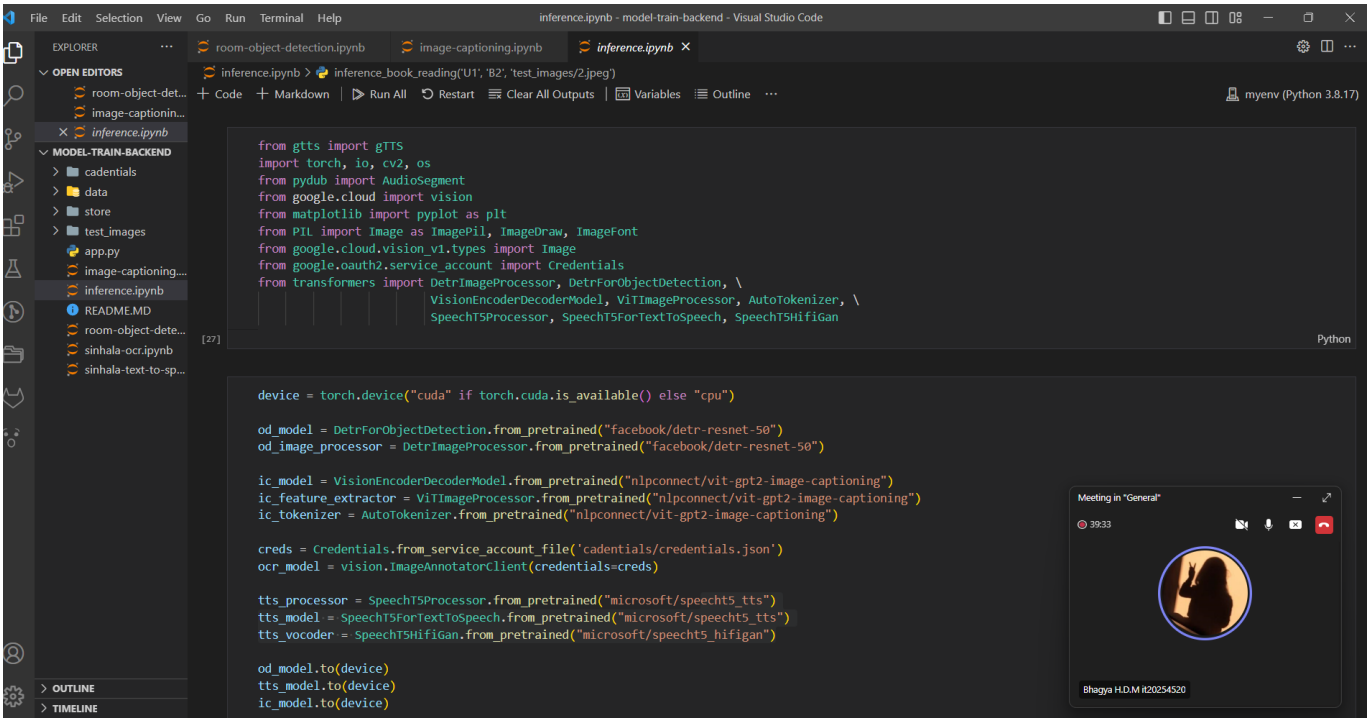
Fig. 1. Overall System Diagram

The subsequent phases involve the integration of OCR technology adept at processing the intricacies of the Sinhala phone.

An image detection software for blind students uses computer vision to recognize and describe objects and scenes in real-time. The software takes pictures of the user's surroundings using a smartphone or tablet camera. Objects and their characteristics are identified using image processing techniques like edge detection, color analysis, and feature extraction. Machine learning models, such as convolutional neural networks (CNNs), are used to identify and categorize objects in real time. These models are trained on massive datasets of annotated photos. The main purpose of Text to speech (TTS) technology is to enable blind individuals to access written text from Sinhala books, allowing them to easily listen to the valuable content. TTS technology enables the reading of Sinhala books in a natural-sounding voice, aiding visually impaired individuals. It uses computer algorithms to analyze the text, generate pronunciation, intonation, and rhythm for each word and sentence, and translate it into speech. The app also allows users to read time in a background process. The camera should start automatically when the user launches it, focus on the paper, alert the user, and capture images when all documents are within the capture frame. The device's storage should be used for capturing images, and the system must detect and correct skew before submitting information to the OCR system.

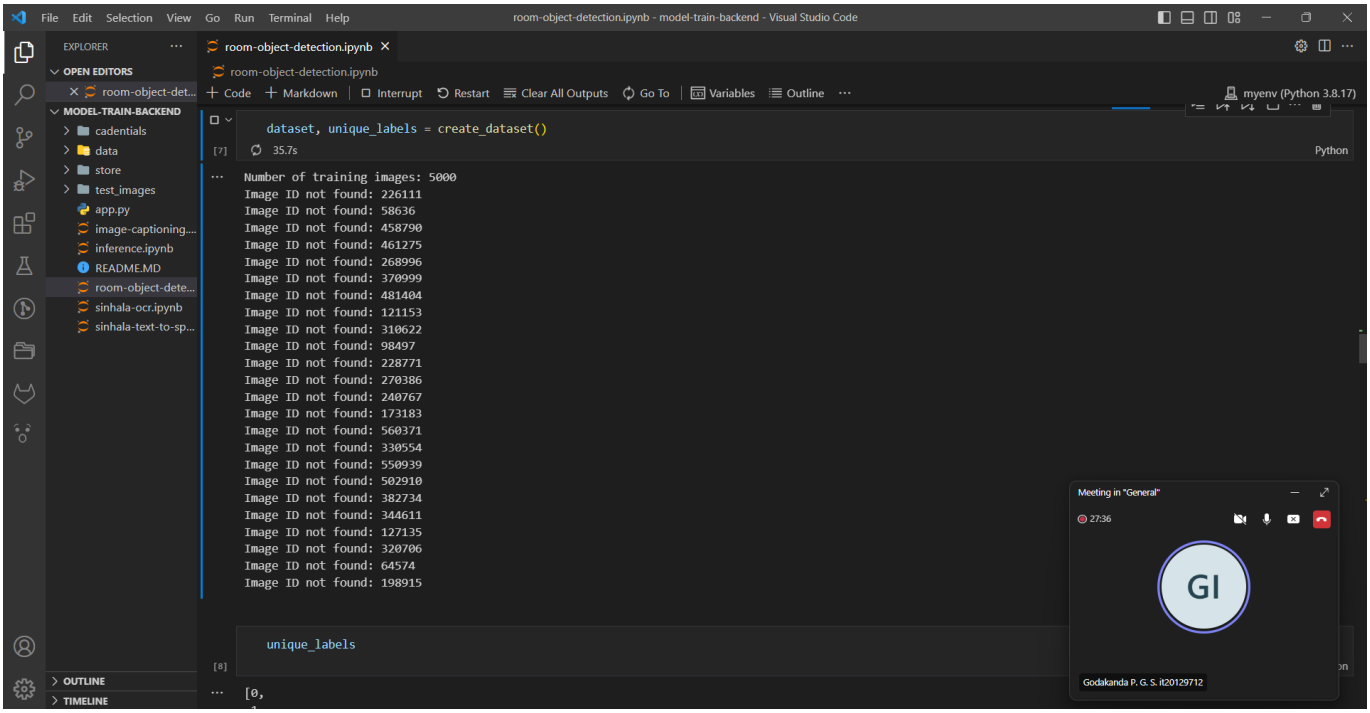
B. Text-to-Speech(TTS)

A high-quality Text-to-Speech (TTS) system is defined by its ability to accurately mimic the subtleties of the human



```
from gtts import gTTS
import torch, io, cv2, os
from pydub import AudioSegment
from google.cloud import vision
from matplotlib import pyplot as plt
from PIL import Image as ImagePil, ImageDraw, ImageFont
from google.cloud.vision_v1.types import Image
from google.oauth2.service_account import Credentials
from transformers import DetrImageProcessor, DetrForObjectDetection, \
                         VisionEncoderDecoderModel, ViTImageProcessor, AutoTokenizer, \
                         SpeechT5Processor, SpeechT5ForTextToSpeech, SpeechT5HifiGan
```

```
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")

od_model = DetrForObjectDetection.from_pretrained("facebook/detr-resnet-50")
od_image_processor = DetrImageProcessor.from_pretrained("facebook/detr-resnet-50")

ic_model = VisionEncoderDecoderModel.from_pretrained("nlpconnect/vit-gpt2-image-captioning")
ic_feature_extractor = ViTImageProcessor.from_pretrained("nlpconnect/vit-gpt2-image-captioning")
ic_tokenizer = AutoTokenizer.from_pretrained("nlpconnect/vit-gpt2-image-captioning")

creds = Credentials.from_service_account_file('cadentials/credentials.json')
ocr_model = vision.ImageAnnotatorClient(credentials=creds)

tts_processor = SpeechT5Processor.from_pretrained("microsoft/speecht5_tts")
tts_model = SpeechT5ForTextToSpeech.from_pretrained("microsoft/speecht5_tts")
tts_vocoder = SpeechT5HifiGan.from_pretrained("microsoft/speecht5_hifigan")

od_model.to(device)
tts_model.to(device)
ic_model.to(device)
```



