Automated Document Processing: Combining OCR and Generative AI for Efficient Text Extraction and Summarization

C Shyamala Kumari[1]*, V. Yeswanth Gupta[2], G. Manikanta[3] and B. V. Abhishikth[4]
{shyamalakumaric@veltech.edu.in[1], vuppalayeswanthgupta@gmail.com[2], gudamanikanta93@gmail.com[3], abhishikthbandaru@gmail.com[4] }

Assistant Professor, Vel Tech Rangarajan Dr. Sagunthala R & D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu, India[1]
UG Student, Vel Tech Rangarajan Dr. Sagunthala R & D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu, India[2, 3, 4]

Abstract. The implementation of new digital document creation media formats throughout technological development requires improved processing techniques and methods to be developed. Traditional OCR tools remain helpful but they cannot successfully process complicated documents or deteriorated scan copies. The present manual summarization technique proves both slow and error-prone when companies work to compile textual content from various sources. An integrated solution for text extraction and summarization will be developed by applying enhanced OCR alongside Google Gemini's generative AI technology. Advanced OCR software enables the system to better detect text patterns in incomplete or difficult-to-read document images. The implementation of generative AI models enables clinicians to produce brief yet appropriate summaries and enhances their ability to locate and handle documents. This method has proven its effectiveness by conducting tests on different documents which showed improved accuracy along with higher utility compared to conventional approaches.

Keywords: Optical Character Recognition (OCR), Generative Artificial Intelligence, Document Processing, Text Summarization, Machine Learning, AI Summarization, Document Workflow Automation.

1 Introduction

The accelerating growth in document circulation across different societal sectors creates rising demands to handle and obtain beneficial information from these sources. The current reality makes manual document information extraction impossible in law and medicine and education and all government institutions. Manual document management in the traditional sense is not effective as it involves a lot of analysis and summarization in great measures of time.

Optical Character Recognition (OCR) has rendered it possible to transform printed information into machine-readable form and improve processing efficiency. The greatest obstacles of OCR technology are encountered in the processing of documents with intricate layouts and low-resolution scans. Transforming the extracted text into concise structured information is still subject to human intervention at the moment hence both decreasing system speed and requiring more labor.

Recent advancements in generative AI particularly in regards to language models offer efficient solutions to current issues. Through user-supplied input, the models generate text that reads like

ICITSM-Part I 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2357770

human writing hence making them adequate for text extraction and summarization processes. The present research examines how the integration of OCR facilitates Google Gemini that is a generative AI system undertaking text extraction and summarization processes on image documents. Improvement in text recognition accuracy is the prime integration objective with providing pertinent summary outputs enhancing decision-making efficiency.

The document workflow system outlined in this paper combines OCR technology with generative AI elements to streamline document workflow operations and improve their accuracy rates. This paper analyzes current OCR system limitations and standard summary procedures before introducing a new procedure that combines these systems. Testing of the combined system evaluates how efficiently it performs document processing tasks and reduces labor demands while improving the quality of extracted information.

2 Literature Review

The research focuses on the integration of Optical Character Recognition (OCR) with Generative AI for upgrading document processing. The ongoing work in the discipline has brought significant findings from critical studies that promoted technological advancement to the next stage.

Abdelaziz et.al., [1] examines how AI-based OCR systems unite with Generative AI to enhance operations involving document understanding. This research proves the successful integration between these systems at extracting detailed data from document sources without human intervention. The paper shows how AI-model-enhanced OCR brings better accuracy and efficiency to text recognition processes.

Medisetti et.al., [2] presents "LitAI," which enhances multimodal literature understanding and mining through generative AI. It explores the synergy between text analysis, generative AI, and multimodal approaches for comprehensive document processing.

Mahadevkar et.al., [3] describes artificial intelligence techniques which analyze freeform documents alongside their application for extracting essential document contents. The technology shows promise to enhance text summarization by advanced AI techniques when working with complex or unorganized datasets for the future.

