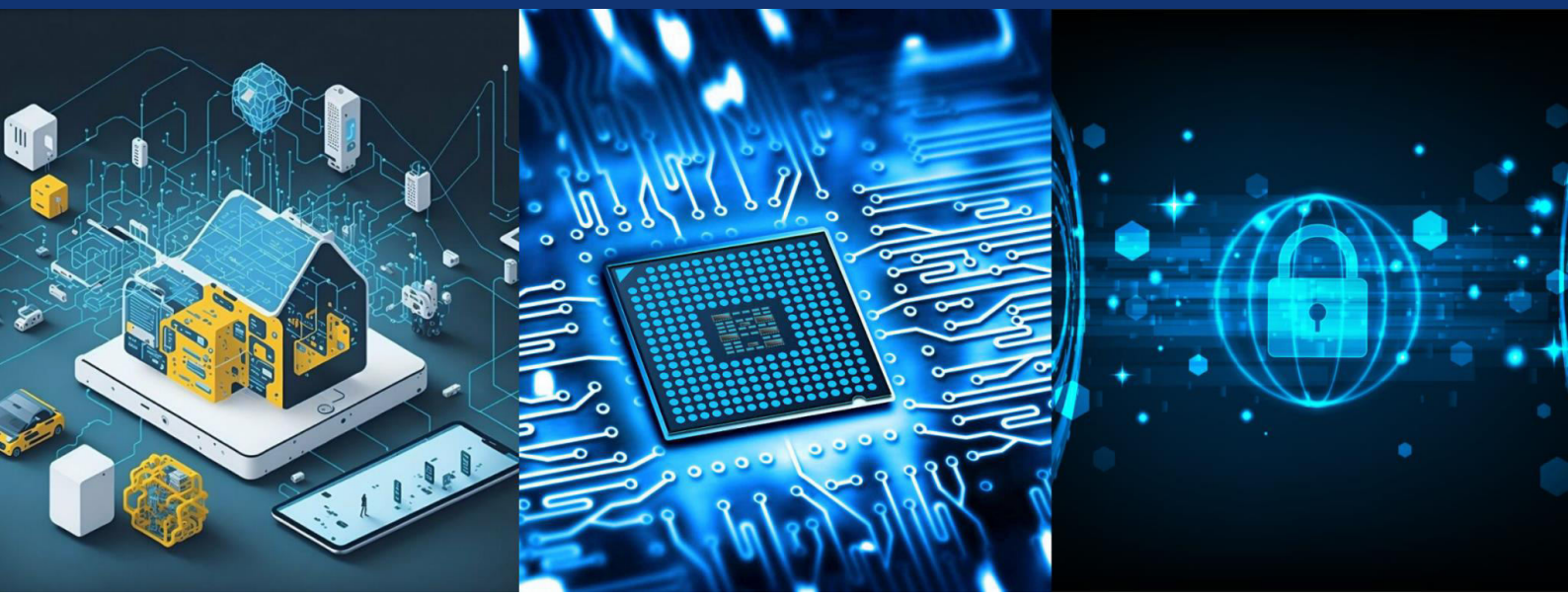
IJIRCCE

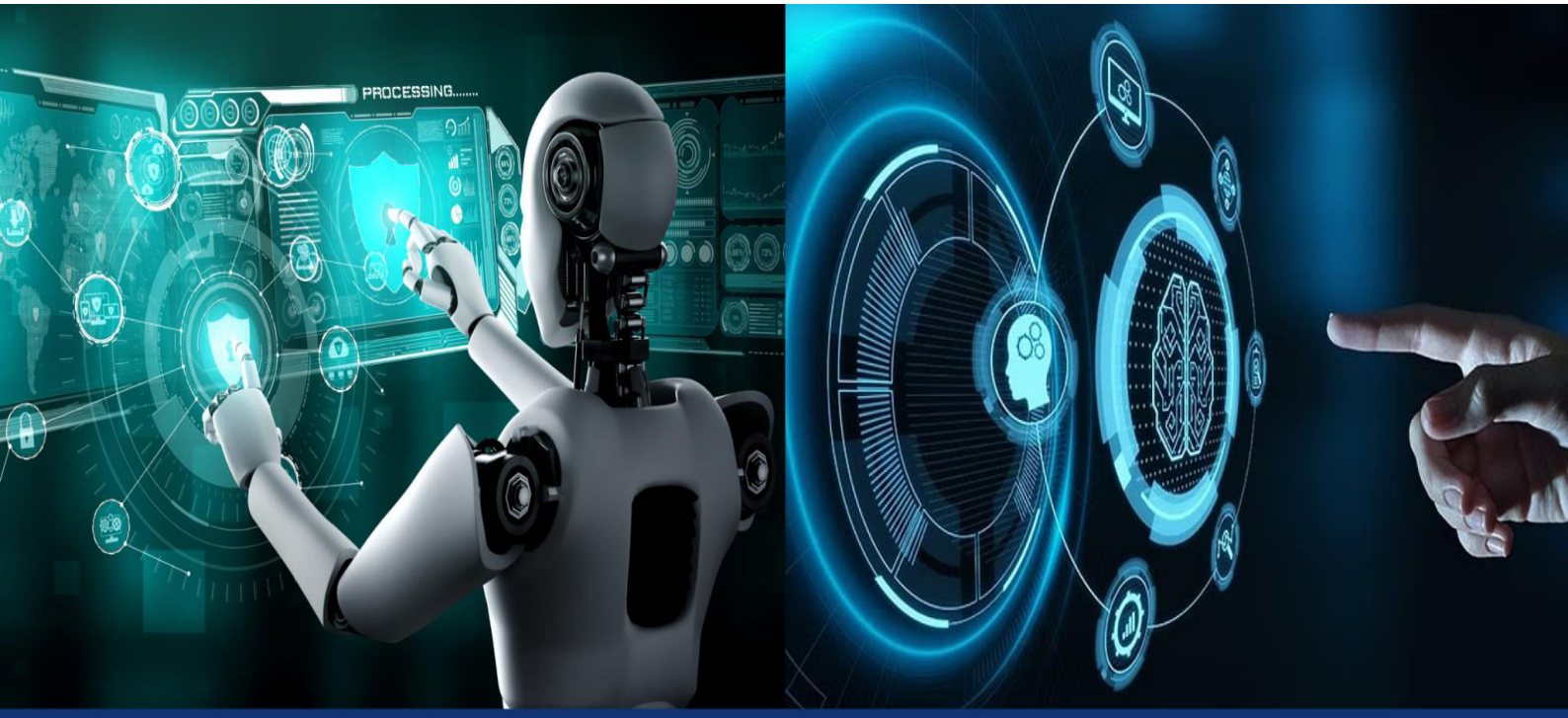
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 13, Issue 5, May 2025

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305162

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Android Text-To-Speech OCR Converter System

Prof. Reshma Lohar, Ammar Rahmatullah Sayed, Faazullah Vasiullah Khan, Saad Shaikh, Husain Nagori

Dept. of Computer Engineering, Rizvi College of Engineering, Mumbai, India

ABSTRACT This paper presents the development and evaluation of an Android-based Text-to-Speech Optical Character Recognition (TTS OCR) Converter System designed to assist visually impaired users in interacting with textual information within their environments. Visual impairment creates significant barriers to accessing written content found in various everyday settings, such as printed materials, signage, and electronic displays. Our proposed solution leverages Optical Character Recognition (OCR) to convert text from captured images into readable digital text, which is subsequently synthesized into speech through Text-to-Speech (TTS) technology. By transforming printed text into audio, this system empowers users to access textual information hands-free and independently. This application integrates image pre-processing techniques to improve text clarity, especially in challenging lighting conditions. The core OCR technology is powered by Tesseract, a robust open-source library, while the TTS functionality uses customizable settings to provide synthesized speech in multiple languages.

KEYWORDS: Android-based assistive technology, Text-to-Speech (TTS), Optical Character Recognition (OCR), visual impairment accessibility, real-time text recognition, Tesseract OCR engine, multilingual support, image pre-processing, user independence, mobile accessibility solutions.