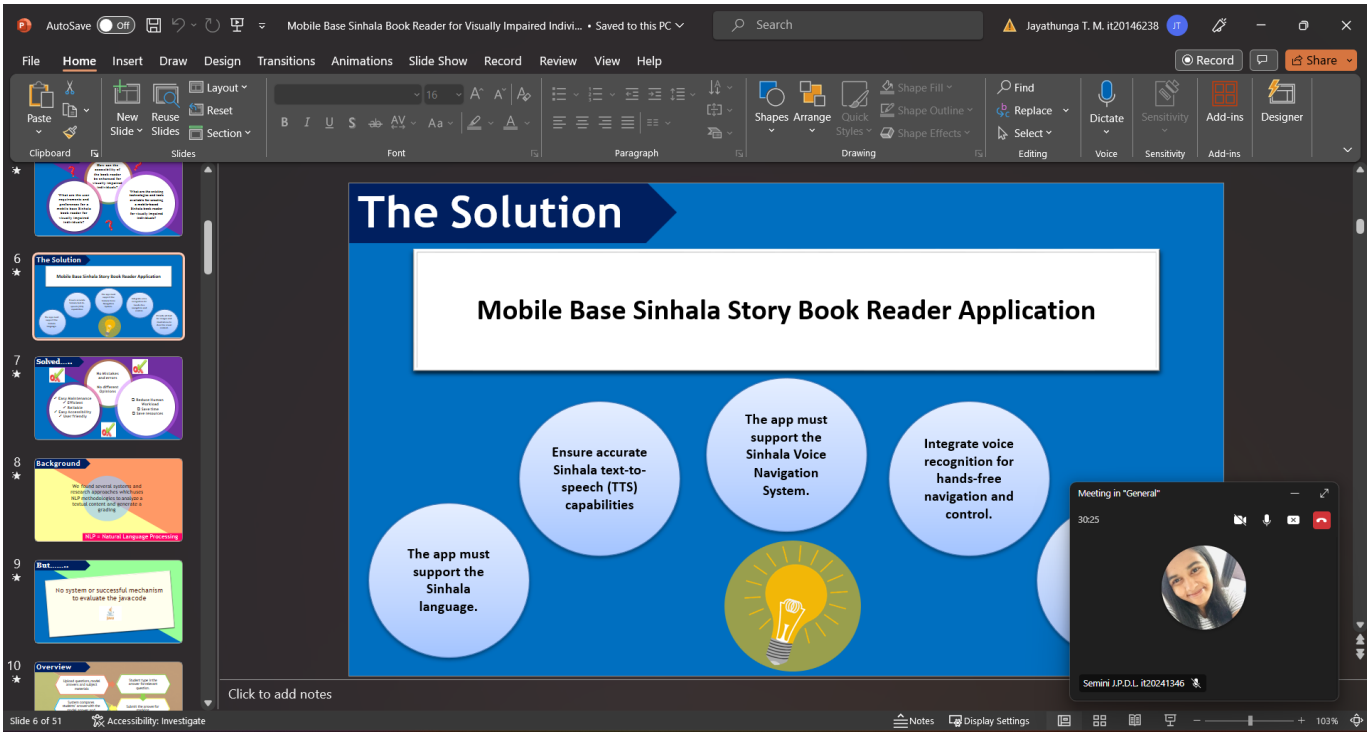
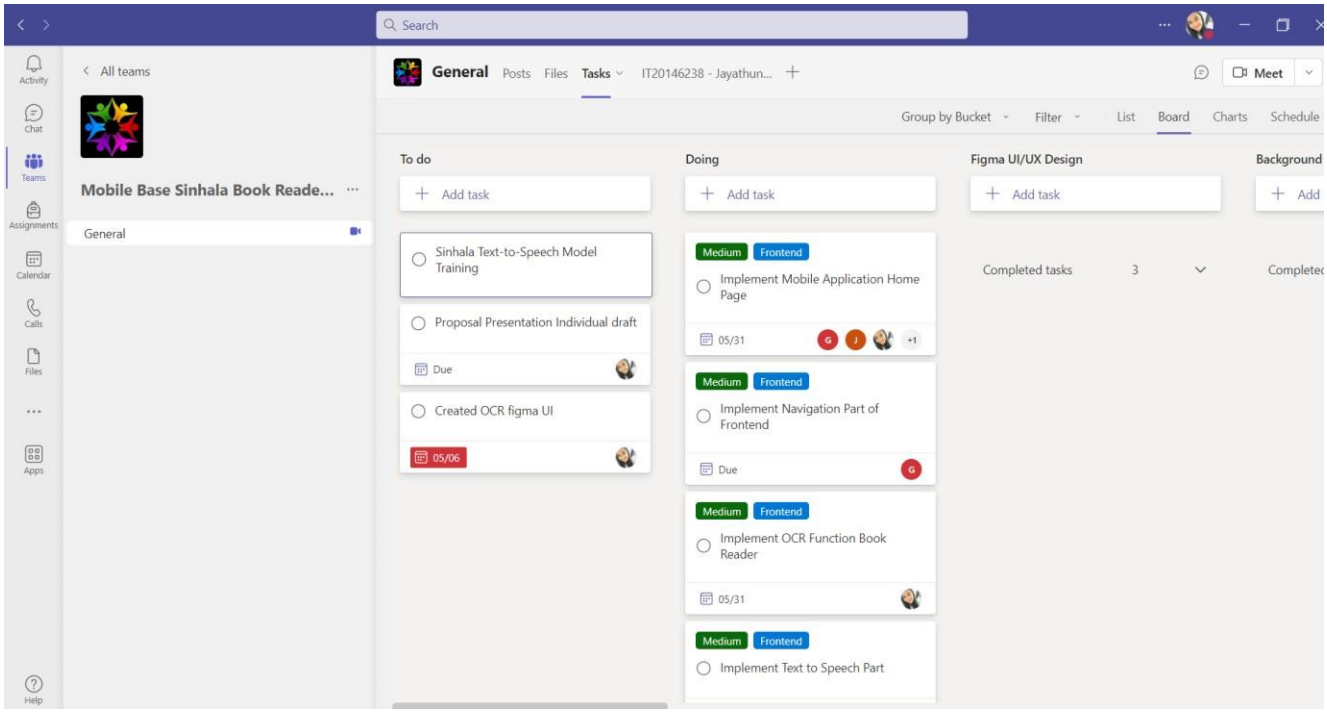
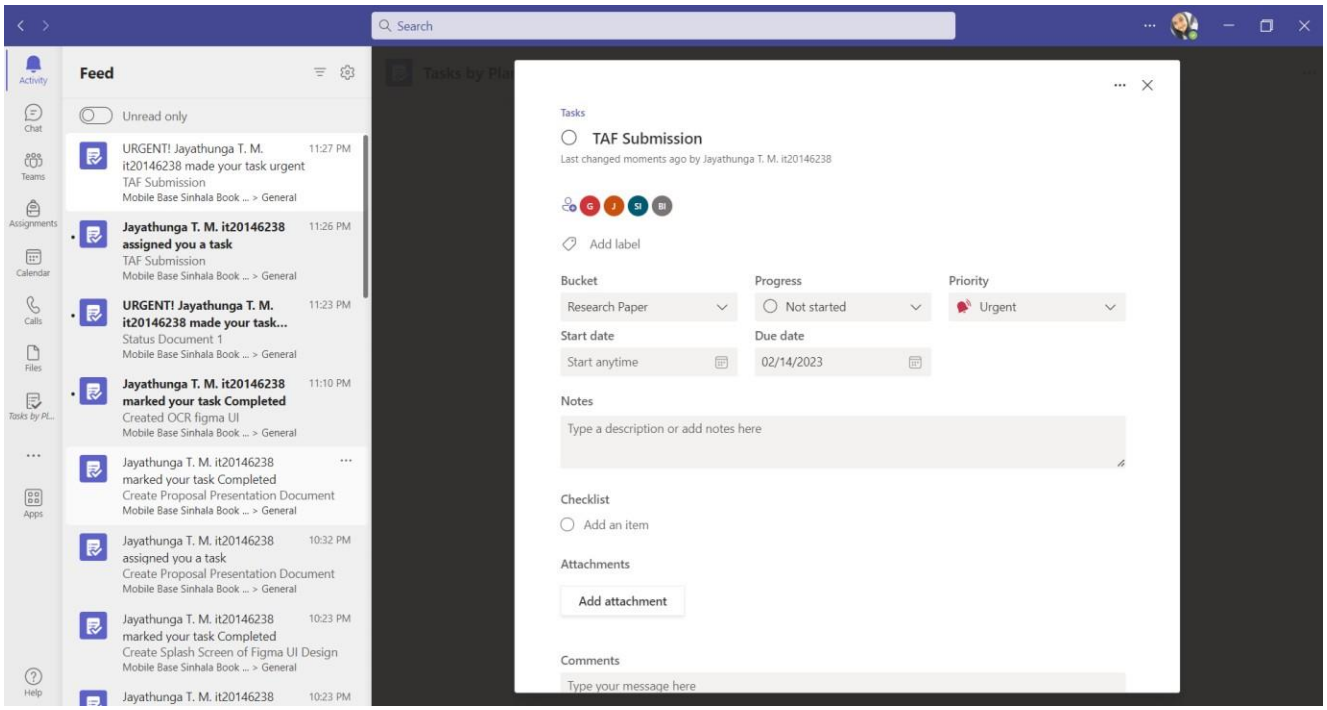


