ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun156

Online PDF to Text and Audio Converter and Language Translator Using Python

Ritika Dhole[1]; Meghana Singh[2]; Vedantika Dhumal[3]; Megha Dhotay[4]

[1, 2, 3, 4] Department of Computer Science and Engineering MIT World Peace University Pune, India

Publication Date: 2025/07/14

Abstract: "Python" aims to simplify document processing by offering an all-in-one solution for text extraction, audio conversion, and language translation. Users can upload PDF files to extract editable text, which can then be converted into audio using text-to-speech functionality, making the platform highly accessible, particularly for visually impaired individuals.

In addition, the system provides multilingual support, enabling users to translate extracted text into multiple languages for wider usability. Developed using Python, the project utilizes libraries such as PyPDF2 (Python PDF Toolkit 2) for text extraction, gTTS (Google Text-to-Speech) for audio generation, and Google Translate API for translations. This tool is designed to be user-friendly, accurate, and efficient, catering to the needs of students, researchers, and professionals, while promoting inclusivity and enhancing productivity.

Keywords: *Document Processing, Text Extraction, Audio Conversion, Language Translation, Text-to-Speech.*

How to Cite: Ritika Dhole; Meghana Singh; Vedantika Dhumal; Megha Dhotay (2025). Online PDF to Text and Audio Converter and Language Translator Using Python. *International Journal of Innovative Science and Research Technology*, (RISEM–2025), 17-24. https://doi.org/10.38124/ijisrt/25jun156

I. INTRODUCTION

In an increasingly digital world, information is more accessible than ever, yet new challenges have emerged, particularly around how we consume this information. Digital documents have become an essential part of professional, academic, and personal settings, with PDF (Portable Document Format) files being the standard format for sharing information due to their ability to preserve layout and formatting across different devices and platforms. PDFs are commonly used for official documents, research papers, eBooks, manuals, and other content that needs to maintain its original structure.

However, despite their benefits, PDFs are not easily accessible to everyone, particularly those with visual impairments or learning disabilities. According to the World Health Organization (WHO), approximately 2.2 billion people worldwide have some form of visual impairment. For these users, reading text on a screen can be challenging, often limiting their access to critical information and hindering their ability to engage fully in various contexts. Another critical demographic that stands to benefit from this project includes individuals with learning disabilities, such as dyslexia. Dyslexia affect around 10% of the global population and can make reading a laborious and frustrating process. By transforming text into an auditory format, this project makes digital information more accessible to people with reading challenges.

Reading digital documents on screens is not just challenging for people with disabilities; it can be an exhausting task even for the average user. Reading large amounts of text on a screen can lead to eye strain, fatigue, and discomfort. Additionally, in today’s fast-paced world, people often seek ways to multitask and make better use of their time. For instance, someone may want to listen to a document while commuting, doing household chores, or exercising, making the traditional format of digital reading impractical in such contexts. This project addresses these challenges by proposing an Online PDF to Text and Audio Converter and Language Translator, which allows users to listen to the contents of a PDF rather than reading it. This solution not only eases the strain of digital reading but also supports people with disabilities and those with busy lifestyles who need flexible ways to consume information. The proposed system integrates multiple technologies to provide a user-friendly and efficient solution. Using libraries like PyPDF2 (Python PDF Toolkit 2), the system extracts text from PDFs, which is then converted to speech using gTTS (Google Text-to-Speech) for enhanced accessibility, especially for visually impaired individuals. Additionally, the text can be translated into various languages using the Google Translate API, catering to a diverse audience.

II. METHODOLOGY

The Online PDF to Text & Audio Converter & Language Translator system employs a combination of advanced Python libraries to deliver an efficient, accessible, and inclusive document processing solution. Document processing involves extracting, transforming, and converting data from digital files into accessible formats. Python, as a versatile programming language, provides several powerful libraries to handle these tasks efficiently. The system leverages PyPDF2, a widely used library for reading and manipulating PDF files, to enable accurate text extraction. The extracted text then serves as the foundation for further processing. To enhance accessibility, the system utilizes gTTS, which converts extracted text into natural-sounding speech. This feature is particularly useful for visually impaired individuals, allowing them to consume digital content audibly.

Additionally, it provides convenience for users who prefer a hands-free mode of interaction with their documents. To break language barriers, the system integrates the Google Translate API, which enables multilingual translation of extracted text with high accuracy. Supporting a wide range of languages, this feature is especially beneficial for students, researchers, and professionals working with multilingual documents in academic and corporate environments. A user-friendly interface ties these functionalities together, ensuring that individuals with limited technical expertise can navigate the platform effortlessly. Moreover, automation plays a crucial role in enhancing efficiency, enabling students, researchers, and professionals to handle large volumes of documents quickly and accurately.

By integrating advanced text extraction, text-to-speech conversion, and multilingual translation, the system offers a highly efficient and accessible document processing solution. It ensures inclusivity by catering to visually impaired individuals, breaking language barriers, and providing a user-friendly experience for non-technical users. By emphasizing accessibility, accuracy, and user-friendliness, the system ensures that it meets the diverse needs of its users, empowering them to interact with digital documents in new and meaningful ways.

III. LITERATURE REVIEW

Below we explore existing systems and techniques for converting documents (such as PDFs and images) into text and audio formats, alongside language translation methods. By reviewing advances in Optical Character Recognition (OCR), text-to-speech systems, and multilingual translation techniques, this survey establishes a foundation for developing a versatile and efficient PDF-to-audio and language translation tool using Python.