I.INTRODUCTION

Visually impaired individuals encounter significant challenges when accessing textual information in their daily lives, as text is omnipresent in various forms such as printed documents, product labels, signboards, and digital screens. Limited access to this essential information can hinder their independence and limit opportunities for full social and professional participation. Although assistive technologies such as Braille and audio based navigation tools exist, they are not universally available and often lack the flexibility to address real-time needs in diverse environments. There is a growing need for portable, accessible solutions that empower visually impaired individuals to independently access text in their immediate surroundings. The convergence of Optical Character Recognition (OCR) and Text-to-Speech (TTS) technologies presents a promising solution for addressing this accessibility gap. OCR technology enables machines to interpret and convert printed or handwritten text into digital text by analysing visual data captured through images. When combined with TTS, this digital text can be converted into synthesized audio output, allowing visually impaired users to “read” text via auditory feedback. With recent advances in mobile computing power and the growing ubiquity of smartphones, it has become feasible to deploy OCR and TTS functionalities on mobile devices, providing an affordable, user friendly solution for real-time text interpretation. This research paper details the design and development of an Android-based TTS OCR Converter System. The system captures images of text using the device's camera, processes the captured data to recognize and extract text, and then converts the text into speech, which is read aloud to the user. This Android-based platform was selected for its extensive user base, cost-effectiveness, and capability to support the processing requirements of OCR and TTS operations.

II.LITERATURE SURVEY

The field of assistive technology has advanced significantly in recent years, with numerous studies exploring the integration of Optical Character Recognition (OCR) and Text-to-Speech (TTS) systems to aid visually impaired users. OCR technology has been widely studied for its ability to convert printed and handwritten text into machine readable formats. Research has demonstrated the effectiveness of OCR in real-world applications, such as document digitization and automated text analysis. However, its application in assistive technology, particularly for mobile platforms, poses unique challenges related to environmental conditions, text clarity, and processing requirements. Recent developments have emphasized the utility of OCR in accessibility tools for visually impaired users. For instance, studies have shown that OCR can achieve high accuracy when processing printed text under controlled lighting conditions. However, variable lighting, diverse text fonts, and complex backgrounds often reduce OCR accuracy in real-world environments. Open-source OCR solutions like Tesseract have become popular in mobile applications due to their adaptability and

support for multiple languages, making them suitable for multilingual use cases. Moreover, Tesseract's extensibility allows it to be customized for real-time mobile applications, further broadening its utility in assistive technologies. In parallel, TTS technology has evolved as a critical tool for accessibility, converting textual information into synthesized speech that users can easily understand. Studies on TTS systems indicate that these tools not only enhance accessibility but also improve user experience by providing auditory information in a customizable format. With advancements in mobile processors and cloud-based TTS solutions, researchers have successfully implemented TTS on mobile devices, thereby making audio-based information more accessible to visually impaired individuals. Customizable features, such as language selection, speech rate adjustment, and voice selection, have further enhanced TTS technology's adaptability to user preferences. The integration of OCR and TTS technologies on mobile platforms has been explored in recent research, primarily for visually impaired and elderly users. Mobile applications utilizing these technologies can provide users with real-time text interpretation capabilities, allowing them to read signs, labels, and other textual information in their environment. Existing solutions, however, often face limitations in terms of processing speed, accuracy in diverse conditions, and the flexibility to handle multilingual text.

III. ANDRIOD OCR TTS CONVERTER

A. Existing System

Currently, individuals with visual impairments face numerous challenges when it comes to accessing printed or handwritten text. Existing assistive technologies are either **expensive, bulky, or lack real-time performance**. Some of the key limitations include:

- **No Real-Time OCR Integration**: Most systems do not extract text from images or the environment in real time.
- **Limited Language Support**: Many available tools support only one or two languages, making them inaccessible for multilingual users.
- **Lack of Offline Functionality**: Most OCR-TTS systems require a constant internet connection to function properly, which is not feasible in remote or low-connectivity areas.
- **Complex Interfaces**: Some applications are not user-friendly for visually impaired users and require assistance to operate.
- **Inaccurate Text Recognition**: Low-quality OCR engines lead to poor text extraction, especially for regional scripts or low-resolution images.
- **No Integration with Navigation or Environment Awareness**: Existing models do not combine OCR and TTS effectively to describe surroundings or handle documents with complex formatting.

