Journal of Computer and Electrical Sciences (JCES)- Volume-1 Issue-4, April 2026
ISSN: 3049-2602

COGNITIVE AI-BASED MULTI-DOCUMENT TEXT SUMMARIZATION WITH VISUAL INSIGHT EXTRACTION USING LIGHTWEIGHT NLP TECHNIQUES

K. Vinay Kumar[1], K. Sivaramakrishna[2]

[1]Student, Department of Computer Science and Engineering (Artificial Intelligence & Machine Learning)

[2]Associate Professor, Department of Computer Science and Engineering (Artificial Intelligence & Machine Learning)

Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, India

Email id: vinay.nani919@gmail.com

***Abstract**: Text summarization has become an essential task in Natural Language Processing (NLP) to efficiently process and understand large volumes of textual data. This project presents a Cognitive AI-Based Text Summarization System that generates meaningful summaries from documents using intelligent text processing techniques. The system supports multiple input formats, including PDF files, images, and direct text input, making it versatile and user-friendly. It extracts textual content using PDF parsing and Optical Character Recognition (OCR) techniques and applies preprocessing steps such as cleaning and normalization to prepare the data for analysis. A frequency-based summarization approach is used to generate both page-wise summaries and a comprehensive final summary, ensuring important information is retained while reducing redundancy. In addition to text summarization, the system incorporates visual content analysis to identify and interpret figures and tables, providing meaningful insights to enhance document understanding. The implementation uses JavaScript, PDF.js, and Tesseract.js with a web-based interface, and experimental results demonstrate that the system effectively summarizes documents, handles multiple input formats, and produces accurate and concise outputs.*

***Keywords** — Text Summarization, NLP, Cognitive AI, PDF Processing, OCR, Visual Analysis, Document Summarization, Web Application*

1. INTRODUCTION

With the rapid growth of digital information, extracting meaningful insights from large textual documents has become a challenging task. Manual reading and analysis of documents such as research papers, reports, and articles is time-consuming and inefficient.

Text summarization, a key area of Natural Language Processing (NLP), aims to generate concise and meaningful summaries while preserving essential information. Traditional approaches often fail to capture context effectively or handle diverse input formats.

To address these challenges, this project proposes a **Cognitive AI-Based Text Summarization System** that supports multiple input formats such as PDF documents and images. The system integrates intelligent text processing techniques with visual content analysis to improve understanding and usability.

2. LITERATURE SURVEY

Existing research in text summarization highlights that extractive methods based on statistical techniques are simple and efficient but often fail to capture deep contextual meaning. These methods rely on frequency and sentence ranking, which may lead to loss of semantic understanding in complex documents.

https://doi.org/10.5281/zenodo.19606084

Advanced approaches using deep learning models such as LSTM and Transformer-based architectures improve summarization quality by learning contextual relationships within text. However, these models require large datasets, high computational resources, and are often difficult to interpret.

Recent systems also focus on document understanding by incorporating multimodal analysis, but most existing solutions lack the ability to handle multiple input formats such as PDFs and images effectively. Additionally, many systems do not provide insights into visual elements like figures and tables, limiting overall document comprehension.

This project addresses these limitations by implementing a hybrid approach that combines efficient frequency-based summarization with visual content analysis. It supports multi-format inputs using PDF parsing and OCR techniques while providing structured outputs such as page-wise summaries, final summaries, and visual insights. The system also emphasizes usability through an interactive web interface and improves overall document understanding.

3. PROPOSED SYSTEM

The proposed system is a **Cognitive AI-Based Text Summarization System** designed to overcome the limitations of traditional summarization approaches by integrating intelligent text processing with visual content analysis. The system combines frequency-based extractive summarization techniques with structured document analysis to generate accurate and meaningful summaries.

Unlike conventional systems that focus only on text summarization, the proposed system also analyzes visual elements such as figures and tables to enhance document understanding. It supports multiple input formats including PDF documents and images, making it more flexible and user-friendly.

