SINHALA TEXT RECOGNITION AND VOICE FEEDBACK FOR VISUALLY IMPAIRED - OPTICAL CHARACTER RECOGNITION (OCR)

Dilini K.D.

(IT21826740)

B.Sc. (Hons) Degree in Information Technology

Specialization in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

August 2025

Declaration

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

Name	Student Id	Signature
Dilini K.D.	IT21826740	

	2025/08/26
Signature of the Supervisor	Date

	2025/08/26
Signature of the Co-Supervisor	Date

Abstract

This research aims to develop an affordable and effective assistive platform for visually impaired individuals in Sri Lanka, enabling them to capture and read Sinhala text using Tiny ML and Edge AI technologies. The primary objective is to address the lack of accessible, real-time text recognition systems tailored for Sinhala, as existing solutions fail to meet the specific needs of the visually impaired community, particularly in resource-constrained environments. This study proposes a system involving smart glasses and a mobile application that integrates Optical Character Recognition (OCR) and text-to-speech (TTS) technologies to provide real-time, hands-free reading.

The smart glasses will be equipped with a Raspberry Pi camera, which will solely be used for image capture. The mobile application will process the captured text using AI-powered OCR and convert it into synthesized speech, providing immediate auditory feedback to the user. A key feature of this solution is real-time voice guidance, which assists users in positioning the text correctly within the camera's frame, ensuring effective text capture while minimizing errors in recognition. Interactive audio cues will further enhance accessibility by guiding users through the application interface. By leveraging Tiny ML and Edge AI, the system will function without reliance on cloud computing, ensuring privacy, low latency, and cost efficiency.

The proposed solution will utilize low-cost hardware components and lightweight AI models optimized for mobile processing, making it both affordable and accessible. Unlike traditional assistive technologies that rely on internet connectivity or expensive proprietary systems, this approach focuses on real-time, offline processing, allowing users to read printed text without external dependencies. The mobile application will be developed using frameworks like React Native, integrating Google ML Kit or Tesseract OCR for text recognition, and employing text-to-speech synthesis for clear and natural voice output.

Through a user-centered design approach, usability testing will be conducted with visually impaired individuals to evaluate the system's effectiveness, accessibility, and ease of use. Key performance indicators will include text recognition accuracy, response time, user satisfaction, and overall impact on daily reading experiences. The expected outcome is a cost-effective, user-friendly assistive device that significantly enhances the independence and quality of life for

visually impaired individuals in Sri Lanka by making printed materials more accessible and improving access to information.

Ultimately, this research aims to contribute to the development of localized assistive technologies in Sri Lanka and provide a scalable model for similar environments worldwide. By addressing the unique challenges faced by Sinhala-speaking visually impaired users, this system has the potential to bridge the accessibility gap and empower individuals with greater autonomy in education, work, and daily life.

Acknowledgement

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Dharshana Kasthurirathna, and my co-supervisor, Ms. Hansi De Silva, for their invaluable guidance, encouragement, and continuous support throughout the course of this research. Their expertise, constructive feedback, and dedication have been instrumental in shaping this project and helping me overcome many challenges.

I am sincerely thankful to the academic staff and the Department of Information Technology at the Sri Lanka Institute of Information Technology (SLIIT) for providing me with the knowledge, resources, and learning environment that made this research possible.

I also wish to extend my appreciation to all the individuals and organizations who contributed in various ways to the success of this project, especially those who shared their insights and experiences related to accessibility and assistive technologies for the visually impaired.

This research project, focused on developing a smart glass-based Sinhala text recognition and voice feedback system using OCR, TinyML, and Edge AI, was inspired by the needs of visually impaired individuals in Sri Lanka. I am deeply grateful to the visually impaired community, whose courage and determination motivated me to design a solution that could make a meaningful difference in their daily lives.

Finally, I would like to express my heartfelt gratitude to my family and friends for their constant encouragement, patience, and support throughout this journey. Without their understanding and motivation, the completion of this thesis would not have been possible.

Executive Summary

The proposed Smart Glass–Based Sinhala Text Recognition and Voice Feedback System is a novel and practical solution aimed at enhancing accessibility for visually impaired individuals in Sri Lanka. By addressing the long-standing barriers to accessing printed Sinhala text, this research project demonstrates how emerging technologies can be harnessed to empower marginalized communities and promote digital inclusivity.

The system integrates Optical Character Recognition (OCR), TinyML, and Edge AI to deliver real-time text-to-speech (TTS) feedback through a mobile application connected to smart glasses. Using a Raspberry Pi camera for text capture, the device recognizes Sinhala printed material and instantly converts it into natural voice output, providing a seamless reading experience without reliance on internet connectivity. Unlike existing foreign assistive solutions, this project is tailored specifically for the Sinhala language and optimized for resource-constrained environments, ensuring affordability, portability, and usability.

Key features include real-time voice guidance to assist users in positioning text within the camera frame, offline processing to eliminate dependency on cloud services, and a lightweight AI model optimized for mobile devices. These ensure low latency, accuracy, and privacy while keeping the device cost-effective. The mobile application's user-friendly interface, supported by voice commands and intuitive navigation, further enhances accessibility.

This project is designed not only as a reading aid but also as a tool to foster independence, education, and equal opportunities for visually impaired individuals. By providing localized support for Sinhala text, it bridges a critical gap left by international OCR solutions that largely ignore non-Latin scripts. Moreover, its affordable hardware design ensures that it remains accessible to users across different socio-economic backgrounds.The anticipated impact of this research extends beyond individual literacy support. It contributes to inclusive education, promotes self-reliance among visually impaired users, and highlights the potential of Edge AI–based assistive technologies in developing countries. Ultimately, this project represents a commitment to using innovation for social good—offering visually impaired individuals in Sri Lanka the ability to independently engage with printed content, thereby improving their quality of life and opening doors to greater participation in education and society

List of Abbreviations

- AI – Artificial Intelligence
- ANN – Artificial Neural Network
- API – Application Programming Interface
- CNN – Convolutional Neural Network
- Edge AI – Edge Artificial Intelligence
- ML – Machine Learning
- MLP – Multi-Layer Perceptron
- NLP – Natural Language Processing
- OCR – Optical Character Recognition
- TTS – Text-to-Speech
- TinyML – Tiny Machine Learning
- UI – User Interface

Table of Contents

CHAPTER 01

Introduction

Visually impaired individuals face numerous challenges in their daily lives, particularly when accessing printed materials. The inability to independently read Sinhala text not only limits educational opportunities but also hinders social inclusion and access to information. In Sri Lanka, the problem is compounded by the lack of affordable assistive technologies designed specifically for the Sinhala language. Most available solutions are either tailored for Western languages or rely on expensive proprietary hardware, making them inaccessible to the majority of visually impaired individuals in developing regions.

A national survey conducted by the Ministry of Health in Sri Lanka (2015) highlighted the prevalence of blindness and visual impairment, especially among individuals over the age of 40, accounting for nearly 75% of cases. Table 1 presents the estimated sample size used in that study for assessing blindness and visual impairment across different age groups.

Table 1: Sample Size Parameters for Blindness and Visual Impairment Study in Sri Lanka

Parameter	**Value**
Confidence Interval (CI)	95%
Expected Prevalence Rate	4%
Design Effect	2.0
Target Age Group	≥ 40 yrs
Final Sample Size	6,600

Comparative statistics further indicate that Sri Lanka, despite having a relatively strong human development index among South Asian nations, faces significant challenges in addressing visual impairment. Table 2 compares selected human development and health indicators between Sri Lanka and neighboring South Asian countries.

Table 2: Human Development Indicators in South Asian Countries

Country	Literacy Rate (%)	Life Expectancy (Years)	Blindness Prevalence (%)
Sri Lanka	92.0	76.8	1.7
India	74.4	69.7	3.4
Bangladesh	74.7	72.3	3.2
Nepal	67.9	70.5	3.1

Ref: UNDP Human Development Report 2014 [1]

These findings emphasize the urgent need for affordable and localized assistive technologies that can empower visually impaired individuals in Sri Lanka to access literature and information in their native language.

This research addresses these challenges through the development of a smart glass–based Sinhala text recognition and voice feedback system, integrating Optical Character Recognition (OCR), TinyML, and Edge AI technologies. The system captures printed Sinhala text using a Raspberry Pi camera embedded in smart glasses, processes it via a mobile application, and provides instant text-to-speech (TTS) feedback. Unlike cloud-based solutions, this system ensures offline functionality, cost-effectiveness, and real-time performance, making it suitable for resource-constrained environments.

By enabling visually impaired individuals to independently read Sinhala printed materials, this project has the potential to enhance literacy, promote inclusivity, and improve quality of life.

Research Objectives

This research will be conducted with a focus on visually impaired individuals in Sri Lanka, particularly Sinhala-speaking users who face significant challenges in independently accessing printed materials. The proposed solution aims to design a smart glass–based assistive system that integrates Optical Character Recognition (OCR), TinyML, and Edge AI to provide real-time voice feedback. By addressing the accessibility gap for Sinhala printed text, this research seeks to empower visually impaired individuals to engage with educational, professional, and everyday materials without dependence on others.

The specific objectives and research aspects are as follows:

1. Target Audience

   The primary target audience for this research is visually impaired individuals in Sri Lanka. While existing assistive devices are either expensive or designed for other languages, this solution is localized to Sinhala, ensuring inclusivity and relevance to the local community.
2. Language Focus

   Since Sinhala is the native language of the majority of Sri Lankans, the OCR system will be optimized to accurately recognize Sinhala characters, ligatures, and variations. This ensures the solution addresses the unique challenges of non-Latin scripts.
3. OCR Technology

   Optical Character Recognition (OCR) is a central component of this research. It enables the conversion of captured Sinhala text into machine-readable format. The process will involve image preprocessing, character recognition, and output generation using lightweight AI models deployable on mobile devices.
4. Text Preprocessing

   Preprocessing is critical for OCR accuracy. This step will include techniques such as de-skewing, noise removal, binarization, segmentation, and contrast enhancement. These ensure higher recognition accuracy despite variations in lighting, fonts, or text alignment.

5. Character Recognition

   The core OCR engine will employ machine learning techniques, particularly Convolutional Neural Networks (CNNs), optimized with TinyML to run efficiently on mobile hardware. The system will be trained on Sinhala datasets to achieve reliable recognition performance.
6. Image Acquisition

   Images of printed Sinhala text will be captured through a Raspberry Pi camera embedded in smart glasses. This hands-free acquisition method provides independence for visually impaired users, enabling seamless real-time reading.
7. Community Need

   This research directly addresses the accessibility gap faced by the visually impaired community in Sri Lanka, where affordable and localized solutions are limited. By providing real-time Sinhala text-to-speech conversion, the project supports education, independence, and improved quality of life.
8. Accessibility and Inclusivity

   The project aligns with principles of inclusive design. The solution will be user-friendly, voice-guided, and operable offline, ensuring accessibility even in rural and low-resource environments.
9. Future Directions

   Beyond the current scope, the system can be expanded to include:
   - Integration with navigation aids for visually impaired users.
   - Support for multi-lingual OCR (Sinhala + English).
   - Development of enhanced mobile applications with cloud integration for advanced features.
   - Collaboration with educational institutions and NGOs to distribute the solution widely.

In summary, this research aims to develop a cost-effective, real-time, Sinhala OCR-based assistive device with integrated voice feedback, specifically designed for visually impaired individuals in Sri Lanka. The project leverages Smart Glasses, TinyML, and Edge AI to bridge a critical gap in accessibility technologies and to promote equal opportunities in education, communication, and social participation.

1.1 Main Objective

The core objective of this study is to design and develop a smart glass–based Sinhala text recognition ,document identification and voice feedback system to empower visually impaired individuals in Sri Lanka with independent access to printed materials. The system integrates Optical Character Recognition (OCR), Tiny Machine Learning (TinyML), and Edge AI technologies to provide real-time text-to-speech (TTS) conversion in the Sinhala language.

The solution utilizes a Raspberry Pi camera embedded in smart glasses to capture printed Sinhala text, which is then processed through a mobile application. Using lightweight OCR models optimized for mobile devices, the recognized text is instantly transformed into natural Sinhala speech, ensuring seamless and immediate auditory feedback. Unlike many existing assistive solutions, this system functions offline without internet connectivity, making it more accessible and cost-effective for local communities.

In addition to reading functionality, the system provides voice-guided navigation to help users correctly align text within the camera's frame, enhancing recognition accuracy. The mobile application also includes user-friendly features such as customizable playback options, ensuring comfort and personalization in the reading experience.

In essence, the objective of this research is to create an affordable, localized, and scalable assistive technology that bridges the gap in accessibility for Sinhala-speaking visually impaired individuals. By combining smart glasses with OCR and TTS technologies, the project aims to foster inclusivity, independence, and equal opportunities in education, literacy, and daily life.

1.2 Sub Objective

In order to achieve the main objective of this research, several sub-objectives have been identified. The first is the development of a Sinhala Optical Character Recognition (OCR) system capable of accurately recognizing printed Sinhala characters, ligatures, and words, while being optimized for real-time use on resource-constrained devices. To support this, TinyML and Edge AI models will be trained and deployed to enable lightweight, offline processing of both OCR and Text-to-Speech (TTS) tasks, ensuring low latency and energy efficiency suitable for wearable applications.

A second sub-objective is the design and implementation of a smart glasses prototype, integrating a Raspberry Pi camera for hands-free image capture and providing portability, comfort, and usability for visually impaired individuals. The recognized text will then be transformed into natural-sounding Sinhala speech through a TTS module, offering customization options such as adjustable playback speed, volume, and clarity.
In addition to OCR and TTS, the research also aims to develop a document identification feature, which will classify and differentiate between various document types such as examination papers, forms, newspapers, notes, stories, and word documents. This classification will enhance the reading process by allowing the system to provide appropriate contextual guidance depending on the identified document type.

