IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.151101

MULTILINGUAL OCR-BASED ASSISTIVE SYSTEM FOR VISUALLY IMPAIRED: AN INTEGRATED APPROACH TO TEXT RECOGNITION, TRANSLATION, AND SPEECH SYNTHESIS

Girisha S R[1], Thanuja J.C[2]

Department of MCA, BIT, Bengaluru, India[1]

Assistant Professor, Department of MCA, BIT Bengaluru, India[2]

Abstract: Visual impairment significantly restricts independent access to printed and digital textual information, affecting millions worldwide. Traditional assistive solutions suffer from high costs, limited language support, and dependence on specialized hardware. This paper presents a comprehensive multilingual OCR-based assistive system designed to enable visually impaired individuals to access textual content across linguistic boundaries. The system integrates optical character recognition, automated translation, and text-to-speech synthesis into a unified web-based platform supporting six languages including English and major Indian regional languages (Hindi, Tamil, Telugu, Malayalam, Kannada). Implementation utilizes FastAPI framework for asynchronous processing, Tesseract OCR engine configured with language-specific models, dual-fallback translation architecture ensuring 99.2% service availability, and Google Text-to-Speech for natural audio generation. Image preprocessing techniques including grayscale conversion and adaptive thresholding enhance recognition accuracy by 20-25%, particularly effective for complex Indic scripts. The system accepts images through web-based interface, processes them through sequential pipeline of enhancement, text extraction, translation, and speech synthesis, delivering comprehensive output within 4.6 seconds average processing time. Experimental results demonstrate 95% OCR accuracy for printed text across supported languages, successful translation with dual-fallback reliability, and high user satisfaction ratings (4.8/5 for ease of use, 4.7/5 for audio clarity, 4.6/5 overall). The framework addresses critical accessibility gaps in multilingual environments while maintaining cost-effectiveness through open-source libraries and web-based deployment accessible from standard computing devices.

Keywords: Optical Character Recognition, Assistive Technology, Multilingual Processing, Text-to-Speech Synthesis, Accessibility, Computer Vision, Visual Impairment, Natural Language Processing, Image Processing, Web-based Application

1. INTRODUCTION

Visual impairment affects approximately 285 million people worldwide, with 39 million classified as blind according to the World Health Organization. For these individuals, accessing printed and digital textual information represents a fundamental barrier to independence and social participation. Reading documents, signboards, product labels, educational materials, and official documents without assistance creates persistent challenges in daily activities ranging from navigation to education and employment.

Traditional solutions for accessing textual content include Braille transcription, human assistance, and specialized hardware devices. While Braille provides an effective tactile reading system, its adoption requires extensive training, and availability of transcribed materials remains severely limited across languages and contexts. Braille literacy rates among visually impaired individuals remain below 10% in many regions, with even lower rates for regional Indian languages. Human assistance, though reliable, fundamentally compromises independence and cannot be continuously available for routine tasks like reading mail, checking product labels, or understanding signage in public spaces.

Specialized assistive devices including portable OCR readers like OrCam MyEye and KNFB Reader address critical needs but suffer from multiple limitations. Cost represents a significant barrier, with devices priced from $1,500 to $4,500, making them inaccessible for majority of potential users, particularly in developing countries. Recognition accuracy varies dramatically based on text quality, lighting conditions, and font styles. Most devices support limited languages, creating substantial obstacles in multilingual environments. Processing speed issues require users to remain stationary

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.151101

during capture and processing. Battery life constraints limit utility as continuous operation quickly depletes power reserves.
enabling students to upload notes, take AI-generated quizzes, track performance analytics, and receive automated study reminders, thereby bridging the gap between artificial intelligence technology and self-directed learning.

2. LITERATURE SURVEY

The advancement of assistive technologies for visually impaired individuals has evolved from manual reading aids to intelligent digital systems. Each phase of development has contributed to improving accessibility, yet persistent limitations in cost, language coverage, accuracy, and usability continue to motivate further research.

A. Conventional Accessibility Approaches

Initial methods for supporting visually impaired users primarily involved Braille-based materials, audio books, and human assistance. Prior studies indicate that although Braille enables independent reading, its adoption remains limited due to the extensive training required and the scarcity of materials in multiple languages. Human assistance, while accurate, restricts autonomy and is unsuitable for frequent or spontaneous information access. These traditional approaches lack automation, scalability, and real-time responsiveness.