B. Proposed System

The proposed **Android OCR TTS Converter System** aims to provide a **low-cost, real-time, and accessible solution** for visually impaired users. The system uses a smartphone's camera to capture text from various sources and reads it aloud using **Text-to-Speech (TTS)**. The core features include:

- **Real-Time OCR**: Live text recognition using the device's camera via Flutter and Firebase ML Kit/OpenCV/Tesseract OCR.
- **Multilingual Support**: Accurate recognition of English, Hindi, Marathi, Urdu, and Arabic, with optional support for CJK (Chinese, Japanese, Korean) characters.
- **Offline Capability**: Lightweight model available for offline use; high-accuracy version as an online API.
- **Simple, Accessible Interface**: Designed with accessibility standards and voice-based navigation in mind.
- **Batch Text Extraction**: Allows processing of multiple images or documents efficiently.

C. Problem Statement

According to the **World Health Organization (WHO)** and accessibility reports, over **285 million people worldwide are visually impaired**, with a large portion residing in developing countries. In India, millions lack access to **affordable and effective assistive tools** for reading printed materials.

Most assistive devices are either not portable, require high-end hardware, or lack support for **regional languages**. As a result, visually impaired users are often **dependent on others** for reading documents, signs, instructions, or books — reducing their independence and quality of life.

The **Flutter-based Android OCR TTS Converter System** is designed to bridge this gap by offering:

2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305162

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- A **portable and affordable solution**.
- Real-time, **voice-assisted reading** of printed or handwritten text.
- Support for **multiple regional languages**.
- A platform that promotes **independence, accessibility, and ease of use** for the visually impaired.

D. Core Modules

1. Image Capture & Preprocessing:
This module enables users to capture images in real time using the device's camera, as well as import existing images from the gallery. It incorporates various image preprocessing techniques—such as resizing, thresholding, and contrast adjustment—to enhance the quality of the images. These preprocessing steps are crucial for improving the accuracy of both Optical Character Recognition (OCR) and object classification, ensuring that even low-quality images can be effectively processed.

2. Optical Character Recognition (OCR):
The OCR module extracts printed or handwritten text from the captured or imported images using tools like Tesseract OCR, Firebase ML Kit, or OpenCV. It supports multilingual recognition, including languages such as English, Hindi, Marathi, Urdu, and Arabic. The module also includes batch processing capabilities, allowing users to process multiple images or documents in a single operation, thereby improving efficiency and user convenience.

3. Text-to-Speech (TTS) Conversion:
This module converts the recognized text from the OCR process into audible speech using Flutter TTS or a native TTS engine. It reads the text aloud immediately after processing, providing real-time feedback to the user. Users can customize the speech output by adjusting the voice pitch, speed, and language for a more natural and personalized listening experience. The module is fully functional offline using pre-installed language models, ensuring continuous accessibility without internet connectivity.

4. Object Classification:
The object classification module analyzes the captured image to identify objects and categorize them into groups such as home goods (e.g., table, chair, appliance), fashion items (e.g., shirt, shoes, bag), plants and nature, electronics, and more. It employs a lightweight image classification model trained on datasets of common objects. The output includes a simple spoken description of the identified item, such as “This is likely a kitchen appliance,” enhancing the user’s understanding of their surroundings.

5. User Interaction & Accessibility:
This module is tailored specifically for visually impaired users. It features a user-friendly interface with large, high-contrast buttons and intuitive navigation supported by audio prompts. Voice command functionality enables hands-free operation for actions like “Capture,” “Read,” “Repeat,” and “Describe Object.” Additionally, haptic feedback provides tactile confirmation for successful operations or error notifications. The entire system is designed to function fully offline, with the exception of certain advanced classification features that may require an internet connection.

E. Technology Used for OCR + TTS Converter App

1. Hardware Requirements:

- **Processor**: Intel i5 or higher
- **RAM**: 8 GB (minimum)
- **Storage**: 500 GB or more (SSD preferred for faster build and testing)
- **Operating System**: Windows 11 / Linux / macOS (any Flutter-compatible OS)

2. Software Stack

- **Front End:**

The front-end development is carried out using Dart with the Flutter framework, which enables cross-platform mobile application development, with a primary focus on Android devices. Flutter’s built-in accessibility widgets are utilized to design a user interface specifically tailored for visually impaired users, ensuring ease of navigation and interaction.