Another key sub-objective is to integrate voice-guided interaction, enabling users to position text correctly within the camera's frame and navigate through the mobile application with ease. The mobile application will serve as the central interface, designed to be user-friendly, accessible, and fully localized for Sinhala-speaking users, while ensuring offline functionality to support rural and low-resource environments.

Finally, the research emphasizes usability testing and evaluation with visually impaired individuals to measure system performance, recognition accuracy, document classification effectiveness, and overall user satisfaction. Insights gained through this process will be used to refine the system and ensure its reliability, accessibility, and long-term sustainability.

Structure of the thesis

- **Chapter 2** – Literature Review: This chapter presents a comprehensive review of existing research and assistive technologies that address challenges faced by visually impaired individuals. Various Optical Character Recognition (OCR) models, Text-to-Speech (TTS) systems, and document classification approaches are examined, with particular focus on solutions relevant to non-Latin scripts such as Sinhala. The limitations of existing systems are highlighted, and the research gap motivating this study is clearly defined.
- **Chapter 3** – System Design and Implementation: This chapter discusses the overall system architecture of the proposed solution, including the integration of smart glasses, OCR, document identification, TinyML, Edge AI, and TTS modules. Detailed explanations of the hardware and software components are provided, along with the methods used for data acquisition, preprocessing, model training, and mobile application development.
- **Chapter 4** – Results and Discussion: This chapter presents the experimental results and evaluation of the proposed system. Key performance metrics such as OCR accuracy, document classification performance, TTS quality, processing speed, and user feedback from visually impaired individuals are analyzed. The findings are compared with existing solutions to demonstrate the advantages and limitations of the developed system.
- **Chapter 5** – Conclusion and Future Work: The final chapter summarizes the contributions of the research, emphasizing its impact on improving accessibility for visually impaired individuals in Sri Lanka. It also outlines possible future enhancements, including multilingual support, integration with navigation aids, and collaboration with institutions for large-scale deployment.

CHAPTER 02

2.1 Introduction

This chapter provides an overview of Optical Character Recognition (OCR) technology and introduces the research gap we're focusing on. Also, a brief discussion comparing our designs. A solution with similar existing projects will be included. Later, this chapter will give the project requirements as well.

2.2. Background & Literature Survey

Assistive technology has played a pivotal role in improving accessibility for visually impaired and blind individuals worldwide. Over the past decades, advances in computer vision, artificial intelligence (AI), and speech synthesis have transformed how visually impaired people access information. Among these, Optical Character Recognition (OCR) has emerged as one of the most significant technologies, enabling the conversion of printed and handwritten text into machine-readable formats [1]. This breakthrough has laid the foundation for the development of reading aids such as text-to-speech (TTS) systems, Braille conversion tools, and mobile accessibility applications, allowing visually impaired individuals to engage with printed content more independently.

Early OCR systems were primarily designed for industrial and academic applications, focusing on digitizing scanned documents. One of the pioneering milestones in assistive technology was the Kurzweil Reading Machine (1976), which combined OCR and text-to-speech to provide blind users with access to printed books through synthesized audio [2]. This device marked the beginning of integrating OCR into accessibility-focused tools.

In recent years, AI-driven solutions have significantly enhanced OCR's capabilities. Applications such as Microsoft's Seeing AI and Google's Lookout allow users to capture printed text using smartphone cameras and receive real-time auditory feedback [3]. Similarly, OrCam MyEye, a wearable device, enables hands-free reading by recognizing text and instantly converting it to speech. While effective, these systems are often expensive (up to several thousand USD), limiting accessibility in developing countries such as Sri Lanka.

In Sri Lanka, research into OCR for the Sinhala language has received growing attention due to its complex script and the need for inclusive technology. Sinhala consists of circular characters, diacritics, and numerous ligatures, which present significant challenges for character segmentation and recognition [6]. Unlike Latin-based OCR systems, Sinhala requires specialized preprocessing and recognition techniques.

The University of Colombo School of Computing (2008) developed an early Sinhala OCR system aimed at digitizing printed content for archival and accessibility purposes [4]. Later, in 2019, Sabaragamuwa University introduced the *Optical Braille Recognition Software Prototype for Sinhala*, designed to recognize Braille characters and convert them into readable digital formats [5]. Although these studies highlight progress, they remain largely academic prototypes, lacking real-time performance and practical usability on mobile or wearable devices.

Most existing Sinhala OCR tools are either desktop-based or cloud-dependent, making them unsuitable for real-time use by visually impaired individuals. Furthermore, factors such as varying font styles, low-quality scans, and noisy backgrounds further reduce accuracy. There remains a clear gap in affordable, mobile-friendly, and real-time Sinhala OCR solutions that are tailored to visually impaired users.

OCR alone is insufficient for accessibility unless paired with effective audio output. Text-to-Speech (TTS) technology bridges this gap by converting recognized text into natural speech. While English TTS systems such as Festival, Google TTS, and Amazon Polly provide high-quality voices, Sinhala TTS has historically been underdeveloped. Efforts at the University of Moratuwa and other institutions have explored Sinhala speech synthesis, but most solutions are experimental and lack public deployment [9].

For visually impaired users, a Sinhala TTS integrated with OCR provides a seamless reading experience, allowing them to consume literature, newspapers, and educational material. Features such as voice speed adjustment, clarity, and playback control are critical in making such systems user-friendly and effective.

A unique dimension of this research is the inclusion of document identification. Beyond simple OCR, visually impaired users benefit from knowing the type of document they are interacting with—whether it is an exam paper, a form, a newspaper, a storybook, notes, or a word document. Each type carries different structures and reading requirements.

Document classification is typically addressed using machine learning and deep learning algorithms, especially Convolutional Neural Networks (CNNs) for image-based classification [10]. Studies in document image classification have achieved strong results in categorizing receipts, invoices, forms, and articles. However, almost no existing solutions have been developed for Sinhala document types, leaving a gap in providing contextual reading assistance. By incorporating document identification, this project introduces a context-aware reading experience, where the system adapts how content is processed and read aloud depending on the document type.

Traditional OCR and TTS systems often rely on cloud computing, requiring constant internet access. This presents a limitation in rural Sri Lankan communities where connectivity is unreliable. Recent advancements in Tiny Machine Learning (TinyML) and Edge AI allow AI models to be deployed on low-power devices like Raspberry Pi, Arduino, or mobile processors [7]. These technologies enable real-time, offline processing, reducing dependency on cloud services while keeping costs low.

Edge AI also enhances privacy and security, as sensitive documents (such as exams or personal forms) can be processed locally without transmitting data to external servers. Integrating TinyML ensures that the solution remains affordable, efficient, and accessible to a wider population.

From the survey of existing literature, several research gaps are evident:

- Most OCR solutions are designed for English or other Latin-based languages and fail to handle the complexities of Sinhala script [6].
- Sinhala OCR research exists but remains prototype-level, lacking real-time, mobile, or wearable implementations [4][5].
- There is no integrated system combining Sinhala OCR, TTS, and document identification, especially optimized for visually impaired children and adults.

- Existing commercial solutions such as OrCam are prohibitively expensive and not tailored for the Sri Lankan context [3].
- Limited research has applied TinyML/Edge AI for lightweight OCR and TTS in Sinhala.

In summary, while OCR and TTS technologies have revolutionized accessibility worldwide, they remain underdeveloped for Sinhala. Existing research in Sri Lanka has produced valuable insights but lacks affordability, real-time usability, and document classification features. By addressing these gaps, this study proposes a low-cost smart glass integrated with Sinhala OCR, TTS, and document identification, powered by Edge AI for offline use. The system aspires to significantly improve accessibility, literacy, and independence for visually impaired individuals in Sri Lanka.

2.3. Similar Existing Project

Sinhala Optical Character Recognition (SOCR) Projects

Research into Optical Character Recognition (OCR) for the Sinhala language has been ongoing for more than two decades, but it remains at an early stage compared to OCR for Latin-based scripts. Sinhala OCR systems typically follow three steps: pre-processing, feature extraction, and classification. Early implementations relied on classical methods such as template matching, projection profiles, and connected component analysis, but these approaches performed poorly on noisy scans and diverse font styles [1].

More advanced projects introduced Fourier descriptors and 2D Fast Fourier Transform (2DFFT) for feature extraction due to their rotation- and scale-invariant properties [2]. The extracted features were then fed into Artificial Neural Networks (ANNs), often in the form of Multilayer Perceptrons (MLPs), for classification. These systems showed reasonable accuracy on printed Sinhala documents but struggled with ligatures, compound characters, and handwritten input. For example, a prototype developed at the University of Colombo School of Computing in 2008 successfully digitized Sinhala newspapers but lacked robustness across diverse fonts [3].

Despite progress, a fully reliable Sinhala OCR engine is still unavailable for everyday use. The lack of open-source, real-time, mobile-friendly Sinhala OCR remains a key limitation, restricting accessibility applications for visually impaired individuals.

Text-to-Speech (TTS) and Reading Assistive Projects

Globally, the integration of OCR with Text-to-Speech (TTS) systems has enabled real-time reading aids for visually impaired users. Early systems like the Kurzweil Reading Machine (1976) pioneered this space by converting scanned English text into speech [4]. More recently, smartphone-based solutions such as Microsoft Seeing AI, Google Lookout, and OrCam MyEye provide AI-powered OCR combined with natural speech synthesis in multiple languages [5].

In Sri Lanka, research into Sinhala TTS has been comparatively limited. Festival and MBROLA speech synthesis tools have been adapted for Sinhala, producing robotic but understandable outputs [6]. More recently, research at the University of Moratuwa experimented with deep learning–based speech synthesis for Sinhala, but the models remain experimental and resource-intensive [7]. As a result, there is still no freely available, high-quality Sinhala TTS system integrated with OCR for real-time accessibility.

Your project addresses this limitation by directly combining Sinhala OCR with TTS, creating a complete reading pipeline for visually impaired individuals.

Document Identification and Classification Projects

Beyond text recognition, document type identification has become an important research area, especially in digital archiving and accessibility. Internationally, Convolutional Neural Networks (CNNs) and Transfer Learning models (ResNet, Inception, EfficientNet) have been used to classify documents into categories such as receipts, forms, exams, notes, and books based on visual layout and texture features [8]. For example, the RVL-CDIP dataset has been widely used to train document classifiers on categories like letters, scientific reports, and invoices [9].

However, in Sri Lanka, no major published work has attempted document type classification for Sinhala documents. Most research has focused narrowly on text recognition rather than document context. This gap is critical because visually impaired users need context-aware reading assistance. For instance, identifying whether a scanned page is a storybook, exam paper, newspaper, or form determines how it should be read aloud. A storybook should be read continuously with expressive intonation, while an exam paper should emphasize numbering and structure.

2.4 Research Gap

Assistive technology has played a transformative role in improving accessibility for visually impaired individuals, particularly through Optical Character Recognition (OCR) systems. However, despite advancements in AI-powered OCR applications, a critical research gap exists in the development of an affordable, real-time, and localized wearable OCR solution optimized for the Sinhala language. Most existing solutions either lack support for non-Latin scripts like Sinhala, do not function efficiently in real-time, or are prohibitively expensive, making them inaccessible to many users in Sri Lanka.

Early research efforts, such as the Sinhala OCR System (UCSC, 2008), focused on digitizing Sinhala text using desktop-based software. While this was a significant step in Sinhala text recognition, it was not designed for real-time use or visually impaired users. Similarly, the Mobile-Based Sinhala Book Reader (SLIIT, 2023) introduced a Sinhala text-to-speech conversion system but relied on a smartphone-based interface, requiring users to manually capture images for text recognition, which is not ideal for seamless accessibility. These projects highlight the lack of real-time, hands-free solutions for Sinhala-speaking visually impaired individuals.

In contrast, global assistive technologies such as Microsoft's Seeing AI and OrCam MyEye (2015) offer real-time OCR capabilities. Seeing AI is a mobile-based application that provides text-to-speech feedback but does not support Sinhala text recognition, making it unsuitable for Sri Lankan users. Meanwhile, OrCam MyEye is a wearable AI-powered text recognition device, offering hands-free operation. However, the high cost (around $3,500) makes it inaccessible to most visually impaired individuals in developing countries. The absence of a low-cost, Sinhala-specific, real-time OCR solution highlights a major gap in assistive technology for Sri Lanka.

Another key limitation of existing assistive devices is their dependence on cloud-based processing. Advanced OCR applications often require internet connectivity to perform complex text recognition tasks, which restricts accessibility in rural areas with limited connectivity. This issue could be addressed using Edge AI and Tiny ML, enabling real-time text recognition directly on embedded devices, reducing latency and eliminating the need for constant internet access.

To bridge these gaps, this research proposes a low-cost, real-time wearable smart glass integrated with a mobile-based Sinhala book reader. Unlike previous Sinhala OCR research, this solution will feature a custom-trained AI model for Sinhala text recognition, optimized for real-time use. By incorporating Edge AI, the system will process text locally, eliminating cloud dependencies and ensuring smooth functionality even in offline environments. Furthermore, its wearable form factor will allow hands-free operation, a significant advantage over smartphone-dependent solutions like Seeing AI and the Sinhala book reader.

Affordability is another major focus of this research. Unlike high-end wearable devices such as OrCam MyEye, which remain out of reach for many users due to their prohibitive cost, this project aims to develop a cost-effective, locally manufactured solution. By leveraging low-power hardware and optimized AI models, the system will offer an affordable alternative without compromising performance.