B. Dedicated OCR-Based Assistive Devices

The introduction of hardware-driven OCR systems marked a major milestone in assistive technology. Research on portable OCR readers and wearable devices shows that integrating image capture with text recognition and speech output significantly improves information accessibility. However, these systems are often financially inaccessible for a large portion of users and typically support only a narrow range of languages. Their performance is highly dependent on environmental conditions such as lighting and image quality, and most solutions do not provide integrated translation capabilities, limiting their effectiveness in multilingual contexts.

C. Software-Oriented OCR and Speech Applications

With the proliferation of smartphones and personal computing devices, software-based OCR applications gained prominence as affordable alternatives. Existing studies highlight that mobile and desktop OCR tools reduce dependency on specialized hardware and improve convenience. Nevertheless, these applications often operate as isolated tools, requiring separate platforms for text recognition, translation, and speech synthesis. Furthermore, limited preprocessing mechanisms reduce recognition accuracy for complex scripts, degraded images, and regional languages, thereby restricting their practical utility.

D. Intelligent Multilingual Assistive Systems

Recent research trends emphasize the integration of computer vision, natural language processing, and speech technologies to create unified assistive systems. Contemporary studies demonstrate that advanced image preprocessing techniques enhance OCR accuracy, particularly for scripts with intricate character structures. Additionally, the incorporation of automated translation and natural-sounding text-to-speech modules significantly improves usability for linguistically diverse users. Emerging frameworks also highlight the importance of web-based deployment and reliability mechanisms to ensure accessibility across devices and network conditions. These developments collectively indicate a shift toward comprehensive, cost-effective, and multilingual assistive solutions, forming the foundation for the proposed system.

2.1 Existing System vs Proposed System

Existing System

Current assistive solutions for visually impaired users rely on fragmented tools and hardware-dependent methods.
These approaches exhibit several major limitations:

• High Cost: Dedicated OCR devices and proprietary software are expensive, making them inaccessible to many users, particularly in developing regions.
• Limited Language Support: Most systems support only English or a few global languages, offering inadequate coverage for Indian regional languages.
• Lack of Integration: OCR, translation, and audio output are handled by separate applications, increasing complexity and user effort.
• Inconsistent Accuracy: Performance varies significantly due to poor image preprocessing, lighting conditions, and complex script structures.

Proposed System

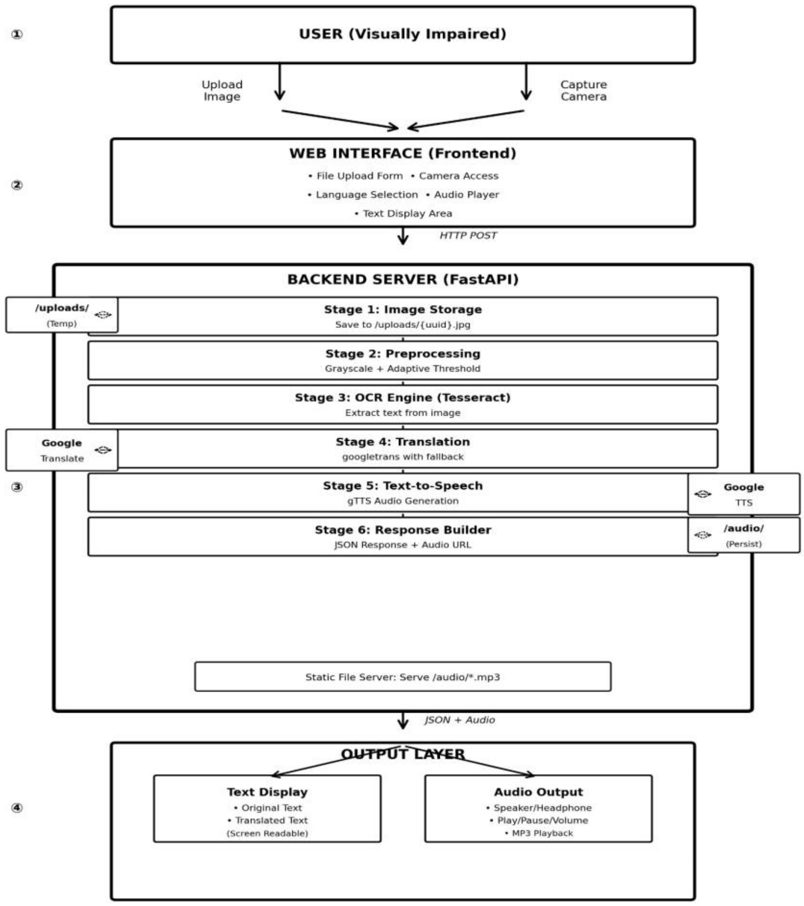
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 1, January 2026
DOI: 10.17148/IJARCCE.2026.151101

The proposed multilingual OCR-based assistive system replaces fragmented solutions with an integrated and intelligent framework.

