International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 6, June 2026)

Swardhwani: Voice-Driven Multilingual Book Reader with Real Time Page Handling.

Prof. V. S. Phatangare[1], Ambre Sakshi Vijay[2], Kolhal Shivam Kailas[3], Mali Omkar Dattatray[4]

[1,2,3,4]*Department of Electronics & Computer Engineering, Amrutvahini College of Engineering, Sangamner*

Abstract— The increasing adoption of digital learning platforms has created a demand for intelligent and accessible reading solutions that cater to diverse user needs. Traditional e-book readers often require manual interaction for navigation and reading, which can be challenging for visually impaired individuals, elderly users, and people who prefer hands-free access to digital content. To address these limitations, this project presents "Swardhwani: Voice-Driven Multilingual Book Reader with Real-Time Page Handling," an innovative system that enables users to interact with digital books through voice commands while providing seamless multilingual reading support.

The proposed system integrates Speech Recognition, Text-to-Speech (TTS) technology, Natural Language Processing (NLP), and real-time page management to create an interactive reading environment. Users can upload digital books in formats such as PDF and EPUB, select their preferred language, and listen to the content through natural-sounding synthesized speech.

Keywords:

Voice-Driven Book Reader, Multilingual Reading System, Speech Recognition, Text-to-Speech (TTS), Natural Language Processing (NLP), Real-Time Page Handling, Voice Commands, Digital Accessibility, E-Book Reader, Language Translation, Assistive Technology, Artificial Intelligence.

I. INTRODUCTION

In today's digital age, technology enhances accessibility and human-machine interaction. Yet, reading physical books remains a challenge for the visually impaired, elderly, and physically challenged [2][8]. While e-books and audiobooks exist, many texts are still only in print [4]. The "Swardhwani" project offers a voice-controlled, multilingual book reader with real-time page turning, making physical books more accessible and interactive [1][3][6]. The word Swardhwani is derived from two Sanskrit terms Swar (sound or voice) and Dhwani (tone or resonance) which together represent the essence of the project: converting the written word into meaningful sound. This project aims to automate the process of reading physical books using a combination of computer vision, natural language processing, and electromechanical control systems [1][5]. It eliminates the need for manual page turning and provides a multilingual audio output, thereby offering an inclusive reading experience for everyone [3][8]. At the heart of the system lies the Microcontroller ESP32, a powerful single-board computer capable of executing advanced algorithms for Optical Character Recognition (OCR), Text-to-Speech (TTS), and speech command recognition [1][9]. A camera module captures high-resolution images of each page, which are then processed through OCR software such as Tesseract OCR to convert the image-based text into digital text [3]. This digital text is further processed through a TTS engine like Google Text-to-Speech (GTTS) or pyttsx3 to produce natural-sounding speech output in multiple languages. Thus, the system converts physical printed text into audio that can be easily understood by the user [3].

To ensure hands-free operation, the project employs voice command recognition, allowing users to control the reading flow with simple verbal instructions such as "next page," 2 "pause," "repeat," or "stop." The system continuously listens to the user's voice through a microphone and interprets the command in real time using speech recognition algorithms. This enables uninterrupted and intuitive control Without physical contact, enhancing comfort and accessibility [2][7][8].

II. PROBLEM STATEMENT

Reading physical books remains a significant challenge for visually impaired individuals, elderly people, and users with physical disabilities who have difficulty handling books and turning pages manually. While audiobooks and e-books are

available, a vast amount of educational, academic, and literary content still exists only in printed form. Existing reading solutions often lack multilingual support, real-time voice interaction, and automated page-turning capabilities, limiting their effectiveness and accessibility.

III . OBJECTIVES

- To convert printed book content into speech using OCR and Text-to-Speech technology.
- To enable hands-free book reading through voice command recognition.
- To implement an automatic page-turning mechanism using a servo motor and exhaust fan.
- To provide multilingual reading support for users from different linguistic backgrounds.
- To improve accessibility and independence for visually impaired and physically challenged users.