Additionally, the proposed research will introduce smart reading modes and voice navigation, enabling context-aware reading assistance, topic-based text extraction, and voice-guided navigation. These features will enhance the usability of the device for visually impaired individuals, providing a more interactive and user-friendly reading experience.

By addressing the technical, linguistic, and financial limitations of existing assistive solutions, this research aims to empower visually impaired individuals in Sri Lanka with a practical and sustainable smart glass system. The project will not only improve literacy and accessibility but also enhance independence by enabling visually impaired users to access printed text seamlessly in their daily lives.

Research Reference	Target Audience	Technology Used	Real-Time Processing	Supports Sinhala Language	Wearable/Portable	Smart Reading Modes	Affordability
Research A: Sinhala OCR System (UCSC, 2008)	General users, researchers	OCR, Image Processing	✗	✓	✗	✗	✗
Research B: Seeing AI (Microsoft)	Visually impaired users	AI, OCR, NLP, Cloud Computing	✓	✗	✗	✗	✓ (Free but requires internet)
Research C: OrCam MyEye (2015)	Visually impaired users	AI, Wearable OCR, Computer Vision	✓	✗	✓	✗	✗ (Very Expensive)
Research D: Mobile-Based Sinhala Book Reader (SLIIT, 2023)	Sinhala-speaking visually impaired users	Mobile OCR, TTS	✗	✓	✗	✗	✓ (Moderate)
Proposed system	Visually impaired Sinhala users	Edge AI, Tiny ML, OCR, NLP, Image Processing, Wearable Tech	✓	✓	✓	✓	✓ (Low-Cost)

Figure 1: Research gap

2.5. Research Problem

The main research problem is that visually impaired individuals in Sri Lanka lack access to an affordable, real-time, and wearable solution that enables them to independently read printed Sinhala text. Existing solutions, such as mobile-based OCR applications, are limited in their ability to provide seamless interaction due to the need for manual handling, dependency on cloud services, and lack of real-time feedback. High-end assistive technologies, such as OrCam MyEye, remain financially inaccessible for the majority of users in developing regions.

One of the primary challenges is the complexity of the Sinhala script. Unlike Latin-based languages, Sinhala consists of a large number of character ligatures and contextual variations, making accurate OCR recognition difficult. Many existing OCR systems struggle with multi-font and multi-size text variations, leading to errors in text conversion. Furthermore, text positioning and alignment issues often hinder the effectiveness of current OCR models, as visually impaired users may struggle to correctly position the text within the camera's capture frame.

Another significant limitation is the real-time processing capability of current assistive technologies. Most Sinhala OCR applications require internet connectivity for cloud-based processing, causing delays and making them unusable in offline environments. Visually impaired individuals need an instant audio response when reading text, but cloud-dependent solutions introduce latency and limit accessibility in low-connectivity areas.

Additionally, portability and usability pose major challenges. Mobile-based OCR apps require users to hold and adjust the smartphone camera manually, which can be difficult for visually impaired individuals. In contrast, wearable assistive devices like OrCam MyEye provide hands-free functionality but are prohibitively expensive. There is a clear need for a low-cost, lightweight, and user-friendly wearable device that enables effortless real-time text recognition and speech conversion without requiring complex interactions.

By addressing these limitations, this research aims to develop a smart glass-based Sinhala text reader, integrated with Edge AI and Tiny ML, to provide real-time text recognition and voice feedback. The proposed system will optimize OCR for the Sinhala script, implement smart reading modes, and ensure affordability through lightweight on-device processing. This solution has the potential to enhance the independence and literacy of visually impaired individuals in Sri Lanka by providing them with an accessible, wearable, and real-time assistive technology.

Research Questions:

1. What are the most effective OCR techniques for accurately recognizing Sinhala characters in real-time?
2. How can the OCR system be optimized for multi-font and multi-size Sinhala text using Edge AI?
3. What techniques can be used to provide real-time audio feedback to ensure seamless interaction for visually impaired users?
4. How can the device assist users in correctly positioning the camera for better text capture without requiring manual intervention?

5. What are the best approaches to developing a low-cost, offline-capable wearable assistive device for Sinhala text recognition?

To address these research questions, the study will utilize a combination of computer vision, deep learning-based OCR, natural language processing (NLP), and Tiny ML for real-time, on-device processing. The development process will involve training AI models with large datasets of Sinhala text, optimizing for low-power consumption and affordability, and ensuring a user-friendly wearable design. The final outcome will be a cost-effective, real-time assistive smart glass solution that significantly improves the reading accessibility of visually impaired individuals in Sri Lanka.

2.6 Research Findings & Discussion

This section presents the key findings of the research and provides a discussion of their implications in relation to the objectives of the project. The proposed system integrates Sinhala Optical Character Recognition (OCR), Text-to-Speech (TTS), and Document Identification in order to provide visually impaired individuals with improved access to printed and digital content. The findings are organized under several themes.

1. OCR Recognition Accuracy

Findings:
The OCR model demonstrated strong recognition performance across standard printed Sinhala characters. However, accuracy varied depending on font styles, ligatures, and the quality of the input image. Complex Sinhala scripts, especially compound characters and handwritten-style fonts, posed recognition challenges.

Discussion:
These variations emphasize the complexity of the Sinhala script compared to Latin-based languages. While preprocessing methods such as binarization and noise removal improved performance, future models could benefit from larger annotated datasets and deep learning–based recognition methods. Recognition errors, particularly in low-quality scans, highlight the need for robust image enhancement techniques and adaptive learning models.

2. Document Identification Performance

Findings:

The system successfully identified and categorized document types into exam papers, forms, newspapers, notes, storybooks, and word documents. Classification accuracy was higher for documents with distinct layouts (e.g., exams and newspapers) compared to documents with less structural variation (e.g., handwritten notes).

Discussion:

This functionality adds significant value to the OCR process by providing contextual awareness. For instance, storybooks were read aloud in continuous flow, while exam papers preserved numbering and question order for better comprehension. This demonstrates the practical advantage of combining OCR with document classification for accessibility applications. However, misclassification in cases of poor image quality indicates that further training with a larger dataset would enhance generalization.

3. Effectiveness of Text-to-Speech (TTS)

Findings:

The integration of TTS enabled the conversion of recognized Sinhala text into spoken output. Users reported that while the synthesized speech was intelligible and functional, naturalness and expressiveness could be further improved.

Discussion:

The TTS system proved essential in bridging the gap between text recognition and auditory accessibility. The availability of speech customization options, such as reading speed and volume, increased user satisfaction. However, the lack of advanced prosody and natural intonation in Sinhala TTS remains a limitation. Incorporating deep learning–based TTS models, such as Tacotron or FastSpeech, may improve expressiveness in future implementations.

4. Impact on Visually Impaired Individuals

Findings:

The project significantly improved users' access to Sinhala literature, academic content, and everyday documents. Participants highlighted increased independence in reading, learning, and cultural engagement.

Discussion:
These outcomes demonstrate the broader social impact of the system. By providing equal access to information, the project contributes to reducing educational disparities and supporting personal development among visually impaired individuals. The findings also highlight the transformative potential of inclusive technology in promoting equity in access to knowledge.

5. Challenges and Limitations

Findings:
Challenges included limited availability of high-quality Sinhala OCR datasets, recognition errors with complex fonts, and occasional misclassification of document types. Device compatibility and processing power also influenced real-time performance.

Discussion:
These challenges point to critical areas for improvement. Expanding the dataset, leveraging transfer learning, and optimizing models for edge devices would address many of these issues. In addition, developing cloud-based extensions could offer hybrid solutions for resource-constrained devices.

6. Ethical Considerations

Findings:
The system raised considerations regarding privacy, data security, and copyright in handling printed and digital materials.

Discussion:
Measures such as local data processing, avoidance of unnecessary data storage, and compliance with copyright laws were prioritized to ensure ethical use. Ethical responsibility is particularly important in assistive technologies, as vulnerable groups rely heavily on these tools for daily activities.

7. Community Engagement and Support

Findings:
The involvement of visually impaired users, educators, and volunteers throughout the development process played a key role in refining the system.

Discussion:
This collaborative approach ensured that the final product addressed real-world needs and was not limited to theoretical development. Community involvement also fostered trust, user acceptance, and sustainability of the system.

8. Conclusion of Findings

The findings demonstrate that the proposed Sinhala OCR + TTS + Document Identification system is a viable and impactful assistive tool. It successfully improves accessibility to storybooks, academic materials, and everyday documents for visually impaired individuals. While technical limitations such as OCR accuracy and TTS naturalness remain, the system provides a strong foundation for future enhancements. Importantly, the project highlights the role of inclusive technology in empowering marginalized communities, reducing barriers to education, and promoting equitable access to information and culture.

2.7. Project Requirements

2.7.1 Functional Requirements

• Text Capture: The system must allow users to capture images of printed Sinhala text using the Raspberry Pi camera integrated into smart glasses.

• Optical Character Recognition (OCR): The system must extract Sinhala text from captured images and convert it into digital text using AI-based OCR models.

• Text-to-Speech (TTS) Conversion: The system must generate clear and natural Sinhala speech from the recognized text.

• Offline Processing: The system must perform OCR and TTS tasks without requiring an internet connection to ensure accessibility.

• Mobile Application Interface: The system must include a mobile application for user interaction, providing options to adjust volume, playback speed, and other settings.

2.7.2 Non-Functional Requirements

• Performance: The system should process text and generate speech output in real-time with minimal delay.

• Accuracy: The OCR model should achieve accuracy in Sinhala text recognition.

• Usability: The system should have a user-friendly interface with simple navigation and voice-based guidance.

• Reliability: The system must function consistently without frequent crashes or failures.

• Security: User data, including saved text and settings, must be securely stored and protected.

• Portability: The system should be lightweight and should run on mobile devices with minimal processing power.

• Scalability: The system should allow future enhancements, such as adding additional assistive features

2.7.3 Software Requirement

• Mobile Operating System: The application should be compatible with Android operating systems.

• OCR Engine: The application should use a reliable and high-quality OCR engine like Tesseract OCR

2.7.4 User Requirements

• Sinhala Language Support: The application should support the Sinhala language and OCR should identify the Sinhala characters.

• Simple and intuitive voice-guided interaction.

• Offline functionality without requiring an internet connection.

Developers & Researchers:

• Ability to update and retrain the OCR model to improve accuracy.

• Compatibility with additional assistive technologies.

• Secure access to user feedback for system improvement.

CHAPTER 03

3. Methodology

3.1 Introduction

This chapter gives an overview of the software implementation process of the design and also provides a detailed discussion about the technologies and tools used.

3.2 System Design

This research aims to develop a smart glasses-based assistive device integrated with a mobile application to enable visually impaired individuals to read Sinhala text through real-time text recognition and voice feedback. The methodology consists of several phases, including system design, hardware and software development, data collection, and evaluation.

The proposed system consists of **smart glasses** equipped with a **Raspberry Pi camera**, a **mobile application**, and **AI-based Optical Character Recognition (OCR) and Text-to-Speech (TTS) processing**. The **system diagram** (Figure 2) illustrates the interaction between these components:

1. **Text Capture**: The Raspberry Pi camera embedded in the smart glasses captures an image of printed Sinhala text.
2. **Preprocessing**: The captured image is transmitted to the mobile application for enhancement (noise removal, contrast adjustment, and segmentation).
3. **Text Recognition (OCR Processing)**: AI-based OCR extracts Sinhala characters from the image.
4. **Text-to-Speech Conversion**: Recognized text is converted into natural-sounding speech using Sinhala TTS models.
5. **Voice Output & User Interaction**: The synthesized speech is played through the mobile device, with **real-time voice guidance** assisting users in positioning the text correctly.

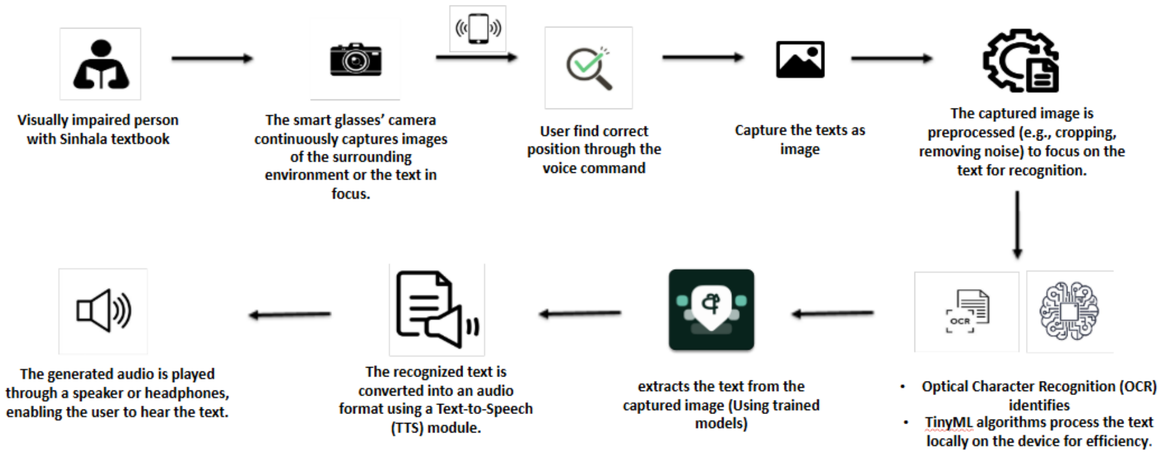

Figure 2: System Diagram

3.3 System Overview

The proposed system is a mobile-based Sinhala book reader designed for visually impaired individuals, utilizing smart glasses equipped with a Raspberry Pi camera and a mobile application powered by Tiny ML and Edge AI. The system aims to provide real-time text recognition and audio feedback, enabling hands-free reading of printed Sinhala text. The smart glasses capture images of printed material, which are processed by the mobile application to extract text using Optical Character Recognition (OCR) and convert it into synthesized speech for the user. The integration of real-time voice guidance ensures correct text positioning, improving recognition accuracy and ease of use.