The system leverages PDF parsing and Optical Character Recognition (OCR) to extract textual content from different sources. A frequency-based scoring mechanism is used to identify important sentences, ensuring efficient and interpretable summarization. Additionally, visual detection techniques are applied to extract captions and provide meaningful explanations for figures and tables.

https://doi.org/10.5281/zenodo.19606084

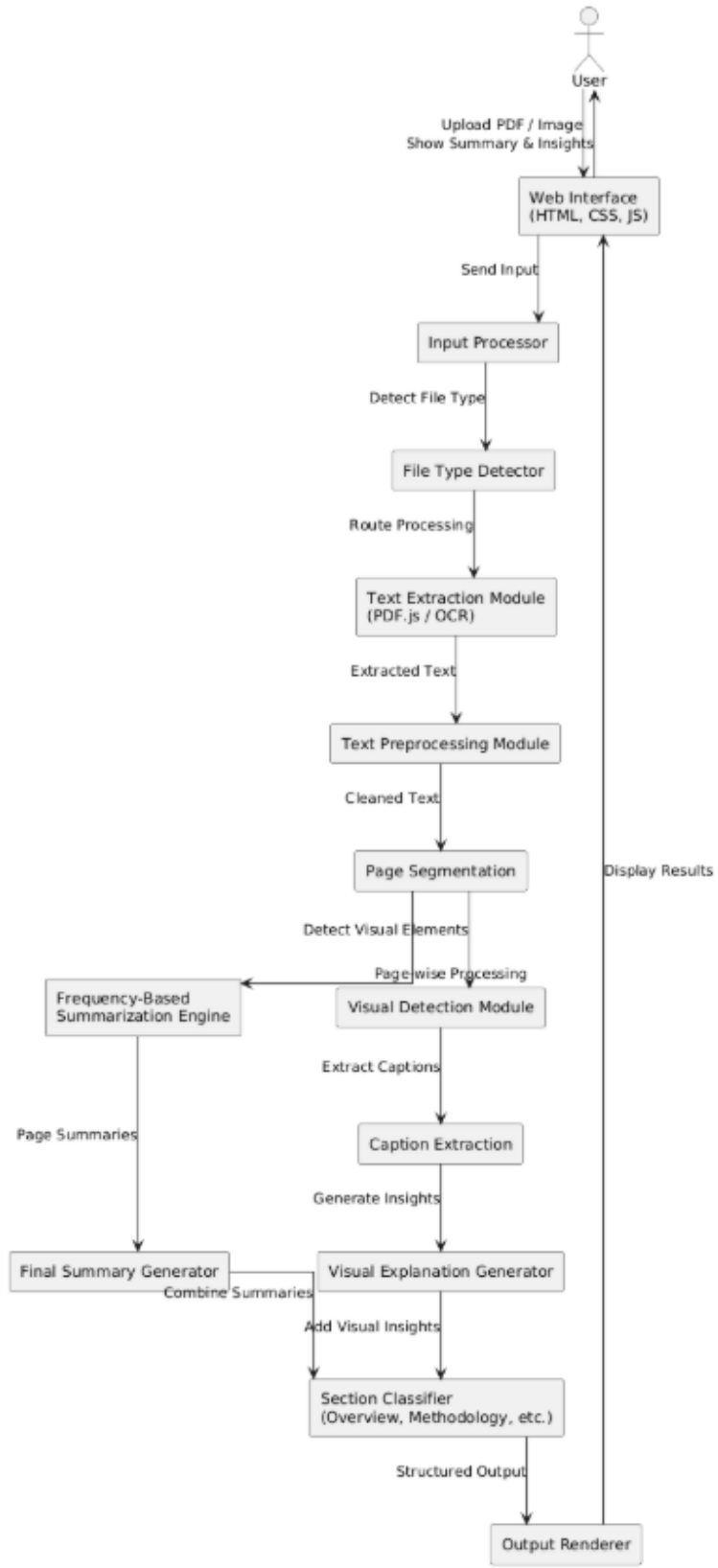

Fig. 1: Proposed System

https://doi.org/10.5281/zenodo.19606084

As shown in Fig. 1, the system follows a structured workflow that integrates text extraction, preprocessing, summarization, and visual content analysis to generate meaningful outputs.

Key Features

• **Multi-format input support:** The system accepts PDF files and images, enabling users to process various types of documents.
• **Page-wise summarization:** Generates summaries for each page individually to provide detailed insights.
• **Final cognitive summary:** Produces an overall summary by combining important information from all pages.
• **Visual content analysis:** Detects figures and tables and provides descriptive insights.
• **Text extraction using PDF.js and OCR:** Ensures accurate extraction from both digital and scanned documents.
• **Noise removal and preprocessing:** Cleans unwanted content such as repeated headers or irrelevant text.
• **Interactive web interface:** Provides a user-friendly platform for uploading files and viewing results.
• **Efficient summarization algorithm:** Uses frequency-based scoring for fast and reliable summary generation.