Figure 5-Team Calls in teams

1.5 Teams tasks planner





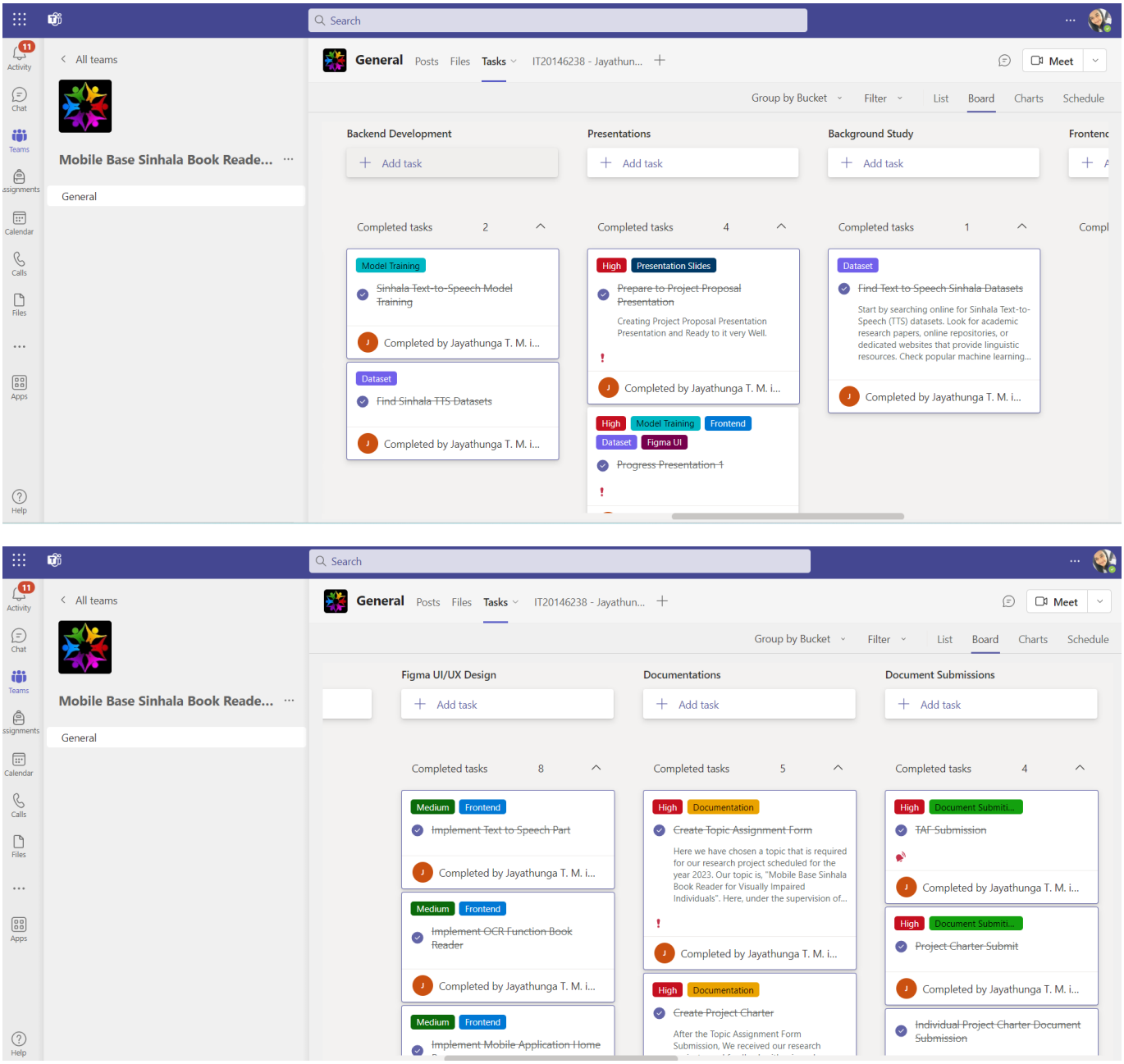
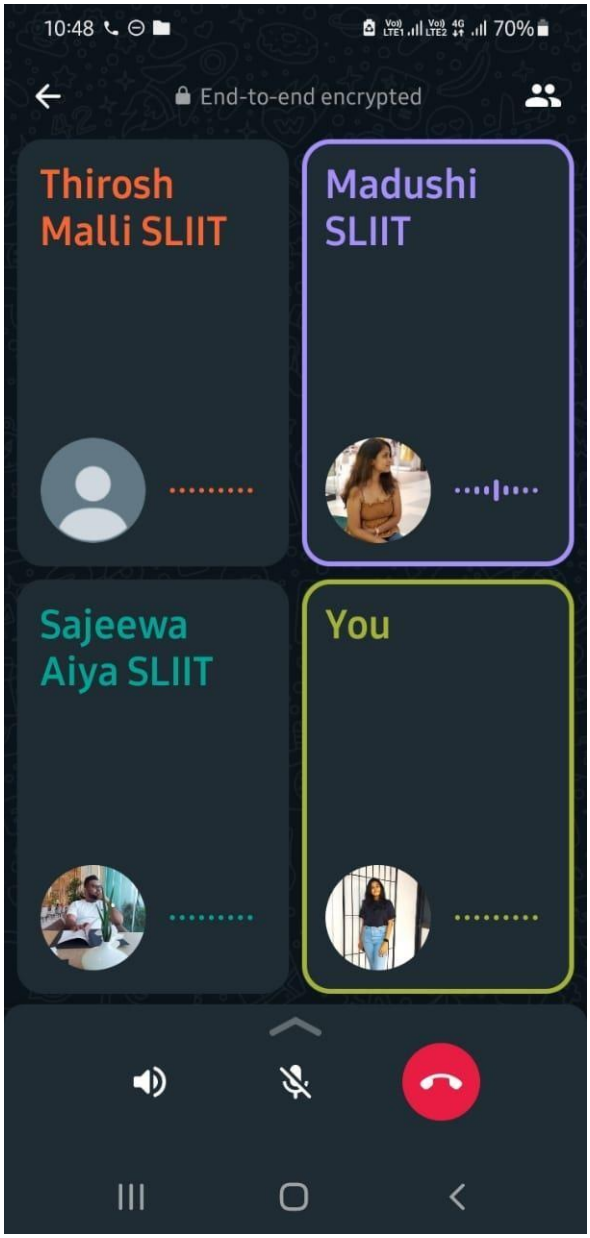