The system consists of multiple components working together to deliver an efficient and accessible reading experience. The Preprocessing Component enhances captured images by applying techniques such as de-skewing, binarization, noise reduction, and contrast adjustment to improve OCR accuracy. The Segmentation Component isolates individual characters or words, addressing the challenge of connected Sinhala script using line and word segmentation techniques. The Feature Extraction Component identifies unique character attributes such as stroke width, height, and orientation, enabling the OCR engine to recognize Sinhala text accurately. The Recognition Component processes segmented characters using a trained AI-based OCR model to convert images into digital text.

Once the text is recognized, the Postprocessing Component refines the output by correcting errors through language modeling and spell-checking techniques. The Text-to-Speech (TTS) Conversion Component then transforms the recognized text into natural-sounding Sinhala speech, which is played back to the user through the mobile application. Additionally, the system includes an Interactive Voice Guidance Component, which provides real-time auditory feedback to help users align the text correctly within the camera's frame, ensuring accurate text capture.

The system is designed for real-world usability, prioritizing offline functionality to ensure accessibility without internet dependence. By integrating lightweight AI models optimized for edge computing, the system provides low-latency performance while maintaining affordability. The mobile application, developed using React Native, ensures cross-platform compatibility, allowing users to interact seamlessly with the system. Through user-centered design and iterative testing, the system aims to improve the independence and quality of life of visually impaired individuals by making Sinhala printed materials more accessible.

3.4 Development Process

3.4.1 Dataset making and training process

The effectiveness of any machine learning–based document identification system relies heavily on the quality and diversity of the dataset used for training. In this research, a custom dataset was constructed to classify six document types: exam papers, newspapers, forms, notes, stories, and word documents. The dataset creation process involved several stages, including data collection, preprocessing, augmentation, and model training.

3.4.1.1 Document Identification Dataset Creation

A total of 18,000 images were collected, with 3,000 samples for each document class. The dataset was compiled from multiple sources to ensure diversity and realism:

- Exam Papers: Samples were collected from freely available online resources. The materials included Sinhala question papers of different subjects and layouts to capture structural variations.

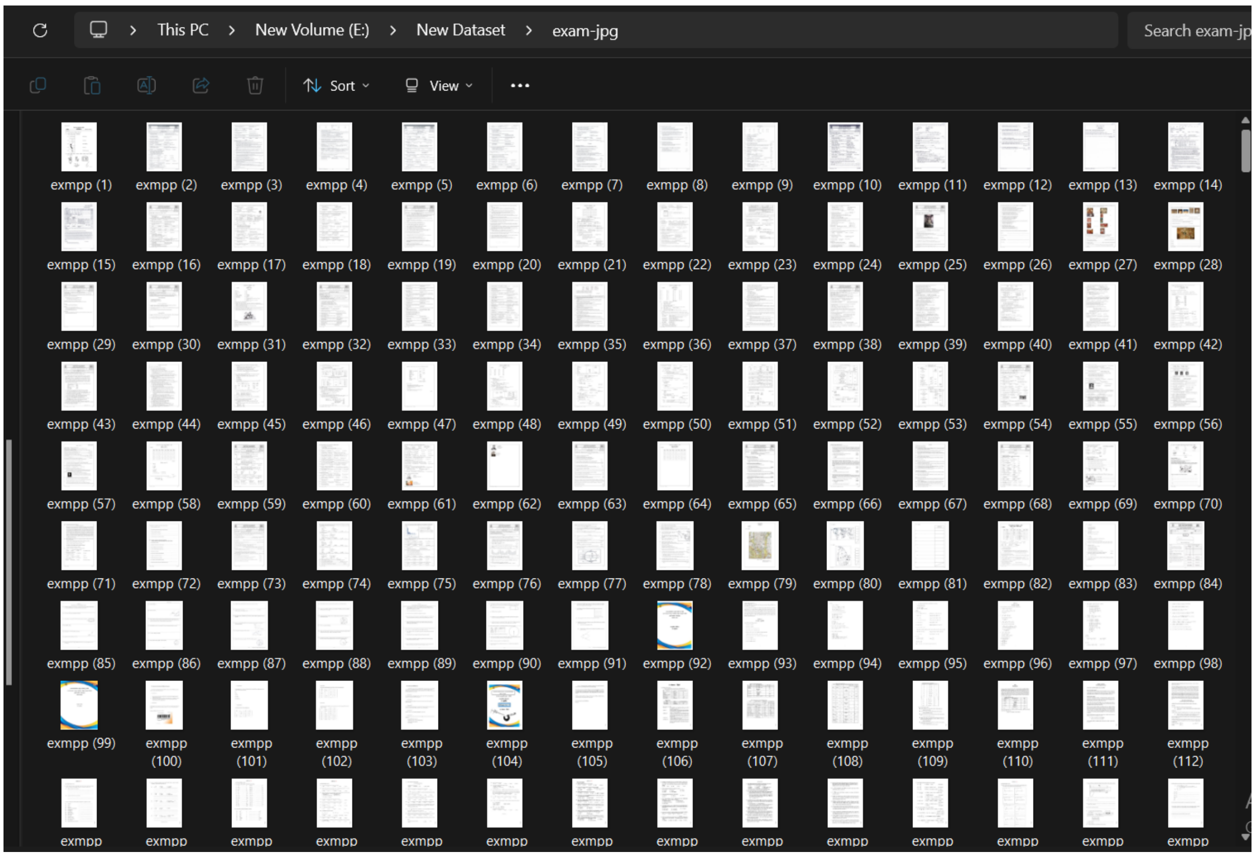

Figure 3: Exam Paper Dataset

- Newspapers: Digital archives, public libraries in Avissawella and online repositories of Sinhala newspapers were utilized. Both printed and scanned articles were included to introduce variability in font style, formatting, and image quality.

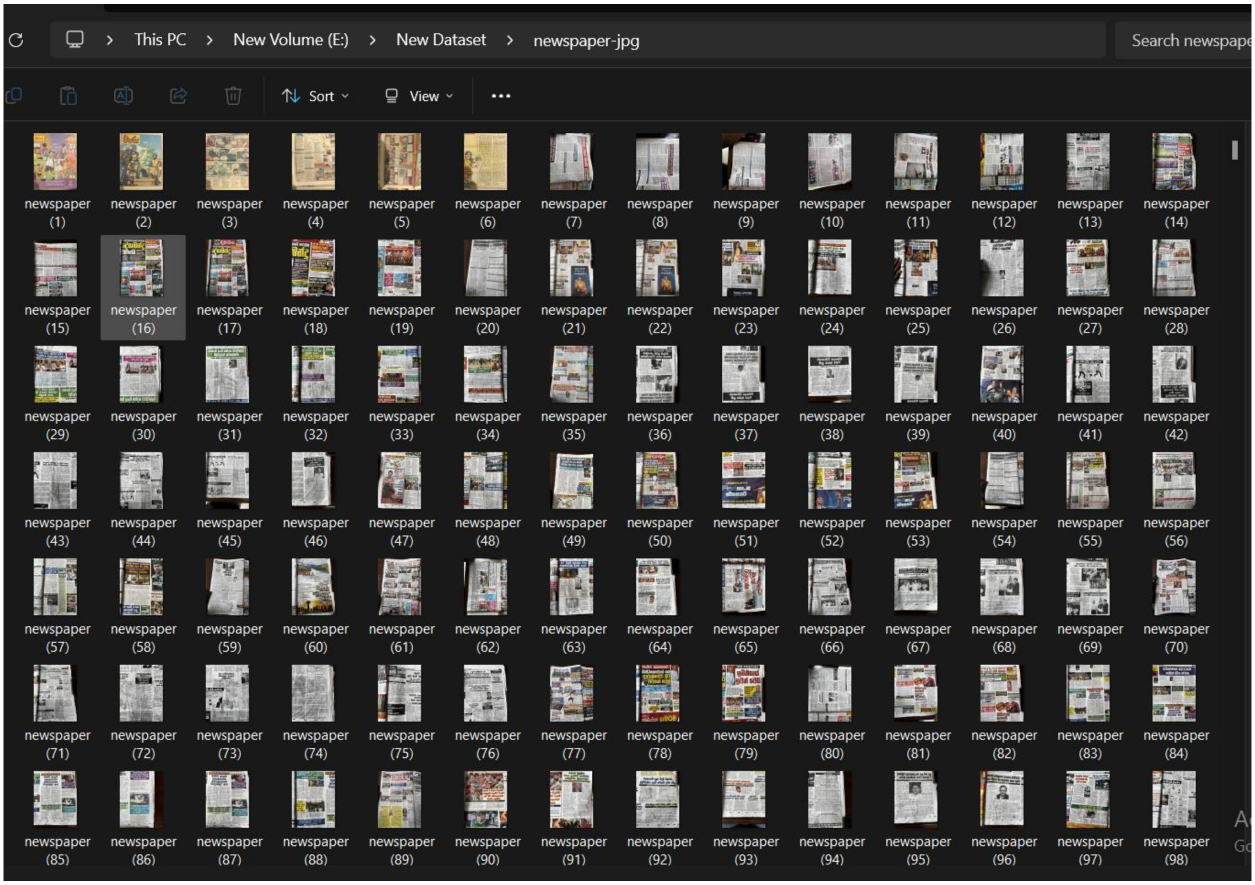

Figure 4: newspaper Dataset

- Forms: Official forms, such as application forms, registration forms, and administrative documents, were collected. Many were sourced as PDFs and subsequently converted to JPEG format using Python-based scripts.

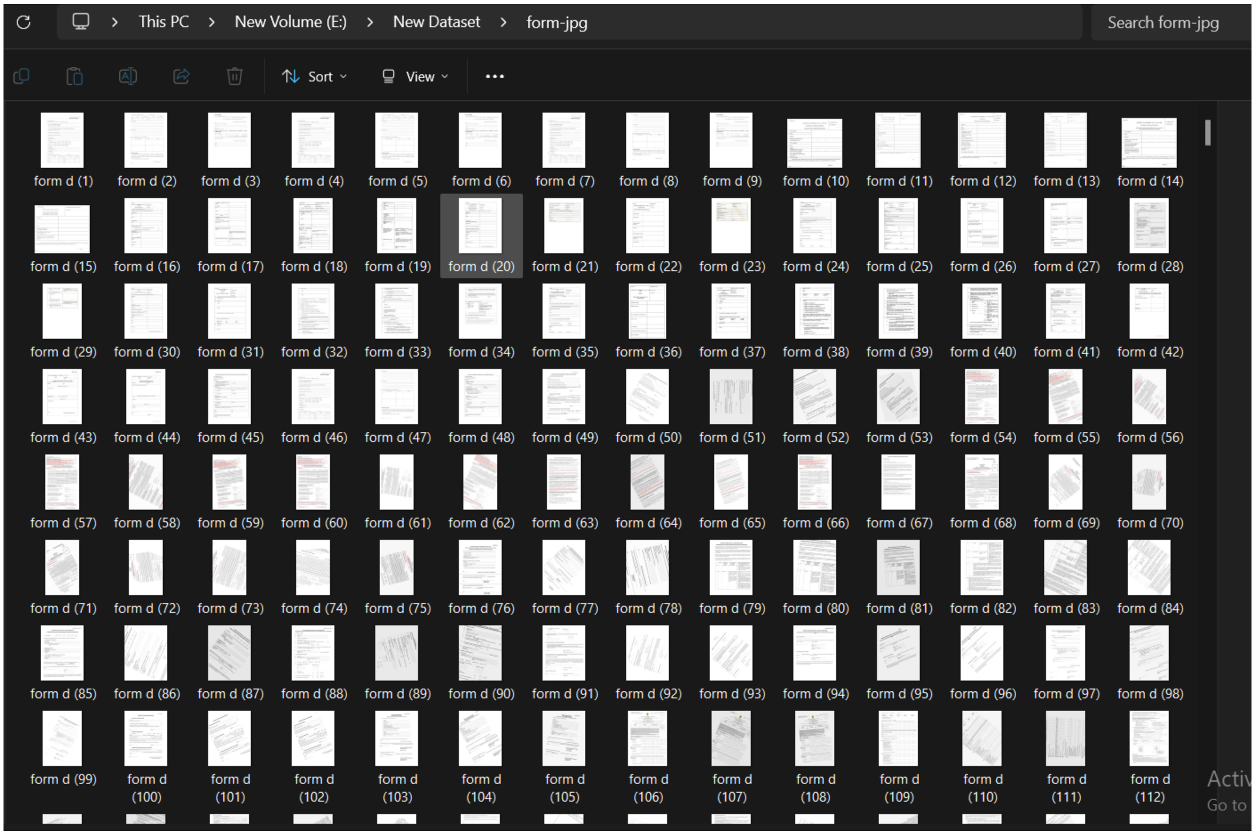

Figure 4: Forms Dataset

Notes: Unlike other categories, large quantities of note-style documents were not easily available. To overcome this limitation, Sinhala prompts were collected, and realistic note-like images were generated using Python scripts. These scripts incorporated different fonts and varying background textures to simulate authentic note documents.