2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305162

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- **OCR Engine:**

For optical character recognition, the system employs Firebase ML Kit, Tesseract OCR, or OpenCV to extract printed or handwritten text from images. These engines provide robust and accurate text recognition capabilities, supporting multiple languages and efficient processing for both single and batch image inputs.

- **TTS Engine:**

The text-to-speech functionality is implemented using the Flutter TTS plugin or the Android native Text-to-Speech API. These tools convert the recognized text into clear and natural-sounding speech, allowing users to hear the content immediately after OCR processing. Offline functionality is supported through pre-downloaded language models.

- **Image Classification:**

For object recognition, the system uses TensorFlow Lite (TFLite) or a custom-trained MobileNet model. These lightweight models are optimized for mobile environments and are capable of identifying a wide range of objects such as home goods, fashion items, plants, and more, providing real-time spoken feedback to the user.

IV. RESULT AND DISCUSSION

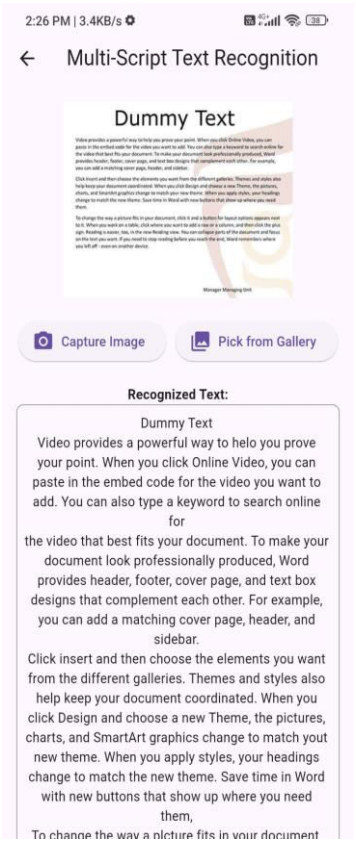


Fig.1 Text-to-Speech Integration

The recognized text is converted to speech using a text-to-speech engine. The system provides a clear and natural auditory output of the extracted text. This feature is particularly beneficial for users with visual impairments. The integration between OCR and TTS is seamless, allowing real-time recognition and speech conversion, as demonstrated in the test interface.

2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305162

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Fig.2 Object Detection Output

The object detection module successfully identifies and localizes multiple object classes within an image using bounding boxes. As shown in the results, the model accurately detects objects such as "person" and "car," demonstrating high precision and reliable bounding box placements. This validates the capability of the system to isolate relevant regions that may contain text or are important for contextual analysis.

Text Recognition Output

The text recognition component extracts textual data from images using an optical character recognition (OCR) engine. The system reliably detects and recognizes text even in the presence of complex backgrounds and varying font sizes. Bounding boxes are correctly aligned with the textual content, ensuring minimal character loss and improved legibility. This confirms the robustness of the OCR model in diverse visual conditions.

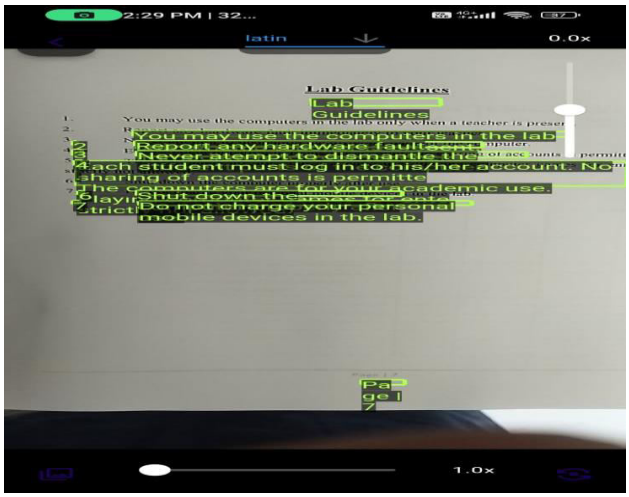


Fig.3 Text Recognition Output

V. CONCLUSION AND FUTURE SCOPE

The development of the **Android-Based OCR and TTS Converter App** marks a significant step toward improving **accessibility and independence for visually impaired users**. By integrating real-time **optical character recognition (OCR)**, **text-to-speech (TTS)**, and **object classification**, the app empowers users to interact more effectively with their environment and printed content.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