Fuad Rahman and Hassan Alam’s [1] study, introduced a novel approach to converting PDF documents into HTML, using Document Image Analysis and the "White Space Rectangle" (WSR) algorithm. This method was designed to preserve both the logical structure and physical layout of PDF documents, addressing challenges related to format conversion. The system's adaptability across different platforms showed promise, but it had limitations. The lack of standardized datasets for benchmarking made performance evaluation difficult, and the algorithm struggled with the proper rendering of complex table structures. Despite these challenges, the study contributed valuable insights into document format conversion and the need for improved handling of tables and layout complexities.

Yue Lu, Li Zhang, and Chew Lim Tan’s [2] paper, focused on improving document retrieval from digital libraries using a technique called Word Image Coding, specifically with Left-to-Right Primitive String (LRPS). This approach was effective in handling noisy datasets and ensuring accurate retrieval, offering improvements over traditional document search techniques. However, the system was computationally intensive, requiring significant processing power. Moreover, it lacked automatic keyword extraction, which could have streamlined the retrieval process and improved efficiency. Despite these limitations, their work paved the way for advancements in document search and retrieval, especially in environments with noisy or fragmented data.

Pankaj Kumar and Sheetal Srivastava [3] presented a syntax-directed translation tool in their paper. They developed a syntax-directed translation tool that utilized Deterministic Finite Automata (DFA) for validating syntax and automating the translation process from English to Hindi. The system’s focus on language translation for rural users was a key strength, promoting accessibility. However, its reliance on predefined translation examples limited its flexibility in handling complex sentences or context-based translations. While the system proved useful in simpler contexts, its inability to dynamically handle more complex linguistic structures highlighted the need for more adaptable translation models.

K. Ragavi, Priyanka Radja, and S. Chithra [4] developed a portable and user-friendly system for visually impaired users by integrating Tesseract OCR and Android’s Text-to-Speech API, aiming to convert text into speech. The system was designed to be affordable and efficient for quick text processing. However, challenges arose with non-standard text formats, and it required external Bluetooth modules in some setups, limiting its versatility. Despite these issues, the study emphasized the significance of creating accessible technologies for the visually impaired. Similarly, Ramakrishna Oruganti’s [5] study combined Tesseract OCR, PyPDF2, and machine learning to convert image-based and PDF documents into editable text, facilitating document digitization. The system supported various file types but struggled with handwritten text recognition and depended on the quality of scans. The study highlighted the need for more advanced techniques to enhance OCR accuracy for handwritten inputs.

Exploring mobile translation tools. Sim Liew Fong, Abdelrahman Osman Elfaki, Md Gapar bin Md Johar, Kevin Loo Teow Aik’s [6] paper, presented a mobile language translation tool designed for translating between English, Bahasa Malaysia, and Bahasa Indonesia, using MIDP and

Object-Oriented Analysis and Design (OOAD). The system was particularly useful for travelers due to its simplicity and efficiency, offering on-the-go translation. However, the tool was constrained by the small screen size of mobile devices, limiting its usability for more complex tasks. Additionally, the system supported only three languages, which reduced its applicability in broader contexts. Despite these limitations, it provided a practical solution for language translation in mobile environments, demonstrating the potential of mobile technology in the translation space.

In the realm of e-learning, Kawal Gill, Rekha Sharma, and Renu Gupta's [7] study, addressed the integration of various assistive tools such as screen readers, audiobooks, and Braille books in the e-learning environment for visually impaired students in higher education. The study emphasized that while assistive technologies have the potential to greatly improve accessibility, their adoption in educational settings faces significant barriers, particularly in terms of affordability and user training. The research highlights the need for greater awareness and more affordable solutions to support visually impaired students in educational environments.

Further, Kevin J. Shannon's [8] paper, explored the implementation of a system that used natural language processing (NLP) to generate structured SQL queries, allowing users to interact with databases using natural language input. While the system simplified query generation, it was limited to basic SQL operations and lacked advanced AI capabilities. The paper suggested that while NLP could greatly enhance user-friendliness, the system's inability to handle complex queries or more sophisticated database interactions demonstrated the need for further advancements in AI and NLP techniques. This research was foundational in understanding how NLP could be used to improve database interactions but also highlighted the challenges of scaling such systems for more complex tasks.

Satoshi Nakamura's [9] paper focuses on translating between English and Asian languages using corpus-based machine translation techniques such as example-based MT and stochastic MT. However, the system faces challenges due to the limited availability of large bilingual spoken language corpora, affecting its ability to translate diverse expressions with high accuracy.

The study on an Android-based language translator application by Roseline Ogundokun and Joseph Awotunde [10], proposed a mobile solution for real-time language translation using Google's translation API and natural language processing with Java. It aimed to bridge communication gaps by translating between major global languages such as English, Spanish, Arabic, Hindi, French, and Chinese, making it particularly useful for tourists and learners. The application leveraged machine translation (MT) techniques, shifting from rule-based to corpus-based methods for better accuracy. Despite its advantages, the system faced challenges with maintaining translation accuracy, handling complex linguistic structures, and ensuring semantic consistency across languages. The research highlighted the growing role of mobile technology in overcoming language barriers and enhancing global communication through AI-driven translation tools.

Yue Lu and Chew Lim Tan [11] introduced an advanced document image retrieval method using partial word image matching to enhance word spotting and similarity measurement. Their approach represents word images as primitive strings and employs inexact string matching to compare them, allowing efficient retrieval despite font variations and touching characters. This method bypasses OCR, addressing challenges in document image databases where text indexing is often absent. However, the technique still depends on accurate word segmentation and does not entirely replace OCR for complex layouts. The study demonstrated improved retrieval performance, showing promise for large-scale document image searches.