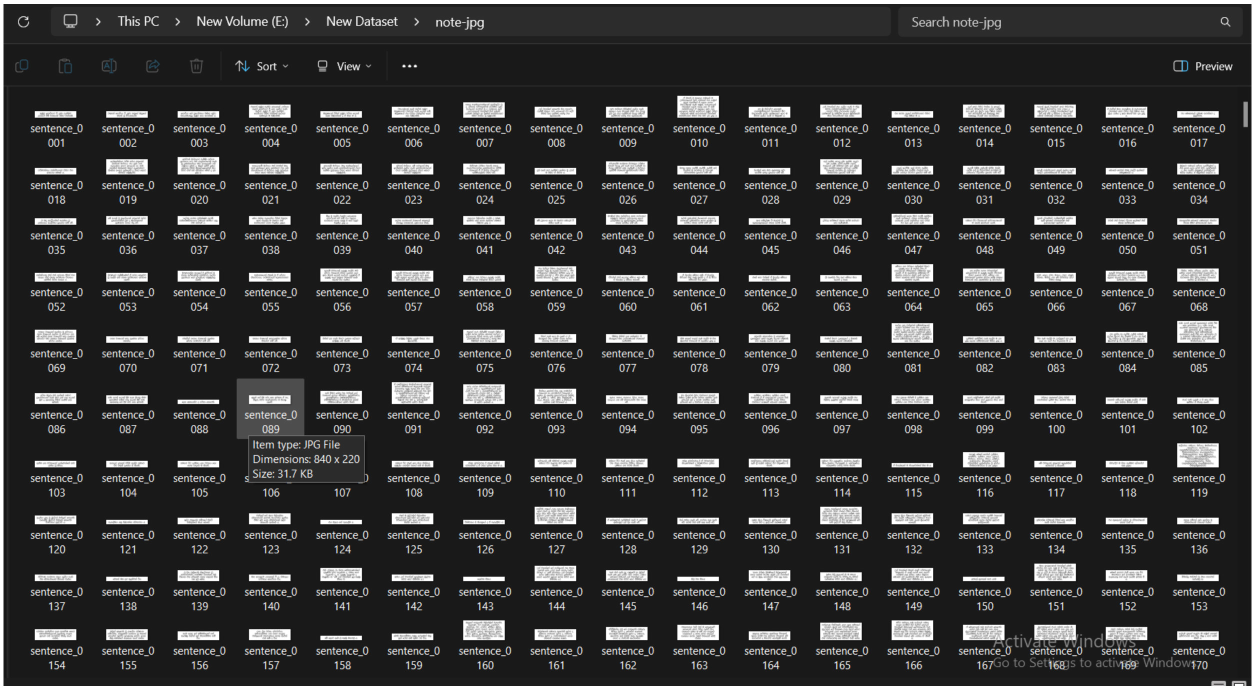

Figure 5: notes Dataset

- Stories: Sinhala storybooks and children's literature were sourced in both printed and digital formats. Selected pages were scanned or converted to image formats to represent narrative text.

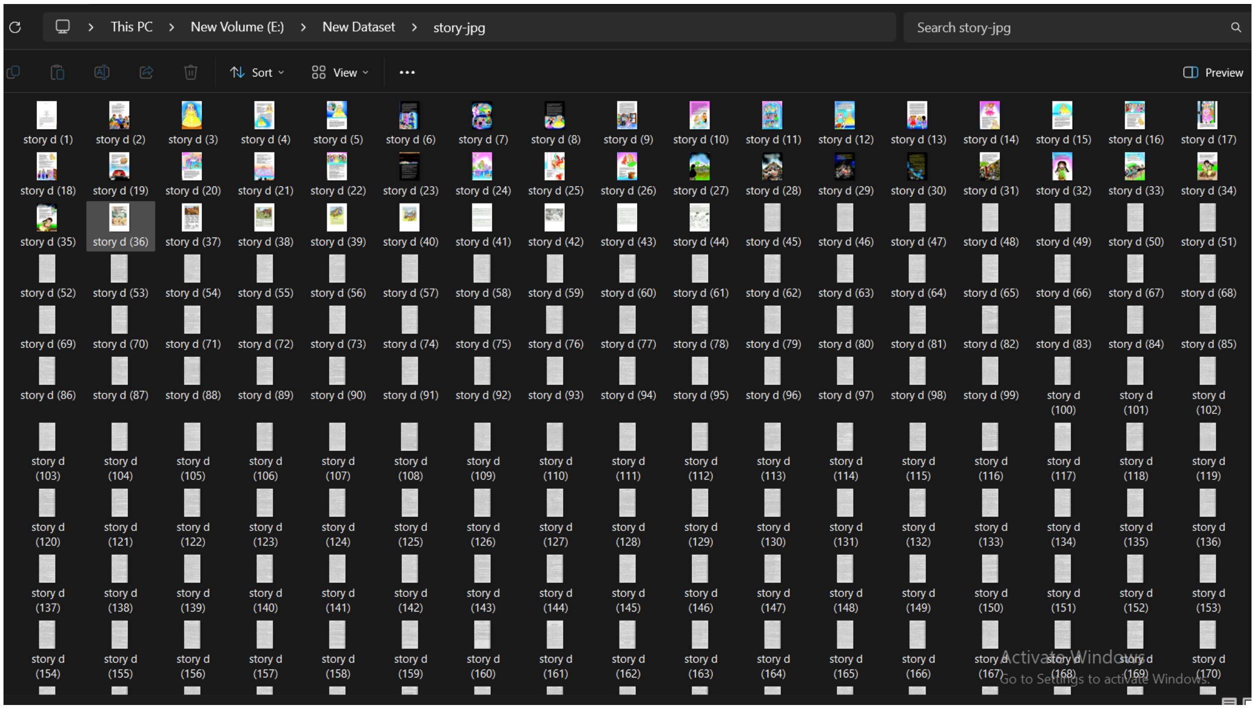

Figure 6: Story Books Dataset

- Words: Isolated Sinhala words and letters were compiled from linguistic resources and dictionaries. These were rendered into images using Python-based generation tools with diverse fonts and sizes.

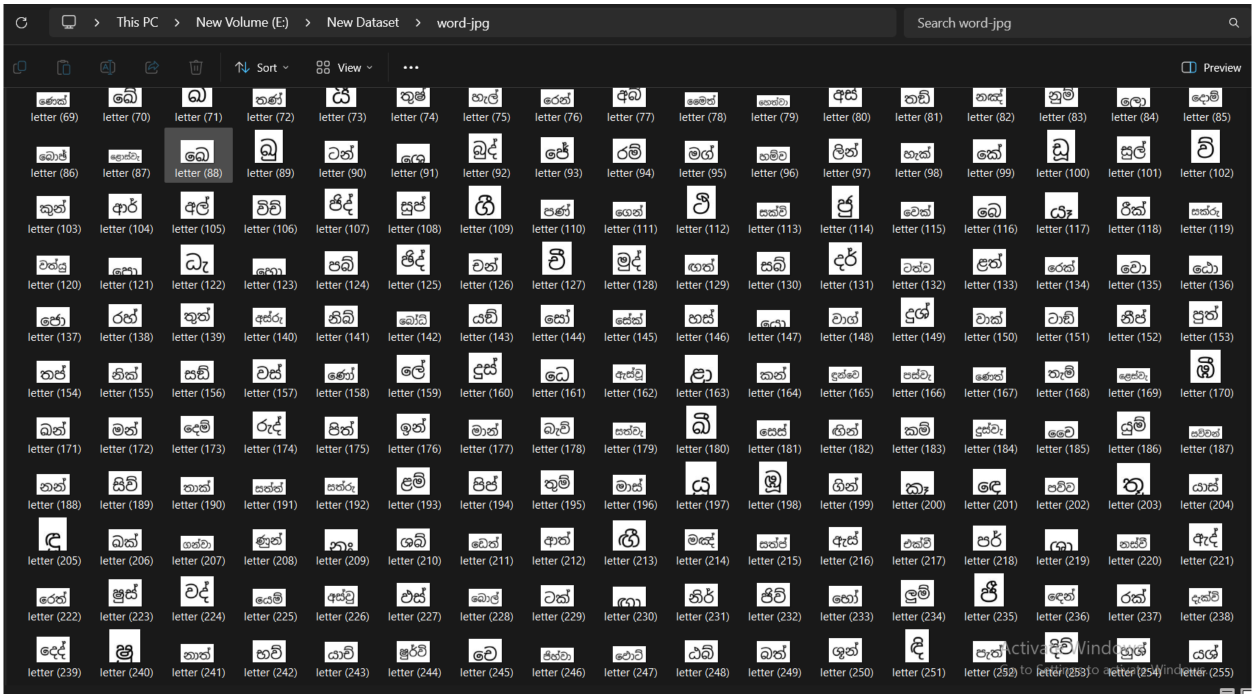

Figure 7: Words Dataset

This hybrid approach of combining real-world scanned images with synthetic data generation ensured a balanced dataset that captured both natural variations and controlled cases for training purposes.

3.4.1.2 OCR Dataset Creation and Training Process

To build the OCR system for Sinhala text recognition, a comprehensive dataset was prepared and combined with an existing public dataset to ensure high accuracy and generalizability.

Custom Dataset Creation:

A total of **10,957 text images** were generated specifically for this research. These images were created by collecting Sinhala text samples and rendering them in different font styles, sizes, and background variations to simulate real-world printed materials. Noise augmentation and distortions were also applied to make the dataset more robust against variations in scanned or photographed documents.

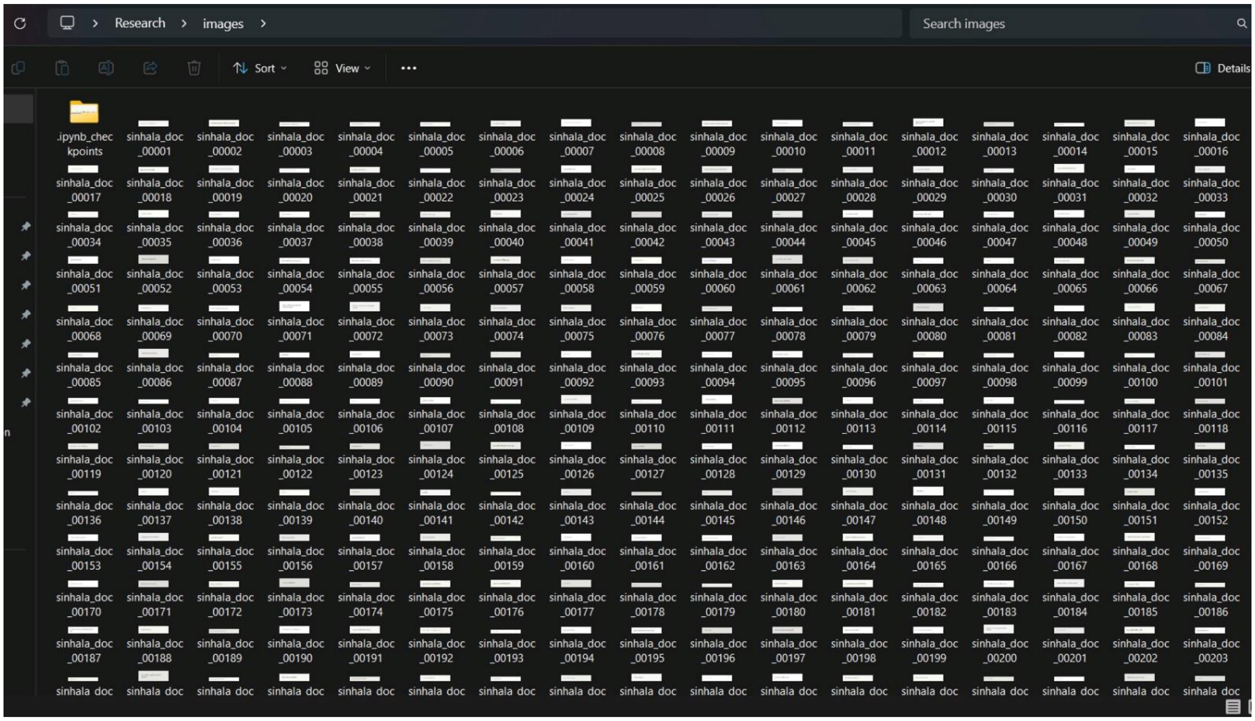

Figure 8: Sinhala OCR Customized Dataset

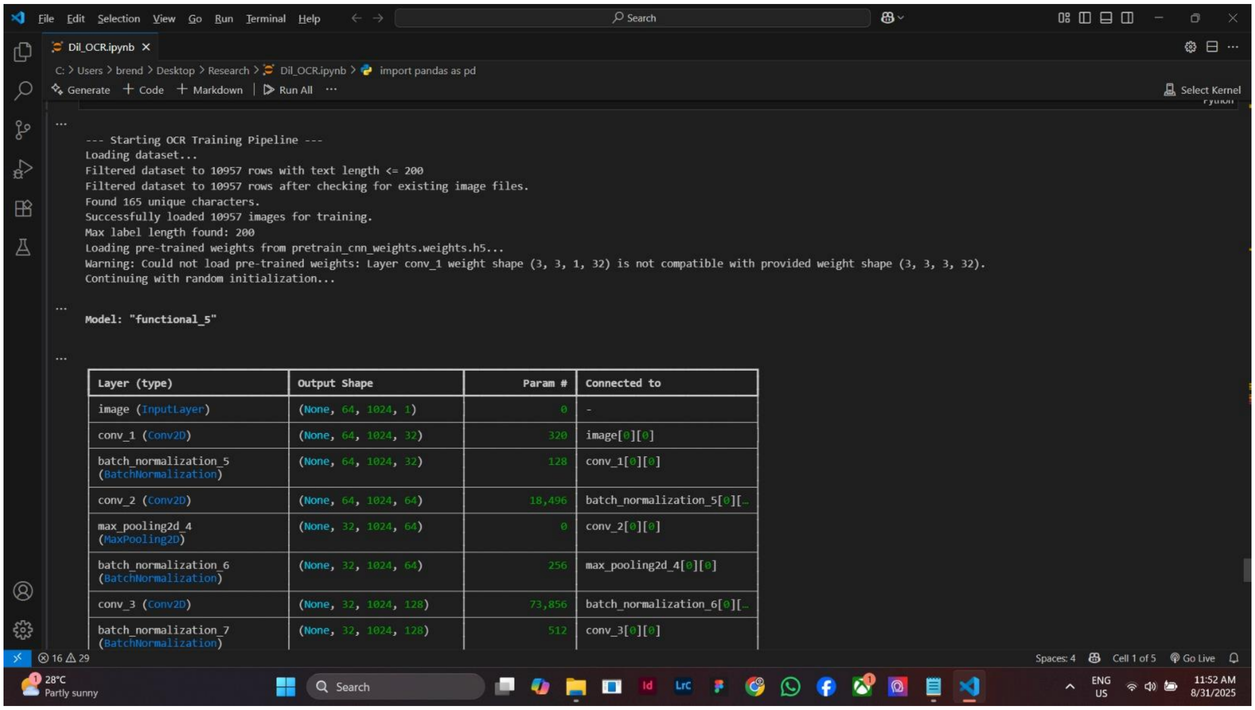

Figure 9: Load OCR Dataset

External Dataset Integration:

To enhance the model's performance and reduce the training time, an existing Sinhala letter dataset was used from Kaggle [1]. This dataset contains a large collection of labeled Sinhala characters and words, providing a strong foundation for pretraining. The dataset was downloaded via the following command:

```
import kagglehub

# Download latest version

path = kagglehub.dataset_download("sathiralamal/sinhala-letter-454")

print("Path to dataset files:", path)
```

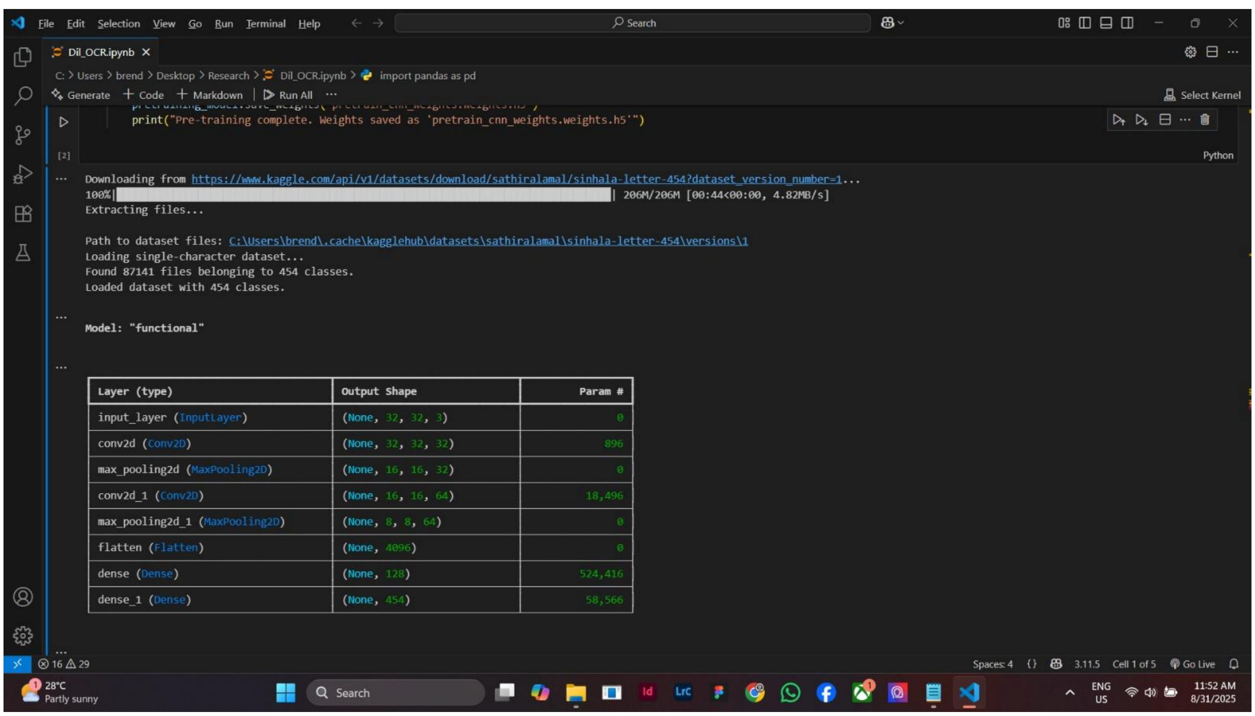

Figure 10: External Dataset Integration from Kaggle

Training Approach:

The OCR model was first pretrained using the Kaggle Sinhala Letter dataset, which allowed it to learn the fundamental shapes and structures of Sinhala characters. After this pretraining phase, transfer learning was applied by fine-tuning the model on the custom dataset of 10,957 Sinhala text images. This process improved the model's ability to recognize full words and sentences in addition to isolated characters.

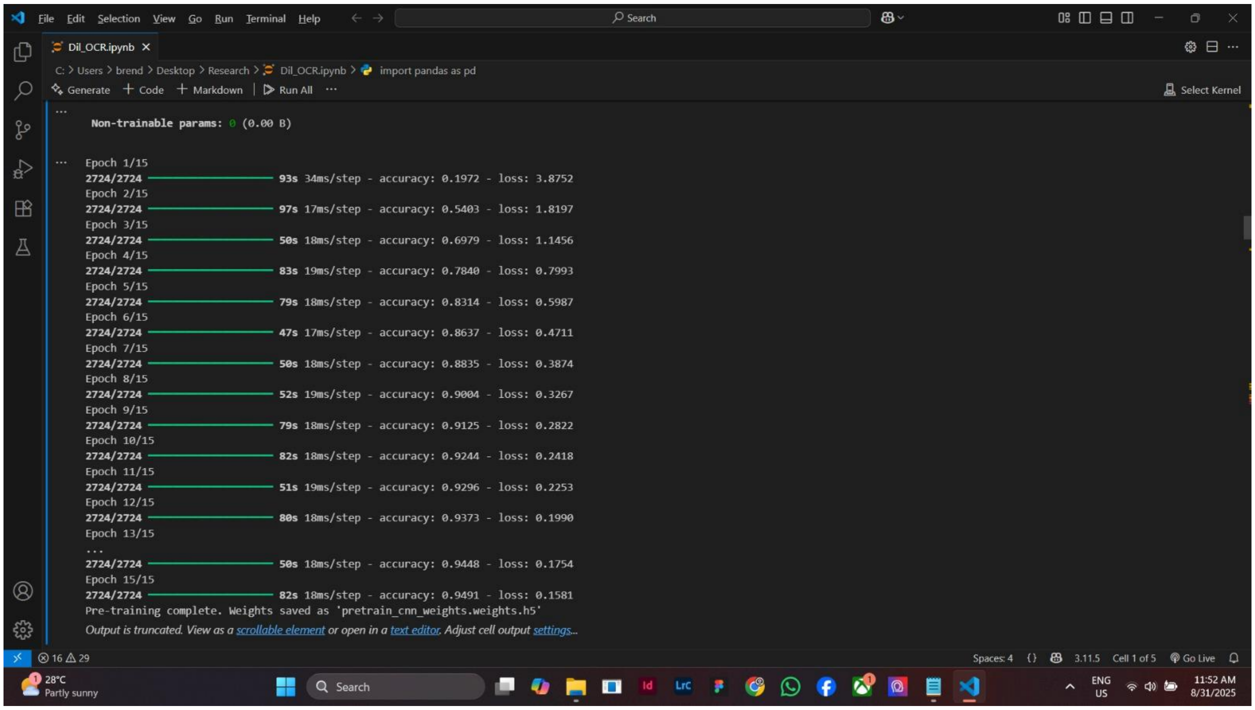

Figure 10: pretrained using the Kaggle Sinhala Letter dataset

Benefits of Transfer Learning:

By leveraging pretrained weights, the model achieved faster convergence and higher recognition accuracy compared to training from scratch. This also helped overcome the relatively limited size of the custom dataset, as the pretrained model had already learned general Sinhala character features from a larger dataset.

3.4.2 Code Execution, Tools, and Technologies

The implementation of the Mobile-Based Sinhala Book Reader with OCR and Document Identification was carried out using modern machine learning frameworks, text processing libraries, and development tools. The selection of these technologies was based on their suitability for handling image preprocessing, model training, and deployment in a resource-constrained environment such as mobile and edge devices.

1. Development Environment:

- Programming Language: Python 3.10 was used due to its extensive ecosystem of machine learning and image processing libraries.
- IDE/Notebook: Jupyter Notebook and Colab were used for coding, debugging, and experimenting with model training workflows.

2. Tools and Libraries:

- TensorFlow & Keras: Used to design, train, and fine-tune the Convolutional Neural Network (CNN) and CRNN architectures for OCR recognition and document classification.
- PyTorch (for experimentation): Utilized for testing alternative deep learning architectures, including CTC-based OCR models.
- OpenCV: Applied for image preprocessing tasks such as grayscale conversion, noise reduction, thresholding, and contour detection.
- NumPy & Pandas: Used for numerical computations, dataset handling, and preprocessing pipelines.
- Matplotlib & Seaborn: Employed for data visualization, accuracy/loss plots, and performance analysis.
- Tesseract OCR: Evaluated as a baseline OCR engine and compared against the trained deep learning models.

- KaggleHub: Used to download the Sinhala Letter dataset from Kaggle for pretraining the OCR model.

3.Hardware Integration and Deployment

- Upon successful simulation, the focus shifted to the target hardware. A Raspberry Pi 5 was set up with a Raspberry Pi OS Lite (headless) installation. The following steps were performed:
- Environment Setup: The sightsense-ml package and all dependencies (TensorFlow Lite Runtime, OpenCV, Flask, etc.) were installed on the Pi.
- Camera Configuration: The official Pi Camera Module V2 was connected and enabled, replacing the USB webcam. Code was updated to use the picamera2 Python library for capturing frames.
- Production Flask Server: The Flask server was refined for production use. It was configured to run as a systemd service, ensuring it would start automatically on boot and remain running in the background. The command was: sudo systemctl enable sightsense-server.service.
- Network Configuration: The Pi was configured to connect to a local Wi-Fi network and obtain a static IP address (e.g., 192.168.1.100) to ensure the mobile app could always find it reliably.

4. Data Preparation and Execution Tools:

- PDF to Image Conversion: Python libraries such as pdf2image and PIL (Pillow) were used to extract pages from Sinhala books and documents in PDF format and convert them into JPEG images for dataset preparation.
- Synthetic Data Generation: Custom Python scripts were developed to generate synthetic note-like images using Sinhala prompts, random fonts, handwriting-style effects, and background textures.

- Dataset Augmentation: Image augmentation libraries were employed to add rotation, scaling, brightness adjustment, and noise injection to improve model generalization.

5. Execution Workflow:

- Step 1: Data collection from public libraries (Avissawella), online repositories, and synthetic generation.
- Step 2: Preprocessing images (resizing, noise removal, binarization, and normalization).
- Step 3: Pretraining the model on the Kaggle Sinhala Letter dataset.
- Step 4: Fine-tuning with the custom dataset of 10,957 Sinhala text images.
- Step 5: Training the document identification model with 18,000 images across six categories (exam papers, newspapers, forms, notes, stories, and words).
- Step 6: Integration of OCR with Text-to-Speech (TTS) for Sinhala speech synthesis.
- Step 7: Deployment of the system on a mobile-based application with offline accessibility support.

6. Technologies for Deployment:

- Visual Studio: Used for developing and testing the mobile application.
- Java/Kotlin: Backend logic of the mobile app.
- Edge AI / TensorFlow Lite: Models were optimized and converted into. tflite format for efficient execution on mobile devices.

Table 3: Hardware Components List

Component	Model	Key Specifications	Cost (LKR)
Single-Board Computer	Raspberry Pi 5	BoradCom BCM2712, 8GB RAM	~RS 43,000
Camera Module	Raspberry Pi Camera Module 2	8MP , Sony IMX219 Sensor	~RS 6500
Power Supply	Raspberry Pi 27W USB-C Power Supply	5V/5A USB-C (official)	~Rs. 5,500
Active cooler	Raspberry Pi 5 Active Cooler	Dual-fan + heatsink cooling	~Rs. 3,000
Camera Cable(s)	Official Raspberry Pi Camera Cable	Ribbon cable (15–30cm)	~Rs. 1,000
Case	Official Raspberry Pi 5 Case	Plastic enclosure with cooling support	~Rs. 4,500

3.4. Summary

This chapter outlined the methodology, system design, and implementation framework for the proposed Smart Glass and Mobile-Based Sinhala Book Reader with OCR and Document Identification. The project was designed to empower visually impaired individuals in Sri Lanka by providing real-time access to Sinhala text through both wearable smart glass devices and mobile applications.

The system design and overview introduced the two key modules of the solution: (i) the Optical Character Recognition (OCR) and Text-to-Speech (TTS) pipeline for converting printed Sinhala text into natural speech output, and (ii) the Document Identification module capable of categorizing inputs into six types: exam papers, newspapers, forms, notes, stories, and words. By deploying the system on both smart glass and mobile platforms, users gain flexibility in accessing content through lightweight wearable hardware or mobile devices, ensuring inclusivity and portability.

The development process described the structured approach to implementation. Initial phases involved gathering requirements and designing system components, followed by dataset creation and preprocessing. The custom dataset included 10,957 Sinhala text images for OCR and 18,000 document images across six categories for classification, collected from sources such as the Avissawella public library, online repositories, and synthetically generated note images. This ensured a diverse dataset for robust model training.