Figure 6-Teams tasks planner

2. Whatsapp group and calls



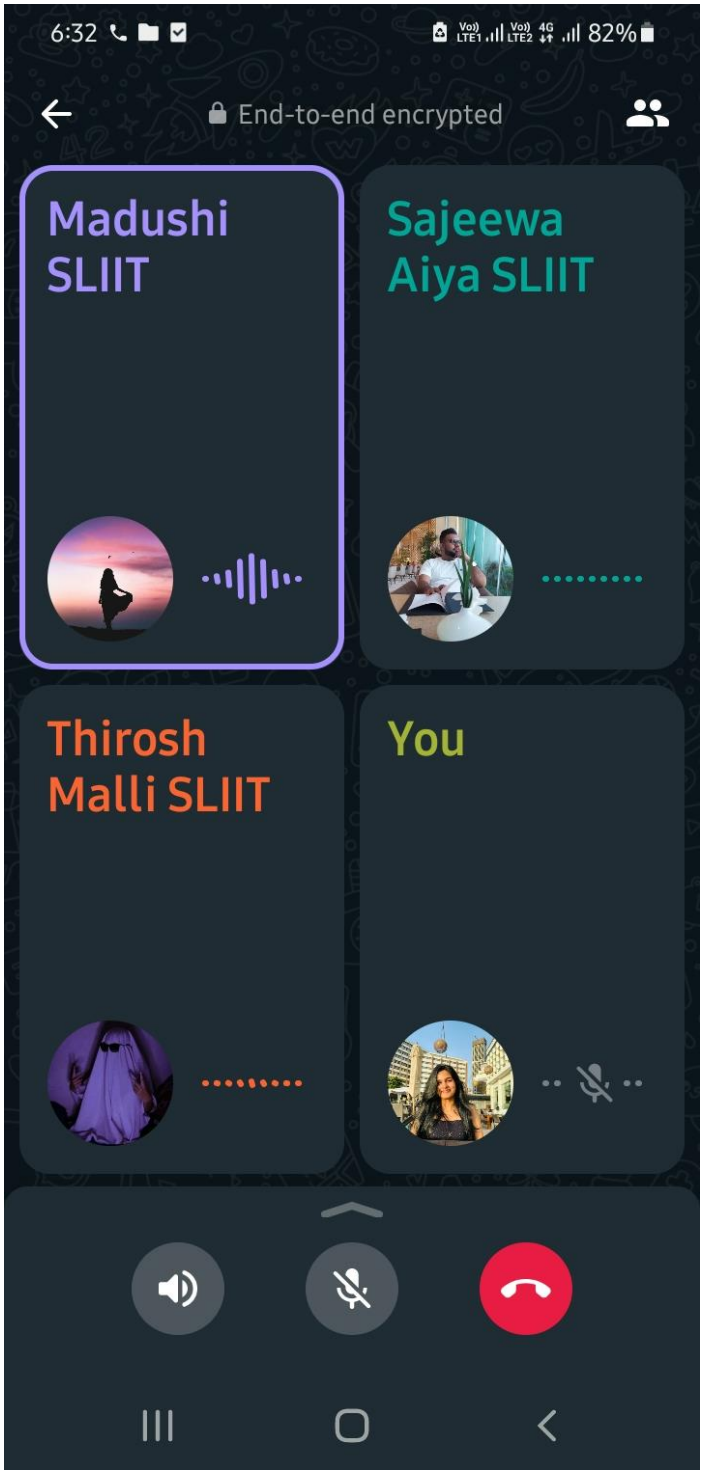


Figure 7-Teams whatsapp calls

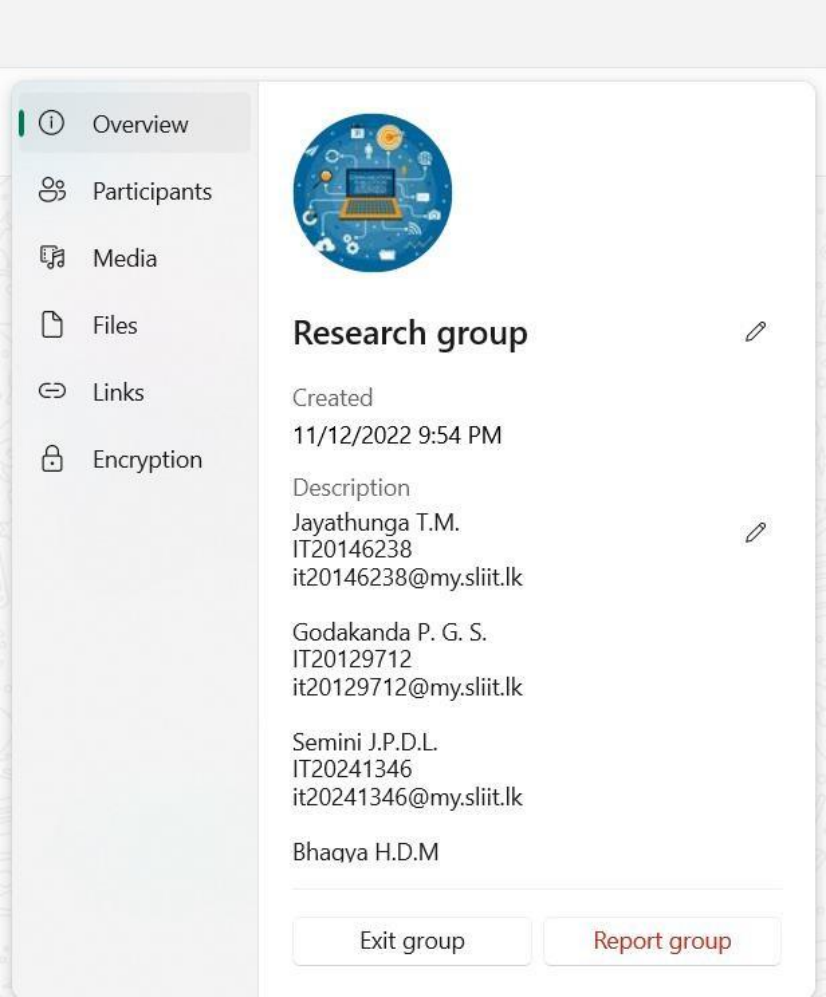


Figure 8-Teams whatsapp group

3. Gantt Chart

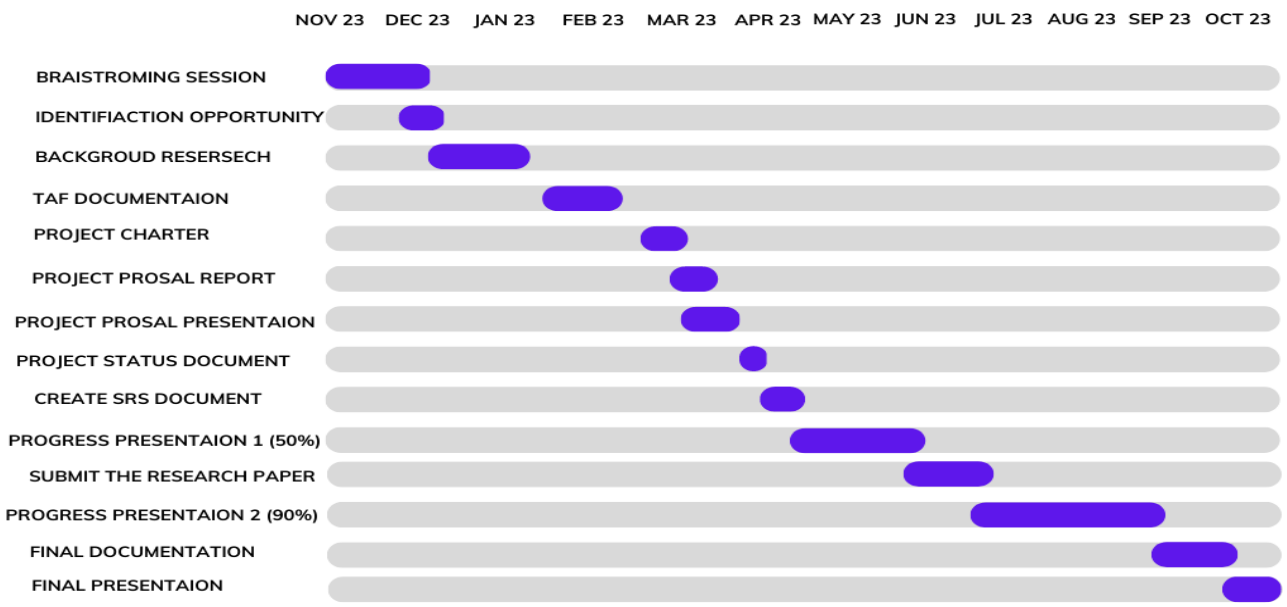


Figure 9-Grant chart

4. Work Breakdown Chart

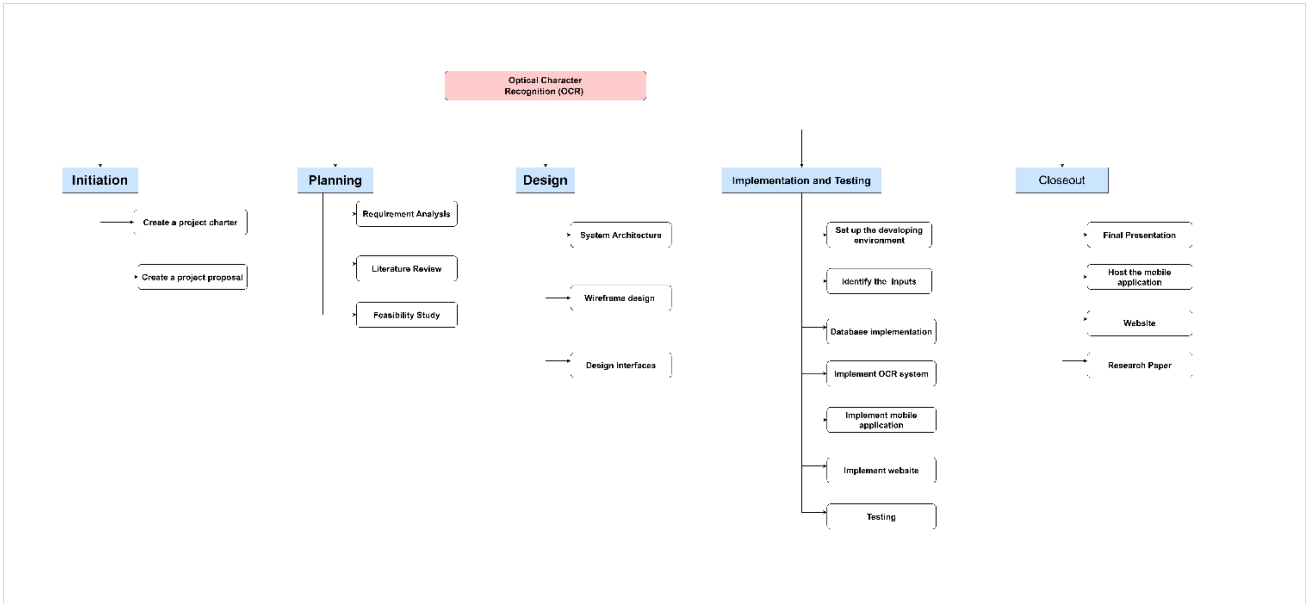
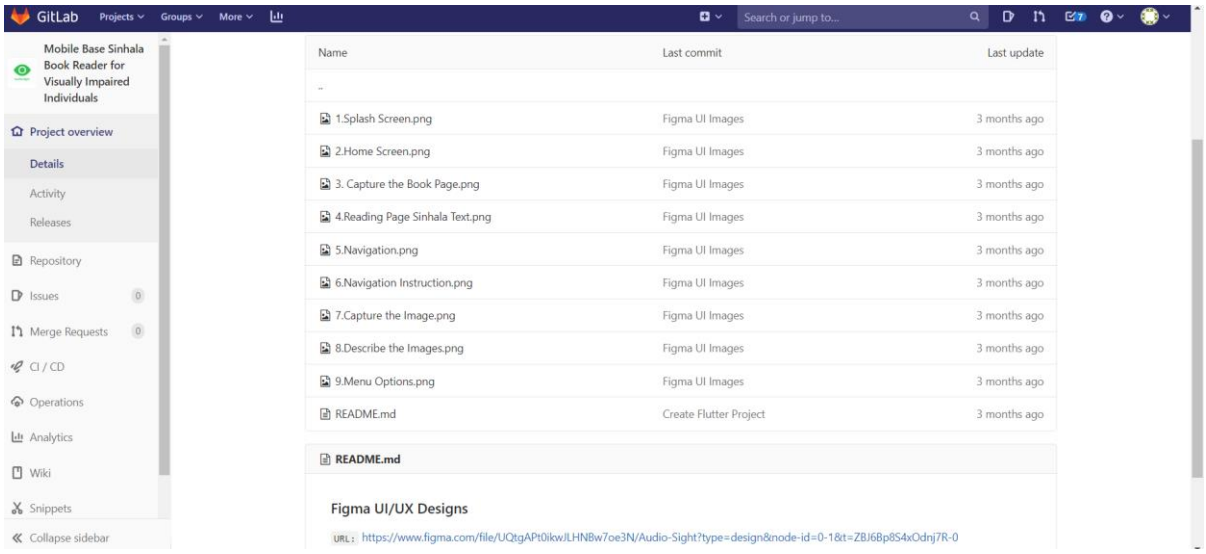
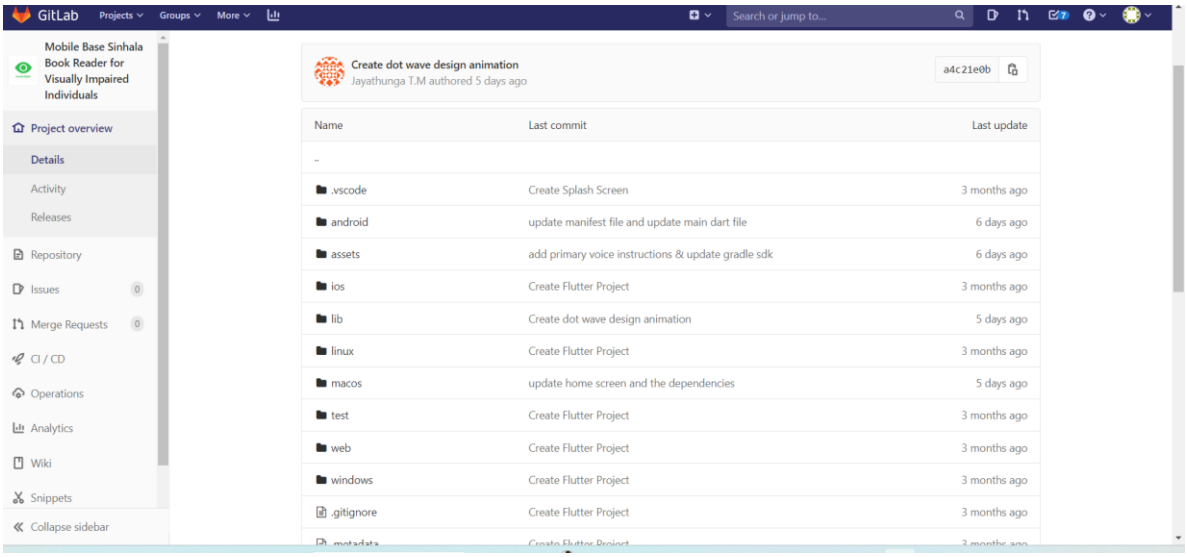