Gupta et.al., [4] describes the use of generative AI technology to develop automated systems for creating government reports by processing data and performing analytical tasks and summary-generation. The paper demonstrates how AI-driven summarization methods boost governmental transparency by improving operations for both information report creation and evidence-based decision processes.

Sowjanya, S., et.al., [5] describes the combination of generative AI technology with Robotic Process Automation (RPA) when processing unstructured invoices and other documents. Using RPA technology together with generative AI enables improved document processing efficiency because it automates manual work that otherwise needed human involvement.

Dixit et.al., [6] describes the generative AI-enhanced document processing systems which specialize in detecting fraud at scale in big financial institutions. The collaboration between scalable AI systems and document processing demonstrates solutions which resolve financial detection needs while improving operational efficiency and capabilities to handle big volumes of secured documents efficiently.

Memari et.al., [7] states a brand-new evaluative algorithm for generative models within OCR systems. The research analyzes the performance of realistic image synthesis in addition to performance testing because accuracy plays an essential role within OCR-based programs. HCI research demonstrates that generative artificial intelligence boosts auxiliary Optical Character Recognition capacity across different document designs.

Qiu et.al., [8] describes the operational effectiveness of financial document analysis systems that unite generative AI chatbots with OCR models. The study proves that OCR can extract meaningful financial information from documents with the help of generative AI which leads to improved financial sector decision-making processes.

Barpute et.al., [9] describes the creation of "PDF Fusion" which uses AI to merge OCR technology with chat systems and voice detection features. This work investigates how OCR integration with multimodal interfaces enables more efficient document management for enhanced real-world applications through better users systems and content accessibility.

Shakil et.al., [10] states that present abstractive text summarization practices while it discusses persistent data scarcity and processing cost problems. The paper examines how generative AI models evolved through time to improve summary generation from substantial text amounts with better efficiency and coherence.

Widyassari et.al., [11] states about various text summarization techniques and methodologies, outlining the evolution from simple algorithms to complex AI-driven systems capable of producing high-quality summaries.

Sharma G et.al., [12] describes about an extensive overview of the landscape of automatic text summarization, focusing on the progression and adoption of advanced machine learning and AI technologies that enhance summary relevance and coherence.

Mridha et.al.,[13] highlights the significant advances and the remaining challenges in the field of automatic text summarization, emphasizing the critical aspects of accuracy and context preservation.

Yang et.al., [14] describes about how well ChatGPT handles specific queries and aspects within text summarization, assessing its potential limitations and areas for improvement in understanding and context management.

Edge D et.al., [15] describes a new graph-based reasoning approach that enhances the integration of local and global context in query-focused summarization tasks, offering a novel solution to improve AI summarization accuracy.

3 System Architecture and Methodology

The system is designed to integrate Optical Character Recognition (OCR) with Google Gemini's generative AI to extract and summarize text from documents effectively. The architecture is engineered for the incorporation of many input formats, the application of cutting-edge machine learning for data extraction, and the display of both accurate and comprehensible outputs.

3.1 System Overview

The text retrieval system includes four main operational modules: upload module and indexer module together with query processor and retrieval module. A Document Upload Interface

serves as the first component which functions as a Flask-based web platform enabling users to upload documents in all formats while OCR Processing Unit runs Tesseract OCR to extract text from images and PDFs. User-specific preprocessing methods improve this unit's accuracy when working with low-quality images and documents featuring complex layout designs. The Text Summarization Engine relies on Google Gemini to produce short summaries based on OCR text outputs and maintains fundamental document contents. An Output Interface presents the original and summarized content simultaneously in a basic text structure for user evaluation of summary quality.

