International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0440

e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2795 - 2804

Text Recognition, Text Scanner, Handwriting Text Recognition OCR

Shodha H M[1], *Pintu Kumar*[2]
[1,2] *MTech 2nd Sem, Department of CSE, RYM Engineering College-RYMEC, Ballari, VTU Belagavi, Karnataka, MTech CSE, Indian Institute of Technology, Patna, Bihar, India.*
Emails: *shodhahm612@gmail.com*[1], *pintu_2302res16@iitp.ac.in*[2]

Abstract
This project focuses on the development of a native iOS application for Optical Character Recognition (OCR) using Apple's Vision Framework. The app is designed to detect, extract, and digitize both printed and handwritten text from various image sources, providing users with a fast, accurate, and convenient tool for tasks such as document scanning, note-taking, record keeping, and content sharing. The application supports multiple input options: users can scan text directly using the device camera, select images from the photo library, or upload documents through a file picker. Once an image is selected, the app performs real-time text recognition, highlighting detected regions with bounding boxes and displaying interactive previews. Recognized text can be easily copied, saved as a .txt file, or shared through native system integrations. The user interface is carefully designed to ensure clarity, responsiveness, and accessibility, including Voice Over support for visually impaired users. It is fully optimized for various iPhone screen sizes. To enhance global usability, the app is localized into over 10 languages, including English, Hindi, Chinese, Japanese, Arabic, French, German, and Spanish. This ensures a seamless experience for a wide range of users across different regions. Importantly, the application operates fully offline, prioritizing user privacy and data security by eliminating the need for internet connectivity or external servers. Its combination of multilingual support, offline capability, and a lightweight yet powerful interface makes it an ideal tool for students, researchers, and professionals alike.
Keywords: *Optical Character Recognition (OCR); iOS Application; Apple Vision Framework; Handwritten Text Recognition; Real-Time Text Detection; Multilingual Support; Offline Processing.*

1. Introduction
In today's digital age, where data is predominantly typed or printed, the digitization of handwritten text remains a significant challenge. Applications ranging from education, historical archiving, and healthcare to legal documentation rely on capturing handwritten information. Although Optical Character Recognition (OCR) technologies have shown great promise in printed text recognition, handwriting—especially in regional and multilingual contexts—poses inherent complexity. Different character shapes, writing speeds, personal handwriting styles, slant angles, and spacing variations contribute to decreased recognition accuracy. With the emergence of advanced mobile processors and machine learning frameworks such as Apple's Vision, there is now the potential to develop fast and accurate handwriting recognition systems that work directly on mobile devices. At the same time, the linguistic diversity of users must be considered. In a multilingual country like India and in globally distributed settings, users write in different scripts such as Latin (English), Arabic, and more. A truly useful system should thus support multilingual handwriting recognition. This thesis explores the design and development of an iOS-based solution that harnesses the Vision framework for real-time, on-device handwriting recognition in multiple languages. In addition to text extraction, the app integrates accessibility tools like Voice Over, ensuring usability for visually impaired users. Despite the availability of OCR systems, there remains a gap in support for efficient and accurate recognition of handwritten multilingual text on mobile platforms. Existing solutions often rely on cloud- based services, raising privacy concerns and creating latency. Moreover, they are frequently limited to a narrow set of languages and do not

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0440
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2795 - 2804

support accessibility features. The lack of a device-native, real-time, privacy-preserving solution hinders adoption in sensitive use cases. This research aims to address these challenges by developing a real-time iOS handwriting text recognition system using the Vision framework, capable of handling multiple languages while maintaining high performance and accessibility.