Deliang Jiang and Xiaohu Yang [12] proposed a method for converting PDF documents into HTML while maintaining the original layout. Their approach utilized the PDFBox Java library to extract text and graphical data, enabling structured content conversion. The method identified text segments using a refined vertical gap detection algorithm, ensuring accuracy in multi-column PDFs. However, the system faced challenges in handling complex layouts and non-standard formatting, requiring further improvements in segment detection and layout preservation techniques. Their study highlighted the importance of precise text extraction for effective document conversion.

Md. Rafiqul Islam, Ram Shanker Saha, Ashif Rubayat Hossain's [13] study presented a Bangla PDF to speech synthesizer using a rule-based concatenative synthesis method to generate natural speech from Bangla text. The system operates in two phases: first, converting PDF text to Unicode, followed by the transformation of Unicode text into speech using normalization and parsing rules. The approach addresses unique challenges in Bangla pronunciation, such as phonetic variations and short forms, and applies specific normalization rules to produce accurate speech. However, the paper highlights that while the method improved the efficiency of Bangla text-to-speech conversion, there is potential for further enhancement in accuracy and naturalness.

Lastly, Maganti Venkatesh, S. V. Chiranjeevi, M. Siva Kumar, S. Shiek Alam, Ganesh Davanam & Sunil Kumar Malchi's [14] study, presented a multilingual OCR algorithm aimed at converting text from images and PDFs, integrated with text preprocessing and Text-to-Speech (TTS) models. This approach provided multilingual accessibility, supporting a broad range of languages. However, the system faced performance issues when processing low-quality inputs, and the integration of complex techniques made it resource-intensive. Despite these drawbacks, the study demonstrated the potential of multilingual OCR in expanding accessibility, particularly in environments with diverse linguistic needs. The research emphasized the importance of improving OCR accuracy and optimizing performance to handle low-quality documents.

Collectively, these studies reveal a comprehensive landscape of technologies and approaches essential for developing an effective Online PDF to Text and Audio Converter and Language Translator. These studies collectively highlight the fragmented nature of existing solutions, emphasizing the need for a unified tool that integrates PDF-to-text conversion, multilingual translation, and TTS functionalities into a seamless system.

IV. PROPOSED SYSTEM

This project, the Online PDF to Text and Audio Converter and Language Translator, comprises several modules as shown in Fig. 1, each designed to serve a distinct function and contribute to the overall user experience. Below is a detailed description of each module:

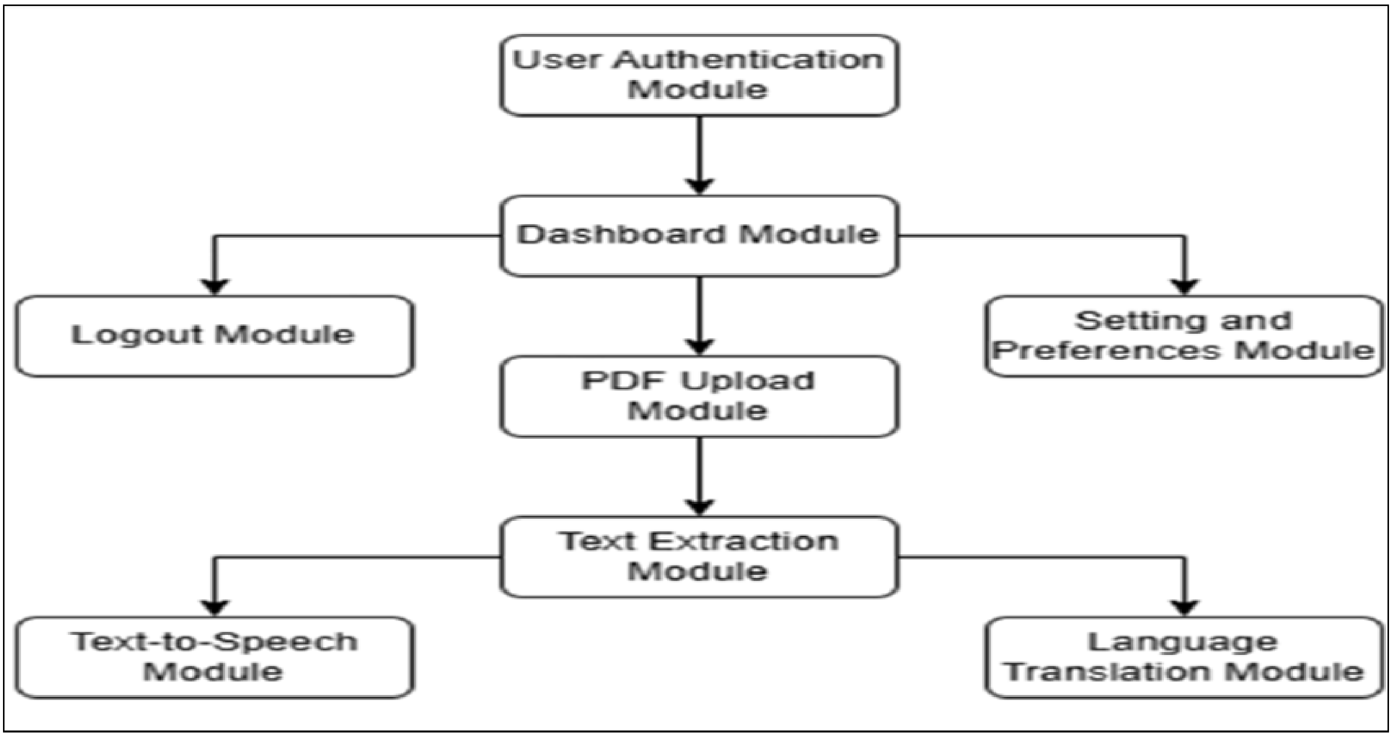


Fig 1 Layered Architecture Model

- *User Authentication Module*

It controls the application's security and user access. By authenticating users through the signup and login panels and logging them out as necessary, this module manages access to the remainder of the application.