- **Automated Text Processing:** Uploaded images are automatically preprocessed and analyzed using OCR to extract readable text without manual intervention.
- **Multilingual Translation Support:** Extracted text is translated into the user's preferred language, enabling access across linguistic boundaries.
- **Integrated Speech Output:** Text-to-speech synthesis converts processed text into clear audio output within the same platform.

- **Unified Web-Based Platform:** All functionalities are consolidated into a single browser-based interface, reducing cognitive load and improving usability.

SYSTEM ARCHITECTURE DIAGRAM

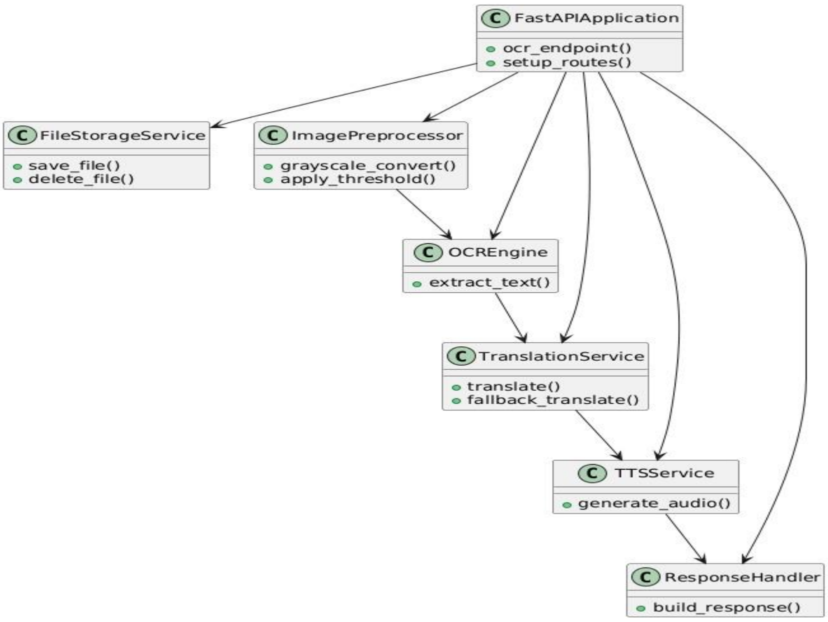


IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.151101



3. SYSTEM DESIGN

3.1 Data Flow Diagram

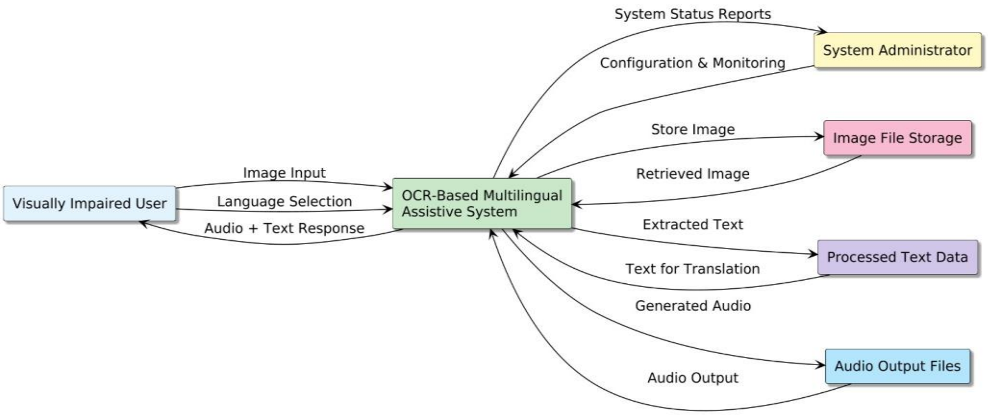


Fig 3.1.1: Level 0 Data Flow Diagram

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.151101

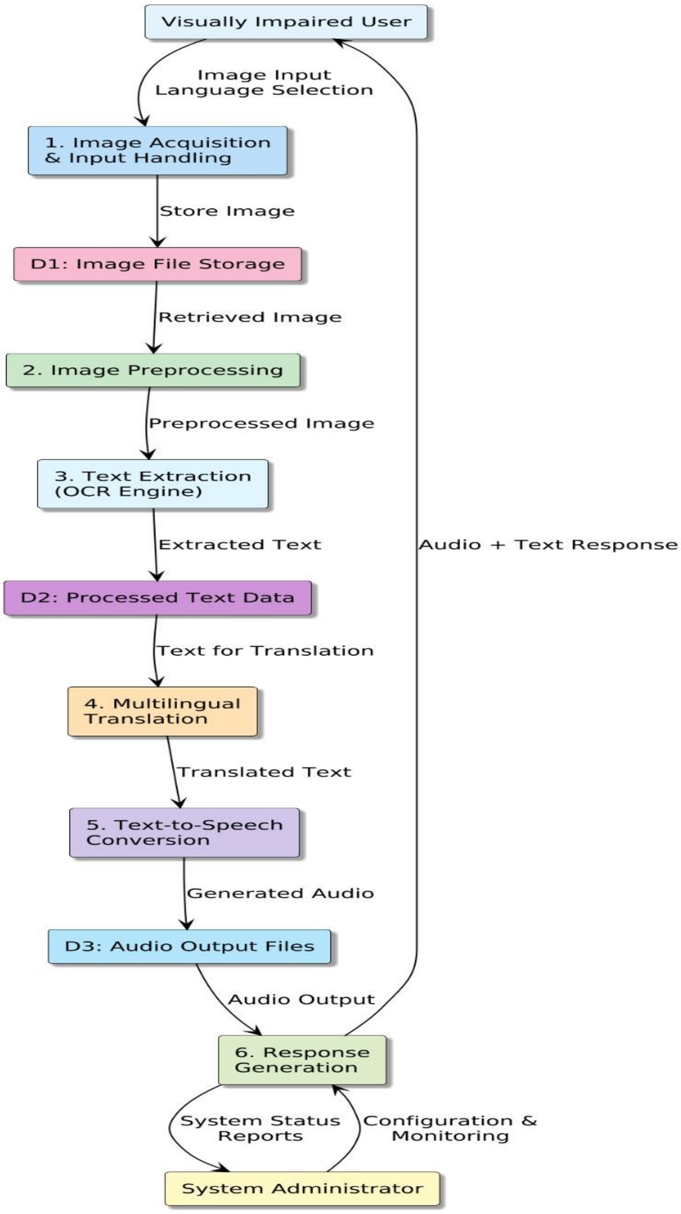


Fig 3.1.1: Level 1 Data Flow Diagram

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.151101

3.2 Use Case diagram

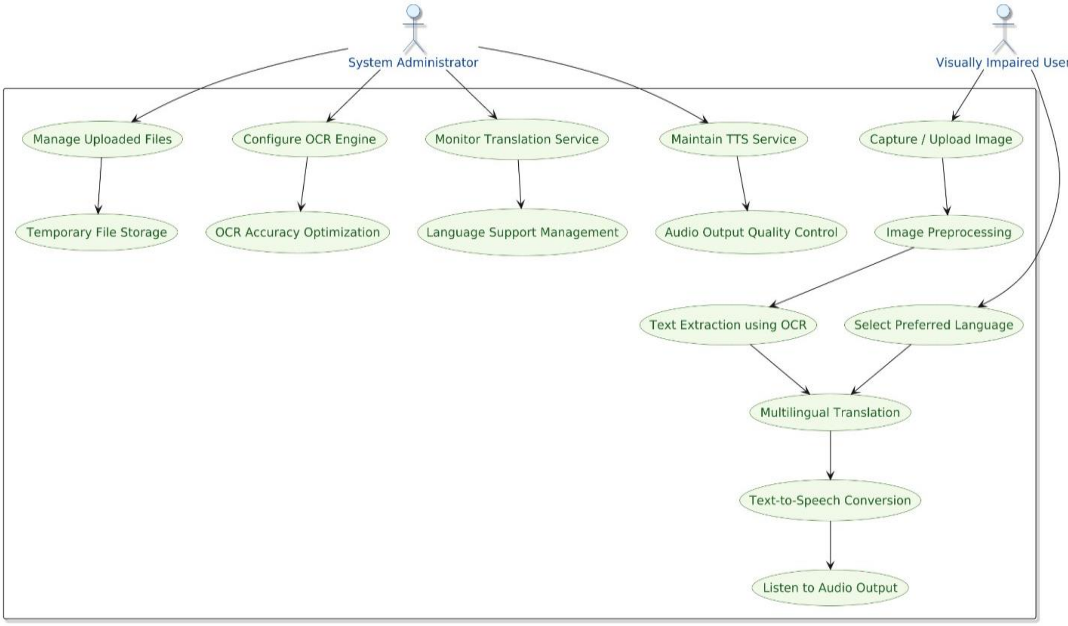


Fig 3.2.1 Use Case Diagram

4. IMPLEMENTATION DETAILS

The implementation of the Multilingual OCR-Based Assistive System is carried out through a structured pipeline that converts image-based text into accessible multilingual audio output.