Working of the System

The system follows a structured pipeline:

1. The user uploads a document (PDF or image) through the web interface.
2. The system processes the input and identifies the file type.
3. Text is extracted using PDF.js for PDFs or OCR techniques for images.
4. The extracted text is cleaned by removing noise and unnecessary content.
5. The document is divided into pages and processed individually.
6. A frequency-based summarization algorithm is applied to generate page-wise summaries.
7. Visual elements such as figures and tables are detected using pattern matching techniques.
8. Captions are extracted and meaningful explanations are generated for visual content.
9. A final cognitive summary is created by combining key information from all pages.
10. The system organizes summaries into structured sections such as Overview, Methodology, Results, and Conclusion.
11. The results, including page-wise summaries, final summary, and visual insights, are displayed to the user through the interface.

4. METHODOLOGY

The methodology of the proposed system is based on an intelligent text processing approach that combines extractive summarization techniques with visual content analysis. The system processes documents through multiple stages including text extraction, preprocessing, summarization, and visual interpretation to generate meaningful outputs.

4.1 Text Extraction

The text extraction module is responsible for retrieving textual content from different input formats such as PDF files and images.

https://doi.org/10.5281/zenodo.19606084

During processing:
• PDF documents are processed using PDF.js to extract text content
• Image-based documents are processed using Optical Character Recognition (OCR)
• Extracted text is arranged in proper reading order
During execution:
• The system identifies the file type
• Appropriate extraction technique (PDF or OCR) is applied
• Text from each page is stored for further processing
Advantages:
• Supports multiple input formats
• Accurate extraction from digital and scanned documents
• Enables flexible document handling
Limitation:
• OCR accuracy depends on image quality
• Complex layouts may affect extraction order

4.2 Text Preprocessing and Summarization

This module focuses on cleaning and summarizing the extracted text using a frequency-based approach.
Working:
• Text is cleaned by removing noise such as repeated headers and irrelevant content
• Sentences are extracted and filtered based on length and relevance
• Word frequency is calculated to identify important terms
• Sentences are scored based on frequency
• Top-ranked sentences are selected as summaries
The summarization process includes:
• Page-wise summary generation
• Final summary generation by combining key content
Advantages:
• Fast and computationally efficient
• Easy to implement and interpret
• Reduces redundancy in text
Limitation:
• Does not fully capture deep semantic meaning
• Depends on frequency rather than context

4.3 Visual Content Analysis

The system includes a visual analysis module to enhance document understanding by detecting and interpreting figures and tables.
Working:
• Pattern matching techniques are used to detect figures and tables
• Captions are extracted from the document
• Descriptive explanations are generated for each visual element
• Duplicate visual detections are filtered
Features:
• Identifies figures and tables from text patterns
• Provides meaningful explanations for visuals
• Enhances overall document comprehension
Advantages:
• Improves understanding of graphical content
• Adds contextual insights beyond text
• Supports academic and technical documents

https://doi.org/10.5281/zenodo.19606084

Limitation:
• Depends on proper caption formatting
• Limited handling of complex visual layouts

4.4 Section Classification and Output Generation

This module organizes the summarized content into structured sections and presents the final output.
Working:
• The combined text is divided into sections such as Overview, Methodology, Results, and Conclusion
• Summaries are generated for each section
• Page-wise summaries and visual insights are integrated
• Results are displayed through an interactive interface
Advantages:
• Provides structured and readable output
• Improves user experience
• Enables quick understanding of documents

This methodology ensures that the system efficiently processes documents, generates accurate summaries, and enhances understanding through visual content analysis. The integration of multiple modules makes the system robust, scalable, and suitable for real-world applications.

5. RESULTS AND DISCUSSION

The proposed system was evaluated using multiple documents, including research papers, reports, and mixed-format files containing both text and visual elements. The results demonstrate the effectiveness of the system in generating accurate summaries and extracting meaningful insights.

Summarization Scenarios

• **Text-rich pages:**
Pages containing sufficient textual content produce clear and informative summaries. Important sentences are accurately identified using frequency-based scoring, resulting in high-quality page-wise summaries.

• **Moderate content pages:**
Pages with moderate text generate summaries that capture key points but may include limited contextual detail. The summaries remain relevant but slightly less comprehensive.

• **Visual-dominant pages:**
Pages with figures, tables, or diagrams and minimal text are identified as visual pages. The system generates descriptive insights based on detected captions, ensuring meaningful interpretation even when textual data is limited.