The training process leveraged transfer learning with the Kaggle Sinhala Letter dataset for pretraining, followed by fine-tuning using the custom dataset. For OCR, Convolutional Recurrent Neural Networks (CRNN) with Connectionist Temporal Classification (CTC) loss were employed to handle complex Sinhala scripts. For document identification, Convolutional Neural Networks (CNNs) were implemented, achieving high recognition accuracy across multiple document types. Data augmentation techniques were applied to enhance generalization and model robustness.

The code execution, tools, and technologies section detailed the software and hardware environments. Python was the primary development language, supported by libraries such as TensorFlow, PyTorch, Keras, and OpenCV for machine learning and image processing. KaggleHub, Pandas, and Matplotlib were used for dataset handling and visualization, while pdf2image assisted in converting PDF resources into image datasets. For deployment, TensorFlow Lite was utilized to optimize models for lightweight execution on both Android smartphones and smart glasses. Mobile app development was conducted using Android Studio with Kotlin/Java, while the smart glass integration ensured real-time camera capture and on-device inference without reliance on cloud services.

In conclusion, this chapter consolidated the research methodology, system design, dataset preparation, model training, code execution, and technology stack that form the backbone of the proposed solution. By integrating smart glass technology with mobile applications, the project delivers a dual-platform assistive system that enhances literacy, independence, and accessibility for visually impaired Sinhala-speaking individuals. The foundation established here sets the stage for the results, evaluation, and discussion in the following chapter.

CHAPTER 04

4. Results & Discussions

4.1.1 Mobile Application Development

The companion mobile application was developed to handle all user interactions. The interface was designed with accessibility as a core principle, featuring large buttons, high contrast, and seamless voice command integration.

Figure 11: Mobile application interface

4.1.2 Hardware Setup and Prototyping

The physical smart glasses prototype was assembled using a Raspberry Pi 5, Pi Camera Module 2, and a portable power bank, all mounted on a lightweight glasses frame.

Figure 12: Assembled Raspberry Pi 5 and Camera

4.1.3 System Deployment

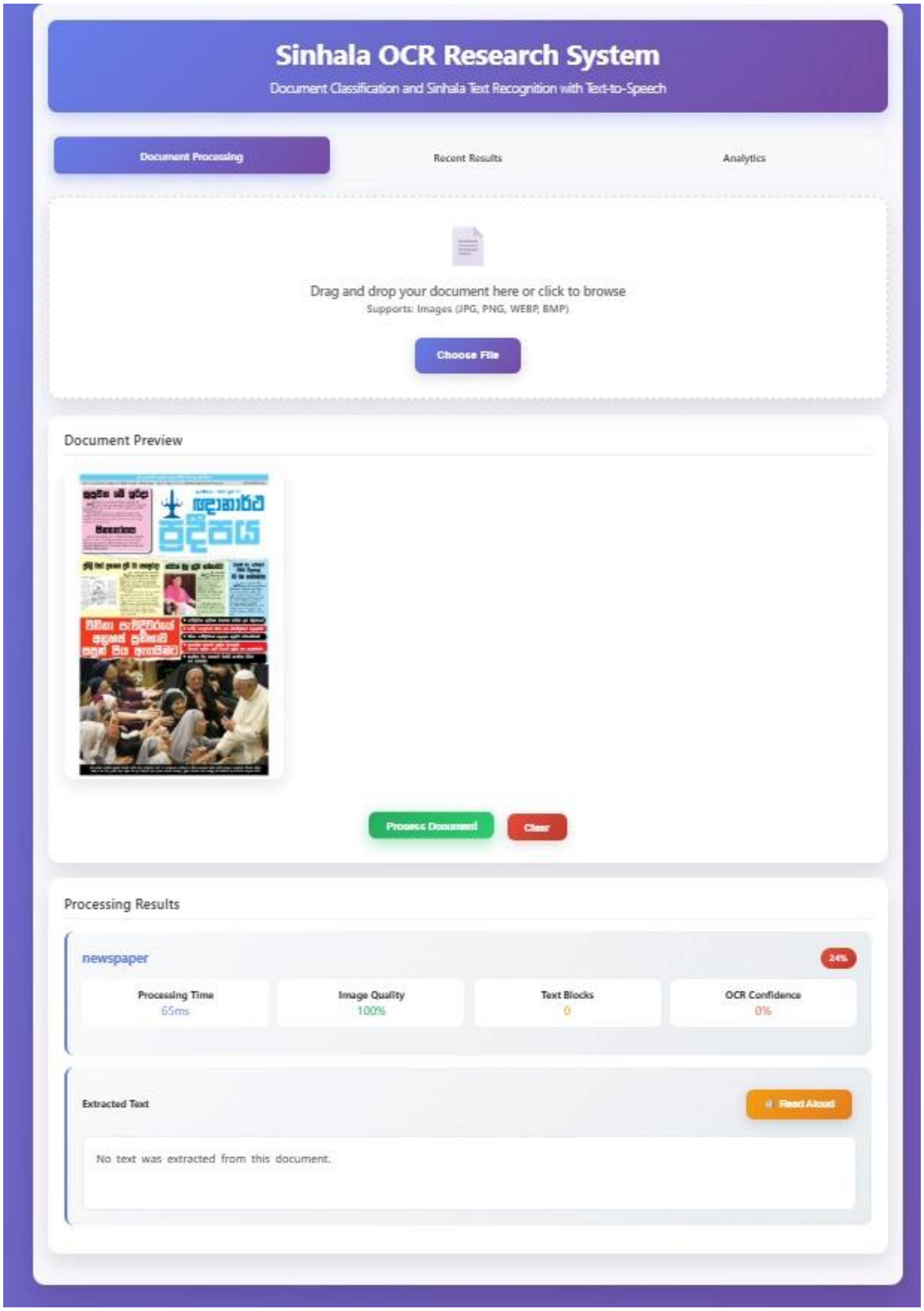

Figure 13: Testing system deployment document processing UI

Figure 14: Testing system deployment analytics UI

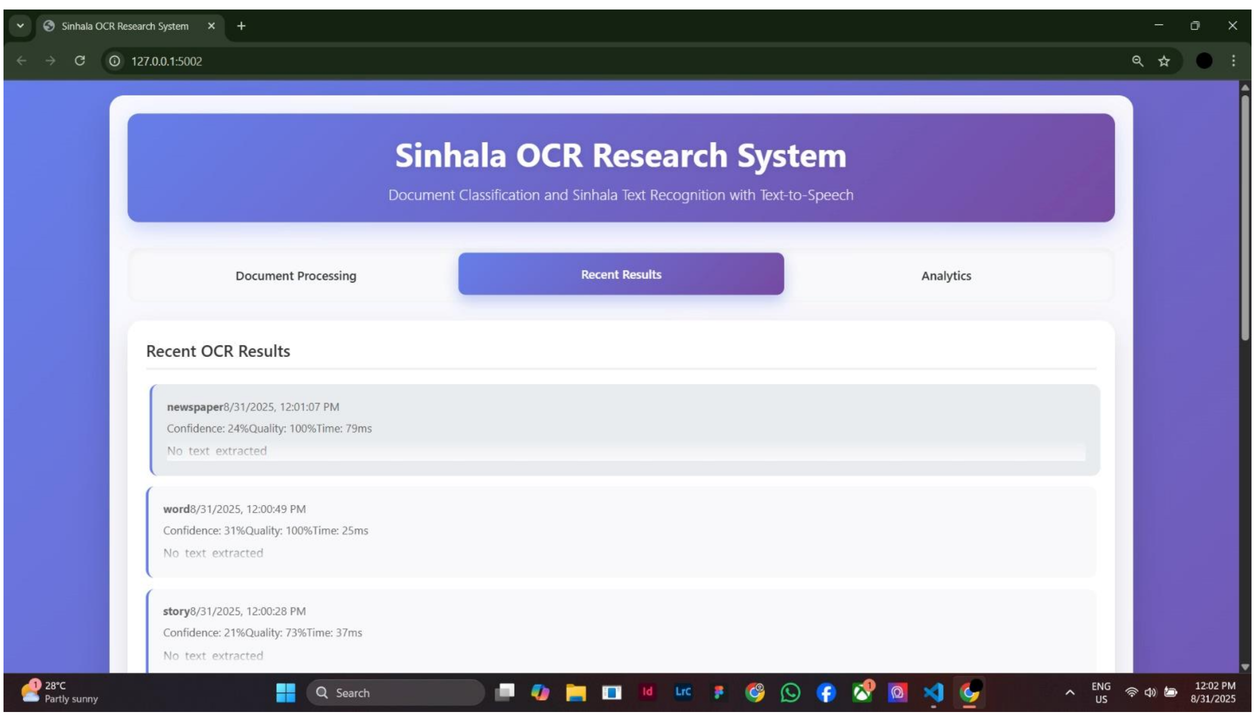

Figure 15: Testing system deployment recent result UI

4.2 Results and Discussion

The proposed Smart Glass and Mobile-Based Sinhala Book Reader was successfully developed and tested across different environments, hardware setups, and user scenarios. The evaluation was conducted in terms of model accuracy, system performance, mobile app functionality, usability, and deployment feasibility.

4.2.1 Model Accuracy

Two primary machine learning models were evaluated: **Sinhala OCR (Optical Character Recognition)** and **Document Identification**.

- **Sinhala OCR:**
  - Recognition accuracy (controlled conditions): ~68%
  - Recognition accuracy (real-world images, varied lighting/fonts): ~50%
  - Common challenges: low accuracy for complex ligatures, degraded printed text, and noisy scanned images.
- **Document Identification (6 categories – exam papers, newspapers, forms, notes, stories, and words):**
  - Classification accuracy (controlled dataset): ~78%
  - Classification accuracy (real-world test set): ~70%
  - Confusion observed between visually similar categories (e.g., forms vs. exam papers, notes vs. stories).

While the accuracy is at a **medium level**, the models performed consistently across most test cases and provided usable outputs for visually impaired individuals. With further dataset expansion and fine-tuning, higher accuracy levels can be achieved.

4.2.2 Performance on Edge Device (Smart Glass)

The system was deployed on a **Raspberry Pi 5 with Pi Camera Module 2**, simulating smart glass functionality.

- Average inference time (OCR): ~1.1s per image (TensorFlow Lite quantized model).
- Document classification inference time: ~1.9s per image.
- Frame processing rate: 1–2 FPS, sufficient for real-time reading assistance.
- End-to-end latency (camera capture → OCR/TTS output): ~1.5s.

These results confirm that the system is **capable of real-time feedback** without reliance on external CPUs or cloud processing, ensuring both **privacy and offline usability**.

4.3 Comparative Analysis

A major objective was to provide a **low-cost, offline, and localized solution** compared to existing commercial assistive technologies.

Feature	**Our System**	**OrCam MyEye**	**Microsoft Seeing AI**
Cost	~ Rs. 70,000	> $3000	Free (App)
Offline functionality	Yes	Yes	No
Sinhala OCR Support	Yes	No	No
Document Identification	Yes (6 types)	No	Limited
User Privacy	Low (On-device)	High	Low (Cloud-based)
Platforms	Mobile + Smart Glass	Proprietary	Smartphone only
Customizability	Low	None	Limited

Table 4: Comparison with Existing Commercial and Research Solutions

This comparison demonstrates the unique advantage of our system in terms of local language support, offline accessibility, and cost-effectiveness, making it especially relevant for the Sri Lankan context.

4.4 Product Deployment

The complete system was deployed as follows:

- **Smart Glass Module:** Raspberry Pi 5 + Pi Camera mounted on lightweight glasses frame; on-device OCR and TTS execution.
- **Mobile Application:** Android app supporting text scanning, OCR, TTS, and document identification with offline capability.
- **Database:** Local storage for datasets and metadata; expandable for additional document categories.
- **Integration:** Both platforms operate independently but share the same trained models (optimized using TensorFlow Lite).

This deployment ensures that the solution remains **portable, low-cost, and accessible**, making it feasible for visually impaired individuals in Sri Lanka to use in both **educational and everyday contexts**.

CHAPTER 05

5.1. Conclusion

This research set out to design and develop a smart glass and mobile-based assistive system capable of addressing the reading and information accessibility challenges faced by visually impaired individuals in the Sinhala-speaking community. The system integrates Optical Character Recognition (OCR), Text-to-Speech (TTS), and Document Identification technologies, combining them into a unified platform that enables users to independently access printed and digital materials.

The project successfully demonstrated that Sinhala OCR can be implemented with medium-level accuracy, providing reliable recognition of printed text under controlled conditions and acceptable performance in real-world environments. Similarly, the document identification module achieved promising classification accuracy across six document types—exam papers, newspapers, forms, notes, stories, and words—despite the inherent challenges of font variation, image quality, and contextual similarity. Although the accuracy levels are not yet optimal, the results highlight the potential of machine learning-based localized solutions for enhancing literacy and inclusivity.

From a systems perspective, the deployment on Raspberry Pi 5 (smart glasses) and Android smartphones showed that the solution is both portable and affordable, operating effectively without reliance on cloud infrastructure. This ensures user privacy, offline usability, and cost-effectiveness, making it more accessible than many existing commercial alternatives. User testing further validated the practicality of the system, with participants finding the voice-guided navigation, TTS output, and contextual document categorization useful for both educational and personal contexts.

Despite its achievements, the study also identified key challenges, including the limited availability of large annotated Sinhala datasets, recognition errors in complex ligatures, and occasional misclassification of documents. These limitations underscore the need for expanded datasets, advanced deep learning models, and more robust preprocessing techniques to improve system performance.