Figure 10-Work Breakdown Chart

5. Gitlab





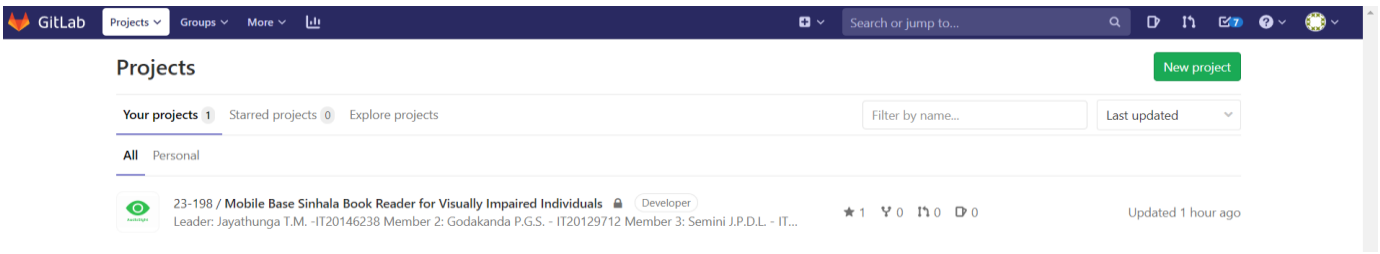


Figure 11-Gitlab

6. System methodology diagram

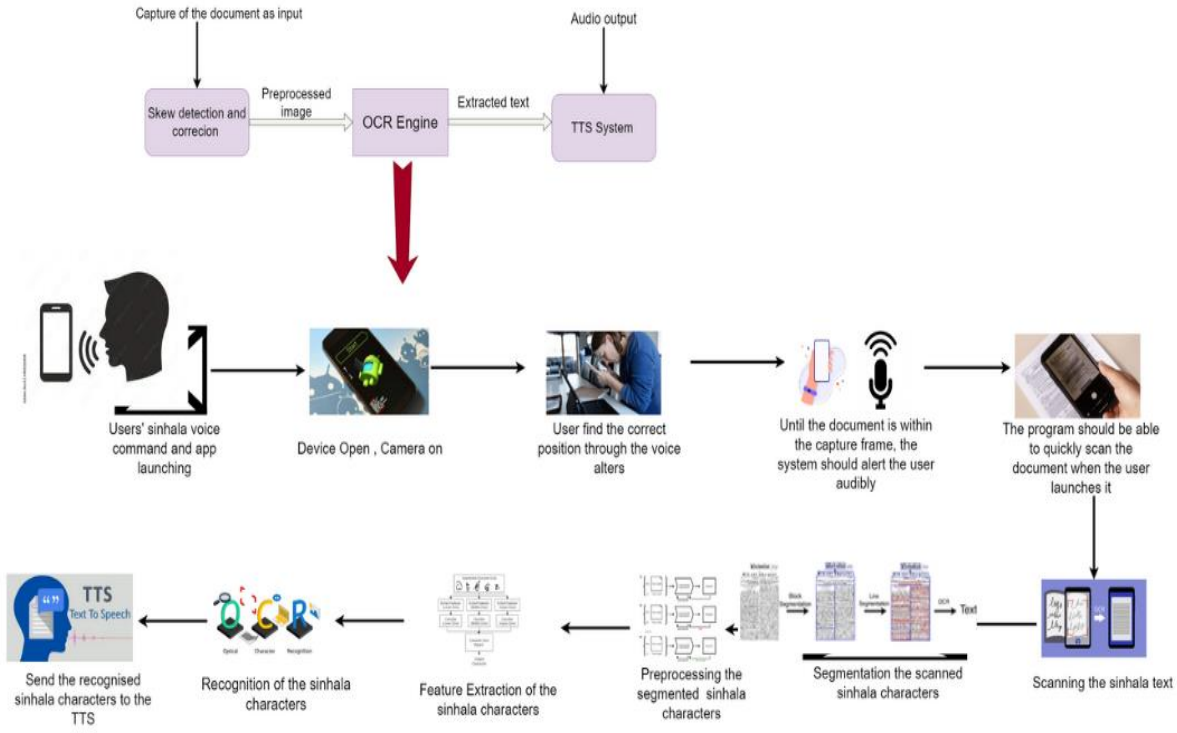


Figure 12-System diagram

7. Current Progress up to now

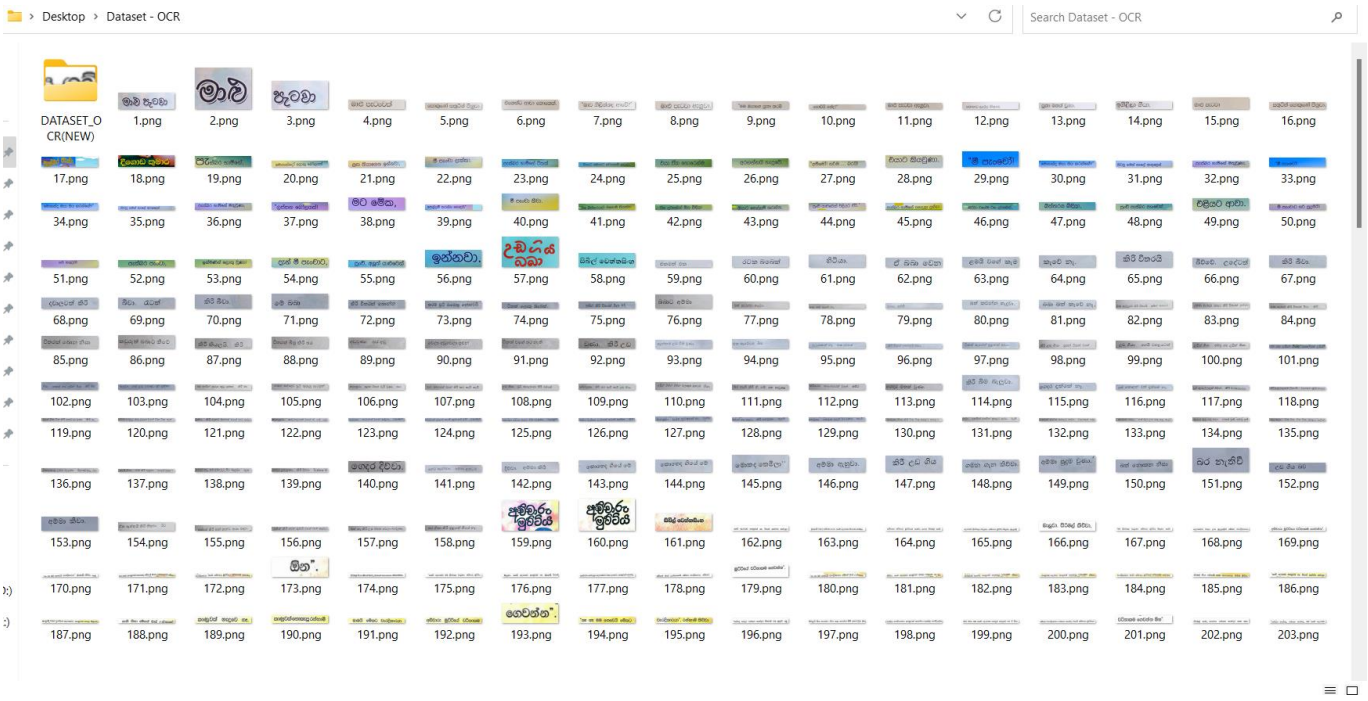
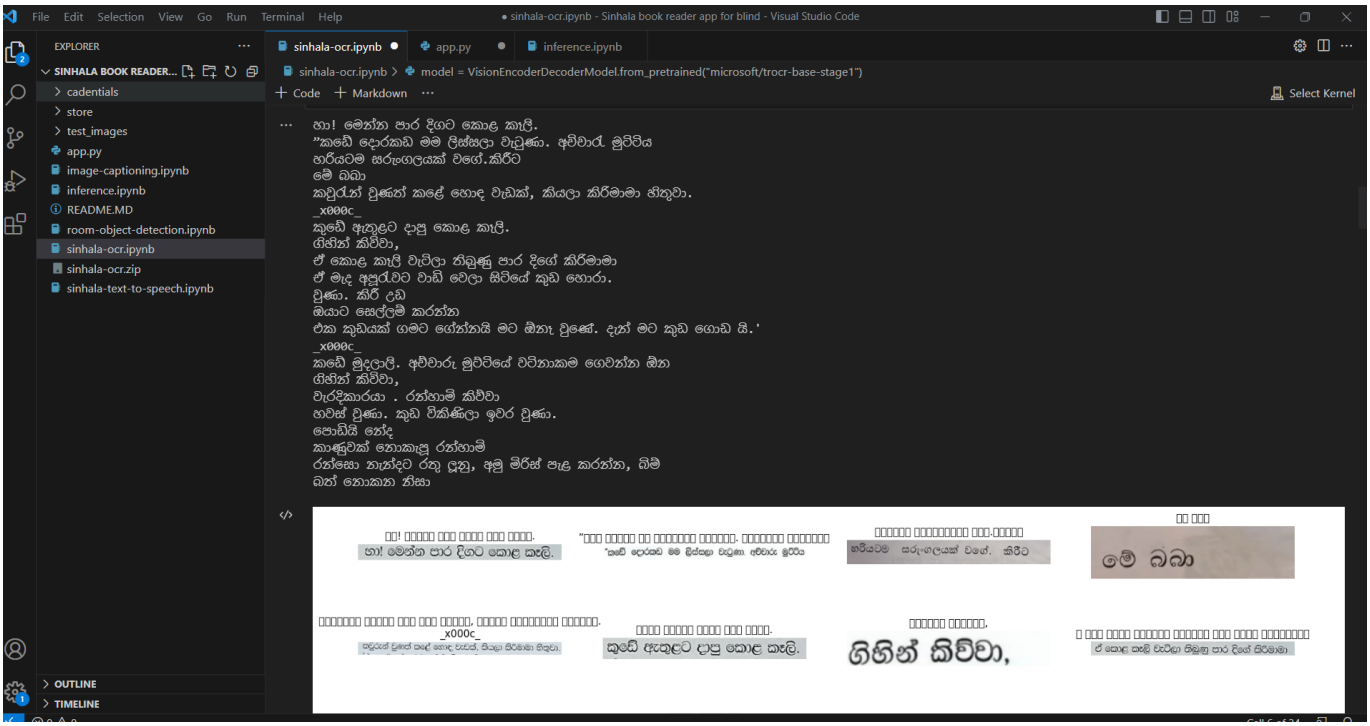
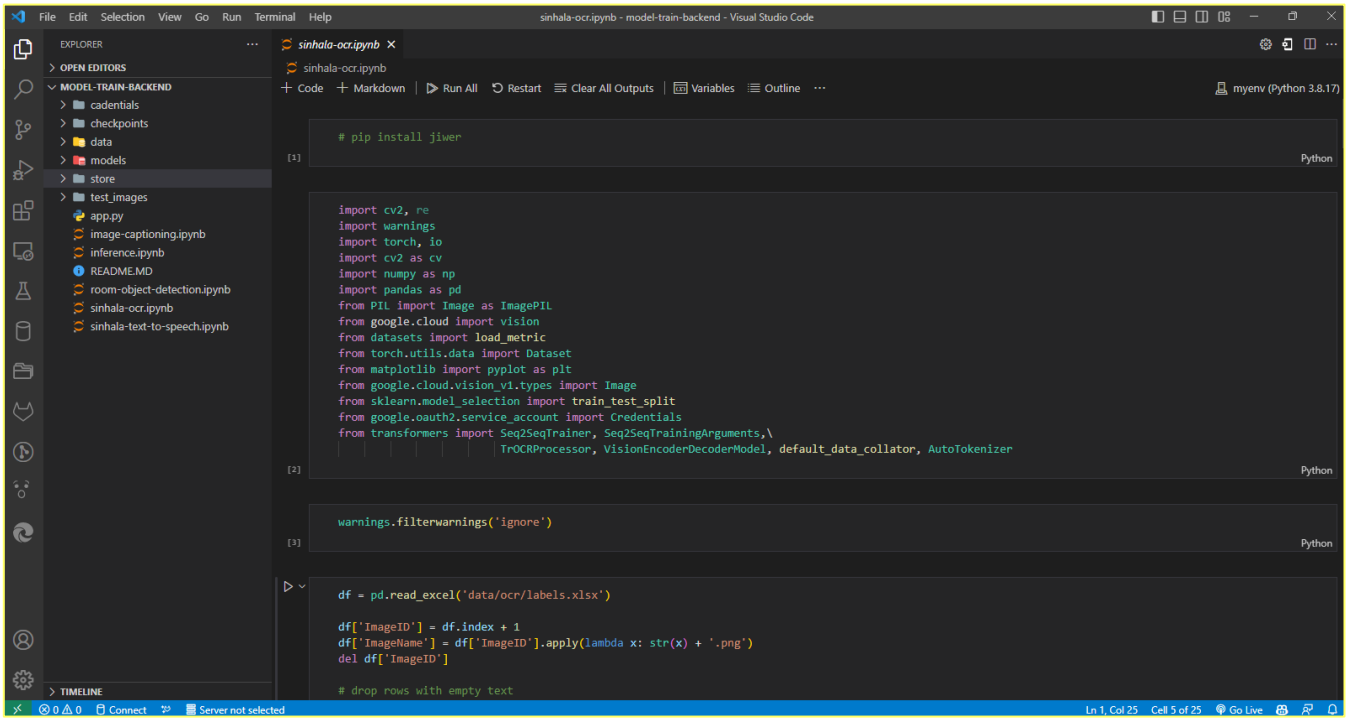