IV. LITERATURE SURVEY

1. Title: Book Reader Using Raspberry Pi for Visually Impaired (2018) Authors: S. Aditi1, SP. Annapoorani, A. Kanchana Findings: This paper aims at making an open-source audio book software to build a book reader with raspberry pi controls. Here in this paper, we present the Pi Book reader which can read a real book aloud and also turn the pages of the book. The overall process of the project involves Image to Text conversion and then Text to Speech conversion. The image to text conversion is carried out with the help of OCR [Optical Character Recognition]. The Optical Character Recognition Technology can be used to convert various kinds of documents like images, scanned documents and PDF files. The OCR algorithm involves various stages like Scanning, Preprocessing, Feature Extraction, Classification and Recognition

2. Title: AI-Powered Multilingual Voice Assistant for Real-Time Tasks and Content Creation (2025) Authors: Manan Aware, Gaurav Sontakke, Ujwal Tambade, Prof. Shraddha Utane Findings: This project introduces an AI-based real-time chatbot assistant that operates via voice commands, enabling users to perform a variety of tasks in multiple languages. The assistant can open and close applications, conduct real-time searches on platforms like Google and YouTube, and gather up-to-date information. Additionally, it is equipped with natural language processing capabilities to generate content such as stories, songs, blogs, and more. By supporting various languages, this assistant provides a versatile, intuitive, and hands-free experience, making it suitable for a global audience. The integration of real-time execution with multilingual support aims to enhance user productivity and simplify digital interactions across different linguistic backgrounds.

3. Title: Smart Reader for Blind with1A4uto Page Turn (2022) Authors: Mahendra H N, Pavankumar S Gowda, Pratheek M Poojary, Preetham Nirvan H S Findings: In an abstract either forward or backward mode, the computerized page-flipper flips the page of books, periodicals, or three-ring notebooks that have been bound. This electromechanical technology allows someone who is limited to using the outside world to communicate through a head switch installed on a wheelchair for independent 19 reading of any text. This presence OCR performs the automatic turning of numerous pages of a book according to the user's wishes, which is controlled through voice commands, and also seeks to develop a book reader module, which includes text extraction and segmentation, as well as speech conversion.

4. Title: AI-Powered Library Management System (2025) Authors: Pratik Palve, Prof. Gaurav Arora Findings: The AI Library Management System is an innovative solution designed to automate and optimize traditional library operations using artificial intelligence. This system leverages AI-driven algorithms to enhance book catalogue, user management, and resource tracking while ensuring a seamless experience for both librarians and patrons. Key features include an intelligent search engine that enables users to find books based on keywords, authors, or subject relevance. AI- based recommendation systems suggest books based on reading history and preferences, improving user engagement. Additionally, automated book issuance and return processes.

.

IV. WORKING OF SYSTEM

The block diagram of the Swardhwani: Voice-Driven Multilingual Book Reader with RealTime Page Turning project illustrates the integration of multiple hardware and software modules working together to automate the process of reading a physical book. It acts as the brain of the system, coordinating the operations of the camera module, OCR processing, text-to-speech conversion, voice recognition, and the mechanical page-turning subsystem. Each block in the diagram represents a critical function that contributes to the system's intelligent automation and real-time interactivity.

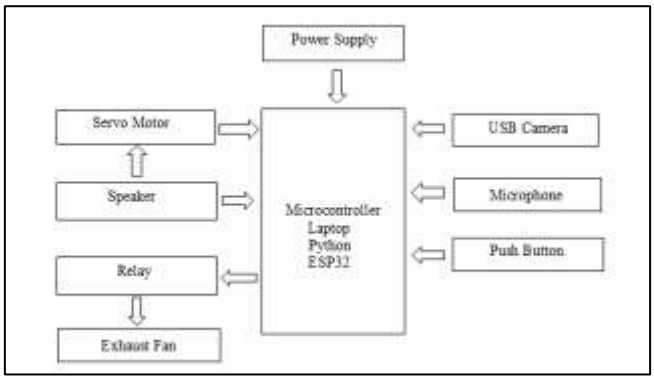

Fig 1: Block Diagram

A. Input Sensing Section

The input sensing section consists of the USB Camera, Microphone, and Push Button, which are responsible for collecting information and user commands. The USB camera captures high-resolution images of book pages for text extraction using OCR. The microphone continuously listens for voice commands such as *"next page," "pause," "repeat,"* and *"stop."* The push button is used to initiate or control the reading process. These input devices provide real-time data to the ESP32 microcontroller for further processing.