This system simplifies the process of reading and understanding printed or handwritten text, even in complex or low-light situations. Leveraging technologies like **Flutter** for cross-platform UI, **Firebase ML Kit** or **Tesseract** for OCR, and **Flutter TTS** for natural speech output, the app ensures both performance and ease of use. The addition of **object classification** helps users understand the category of items they encounter in daily life, such as identifying whether an item is a home appliance, fashion item, or plant.

With a **voice-controlled interface**, **haptic feedback**, and offline functionality, the application prioritizes accessibility and inclusiveness. The intuitive design allows visually impaired users to operate the app with minimal guidance, helping bridge the gap between technology and real-world accessibility challenges.
The app has been tested for reliability, accuracy, and ease of use. It runs smoothly on a wide range of Android devices, offering fast image capture, efficient text recognition, and accurate speech synthesis. Its modular design makes it easy to maintain, scale, and enhance with new features.

VI. FUTURE SCOPE

In future versions, the app can be extended and improved with the following enhancements:

1. **Expanded Object Recognition**: Integrate more detailed classification (e.g., specific brands, subcategories) and additional object categories relevant to users' daily lives.
2. **Smart Reading Features**: Implement AI-based **context summarization**, **automatic language translation**, and **semantic text understanding** to improve the reading experience.
3. **Real-Time Scene Description**: Integrate with AI vision models to **describe full scenes**, aiding navigation and situational awareness.
4. **Voice-Based File Navigation**: Add features to **save and organize scanned content**, retrievable through voice commands.
5. **Integration with Smart Glasses or Wearables**: Extend functionality to smart glasses or AR devices to make the system even more seamless and hands-free.
6. **Crowdsourced Dataset Expansion**: Allow users or caretakers to **contribute images** to improve model accuracy across diverse scripts and object types.
7. **Multilingual Feedback & Language Model Switching**: Enable dynamic switching of voice and OCR language based on region or preference.

REFERENCES

[1] Dr. Prashantha H S et al. Image Text to Speech Conversion in Desired Language, International Journal of Creative Research Thoughts (IJCRT), December 2023.
[2] Muhammad Ajmal et al. Image to Multilingual Text Conversion for Literacy Education, December 2018.
[3] H. Waruna H. Premachandra et al. Converting High Resolution Multilingual Printed Document Images into Editable Text Using Image Processing and Artificial Intelligence, March 2022.
[4] Nikolaos Bourbakis et al. Image Understanding for Converting Images into Natural Language Text Sentences, August 2010.
[5]https://ieeexplore.ieee.org/search/searchresult.jsp?action=search&newsearch=true&matchBoolean=true&queryText=(%22All%20Metadata%22:image%20to%20text%20conversion%20for%20visual%20impaired%20people)&highlight=true&returnFacets=%20ALL&returnType=SEARCH&matchPubs=true&pageNumber=2
[6] https://ijsrset.com/paper/5886.pdf
[7]https://www.researchgate.net/publication/286800382_OCR_Application_on_Smartphone_for_Visually_Impaired_People/fulltext/57aab49a08ae7a6420bd3805/OCR-Application-on-Smartphone-for-Visually-Impaired-People.pdf?origin=publication_detail&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uRG93bmxvYWQiLCJwcmV2aW91c1BhZ2UiOiJwdWJsaWNhd%20GlvbiJ9fQ
[8]https://docs.flutter.dev/?ref=phannhatchanh.com&gad_source=1&gclid=Cj0KCQjw05i4BhDiARIsAB_2wfCjusHTMDMeq1uEl0dkeV0xXvvqJ9GP2pbQYMKIEk4Bxgk%203s1xo4mgaAo5XEALw_wcB&gclsrc=aw.ds
[9] https://pub.dev/

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details