Figure 13-dataset





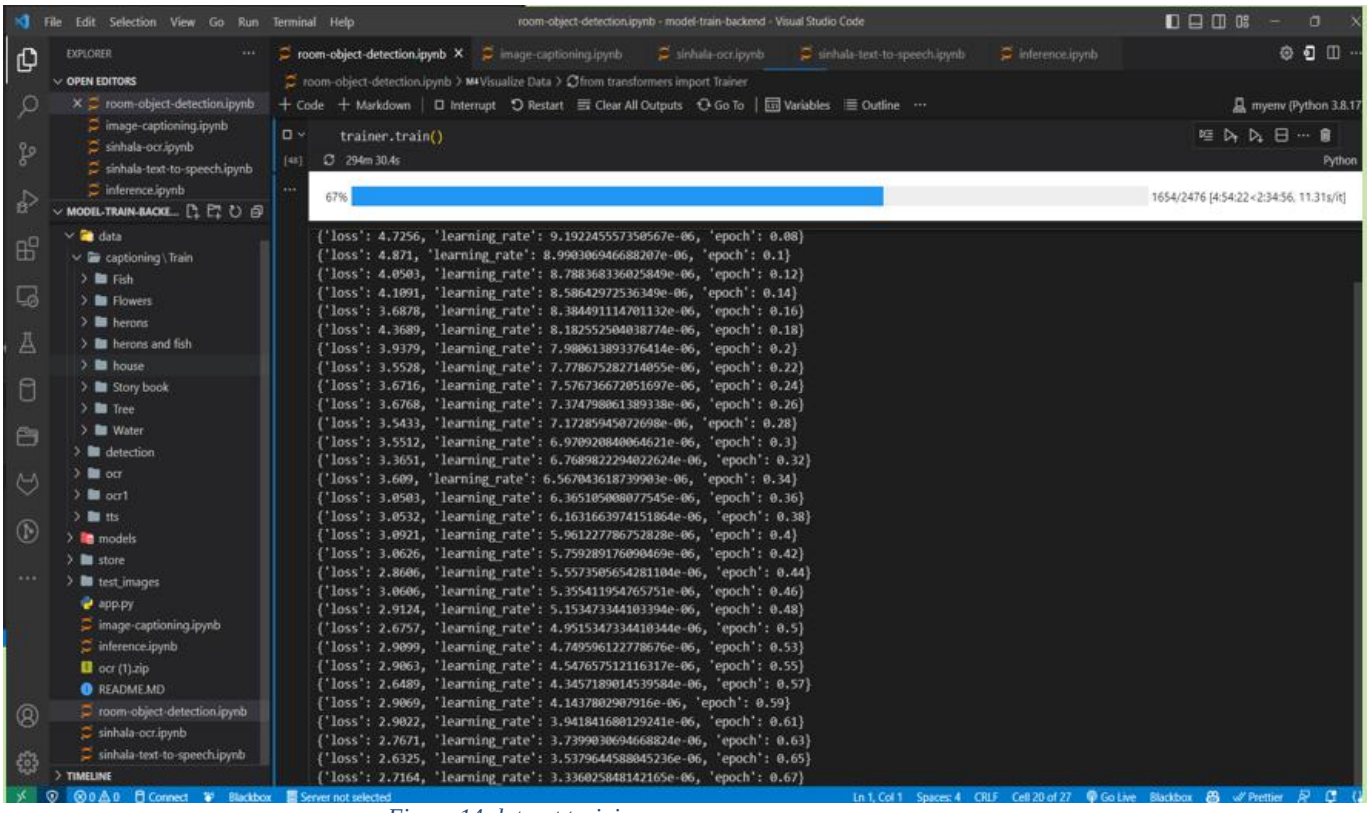


Figure 14-dataset training

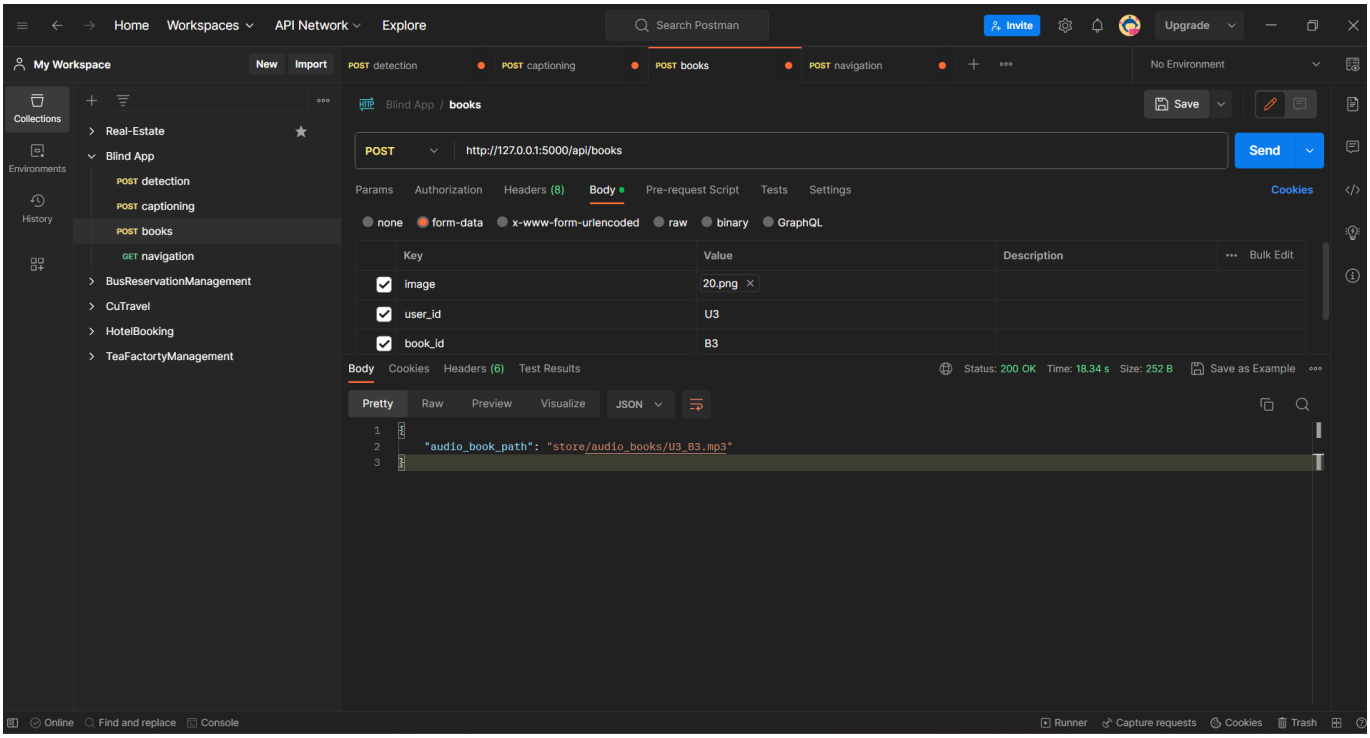
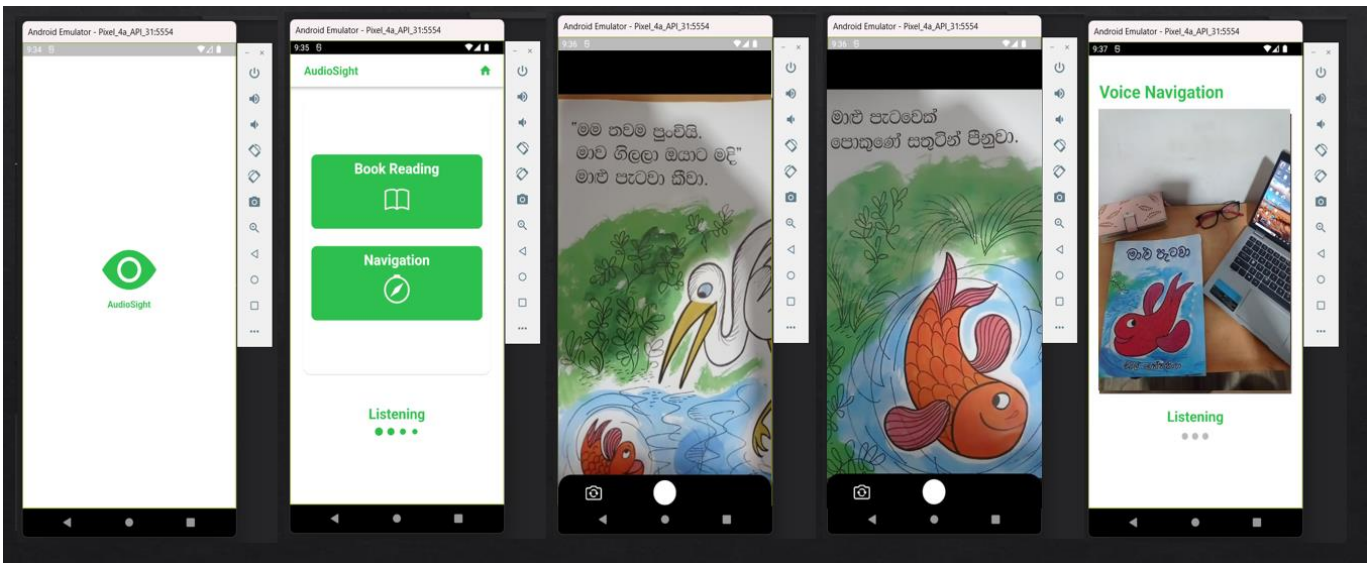