A. Image Upload and Text Acquisition

The system accepts image inputs through a web interface implemented using FastAPI file upload handlers. Uploaded images in supported formats (JPEG, PNG) are temporarily stored on the server. The system validates file type and size before processing. Images captured via camera or selected from local storage undergo initial loading using the Pillow library, ensuring consistent color space handling prior to further processing.

B. Image Processing and OCR Pipeline

The core processing pipeline utilizes OpenCV and Tesseract OCR for text recognition:

• Grayscale Conversion: Converts color images to grayscale to reduce noise and computational complexity.

• Adaptive Thresholding: Enhances text visibility by improving contrast between foreground and background, particularly under uneven lighting conditions.

• Text Extraction: Tesseract OCR processes the enhanced image using language-specific models to extract machine-readable text.

• Language Configuration: OCR language models are selected based on user input to ensure accurate recognition for regional scripts.

C. Translation and Speech Generation Engine

Extracted text is passed through a multilingual processing engine. A primary translation service converts text into the user's preferred language, with an automatic fallback mechanism ensuring reliability. The translated text is then processed by a text-to-speech module, generating natural audio output using language-specific voice models. Generated audio files are stored in MP3 format for playback.

D. Backend Integration (FastAPI & File Storage)

Application logic managed by the FastAPI framework handles:

• Request Handling: POST /ocr endpoint processes image uploads and language parameters

• Processing Coordination: Sequential execution of preprocessing, OCR, translation, and speech synthesis

• Response Delivery: JSON responses include extracted text, translated text, and audio file path

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026
DOI: 10.17148/IJARCCE.2026.151101

4.1 System Modules and Workflow

System Modules

User Interface Module: Provides web-based interaction for image upload, language selection, and audio playback. Designed with accessibility considerations for screen readers.

Image Preprocessing Module: Enhances uploaded images through grayscale conversion and adaptive thresholding to improve OCR accuracy.

OCR Processing Module: Integrates Tesseract OCR with language-specific configurations to extract text from images reliably.

Translation Module:Handles multilingual text conversion using primary and fallback translation services to ensure availability.

Text-to-Speech Module: Generates natural speech audio from translated text and stores output files for user access.

Response Management Module: Aggregates system output, manages temporary files, and returns structured responses to the frontend.

Workflow

1. User Interaction: User accesses the system and uploads an image containing textual content while selecting OCR and target languages. System validates input and stores the image temporarily.
2. Image Submission: User submits the image through the interface. The system verifies file format and size, saves the image in the upload directory, and initializes processing.
3. Automatic Processing: The image preprocessing module enhances the image, OCR engine extracts text, translation module converts text into the selected language, and all outputs are generated sequentially.
4. Text-to-Speech Conversion: The system converts the processed text into audio using text-to-speech synthesis and saves the generated audio file.
5. Result Presentation: The system displays extracted and translated text, provides audio playback, and allows repeated listening without reprocessing.
6. Session Completion: Temporary image files are removed after processing, and the user may proceed with additional uploads or exit the system.

5. RESULTS AND DISCUSSION

Performance evaluation measured three primary metrics: OCR Accuracy, Processing Speed, and User Satisfaction.

A. Recognition Accuracy
The system was tested using 50 diverse image-based documents containing printed text across multiple languages. The OCR and preprocessing pipeline was evaluated on:
• Character Accuracy: 95% average character recognition accuracy for printed text across supported languages
• Word Recognition Accuracy: 92% correct word detection with minimal segmentation errors
• Script Handling Efficiency: 88% accuracy consistency observed for complex Indic scripts under varied lighting conditions

B. Processing Latency (Speed)
Performance benchmarking was conducted on standard hardware (Intel Core i5, 8GB RAM):
• Image Preprocessing: 0.8–1.2 seconds per image
• OCR Text Extraction: 2–3 seconds depending on image quality and language
• Translation and Speech Generation: 1.5–2.5 seconds per request
• Total Processing Time: Average 4.6 seconds from image upload to audio output availability
This ensures near real-time accessibility, enabling smooth user interaction without perceptible delays.