- *Dashboard Module*

It provides a summary of the features and settings in the program and acts as the primary landing page following login. The dashboard allows users to access other modules.

- *Settings and Preferences Module*

Users can alter and control the application's settings with it. Settings can be changed by users, and they are saved and used in subsequent sessions.

- *PDF Upload Module*

Users can upload PDF files for processing. Depending on how the program is configured, uploaded PDFs are either permanently or temporarily kept and are available for additional processing.

- *Text Extraction Module*

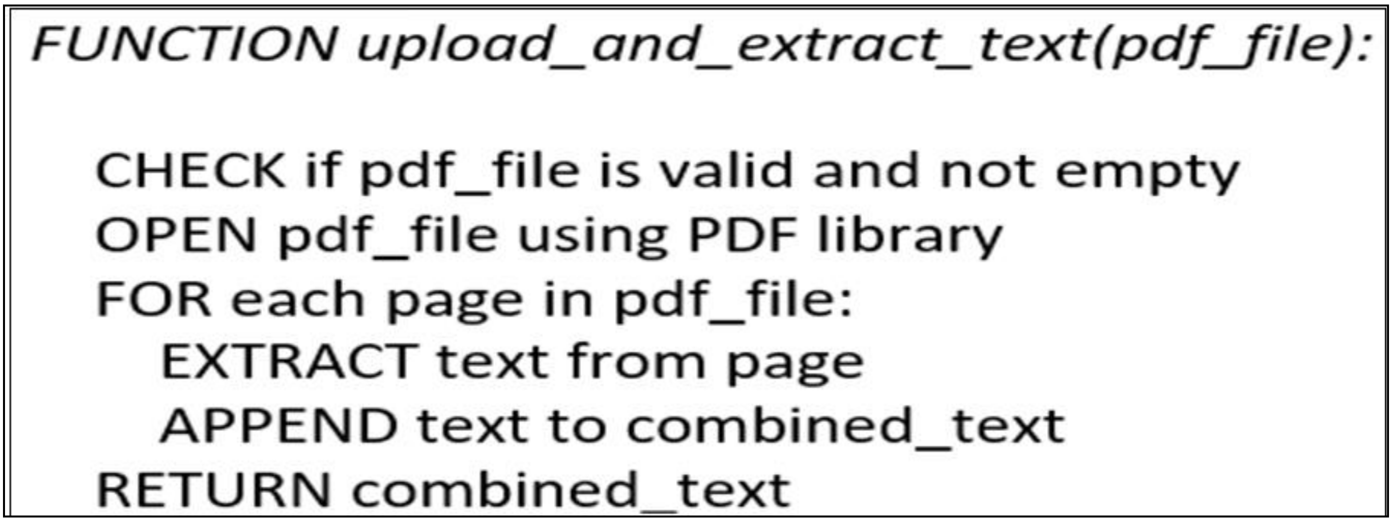
Fig. 2 presents the pseudocode representing the workflow for extracting text from PDF files using python libraries like PyPDF2, PyMuPDF, pdfplumber and pdfMiner, which automates the retrieval of textual data. It extracts text content from uploaded PDF files. Also, the project has further used Tesseract OCR for image-based text recognition. After extraction is finished, the text data is routed to additional modules for text processing, such as the Text-to-Speech or Language Translation Modules.

- *Language Translation Module*

It converts text that has been extracted across languages. extracts text, integrates Google Translate API followed with mBART to translate it according to user preferences, and then makes the content available for display or additional processing. The logic followed during this process is illustrated in the pseudocode (as shown in Fig. 3).

- *Text-to-Speech (TTS) Module*

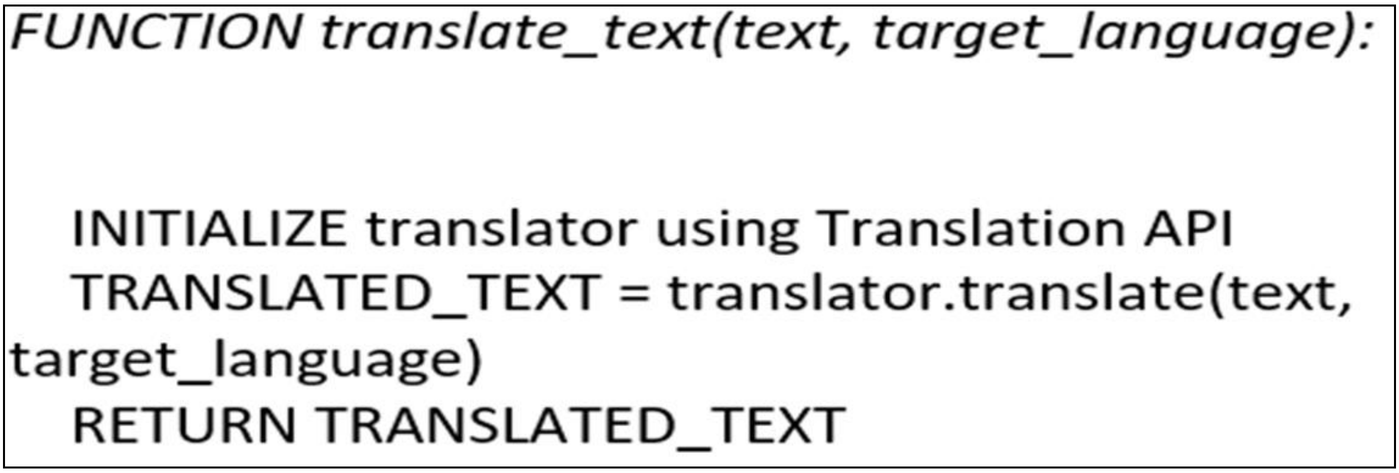
It transforms written material into audio and produces an audio output for accessibility by turning the original or translated text to voice (refer to Fig. 4 for pseudocode logic) utilizing gTTS, pydub and ffmpeg.



```
FUNCTION upload_and_extract_text(pdf_file):

    CHECK if pdf_file is valid and not empty
    OPEN pdf_file using PDF library
    FOR each page in pdf_file:
        EXTRACT text from page
        APPEND text to combined_text
    RETURN combined_text
```