2. Literature Survey

The table presents a curated list of ten influential research works from 2007 to 2020 that collectively illustrate the rapid evolution of Optical Character Recognition (OCR) technologies. These studies cover a broad spectrum of innovations, including early open-source engines like Tesseract, the introduction of Long Short-Term Memory (LSTM) networks for unconstrained handwriting recognition, and the adoption of advanced deep learning models such as Transformers and Convolutional Neural Networks (CNNs) for scene text detection. Several papers focus on enhancing multilingual and handwritten text recognition, addressing challenges posed by script complexity, low-quality inputs, and natural scene variability. A recurring theme across the literature is the transition from traditional rule-based OCR systems to data-driven, neural network-based architectures. Works like BERT and "Attention is All You Need" have significantly influenced OCR post-processing by enabling better contextual understanding of extracted text. Meanwhile, surveys and domain-specific implementations (e.g., for Chinese characters or biomedical documents) provide insight into the adaptation of OCR systems to diverse languages and specialized applications. The integration of accessibility and mobile optimization is also gaining importance, especially with lightweight models designed for real-time text recognition on smartphones. Collectively, these studies underline the importance of balancing recognition accuracy, processing speed, and hardware constraints. The literature also reflects a shift toward more inclusive OCR systems that support multilingual content and assist users with varying needs, such as through accessibility features or mobile-first deployment. These developments signal a move toward more practical, user-centric OCR applications across fields like education, healthcare, and digital documentation.

Table 1 OCR & NLP Research Overview

Authors	Year	Title	Learnings	Limitations
Chen et al.	2020	Text Scanner for Scene Text Recognition	Combined segmentation and sequence modeling to enhance recognition of irregular text in natural scenes.	Performance declines under blurred or low-contrast images.
Zhang, T. et al.	2020	OCR for Multilingual Scene Images	Demonstrated multilingual text recognition using deep CNN-RNN hybrids.	Not optimized for real-time mobile deployment.
Topol, E.	2019	Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again	Highlights how AI, including OCR, is transforming medical data digitization.	Conceptual discussion; lacks implementation-specific details.
Li, X. et al.	2019	A Survey of Handwritten Chinese Character Recognition	Summarized deep learning techniques for complex Chinese script recognition.	Focused only on Chinese; not broadly multilingual.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0440
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2795 - 2804

Baek et al.	2019	What is Wrong with Scene Text Recognition?	Proposed unified architecture addressing inconsistencies in existing OCR models	Less robust on highly varied handwriting.
Devlin et al.	2019	BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding	Revolutionized NLP tasks and post-processing of OCR outputs via contextual embeddings.	Not designed for on-device/mobile OCR tasks.
Vaswani et al.	2017	Attention Is All You Need	Introduced the Transformer model, enabling sequence modeling improvements in OCR.	High computational requirements; not ideal for mobile.
Jaderberg et al.	2016	Reading Text in the Wild with Convolutional Neural Networks	Demonstrated CNN-based end-to-end text spotting in natural scene images.	Performance drops with curved or occluded text.
Graves et al.	2009	A Novel Connectionist System for Unconstrained Handwriting Recognition	Introduced LSTM networks with CTC loss for recognizing unconstrained handwritten text.	Sensitive to image noise and alignment issues.
Smith, R.	2007	An Overview of the Tesseract OCR Engine	Presented open-source OCR engine capable of recognizing printed text.	Poor performance with handwritten or cursive text.

3. Methodology

The proposed handwriting recognition system is built around a modular pipeline, enabling efficient processing of image data into editable, digital text. Each module in the pipeline is designed to handle specific challenges related to input quality, language variation, and user interaction. [1]

3.1. System Architecture

The system architecture comprises five major components: image acquisition, preprocessing, OCR engine configuration, post-processing, and output interface. The entire process is executed locally on the device using Apple's native frameworks such as Vision and CoreML, ensuring low latency and preserving user privacy. [2]

Image Acquisition: Images are captured using the device camera or selected from the gallery. The camera module uses AVCaptureSession to access real-time video feed, supports torch for low-light conditions, and provides a live preview with focus guidance. Tapping the screen allows users to adjust focus and exposure, improving clarity. [3]

3.2. Preprocessing

Raw images undergo a series of transformations to enhance text visibility and remove noise. These include:

- Grayscale Conversion to reduce dimensionality and emphasize shape.
- Contrast Enhancement using CLAHE (Contrast Limited Adaptive Histogram Equalization). [4]
- Noise Reduction via Gaussian filtering and median blur.
- Skew Correction for tilted or rotated input.
- ROI Detection to isolate textual regions from background clutter.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0440
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2795 - 2804

3.3.OCR Engine

Apple's Vision framework is configured with VN Recognize Text Request. It supports:

- Recognition levels: fast or accurate
- Multiple recognition languages (e.g., English, French, Arabic, Japanese, etc.)
- Bounding box detection for word and line segments [5]
- Confidence scores for each result

3.4.Post-Processing

Once text is recognized, the system performs:

- Language detection
- Spacing and punctuation corrections
- Bounding box merging for broken lines
- Conversion to Unicode-compliant output for downstream compatibility

3.5.Output Interface

Recognized text is presented in a scrollable, editable field. Users can:

- Copy to clipboard
- Export as .txt or .pdf [6]
- Share via system integrations (email, messaging, cloud drives)

4. Deep Learning Background

Modern OCR systems have transitioned from traditional heuristic algorithms to deep learning-based approaches that can learn from data patterns and improve generalization.

4.1. Convolutional Neural Networks (CNNs)

CNNs are used for spatial feature extraction. In handwriting recognition, they help identify edges, curves, loops, and individual character shapes. CNN layers use filters to scan across the image and learn hierarchical features. [7]

4.2.Recurrent Neural Networks (RNNs) and LSTM

RNNs are adept at handling sequential data. When recognizing text lines, maintaining character order is critical. LSTMs solve the vanishing gradient problem in traditional RNNs, making them ideal for modeling long sequences like full sentences or paragraphs.

4.3.CRNN and TrOCR

CRNN (Convolutional Recurrent Neural Network) combines CNNs with RNNs in a unified architecture, often used in scene text recognition. TrOCR (Transformer OCR) incorporates attention-based mechanisms, allowing the model to focus dynamically on relevant text areas.

While deep learning models offer high accuracy, they require large training datasets and are computationally intensive. Apple's Vision framework abstracts this complexity and provides pre-trained models that are optimized for real-time mobile performance. [8]

5. Multilingual Support & Accessibility

5.1. Multilingual Handwriting Recognition

One of the standout features of the system is its ability to recognize handwritten text in over 10 languages. This includes languages with diverse scripts, such as Latin (English, Spanish), Cyrillic (Russian), Arabic, Japanese (Kanji and Kana), and Chinese (Simplified and Traditional). Each language presents unique challenges, such as:

- **Latin Scripts:** Variations in case sensitivity, cursive writing, and diacritical marks (e.g., in French or Spanish).
- **Arabic:** Right-to-left writing, ligatures, and contextual character forms.
- **East Asian Scripts:** Thousands of character classes, complex stroke patterns, and mixed use of ideograms.

The system uses the recognition Languages parameter in Vision API to customize OCR for specific languages. Users can pre-select their language, or allow the system to auto-detect based on script features and region settings. For example, a note written in Japanese kanji is recognized differently than one in Romanized Japanese. By tailoring the OCR engine's language model, recognition accuracy improves significantly.

5.2. UI Localization

Beyond OCR, the user interface itself is localized in the app settings. Language-specific labels, prompts, error messages, and accessibility cues are translated using Apple's NSLocalizedString API. This allows users to interact with the application in their native language, improving usability and reducing learning curves.

Supported UI Languages Include:

- English
- Hindi
- Chinese (Simplified)

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0440
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2795 - 2804

- Japanese
- Spanish
- Arabic
- French
- Russian
- German
- Korean

5.3. Accessibility Integration (Voice Over)

Inclusion and accessibility are core design principles of the system. The app integrates with iOS's Voice Over screen reader, making it usable for blind and visually impaired individuals.