• **Noisy or irregular content:**
Pages containing unwanted elements such as repeated headers or formatting noise are cleaned during preprocessing. The system effectively removes irrelevant content and produces readable summaries.

https://doi.org/10.5281/zenodo.19606084

System Performance

The system successfully:

• Generated accurate page-wise and final summaries for multi-page documents
• Extracted text effectively from PDF files using PDF.js
• Handled visual content by detecting figures and tables and providing explanations
• Removed noise such as repeated or irrelevant text during preprocessing
• Processed multiple pages efficiently with consistent performance
• Displayed structured outputs including summaries and visual insights through an interactive interface

Discussion

The results indicate that the proposed system performs effectively across different document types and content scenarios.

• The frequency-based summarization method works well for extracting important information from text-heavy documents.
• The visual analysis module enhances document understanding by interpreting graphical elements.
• The combination of text summarization and visual insight generation improves overall output quality.

The system successfully handles various cases including text-rich pages, moderate content, visual-dominant pages, and noisy data. The structured output format further improves readability and usability, making the system suitable for academic and real-world applications.

Overall, the system demonstrates reliable performance, efficiency, and scalability in summarizing documents and providing meaningful insights.

6. CONCLUSION

The proposed **Cognitive AI-Based Text Summarization System** successfully addresses the limitations of traditional summarization techniques by integrating efficient text processing with visual content analysis. The system generates accurate, concise, and structured summaries, making it suitable for real-world document analysis applications.
The inclusion of multi-format input support enhances usability, allowing users to process PDF documents and images seamlessly. The use of PDF parsing and OCR techniques ensures effective text extraction from both digital and scanned documents.

A significant contribution of this work is the incorporation of visual content analysis. By detecting figures and tables and providing meaningful explanations, the system enhances overall document understanding beyond plain text summarization.
The frequency-based summarization approach ensures fast and reliable performance, while preprocessing techniques improve the quality of extracted content by removing noise and irrelevant data. The system also presents outputs in a structured format, including page-wise summaries, final summaries, and visual insights, improving readability and user experience.

Overall, the system demonstrates strong performance, efficiency, and scalability. It can be further extended by integrating advanced deep learning models, supporting multilingual summarization, and enabling real-time applications such as automated report analysis, research paper summarization, and intelligent document processing systems.

https://doi.org/10.5281/zenodo.19606084

REFERENCES

[1] M. Nath, L. Ethirajan, and J. Joshi, "Multilingual Infographics Generator: A Language-Agnostic Visual Summarizer," 2025 Conference on Artificial Intelligence x Multimedia (AIxMM), IEEE, 2025, doi: 10.1109/AIxMM62960.2025.00009.

[2] M. R. Costa-jussà et al., "No Language Left Behind: Scaling Human-Centered Machine Translation," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 31, pp. 1234–1248, 2023.

[3] Y. Chang et al., "A Survey on Evaluation of Large Language Models," ACM Transactions on Intelligent Systems and Technology, vol. 15, no. 3, 2024.

[4] L. Hu et al., "A Survey of Knowledge-Enhanced Pre-Trained Language Models," IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 4, pp. 1413–1430, 2024.

[5] C. Mi and S. Xie, "Language Relatedness Evaluation for Multilingual Neural Machine Translation," Neurocomputing, vol. 570, 2024.

[6] A. Vaswani et al., "Attention Is All You Need," IEEE Conference on Neural Information Processing Systems (NeurIPS), 2017.

[7] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," NAACL-HLT, 2019.

[8] J. Achiam et al., "GPT-4 Technical Report," IEEE/ArXiv Preprint, 2023.

[9] A. Q. Jiang et al., "Mistral 7B: Efficient Large Language Models," 2023.

[10] H. Touvron et al., "LLaMA: Open and Efficient Foundation Language Models," 2023.

[11] R. Anil et al., "Gemini: A Family of Highly Capable Multimodal Models," 2023.

[12] K. Papineni et al., "BLEU: A Method for Automatic Evaluation of Machine Translation," ACL, 2002.

[13] Google, "PDF.js: A General-Purpose Web-Based PDF Processing Framework," 2025. [Online]. Available: https://mozilla.github.io/pdf.js/

[14] Tesseract OCR, "Tesseract.js: JavaScript OCR Engine for Multi-Language Text Recognition," 2025. [Online]. Available: https://tesseract.projectnaptha.com/

[15] D. Jurafsky and J. H. Martin, "Speech and Language Processing," 3rd ed., Pearson, 2023.

https://doi.org/10.5281/zenodo.19606084