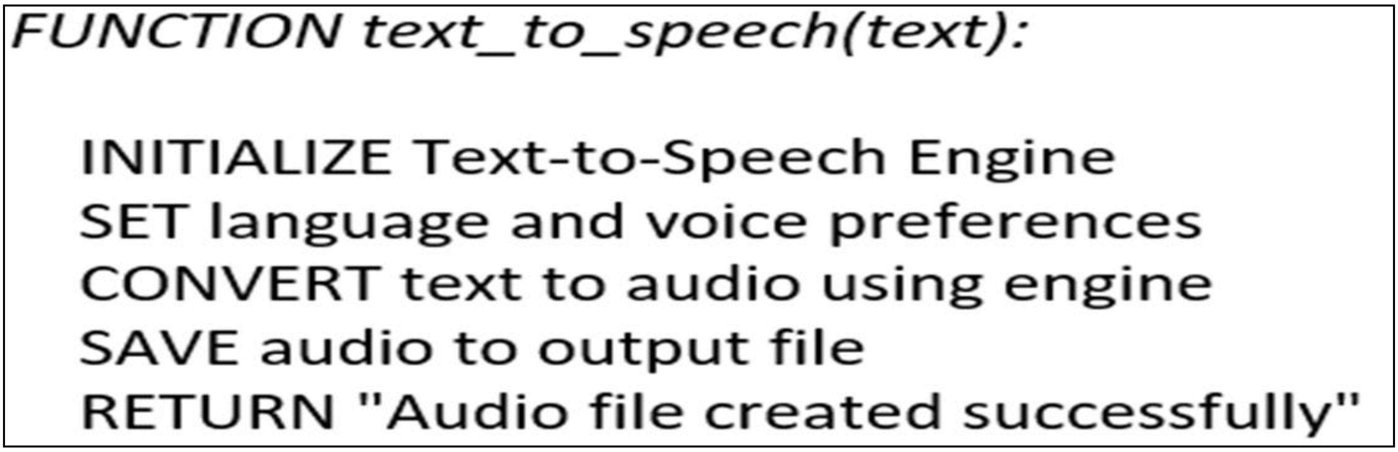
Fig 2 Pseudocode for PDF Text Extraction.



```
FUNCTION translate_text(text, target_language):


    INITIALIZE translator using Translation API
    TRANSLATED_TEXT = translator.translate(text,
target_language)
    RETURN TRANSLATED_TEXT
```

Fig 3 Pseudocode for Language Translation.



```
FUNCTION text_to_speech(text):

    INITIALIZE Text-to-Speech Engine
    SET language and voice preferences
    CONVERT text to audio using engine
    SAVE audio to output file
    RETURN "Audio file created successfully"
```

Fig 4 Pseudocode for Audio Conversion.

It begins with the user launching the application and registering or logging in as shown in Fig. 5. The user is taken to the dashboard after successful authentication, where they can access settings and choices or upload a PDF. The process of processing uploaded PDFs to extract text automates the retrieval of textual material, saving time and decreasing manual labor. The extracted text is then converted into speech, providing a different means of consuming digital content. This feature enhances accessibility for visually impaired users and offers convenience for individuals who favor listening over reading. Additionally, it allows users to translate extracted text into multiple languages, ensuring that information is not hindered by linguistic barriers. In professional and academic settings, where documents frequently need to be accessed in multiple languages, this translation feature is extremely helpful. Users can change their preferences or log out to end their session in the settings section. This streamlined process ensures ease of use for document conversion and translation tasks.