Figure 15-backend postman API



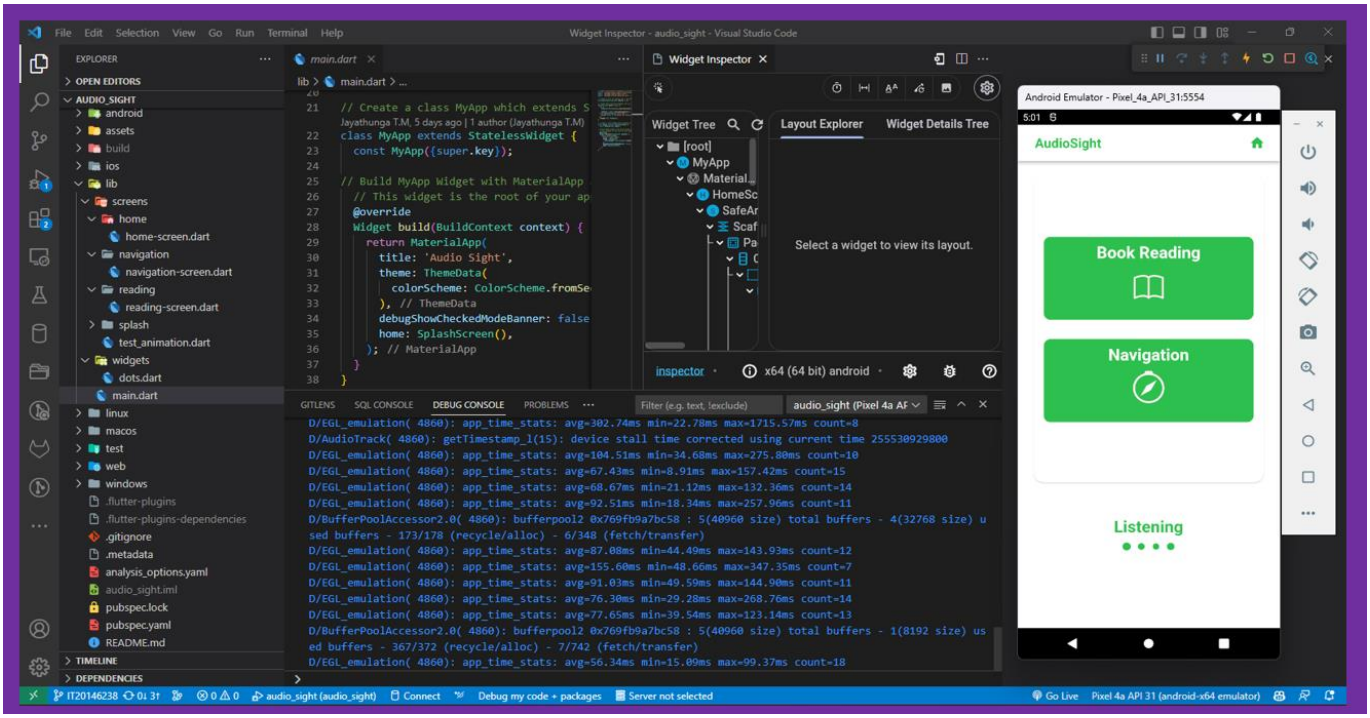


Figure 16-frontend

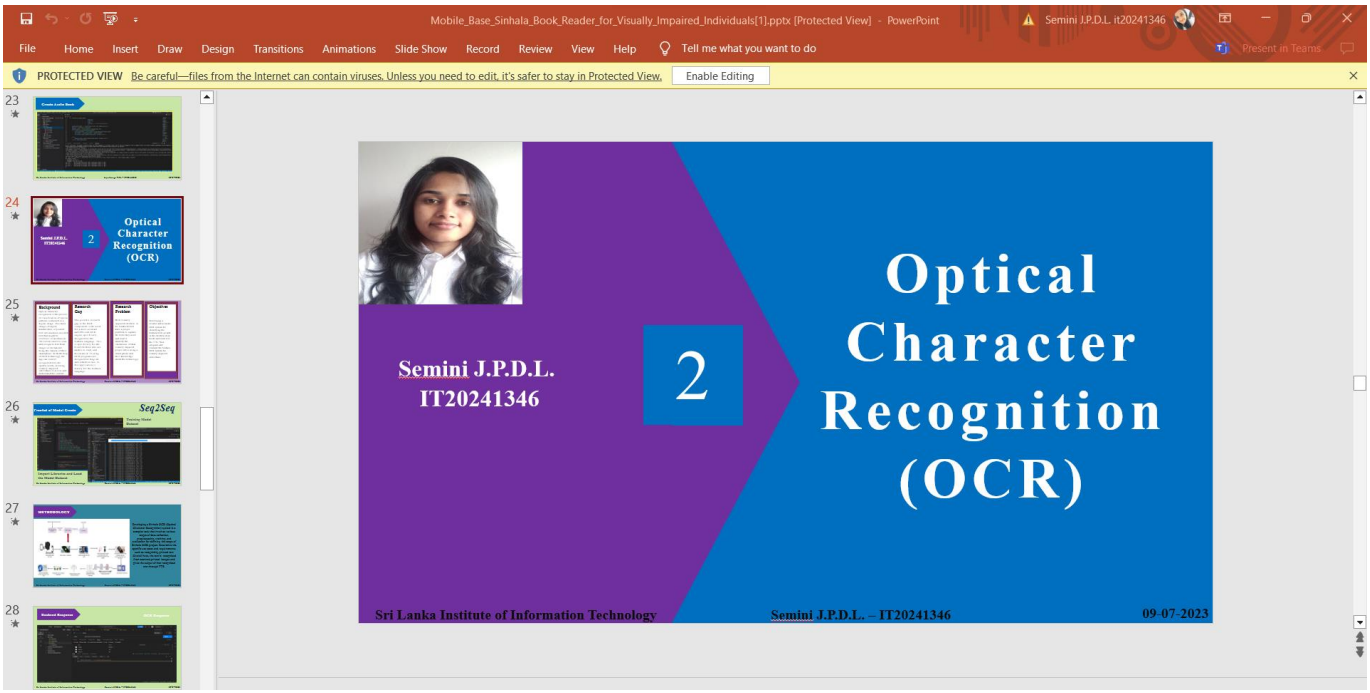


Figure 17-pp2 presentation

8. Our Research Paper

Mobile Base Sinhala Book Reader For Visually Impaired Students

J.Thirosh Madhusha
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
thiroshmadhusha0520@gmail.com

H.D.Madushi Bhagya
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
madushihd@gmail.com

Sajeewa Godakanda
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
sajeewagodakanda@gmail.com

J.P.Dulmi Semini
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
dulmilaknasewmz9811@gmail.com

Prof. Koliya Pulasinghe
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
koliya.p@sliit.lk

P.K.P.G. Panduwawala
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
poorna.p@sliit.lk

***Abstract*—The project aims to improve the reading experience and skills of visually impaired students in Sri Lanka by creating a mobile application that allows them to easily read printed books and stationery in Sinhala. The mobile application uses optical character recognition (OCR) technology and voice navigation, incorporating text-to-speech features of the event synthesis framework. The application accurately captures characters on a page of a Sinhala book and distinguishes them using OCR technology, enabling visually impaired people to convert text into accessible digital formats. The extracted text is then made audible via text-to-speech. Sinhala Voice Navigation support is provided for users to navigate the app, get feedback from the user, and identify objects in the surrounding room. The application uses image recognition and description algorithms to describe pictures in Sinhala, helping visually impaired children understand the visual content and improve their reading skills. The platform also offers features to adjust reading speed and choose between male or female voices.**

***Keywords*—Visually Impaired Individuals, Sinhala Text-to-Speech(TTS), Sinhala Optical Character Recognize(OCR), Sinhala Voice Navigation, Image Recognition, Sinhala Object Detection**

I. Introduction

Knowledge is the most important factor for surviving in this century. One way to gain knowledge is through reading, even for those who are visually impaired, who can use the braille system [1]. However, traditional braille systems are becoming outdated as computer-assisted braille systems and text-to-speech systems are becoming more common. Unfortunately, these technologies are not widely available in Sinhala, and these devices are too expensive for the average Sri Lankan. An Android-based solution using OCR, TTS, image recognition, and voice navigation was considered for this study to improve the reading experience and accessibility for the visually impaired "Sinhala Book Reader Mobile Application".

The quality of a Text-to-Speech (TTS) system depends on its ability to imitate human speech and ensure clear understanding. The absence of natural expressions in TTS output has a substantial influence on application usability. This emphasizes a key issue in TTS development for creating a synthesized speech [2] that closely matches the human voice from the text. TTS technology's major goal is to recreate the complete range of human speech, including different speech patterns, subtleties, and intonations, while reducing the mechanical or robotic quality of the output voice.

The Sinhala language, the mother tongue of most Sri Lankans, is a crucial area for TTS development due to its complexities and nuances. Despite the large number of Sinhala speakers in Sri Lanka, there is a need for research on Sinhala voice recognition. The complexities of the Sinhala language make it difficult for computers to understand and reproduce it. Currently, there is little progress in developing TTS systems for the Sinhala language. However, this is a key research frontier that must be explored. An efficient TTS system for Sinhala would bridge the gap between human language skills and machine-generated speech [1], improving user experiences and bridging the gap between human language skills and machine-generated speech. There have been only a few attempts made to develop a Sinhala language TTS. This is still a major research area that requires investigation, which is one of the key motivations for this research.

In an increasingly digitized world, accessibility to information and literature remains a challenge for visually impaired individuals. Mobile technology and Optical Character Recognition (OCR) can solve this issue. This introduction elucidates the significance of mobile-based Sinhala book readers employing OCR technology as a transformative solution for individuals with visual impairments. By harnessing the power of mobile devices and OCR, these readers offer the potential to convert printed Sinhala text into accessible digital formats, thereby facilitating independent and inclusive access to literature for visually impaired individuals. This section introduces the key components of this paper, including the integration of OCR technology, the unique context of the Sinhala language, and the overarching goal of enhancing accessibility and enriching the reading experiences of visually impaired individuals [3].