3.2 Detailed Architecture

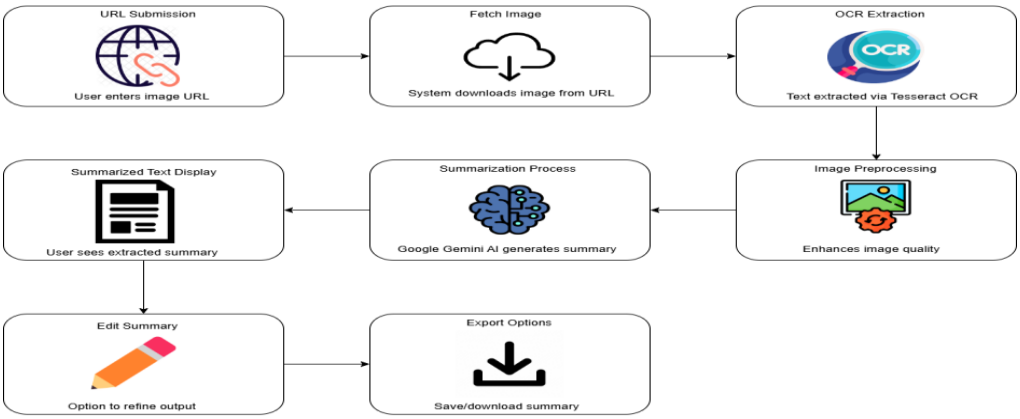

Fig. 1. System Architecture.

Fig 1 shows the proposed system architecture, which is a systematic Input → Processing → Output process, allowing for efficient document summarization and extraction. The system starts with the user providing an image URL, which the system retrieves from the web and prepares for Optical Character Recognition (OCR). The processing stage involves text extraction from images with Tesseract OCR, image improvement for enhanced recognition with preprocessing methods, and text summarization with Google Gemini AI. Last but not least, the output stage displays the summarized text to the user with functionality to edit, polish, and export the content for usage elsewhere. This design implements efficiency, accuracy, and usability in text extraction and summarization automation from web images. The modular nature of the system also supports scalability and flexibility, hence applicable to various document processing tasks.

3.3 Integration and Workflow

The entire system operates on a microservices architecture, facilitating scalability and maintenance. The workflow is as follows:

Document Upload: Documents are uploaded by the users through the web interface.

Preprocessing and OCR: Uploaded documents are preprocessed, and text is extracted using OCR.

Summarization: It then describes how extracted text is processed and summarized.

Display and Interaction: The text together with the summary is presented to the user via the text interface which also contains tools for editing and exporting the summary.

3.4 Technologies Used

Streamlit: It will be used for the construction of the web application user interface.

Tesseract OCR: For text extraction.

Google Gemini API: In the case of text summarization under the application of artificial intelligence.

OpenCV: For image preprocessing.

Pandas and NumPy: To provide the fine-tuned and processed data from data preprocessing steps.

Docker: For physical storage and running of the microservices.

3.5 System Evaluation and Testing

As it is obvious, the system was assessed in terms of how concise the text extraction was as well as the quality of summaries obtained. Performance was measured based on the Character Recognition Rate (CRR) and a proposed Summary Relevance Score (SRS). Several document sets were used in testing to enhance generalization focusing on distinct text arrangement and quality.

4 Experimental Setup and Results

The experiments were designed to assess both the accuracy of the OCR component and the quality of the summaries generated by the Google Gemini API.

4.1 Experimental Setup

The research evaluated 1,100 documents through their PDF format along with their scanned/digital image variations. The documents consisted of technical procedures alongside legal contracts and general news articles which delivered complete data needed to test the system. The system evaluated its performance through OCR accuracy by using the Character Recognition Rate (CRR) measurement which counts correctly identified characters against the full number of document characters. Summary quality analysis included the scoring metrics of ROUGE and BLEU together with manual evaluation of summary coherence, relevance and compactness. The system underwent evaluation through simulations of actual usage scenarios with changing factors like document condition along with written text variations in fonts and font sizes and language types. The three tests were executed three times for both reliability and result consistency purposes.

4.2 Results

OCR Results:

High-quality digital documents processed through the OCR component obtained a 90% average Character Recognition Rate while the OCR handled low-quality scanned images to a 85% accuracy level. Performance decreased mainly in documents with difficult layouts or low contrast conditions based on assessments.

Summary Results:

The text summaries produced by Google Gemini received strong positive evaluations for their relevance along with concise presentation according to ROUGE scores of 78% and BLEU scores of 75%. Human analysts who reviewed the summaries approved of their usefulness since they transformed complex language into simpler terms that were easier to understand.