B. Processing and Control Section

The ESP32 Microcontroller running Python-based control algorithms acts as the central processing unit of the system. It receives image data from the camera and voice inputs from the microphone. The controller performs image processing, OCR text extraction, speech recognition, and decision-making operations. Based on the received commands, it controls the speaker, servo motor, relay, and exhaust fan to execute the required actions. This section serves as the brain of the entire system.

C. Image Processing and Text Recognition Section

The USB camera captures images of the current book page and sends them to the ESP32. The captured images undergo preprocessing techniques such as grayscale conversion, noise reduction, and thresholding. The processed image is then passed through the OCR engine, which extracts printed text and converts it into machine-readable digital text. This extracted text forms the basis for audio generation and multilingual reading.

D. Audio Output and User Interaction Section

The speaker is connected to the ESP32 and provides audio output to the user. After text extraction, the Text-to-Speech (TTS) engine converts the digital text into natural-sounding speech. The generated voice is played through the speaker, allowing users to listen to the book content. This section enables visually impaired and physically challenged users to access printed books through audio interaction.

E. Mechanical Page Turning Section

The mechanical page-turning section consists of the Servo Motor, Relay, and Exhaust Fan. When the user gives a voice command such as *"next page,"* the ESP32 activates the relay, which powers the exhaust fan. The exhaust fan generates suction to gently lift the page, while the servo motor performs a controlled movement to flip the page. After the page is turned successfully, the servo motor returns to its initial position, preparing the system for the next reading cycle.

F. Power Supply Section

The power supply unit provides regulated electrical power to all system components, including the ESP32 microcontroller, USB camera, microphone, speaker, relay, servo motor, and exhaust fan. A stable power source ensures reliable operation of image processing, voice recognition, audio generation, and mechanical page-turning functions. Proper voltage regulation protects the hardware and improves the overall efficiency and performance of the system.

V. SYSTEM DESIGN

A. Microcontroller ESP32

The ESP32 Microcontroller serves as the central control unit of the Swardhwani system and coordinates the operation of all hardware and software modules. It receives input from the USB camera, microphone, and push button, processes the collected data, and controls the output devices such as the speaker, servo motor, relay, and exhaust fan. The ESP32 executes the control logic required for image acquisition, voice command processing, OCR integration, and automated page-turning operations. Its built-in Wi-Fi and Bluetooth capabilities provide flexibility for future IoT-based enhancements and remote monitoring features.

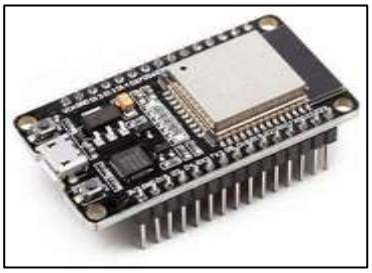

Fig 2: Microcontroller ESP32

B. Web Camera Module

The Web Camera Module is an essential input device in the Swardhwani system, responsible for capturing clear images of book pages for text extraction. It is positioned above the book and continuously acquires high-resolution images of the current page. These images are sent to the ESP32-based processing unit, where image preprocessing techniques such as grayscale conversion, thresholding, and noise reduction are applied to improve text visibility. The processed images are then passed to the OCR engine, which converts the printed content into machine-readable text.

Fig 3: Web Camera Module

C. Servo Motors

The Servo Motor is a key component of the Swardhwani system and is responsible for the automatic page-turning operation. It provides precise angular movement required to flip book pages accurately and smoothly. When the user issues a voice command such as **"Next Page,"** the ESP32 microcontroller sends control signals to the servo motor. Simultaneously, the exhaust fan creates suction to lift the edge of the page, and the servo motor rotates through a predefined angle to turn the page without causing damage..

Fig 4: Servo Motors

D. Motor Driver Module

The Motor Driver Module (L298N) acts as an interface between the ESP32 microcontroller and the mechanical components of the Swardhwani system. Since the ESP32 cannot directly provide the current required to drive motors, the L298N motor driver is used to control and regulate the power supplied to the servo motor and other motorized components. It receives control signals from the ESP32 and converts them into suitable voltage and current levels for motor operation. The module supports bidirectional motor control, speed regulation, and safe switching of electrical loads.

Fig 5: Motor Driver Module

E. Exhaust Fan

The Exhaust Fan is an important component of the Swardhwani system and plays a vital role in the automatic page-turning mechanism. Its primary function is to generate a controlled suction force that gently lifts the edge of a book page before it is turned by the servo motor. When a user gives a voice command such as "Next Page," the ESP32 microcontroller activates the exhaust fan through the relay circuit. The fan draws air from beneath the page, creating a vacuum effect that separates a single page from the stack without causing damage.