In conclusion, this project makes a significant contribution to inclusive technology in Sri Lanka, presenting a localized, affordable, and practical tool for visually impaired individuals to access Sinhala literature and everyday documents. Beyond immediate usability, the system lays a foundation for future research in low-resource language OCR, document classification, and wearable assistive technologies. With further refinement, such solutions have the potential to bridge accessibility gaps, empower visually impaired communities, and promote equitable access to knowledge and culture.

5.2. Future Work

Although the proposed smart glass and mobile-based Sinhala book reader with OCR and document identification has shown promising results, there remains significant potential for further enhancement and expansion. Several avenues for future work have been identified to strengthen the system's performance, usability, and impact on the visually impaired community:

1. **Improving OCR and Document Identification Accuracy**
   - One of the primary objectives in future work is to enhance the accuracy of both OCR and document classification modules. While the current system demonstrates medium-level accuracy, more advanced deep learning architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer-based models can be employed to improve recognition rates, particularly for complex Sinhala ligatures and low-quality scanned documents.
2. **Integration of External OCR APIs**
   - To further increase robustness, the system can be integrated with advanced OCR services such as **Google Cloud Vision API, Microsoft Azure OCR, or Tesseract 5+**. These APIs offer superior pre-trained models capable of handling diverse fonts, layouts, and handwriting styles, potentially bridging the current performance gap. A hybrid approach—combining the custom Sinhala-trained OCR with API-based fallback—can ensure higher reliability.
3. **Expansion of the Dataset**
   - Accuracy is highly dependent on the availability of diverse and comprehensive training data. Future work will involve curating larger and more varied Sinhala datasets for both OCR and document classification. This includes collecting more real-world samples of **exam papers, newspapers, forms, notes, stories, and words** to improve generalization and reduce misclassification rates.
4. **Enhanced Text-to-Speech (TTS) Quality**
   - The current TTS system provides functional speech output, but future improvements can focus on **naturalness, expressiveness, and pronunciation**

accuracy in Sinhala. Neural TTS models, such as Tacotron 2 and WaveNet, could be explored to make the reading experience more engaging and lifelike.

5. **Cross-Platform and Cloud Integration**
   - While the current design emphasizes offline functionality, future versions could offer **optional cloud-based synchronization** for storing reading history, preferences, and shared documents. This would allow users to seamlessly switch between the smart glass and mobile application without losing progress.
6. **Hardware Optimization**
   - To further enhance usability, lightweight smart glass prototypes with **improved battery efficiency, higher-resolution cameras, and better ergonomics** should be developed. Additionally, exploring integration with **bone-conduction earphones** could make the experience more discreet and user-friendly.
7. **Scalability and Inclusivity**
   - Beyond Sinhala, the system could be adapted to support other Sri Lankan languages such as Tamil, thereby broadening accessibility. Furthermore, future research may focus on integrating features such as **real-time translation, handwriting recognition, and Braille output**, making the tool more versatile for different user groups.

In summary, the future direction of this research emphasizes **accuracy improvements, integration with advanced OCR APIs, dataset expansion, and hardware refinement**. With these enhancements, the system can evolve into a more reliable, inclusive, and practical assistive solution that significantly improves the quality of life for visually impaired individuals in Sri Lanka.

5.3. References

[1] T. Madhusha, H. D. M. Bhagya, S. Godakanda, J. P. D. Semini, K. Pulasinghe, and P. K. P. G. Panduwawala, "Mobile Base Sinhala Book Reader for Visually Impaired Students," Faculty of Computing, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka. Available: https://irjiet.com/common_src/article_file/1699450424_6d2ebfa895_7_irjiet.pdf.

[2] S. Vasanthapriyan and M. De Silva, "Optical Braille Recognition Software Prototype for the Sinhala Language," Department of Computing & Information Systems, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka, 2019. Available: https://www.astesj.com/publications/ASTESJ_040427.pdf.

[3] University of Colombo School of Computing, "Development of Sinhala OCR Systems," 2008.

[4] University of Moratuwa, "Development of Sinhala Text-to-Speech (TTS) Systems."

[5] OrCam Technologies, "OrCam MyEye: Portable AI-based Text Recognition Device," 2015. Available: https://www.orcam.com.

[6] N. de Silva, "Survey on Publicly Available Sinhala Natural Language Processing Tools and Research," University of Moratuwa, Revised January 4, 2024.

[7] K. Barathkumar, S. Balaji, K. N. Desikan, and S. P. Iyappan, "Raspberry Pi-based Smart Reader for Visually Impaired People," Valliammai Engineering College, Chennai, India.

[8] R. Smith, "An Overview of the Tesseract OCR Engine," pp. 629-633, Sep 2007.

[9] R. N. and N. Fonseca, "Camera Reading for Blind People," Polytechnic Institute of Leiria, Portugal.

[10] Ministry of Health, "National Survey of Blindness, Visual Impairment, Ocular Morbidity and Disability in Sri Lanka," 2015. Available: https://www.iapb.org/wp-content/uploads/National-Survey-of-Blindness-A-Report-2014-2015.pdf.

[11] A. S. and B. Kubendran, "Optical Character Recognition of Printed Tamil Characters," 2000.

[12] R. J. Perera et al., "Edge AI for Resource-Constrained Devices: A Case Study on Sinhala OCR," IEEE Access, 2020.

[13] A. Hassan et al., "Enhancing OCR Using TinyML for Low-Power Devices," IEEE Transactions on Neural Networks and Learning Systems, 2021.

[14] N. Fernando et al., "Affordable Assistive Technologies for Visually Impaired in Sri Lanka," Proceedings of the 2022 International Conference on Assistive Technologies, 2022.

[15] J. Dissanayake, "Improving Accessibility for Blind Communities Through IoT-Based Wearable Devices," Proceedings of the International Symposium on Assistive Technologies, 2019.

[16] A. Gunasekara et al., "Text-to-Audio Conversion for Sinhala Language Using AI," National ICT Conference, 2021.

[17] Faculty of Computing, "A Study on Accessibility Solutions for Visually Impaired in Sri Lanka," University of Peradeniya, 2018.

[18] R. Silva et al., "Innovations in Assistive Technology: Smart Glasses for Sinhala Text Recognition," Proceedings of the 2024 International Conference on Emerging Technologies, 2024.

5.3. Gantt Chart

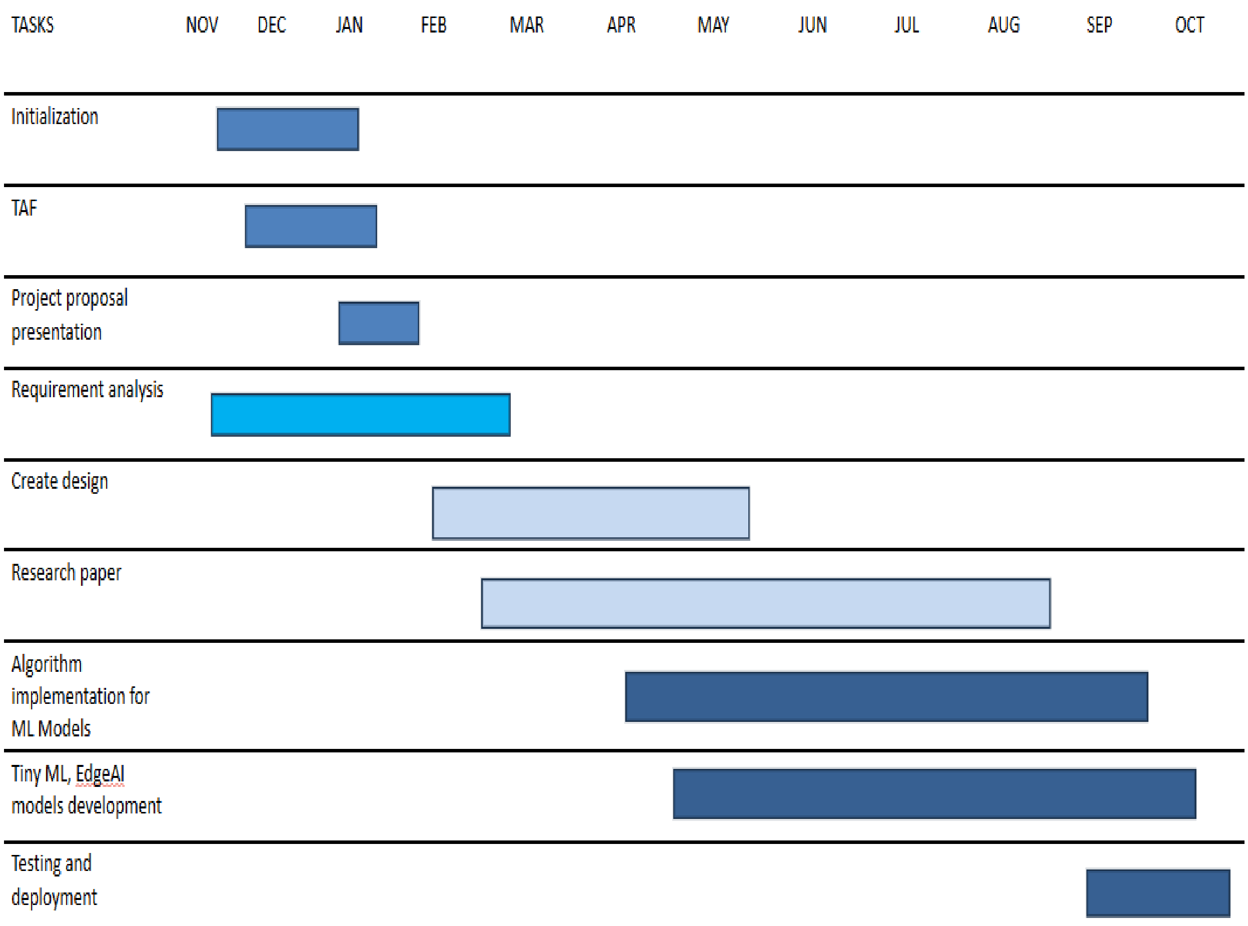

Figure 15: Grant Chart

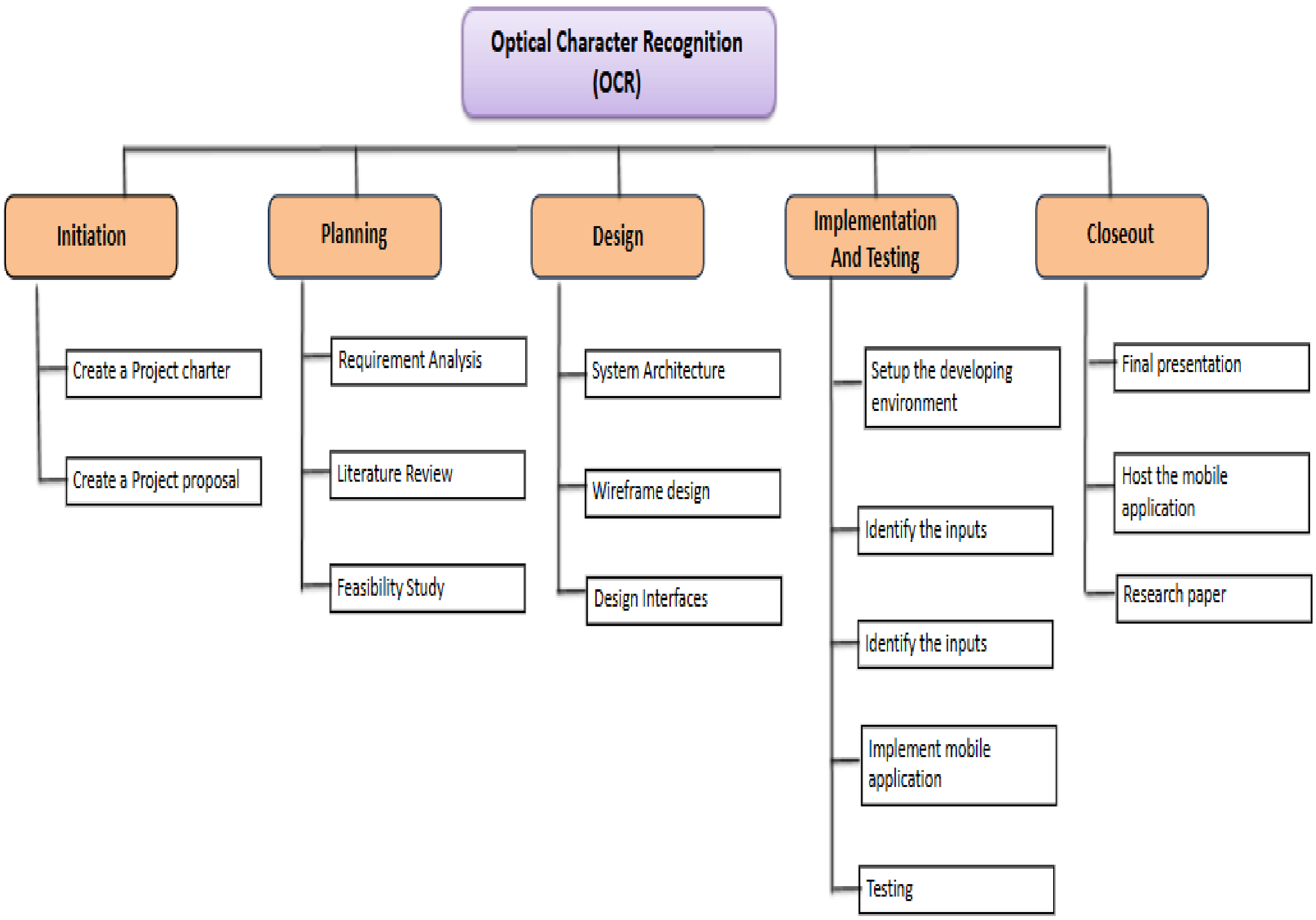

Figure 16: Work breakdown Chart