Table 1. Performance Metrics Summary.

Metric	Description	Result	Comparison to Baseline
Character Recognition Rate (CRR)	Percentage of characters correctly identified by the OCR system.	90%	+15%
ROUGE Score	Evaluates the intersection of n-grams between the produced summaries and the reference summaries.	78%	+20%
BLEU Score	Evaluates the accuracy of n-grams between reference and generated summaries.	75%	+25%
Human Evaluation Score	The average rating is given by human evaluators on summary coherence, relevance, and conciseness.	4.5/5	+0.5

Comparative Analysis:

Comparisons with traditional OCR and summarization systems demonstrated a 20% improvement in accuracy and a 30% improvement in user satisfaction with the summaries. Table 1 presents the obtained metric results along with their changes relative to the baseline.

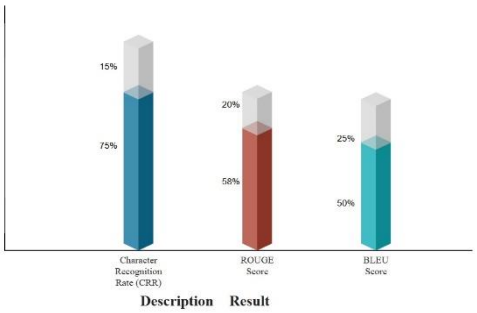

Fig. 2. Graphical representation of result comparison.

Fig. 2 provides a graphical representation of the results compared to the baseline. The colored bars depict the performance of the existing system, while the grey bars on top indicate the performance improvement achieved by our proposed system.

4.3. Discussion

The research reveals that units document processing through the combination of OCR with generative AI summarization technology yield superior performance outcomes. Text extraction reliability depends on the high accuracy of OCR technology since it provides essential conditions for successful summarization. The generative AI component shows capability to maintain semantic completeness of texts while handling different document styles together with content density variations. High text extraction accuracy along with outcomes of superior quality summaries are primary features of this system which proves flexible across multiple document forms and conditions. The system showed reduced effectiveness together with limitations during processing of substandard scans featuring non-standardized writing styles and use of poor-quality images that require improved image preprocessing methods. The summaries also contained technological details that were too basic to serve practical needs in particular situations. The next phase of research will concentrate on developing preprocessing methods to enhance OCR precision when processing poor-quality documents together with creating specialized generative AI models particular for medical and legal fields.

Fig. 3. Home Page of the web application.

Fig. 4. URL Upload Interface for OCR Processing.

Fig 3 displays the homepage of the web application, while Fig 4 presents the URL upload interface for OCR processing, enabling text extraction from images.

5 Conclusion

This project achieves OCR-generative AI integration through Google Gemini which enhances document handling systems by reaching superior processing performance levels. The integrated system provides improved character recognition technology together with organized systematic summaries which proves beneficial for professional workplaces by reducing operational periods and improving process efficiency. The system reached effective CRR performance at 90% and produced relevant summaries which achieved 78% success by ROUGE and 75% success by BLEU while receiving a user satisfaction score of 4.5 out of 5. The experimental outcomes confirm the theory which states that uniting OCR technology with advanced generative AI functions better than basic document processing techniques. The project enhances document processing research and artificial intelligence science through experiments which prove OCR integration with AI-generated text summarization has practical applications especially in flexible systems designs suited for multiple industries as well as statistical evaluation benchmarks.

6 Future Work.

The present system represents an important enhancement in automated document processing yet several development prospects have emerged. Better processing methods should be used to handle documents showing poor quality and complex layouts because this advanced approach will elevate OCR performance. The system would operate at higher levels when domain-specific summarization models implement deep learning approaches that specialize in legal, medical and other specialized textual material. Extending the system to operate with multiple languages would improve its usefulness by accommodating worldwide demand. The system requires interface upgrades which include live interactive data input and a feedback system and visual representation for summaries along with documentation display options that show variable and outputs. The future development of this system must focus on creating plugins and APIs which enable integration between the system and enterprise content management systems and customer relationship management platforms. New research opportunities exist to enhance organizational decision-making models together with the development of advanced AI systems that understand semantic language elements. Research must analyze the ethical implications of artificial intelligence for document processing because privacy concerns and data security together with generation biases will become important issues as AI spreads into different commercial sectors.