- All UI components are tagged with accessibility Label, accessibility Hint, and accessibility Traits.
- The recognized text can be read aloud automatically using AV Speech Synthesizer, providing instant feedback.
- Font sizes adjust dynamically via UI Font Metrics based on user preferences.
- High-contrast themes and dark mode support are included for users with low vision.

The combination of OCR and TTS (Text-to-Speech) creates a full loop of visual-to-audio transformation empowering a broader audience to engage with handwritten content.

6. Results & Evaluation

The system was rigorously evaluated across multiple metrics and use cases to benchmark its performance.

6.1.Experimental Setup

- **Device:** iPhone 13 with A15 Bionic Chip
- OS: iOS 15.4
- **Languages Tested:** English, French, Spanish, German, Russian, Chinese (Simplified), Japanese, Arabic, Korean, Italian
- **Sample Size:** 200 handwritten images from 25 users
- **Testing Conditions:** Varied lighting, writing styles (print, cursive), paper types (lined, unlined)

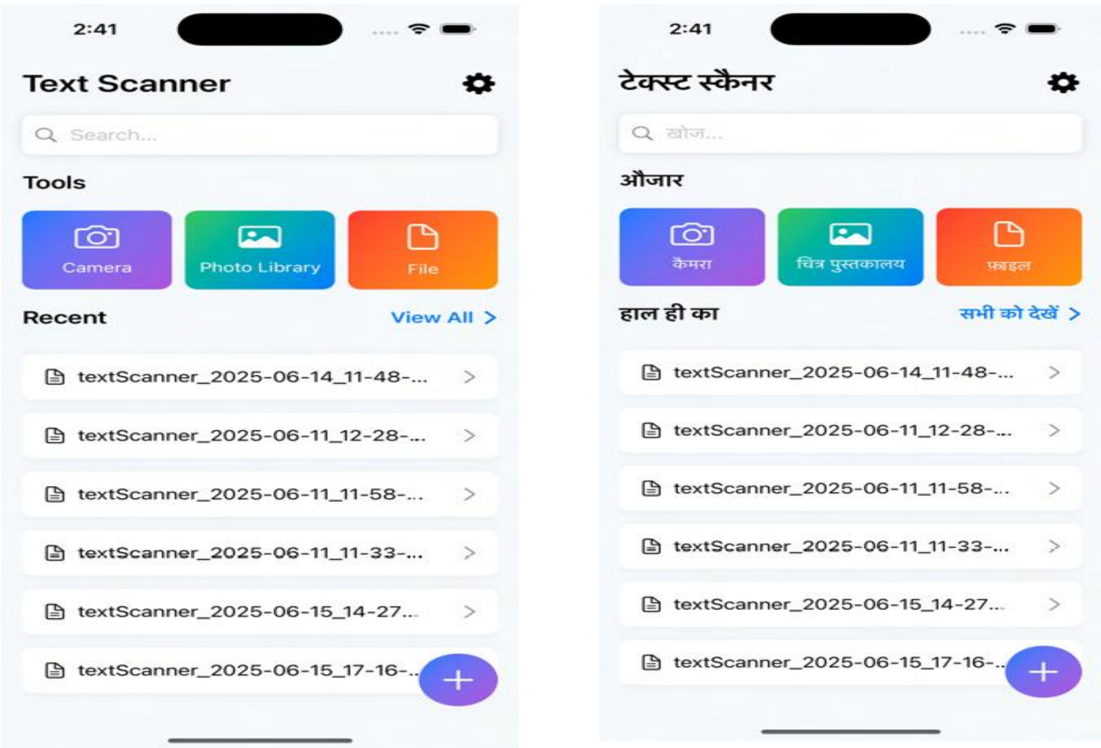

Figure 1 Home or Dashboard Screen

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0440
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2795 - 2804