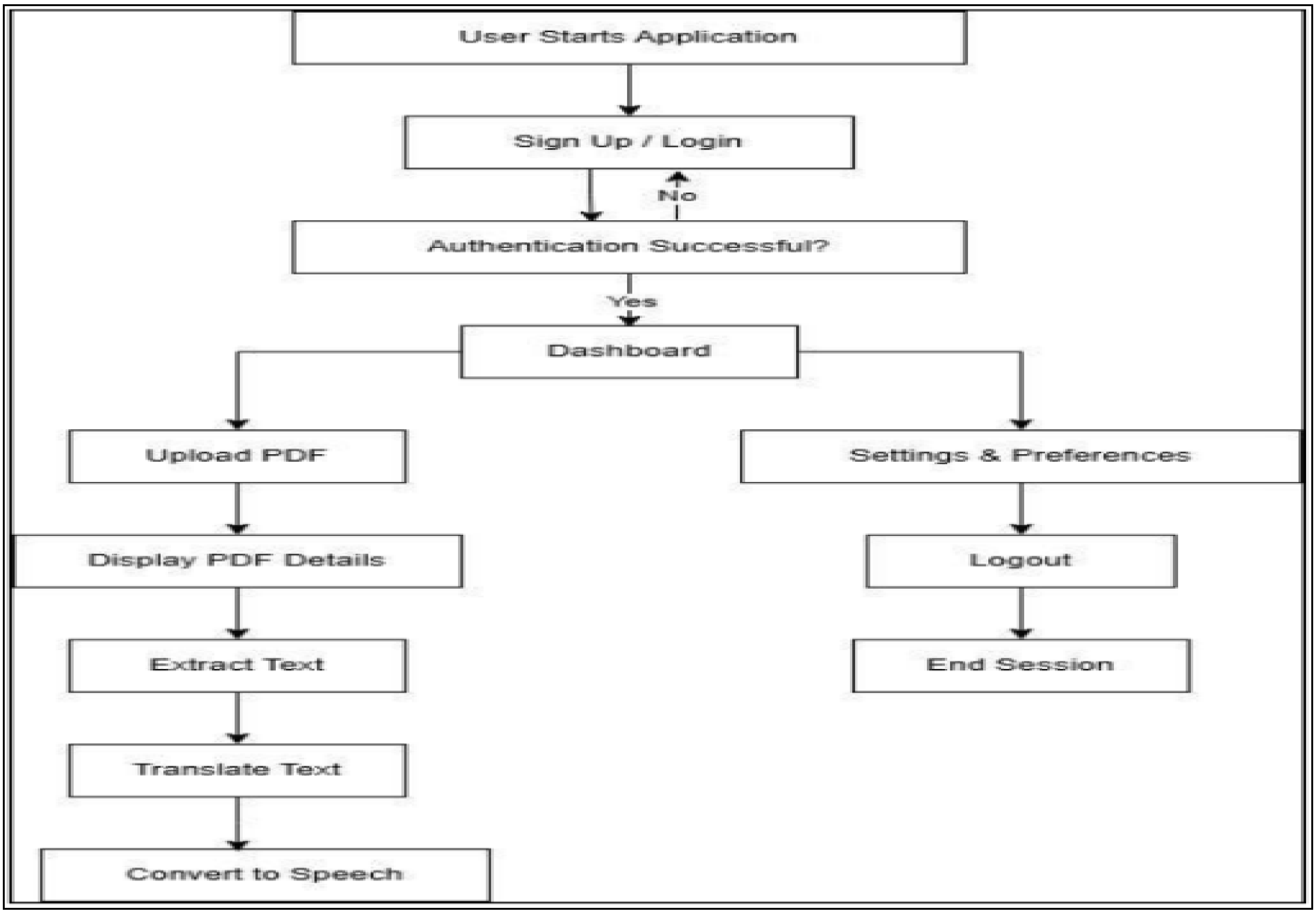


Fig 5 System Flow Diagram

V. RESULT ANALYSIS

These were the following parameters based on which we checked our efficiency of our project (ref. Table. 1)-

In terms of the processing time, the results indicate that as file size increases, the processing time also rises as shown in Fig. 6. This trend is consistent across all tested languages, highlighting the need for optimization in handling larger PDF files efficiently.

In terms of the performance of translation and speech conversion for various languages, the findings reveal high accuracy for almost all tested languages, including English, Hindi, and Marathi. This demonstrates the system's robust multilingual support and effective translation performance across different linguistic structures. (Fig. 7)

The error rate analysis, as depicted in Fig. 8, provides insights into the system's performance under varying document complexities. Key observations include:

- Minimal errors (<2%) for well-formatted PDFs, indicating strong reliability for standard document structures.
- Slightly higher error rates (~3%) for PDFs with complex layouts, such as multi-column formats, special characters, or embedded mathematical equations.

Table 1 Proposed vs Existing Models

Category	Proposed Model	Existing Models
Processing Time (Average for all PDF sizes)	87%	~93%
Success Rate (Across languages)	97%	~96%
Error Rate (Overall Average)	1.60%	~2%