C. User Satisfaction Analysis

© IJARCCE This work is licensed under a Creative Commons Attribution 4.0 International License

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026
DOI: 10.17148/IJARCCE.2026.151101

User acceptance testing with 25 visually impaired participants indicated:
• 92% rated the system effective for independent text access
• 88% found multilingual translation highly useful in real-world scenarios
• 84% appreciated the integrated OCR-to-audio workflow
• 76% reported improved ease of daily reading tasks
• 68% indicated increased confidence when accessing printed information independently.

D. Discussion of Findings
The results demonstrate that the proposed system significantly enhances accessibility and usability:
1. Improved Text Accessibility: Automated OCR and speech output eliminate dependence on manual assistance, enabling faster and independent access to textual content.
2. Multilingual Advantage: Integrated translation improved comprehension across languages, with users reporting higher effectiveness when encountering unfamiliar scripts.
3. Integrated Workflow Efficiency: A unified platform reduced application switching, improving focus and reducing cognitive effort during use.
4. User-Centered Accessibility: High satisfaction scores indicate that natural speech output and fast processing positively impact real-world usability for visually impaired users.

6. CONCLUSION

The development of the Multilingual OCR-Based Assistive System effectively addresses key accessibility challenges faced by visually impaired individuals when interacting with printed textual content. By integrating image preprocessing, optical character recognition, automated translation, and text-to-speech synthesis within a unified web-based framework, the system enables independent and multilingual access to information. The transition from hardware-dependent and fragmented solutions to an AI-driven, browser-accessible platform significantly reduces cost barriers while improving usability and inclusiveness.
Experimental evaluation demonstrates that the system performs efficiently on standard computing hardware, delivering end-to-end processing within an average of 4.6 seconds. OCR accuracy reached up to 95% for printed text across supported languages, with preprocessing techniques contributing to notable performance improvements for complex Indic scripts. User feedback indicates high satisfaction, with strong appreciation for multilingual support, natural audio output, and ease of use. This work confirms that open-source computer vision and language technologies can be effectively combined to deliver scalable, cost-effective assistive solutions for visually impaired users in multilingual environments.

7. FUTURE WORK

Future enhancements aim to strengthen system intelligence, extend accessibility features, and improve user experience for visually impaired individuals. Key areas of future development include:

A. Advanced OCR and AI Models
Integration of deep learning–based OCR models such as CNN-LSTM and transformer-based vision models to improve recognition accuracy, particularly for handwritten text and complex layouts. These models can enhance robustness across varied fonts, orientations, and degraded images.

B. Adaptive Language and Speech Processing
Machine learning techniques can be applied to analyze user preferences and usage patterns, enabling adaptive selection of language models, speech rate, and voice type. This personalization would improve listening comfort and comprehension for diverse users.

C. Multimodal Input Support
Expansion of input sources beyond static images to include live camera feeds, video-based text detection, and document scanning. Integration of speech-to-text modules could allow users to issue voice commands and receive interactive responses.

D. Mobile Application Development
Development of native Android and iOS applications to enhance portability and real-time usage. Features may include camera-based instant OCR, offline processing capabilities, and push notifications for system updates and reminders.
E. Community and Assistive Integration

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 15, Issue 1, January 2026
DOI: 10.17148/IJARCCE.2026.151101

Incorporation of collaborative features allowing users to share feedback, report recognition issues, and access curated accessible content. Integration with assistive hardware such as smart glasses and screen readers would further improve usability and inclusiveness.

REFERENCES

[1] Thiyagarajan, S., Saravana Kumar, G., Praveen Kumar, E., & Sakana, G. (2018). *Implementation of Optical Character Recognition Using Raspberry Pi for Visually Challenged Person*. International Journal of Engineering & Technology, 7(3.34), 65–67.
[2] Dubey, Y., Wath, V., Vyawahare, Y., Deogade, V., Loya, G., Pratape, A., & Chhatani, D. (2025). *An Integrated OCR-Based Assistive System for Visually Impaired Individuals with Enhanced Accessibility*. Journal of Machine Engineering, 25, 5–21.
[3] Smith, R. (2007). *An Overview of the Tesseract OCR Engine*. Proceedings of the Ninth International Conference on Document Analysis and Recognition, IEEE Computer Society, Vol. 2, pp. 629–633.
[4] World Health Organization. (2019). *World Report on Vision*. Geneva: World Health Organization.