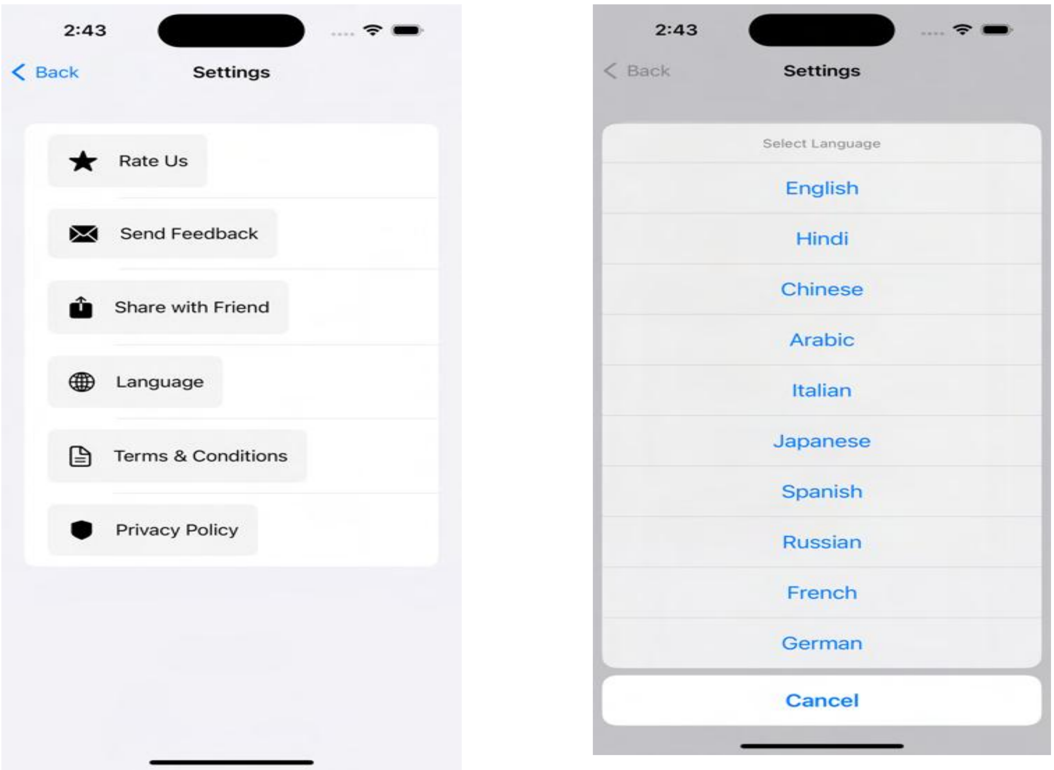

Figure 2 Settings or Language Selection Screen

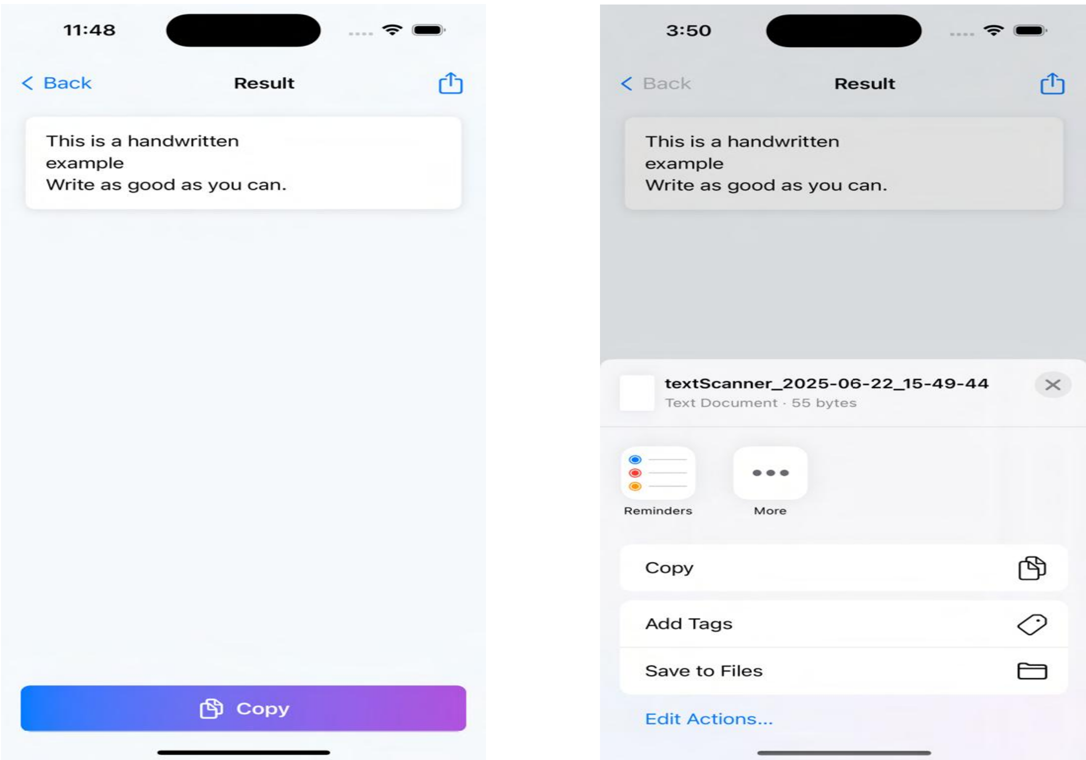

Figure 3 Export

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0440
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2795 - 2804

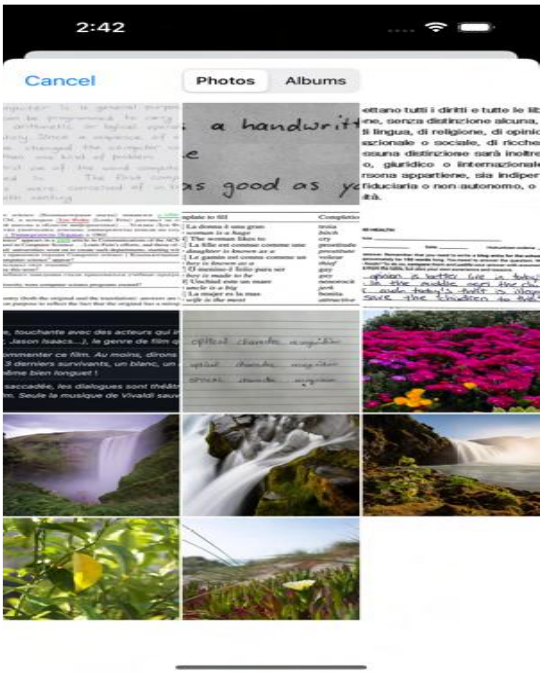

Figure 4 Image Selection

6.2.Performance Metrics

Table 2 Metric & Value

Metric	Value
Line-Level Accuracy	94.2%
Word-Level Precision	95.6%
Word-Level Recall	93.1%
Inference Time per Image	0.8 seconds (avg)
User Satisfaction Rating	4.6/5

6.3.Observations

- **Best Results:** English, Spanish, French (clean Latin scripts)
- **Most Challenging:** Arabic (ligatures) and Chinese (dense characters)
- **Influencing Factors:** Poor lighting reduced accuracy by ~6%
- Cursive handwriting was 7–10% less accurate than print
- High-resolution images improved bounding box precision

Figure 5 Before and After OCR Result

7. Case Studies

To demonstrate practical utility, the app was evaluated in three real-world case studies.

7.1. Student Use Case

A postgraduate student used the app to scan lecture notes written in English and Portuguese. The handwritten equations and bullet points were accurately recognized and converted to editable text. The student could quickly share notes in .txt format, improving revision efficiency.

7.2. Journalist Scenario

A journalist working on a multilingual story used the app to scan Arabic headlines from printed newspapers. Despite the script complexity, the OCR engine identified the text accurately, and the VoiceOver feature helped verify content before exporting.