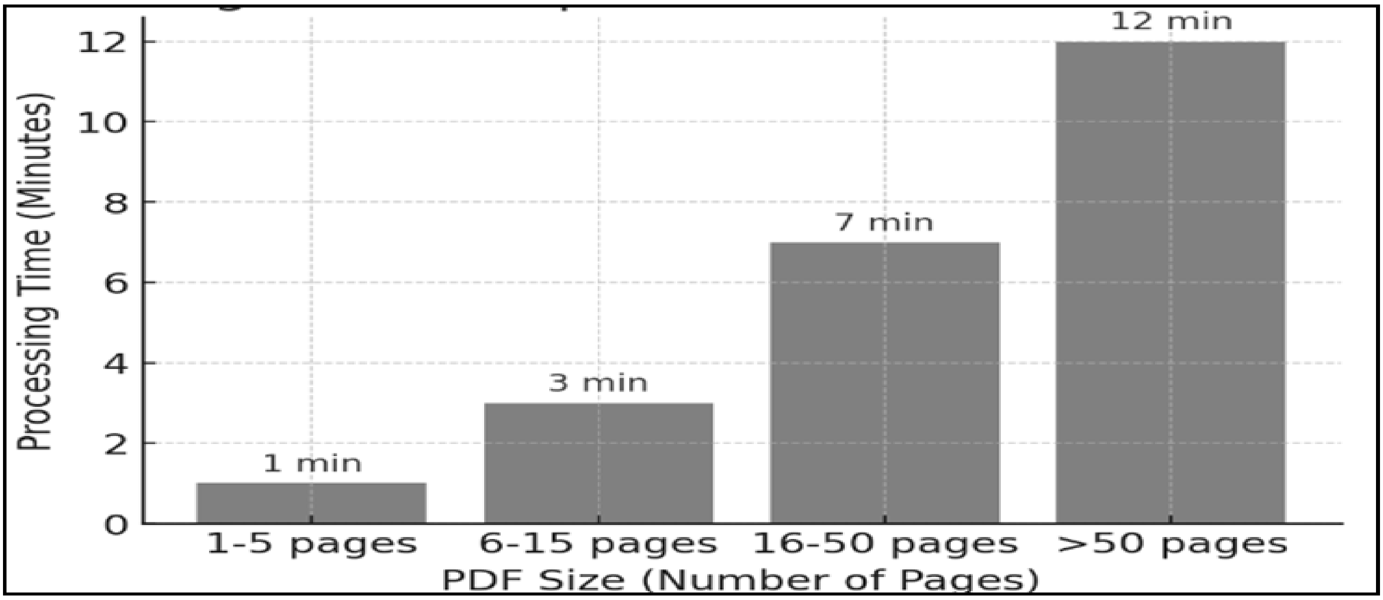


Fig 6 Processing Time v/s File Size.

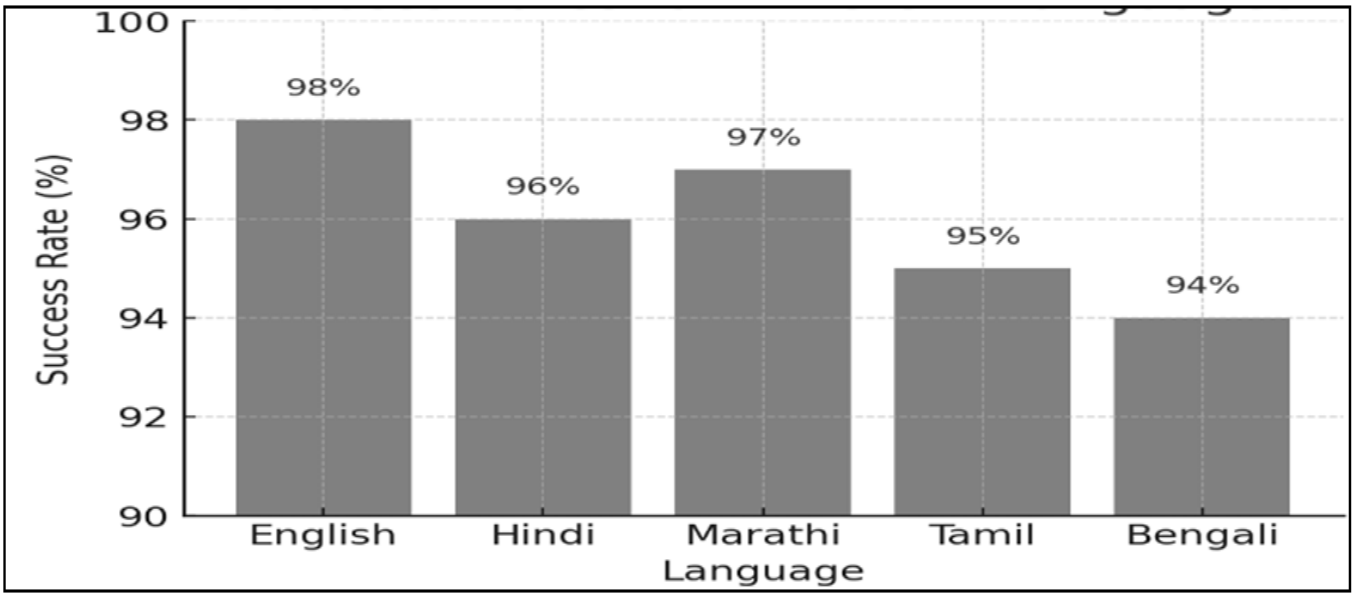


Fig 7 Success Rate v/s Languages.