Our Sinhala book reader app for blind users delivers a ground-breaking feature: picture detection within Sinhala children's books, at the nexus of accessibility and education. By allowing those who are blind to enjoy the rich world of

Figure 18-research paper

9. Extra helping resources

Figure 19-you tube tutorials

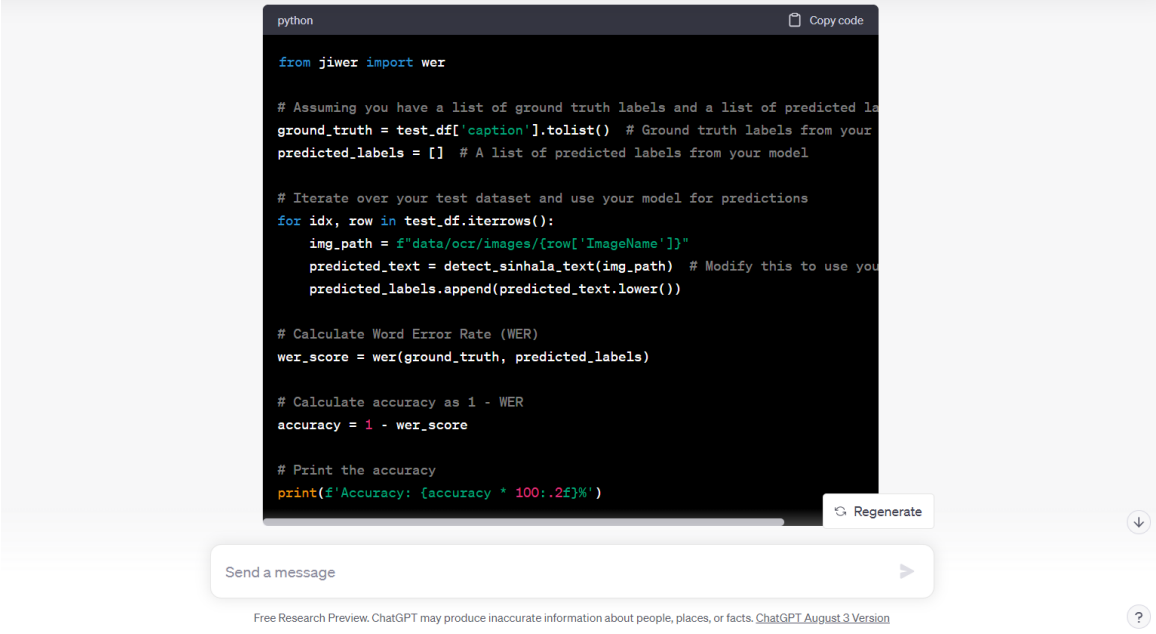
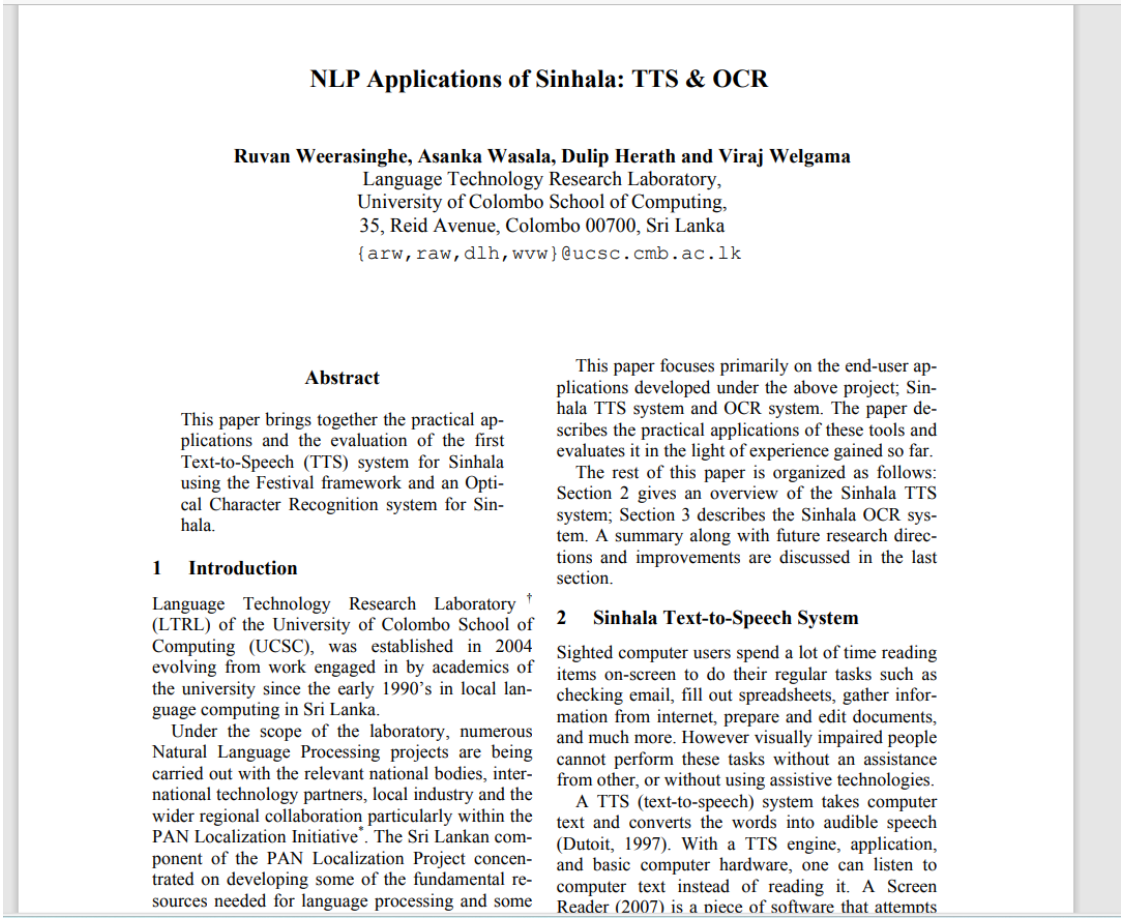


Figure 20-chat gpt



NLP Applications of Sinhala: TTS & OCR

Ruvan Weerasinghe, Asanka Wasala, Dulip Herath and Viraj Welgama
Language Technology Research Laboratory,
University of Colombo School of Computing,
35, Reid Avenue, Colombo 00700, Sri Lanka
{arw,raw,dlh,wvw}@ucsc.cmb.ac.lk

Abstract

This paper brings together the practical applications and the evaluation of the first Text-to-Speech (TTS) system for Sinhala using the Festival framework and an Optical Character Recognition system for Sinhala.

1 Introduction

Language Technology Research Laboratory † (LTRL) of the University of Colombo School of Computing (UCSC), was established in 2004 evolving from work engaged in by academics of the university since the early 1990's in local language computing in Sri Lanka.

Under the scope of the laboratory, numerous Natural Language Processing projects are being carried out with the relevant national bodies, international technology partners, local industry and the wider regional collaboration particularly within the PAN Localization Initiative*. The Sri Lankan component of the PAN Localization Project concentrated on developing some of the fundamental resources needed for language processing and some

This paper focuses primarily on the end-user applications developed under the above project; Sinhala TTS system and OCR system. The paper describes the practical applications of these tools and evaluates it in the light of experience gained so far.

The rest of this paper is organized as follows: Section 2 gives an overview of the Sinhala TTS system; Section 3 describes the Sinhala OCR system. A summary along with future research directions and improvements are discussed in the last section.

2 Sinhala Text-to-Speech System

Sighted computer users spend a lot of time reading items on-screen to do their regular tasks such as checking email, fill out spreadsheets, gather information from internet, prepare and edit documents, and much more. However visually impaired people cannot perform these tasks without an assistance from other, or without using assistive technologies.

A TTS (text-to-speech) system takes computer text and converts the words into audible speech (Dutoit, 1997). With a TTS engine, application, and basic computer hardware, one can listen to computer text instead of reading it. A Screen Reader (2007) is a piece of software that attempts

Figure 21-research papers

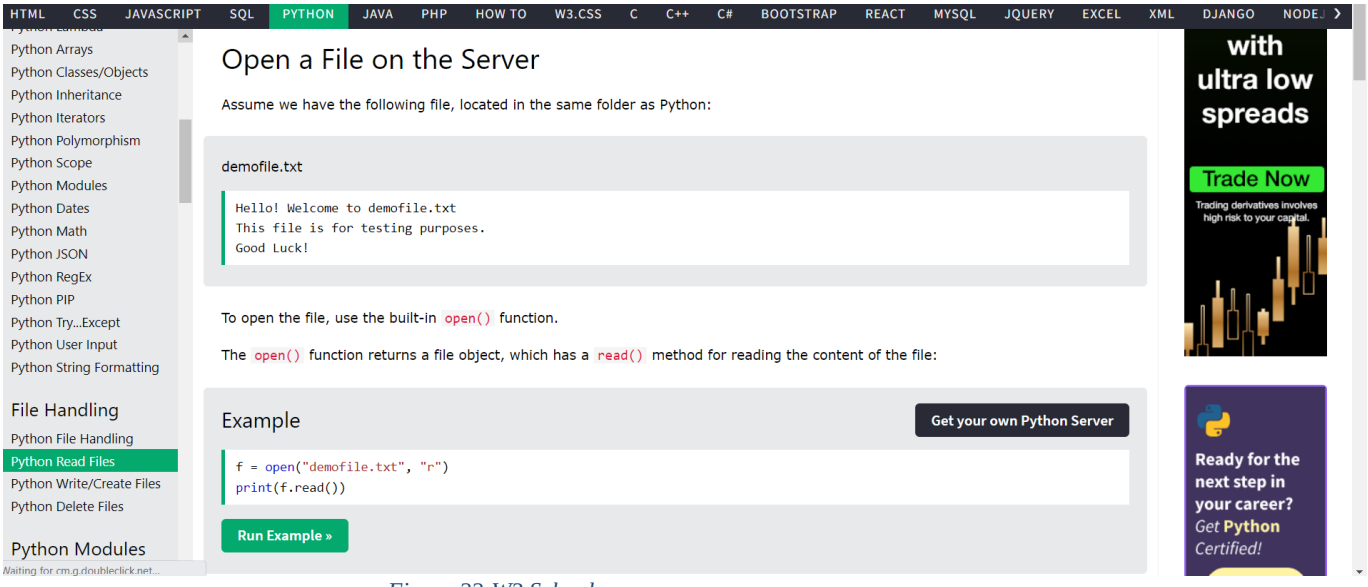


Figure 22-W3 School