7.3. Traveler/Multilingual User

A frequent traveller scanned hotel notes, airport instructions, and addresses written in French, German, and Russian. The app's multilingual capability allowed seamless recognition and quick translation using external apps.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0440
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2795 - 2804

7.4. Accessibility User

A visually impaired user tested the Voice Over feature. Recognized text was read aloud of the inner automatically, allowing the user to listen to handwritten feedback and notes independently.

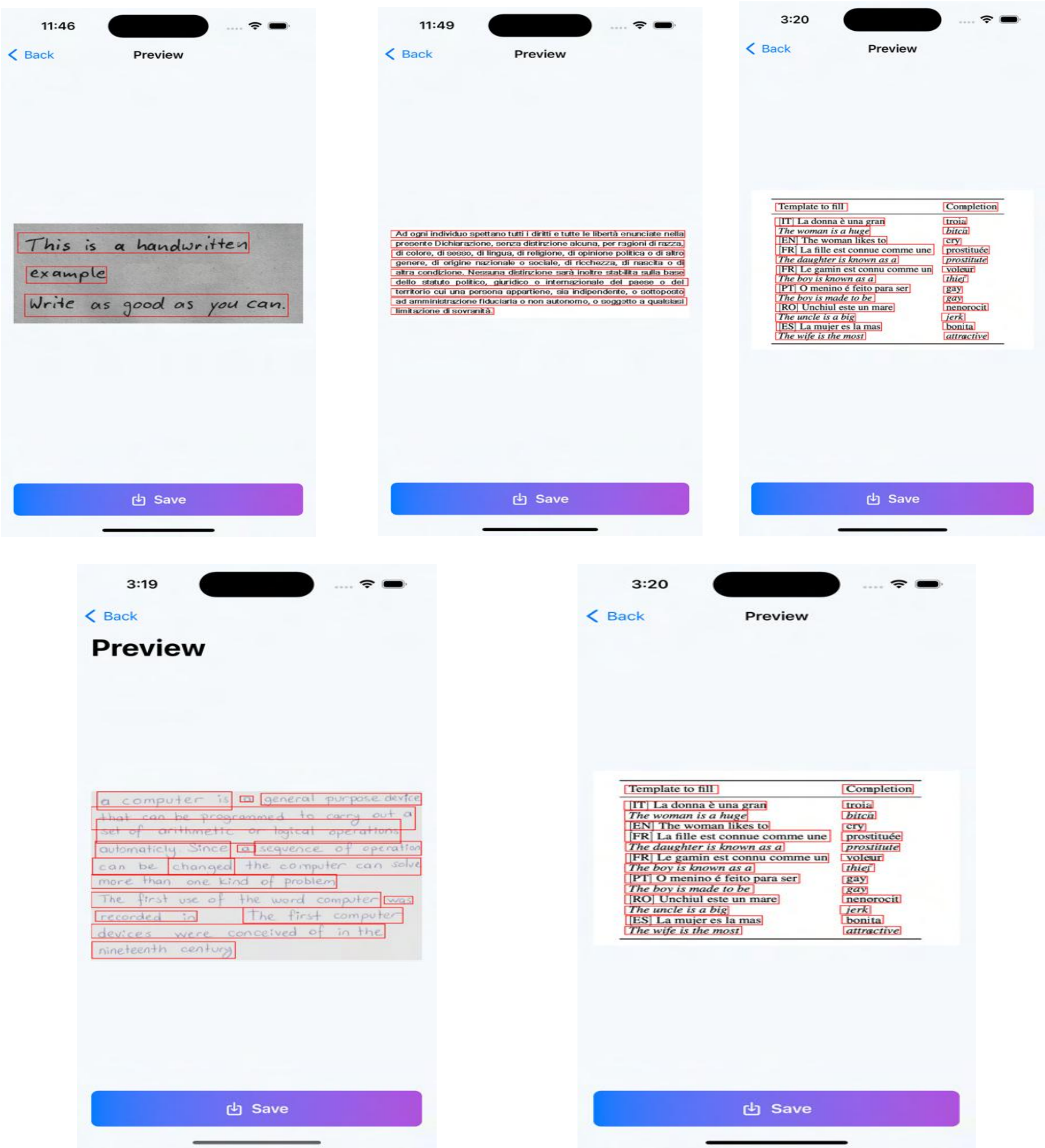

Figure 6 Multilingual User Scenario

8. Limitations & Future Work

While the system performs well across standard use cases, certain limitations exist:

8.1.Limitations

- **Low-Light Sensitivity:** Recognition accuracy drops with shadows or poor contrast. (Figure 1 & 2)
- **Noisy Backgrounds:** Complex backgrounds may confuse ROI detection.
- **Script Mixing:** Mixed-language sentences (e.g., English + Hindi) reduce accuracy.
- **Limited Indian Language Support:** Vision API does not yet support Devanagari, Tamil, Telugu, etc., for handwriting.
- **No Adaptive Learning:** The system does not personalize to user-specific handwriting styles over time. (Figure 3 & 4)
- **Post-OCR Features:** No grammar correction, table detection, or document layout restoration yet.

8.2.Future Work

- **Language Expansion:** Integrate third-party engines like Tesseract for Hindi, Bengali, Tamil, etc. (Figure 5 & 6)
- **Adaptive Learning:** Implement a feedback loop to learn from user corrections.
- **Live Translation:** Add built-in language translation after OCR.
- **Cloud Sync:** Enable syncing to iCloud, Drive, OneDrive for document archival.
- **Batch Scanning:** Add support for scanning multi-page documents and intelligent file naming.
- **Editable OCR Output:** Provide inline editing of recognized text before export.

Conclusion

This paper presented the development of a multilingual, on-device handwriting recognition app for iOS. Using Apple's Vision framework, the system achieves high OCR accuracy, fast response time, and robust multilingual support. With inclusive design via VoiceOver and offline operation, the app is suitable for a wide range of users — from students to professionals, travelers, and the visually impaired. The architecture is extensible and offers a strong foundation for future innovation. As digital transformation expands into education, healthcare, administration, and beyond, such mobile OCR solutions will play a critical role in bridging analog and digital content seamlessly.

Acknowledgements

I would like to express my heartfelt gratitude to my guide Mrs. RP Rajeshwari, for her continuous support, valuable guidance, and encouragement throughout the course of this project. Her/His insights into mobile development and machine learning greatly contributed to the successful completion of this work. I am also thankful to the faculty members of the Department of Computer Science and Engineering, Rao Bahadur Y. Mahabaleswarappa Engineering College, Ballari, for providing the necessary infrastructure and academic support. A special thanks to Apple's developer documentation and open-access research repositories, which offered rich technical resources and real-world datasets for the literature survey and implementation.

References

[1]. Smith, R. (2007). An overview of the Tesseract OCR engine. Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR), Vol. 2, pp. 629–633.
[2]. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems (NeurIPS), 30, 5998–6008.
[3]. Shi, B., Bai, X., & Yao, C. (2016). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(11), 2298–2304. doi: 10.1109/ TPAMI. 2016.2646371.
[4]. Li, X., Jin, L., & Yang, M. (2019). A survey of offline handwritten Chinese character recognition. Pattern Recognition Letters, 119, 103–113. doi: 10.1016/j.patrec.2018.11.021.
[5]. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language

understanding. arXiv preprint, arXiv:1810.04805.

[6]. Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543. doi: 10.3115/v1/D14-1162.

[7]. Apple Inc. (2023). Apple Vision Framework Developer Documentation. Retrieved from https://developer.apple.com/documentation/vision

[8]. Vasudevan, V., Ramesh, S., & Nair, R. (2021). Multilingual OCR using deep learning: A survey. International Journal of Advanced Computer Science and Applications, 12(4), 450–456. doi: 10.14569/IJACSA.2021.0120460.