Fig 6: Exhaust Fan

F. Open CV

(Open Source Computer Vision Library) is a powerful open-source computer vision and image processing library used in the Swardhwani system for capturing, processing, and analyzing images of book pages. It plays a crucial role in improving the quality of images before they are passed to the OCR engine for text extraction. OpenCV performs various image preprocessing operations such as grayscale conversion, noise reduction, thresholding, edge detection, image enhancement, and skew correction

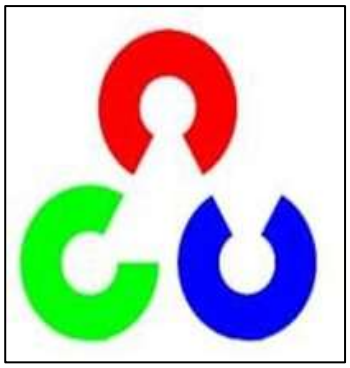

Fig 7: Open CV

G: Python

Python is a high-level, interpreted, and versatile programming language used as the primary software platform in the Swardhwani system. It is responsible for implementing the core functionalities of the project, including image processing, Optical Character Recognition (OCR), Text-to-Speech (TTS) conversion, voice command recognition, and hardware control. Python provides a rich collection of libraries such as OpenCV for image processing, Tesseract OCR for text extraction, gTTS for speech generation, SpeechRecognition for voice command processing, and GPIO libraries for interfacing with hardware components

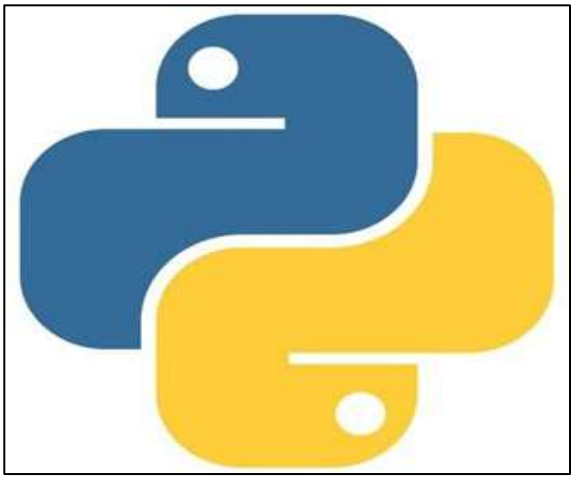

Fig 8: Python

VI. RESULTS

The Swardhwani: Voice-Driven Multilingual Book Reader with Real-Time Page Handling system was successfully designed and implemented to provide an automated and hands-free reading experience. The system effectively integrated a Web Camera, ESP32 Microcontroller, OCR Engine, Text-to-Speech Engine, Microphone Module, Servo Motor, Exhaust Fan, and Speaker to perform real-time book reading and page navigation. The camera successfully captured book pages, and the Tesseract OCR engine accurately extracted text from printed documents. The extracted text was converted into clear and understandable speech using the GTTS Text-to-Speech engine.

The voice recognition module successfully detected user commands such as “Next Page,” “Pause,” “Repeat,” and “Stop” and responded appropriately. The automatic page-turning mechanism, consisting of the exhaust fan and servo motor, performed smooth and reliable page flipping with minimal user intervention. The system demonstrated satisfactory performance under normal indoor lighting conditions and maintained real-time synchronization between text extraction, speech generation, and page navigation.

Parameter	Obtained Result
OCR Accuracy	85–95%
Voice Recognition Accuracy	80–90%
Average OCR Processing Time	2–4 Seconds per Page
Audio Generation Delay	Approximately 1 Second
Voice Command Response Time	1–2 Seconds
Page Turning Response Time	4–6 Seconds
Supported Languages	English, Hindi, Marathi and Others
System Operation	Real-Time Voice Driven

The practical evaluation showed that the system achieved high OCR accuracy in well-lit environments, reliable speech output, and smooth page-turning performance for standard printed books. The project successfully met its objectives of accessibility, multilingual support, automation, and hands-free operation, making it a valuable assistive technology for visually impaired users, elderly individuals, students, and book readers

VIII. FUTURE SCOPE

The Swardhwani: Voice-Driven Multilingual Book Reader with Real-Time Page Handling system has significant potential for future enhancements. The current implementation can be extended by integrating Artificial Intelligence (AI) and Machine Learning (ML) techniques to improve OCR accuracy, voice recognition performance, and personalized reading experiences. Advanced deep learning-based OCR models can be incorporated to recognize handwritten text, damaged documents, and low-quality printed materials with higher accuracy.