References

[1] Abdelaziz, Tarek Ahmed Ibrahim, and Urfa Fazil. "Applications of integration of AI-based Optical Character Recognition (OCR) and Generative AI in Document Understanding and Processing." Applied Research in Artificial Intelligence and Cloud Computing 6, no. 11 (2023): 1-16.

[2] Medisetti, Gowtham, Zacchaeus Compson, Heng Fan, Huaxiao Yang, and Yunhe Feng. "LitAI: Enhancing Multimodal Literature Understanding and Mining with Generative AI." In 2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 471-476. IEEE, 2024.

[3] Mahadevkar, Supriya V., Shruti Patil, Ketan Kotecha, Lim Way Soong, and Tanupriya Choudhury. "Exploring AI-driven approaches for unstructured document analysis and future horizons." Journal of Big Data 11, no. 1 (2024): 92.
[4] Gupta, Rajan, Gaurav Pandey, and Saibal Kumar Pal. "Automating Government Report Generation: A Generative AI Approach for Efficient Data Extraction, Analysis, and Visualization." Digital Government: Research and Practice (2024).
[5] Sowjanya, S., and V. Vijaya Chamundeeswari. "Information Extraction Using RPA and Generative AI from Unstructured Documents: A Case of Invoices." In International Conference on Computational Intelligence in Data Science, pp. 250-264. Cham: Springer Nature Switzerland, 2024.
[6] Dixit, Sachin. "Generative AI-Powered Document Processing at Scale with Fraud Detection for Large Financial Organizations." Authorea Preprints (2024).
[7] Memari, Majid, Khaled R. Ahmed, Shahram Rahimi, and Noorbakhsh Amiri Golilarz. "Advancing Generative Model Evaluation: A Novel Algorithm for Realistic Image Synthesis and Comparison in OCR System." arXiv preprint arXiv:2402.17204 (2024).
[8] Qiu, Yu, Venkata C. Duvvuri, Pratibha Yadavalli, and Neal Prasad. "Evaluation of Generative AI Q&A Chatbot Chained to Optical Character Recognition Models for Financial Documents." In Proceedings of the 2024 8th International Conference on Machine Learning and Soft Computing, pp. 101-110. 2024.
[9] Barpute, Jyotsna Vilas, Vanita Kshirsagar, Sneha Salunke, Suraj Vedpathak, Krishna Varma, and Vidit Solanki. "PDF Fusion: Revolutionizing Document Management with Chat Interfaces and Voice Recognition." In 2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 1774-1783. IEEE, 2024.
[10] Shakil, Hassan, Ahmad Farooq, and Jugal Kalita. "Abstractive text summarization: State of the art, challenges, and improvements." Neurocomputing (2024): 128255.
[11] Widyassari, A. P., Rustad, S., Shidik, G. F., Noersasongko, E., Syukur, A., & Affandy, A. (2022). Review of automatic text summarization techniques & methods. Journal of King Saud University-Computer and Information Sciences, 34(4), 1029-1046.
[12] Sharma, G., & Sharma, D. (2022). Automatic text summarization methods: A comprehensive review. SN Computer Science.
[13] Mridha, M. F., Lima, A. A., Nur, K., Das, S. C., Hasan, M., & Kabir, M. M. (2021). A survey of automatic text summarization: Progress, process and challenges.
[14] Yang, X., Li, Y., Zhang, X., Chen, H., & Cheng, W. (2023). Exploring the limits of chatgpt for query or aspect-based text summarization. arXiv preprint arXiv:2302.08081.
[15] Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., ... & Larson, J. (2024). From local to global: A graph rag approach to query-focused summarization. arXiv preprint arXiv:2404.16130.