The system can be enhanced with AI-based language translation, allowing users to listen to books in their preferred language regardless of the original language of the text. This feature would make educational and literary content accessible to a wider audience across different regions and linguistic backgrounds..

CONCLUSION

The Swardhwani: Voice-Driven Multilingual Book Reader with Real-Time Page Handling system successfully demonstrates an intelligent and automated solution for hands-free reading of physical books. By integrating Optical Character Recognition (OCR), Text-to-Speech (TTS), voice recognition, and an automatic page-turning mechanism, the system provides a seamless reading experience for users. The combination of a web camera, microphone, ESP32 microcontroller, servo motor, and exhaust fan enables real-time text extraction, speech generation, voice-controlled navigation, and automated page handling.

The system is particularly beneficial for visually impaired individuals, elderly users, students, and people with physical disabilities, as it reduces dependence on manual book handling and promotes independent access to printed information.

REFERENCES

1) S. Kumar and A. Sharma, Automated Book Reading System Using Raspberry Pi, International Journal of Computer Applications, Vol. 182, No. 35, 2023.
2) R. Singh and P. Gupta, Voice Controlled Book Reader for Visually Impaired, International Journal of

Innovative Research in Science, Engineering and Technology (IJIRSET), Vol. 12, Issue 4, 2023.
3) A. Patel and M. Joshi, Multilingual Text-to-Speech System for Accessibility Applications, IEEE Access, 2022.
4) J. Lee and H. Kim, Optical Character Recognition Techniques for Printed and Handwritten Text, International Journal of Computer Vision, 2022.
5) P. Sharma and R. Verma, Raspberry Pi Based Assistive Reading Devices, International Journal of Advanced Research in Computer Science, 2023.
6) K. Mehta and S. Joshi, Mechanical Page Turning System Using Servo Motors and Fans, International Journal of Mechanical Engineering and Robotics Research, 2022.
7) L. Wang and Y. Zhao, Real-Time Voice Recognition for Human-Computer Interaction, IEEE Transactions on Human-Machine Systems, 2023.
8) S. Reddy and P. Kaur, Hands-Free Reading Assistance Systems for Visually Impaired, International Journal of Assistive Technologies, 2023.
9) H. Patel and A. Joshi, Integration of OCR and TTS for Automated Document Reading, Journal of Embedded Systems and Applications, 2022.
10) R. Kumar and S. Meena, IoT-Based Smart Reading System Using Raspberry Pi, International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2023.
11) S. Aditi, S.P. Annapoorani and A. Kanchana, Book Reader Using Raspberry Pi for Visually Impaired, International Journal of Engineering Research and Technology, 2018.
12) M. H. N., P. S. Gowda, P. M. Poojary and P. N. H. S., Smart Reader for Blind with Auto Page Turn, International Conference on Smart Technologies, 2022.
13) B. H. N., K. S. Nayak, K. Sowmya and S. T., Voice Controlled Smart Page Turner for Differently Abled, International Journal of Emerging Technologies, 2018.
14) Y. O. Sharrab, H. Attar, Y. Al-Omary and W. E., Advancements in Speech Recognition: A Systematic Review of Deep Learning Transformer Models, Trends, Innovations, and Future Directions, IEEE Access, 2025.
15) R. Smith, An Overview of the Tesseract OCR Engine, Proceedings of the International Conference on Document Analysis and Recognition (ICDAR), 2007.
16) A. Rosebrock, Practical Python and OpenCV: Case Studies for Image Processing and Computer Vision, PyImageSearch Publications, 2023.
17) Google Developers, Google Text-to-Speech API Documentation, Google Cloud Platform, 2024.
18) OpenCV Foundation, OpenCV Documentation and Computer Vision Library Reference Manual, 2024.
19) Python Software Foundation, Python Programming Language Documentation, Version 3.x, 2024.
20) Espressif Systems, ESP32 Series Microcontroller Technical Reference Manual, Espressif Systems, 2024.