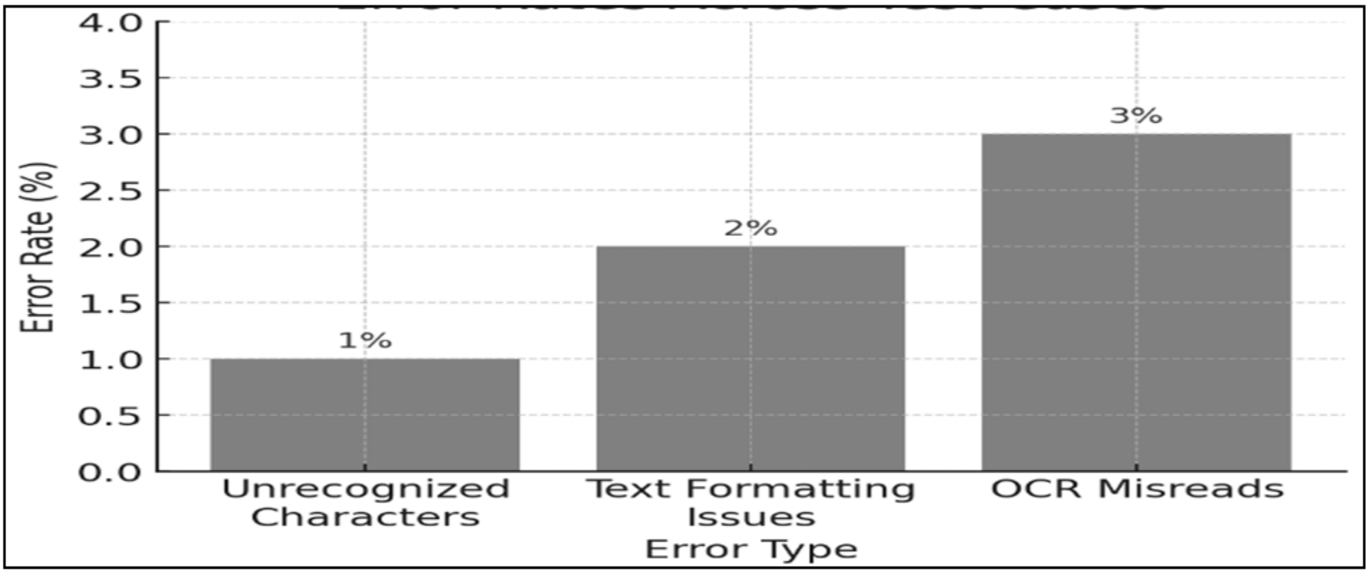


Fig 8 Error Rate v/s Error Type.

VI. CONCLUSION

The development of the "Online PDF to Text and Audio Converter and Language Translator" addresses a critical gap in accessibility for digital documents by integrating text extraction, multilingual translation, and text-to-speech functionalities into a single platform. This tool enhances the accessibility of PDF content for visually impaired users, individuals with reading disabilities, and those navigating language barriers, thereby promoting inclusivity and usability.

Through the implementation of Python-based technologies like PyPDF2, Google Translate API, and gTTS, the system demonstrated efficient performance in text conversion, accurate translation into multiple languages, and high-quality audio generation. Its modular structure ensures scalability and adaptability, making it a practical solution for a wide range of users.

Despite its success, the system has some limitations, such as challenges in handling complex PDF layouts and low-quality scanned images. Future work could focus on optimizing processing time, especially for larger PDFs, to make the system faster and more efficient. Integrating advanced OCR technologies will improve the handling of scanned PDFs, addressing potential misreads and inaccuracies in the text. There is also an opportunity to improve the translation of mathematical equations, symbols, and special characters to better support educational or technical documents. Additionally, enhancing the user interface could further improve user experience.

In conclusion, this project not only contributes to bridging the gaps in document accessibility but also paves the way for future advancements in the integration of natural language processing, machine learning, and accessibility technologies. By making digital content more inclusive and easier to consume, this research demonstrates the transformative potential of technology in addressing modern accessibility challenges.

REFERENCES

[1]. F. Rahman and H. Alam, "Conversion of PDF documents into HTML: a case study of document image analysis," The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, Pacific Grove, CA, USA, 2003, pp. 87-91 Vol.1.

[2]. Yue Lu, Li Zhang and C. L. Tan, "Retrieving imaged documents in digital libraries based on word image coding," First International Workshop on Document Image Analysis for Libraries, 2004. Proceedings., Palo Alto, CA, USA, 2004, pp. 174-187.

[3]. P. Kumar, S. Srivastava and M. Joshi, "Syntax directed translator for English to Hindi language," 2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, India, 2015, pp. 455-459.

[4]. Ragavi, K., Radja, P., Chithra, S. (2016). Portable Text to Speech Converter for the Visually Impaired. In: Suresh, L., Panigrahi, B. (eds) Proceedings of the International Conference on Soft Computing Systems. Advances in Intelligent Systems and Computing, vol 397. Springer, New Delhi.

[5]. Ramakrishna Oruganti, "Transcriber: An Image and PDF to Text Converter - CORE," unpublished.

[6]. S. L. Fong, A. O. Elfaki, M. G. bin Md Johar and K. L. T. Aik, "Mobile language translator," 2011 Malaysian Conference in Software Engineering, Johor Bahru, Malaysia, 2011, pp. 495-500.

[7]. Kawal Gill, Rekha Sharma, Renu Gupta, "Empowering Visually Impaired Students through E-Learning at Higher Education Problems and Solutions," IOSR Journal Of Humanities And Social Science (IOSR-JHSS) Volume 22, Issue 8, Ver. 7 (August. 2017) PP 27-35 e-ISSN: 2279-0837, p-ISSN: 2279-0845.

[8]. Kevin J. Shannon, "Implementing a Natural Language to Structured Query Language Translator - CORE Reader," unpublished.

[9]. S. Nakamura et al., "The ATR Multilingual Speech-to-Speech Translation System," in IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no. 2, pp. 365-376, March 2006

[10]. Roseline Oluwaseun Ogundokun, Joseph Awotunde, "An android based language translator application," 2021, J. Phys.: Conf. Ser. 1767 012032

[11]. Yje Lu and Chew Lim Tan, "Information retrieval in document image databases," in IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 11, pp. 1398-1410, Nov. 2004.

[12]. Deliang Jiang, Xiaohu Yang, "Converting PDF to HTML approach based on Text Detection," ICIS '09: The 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, pp. 982 - 985, November 2009

[13]. M. R. Islam, R. S. Saha and A. R. Hossain, "Automatic Reading from Bangla PDF Document Using Rule Based Concatenative Synthesis," 2009 International Conference on Signal Processing Systems, Singapore, 2009, pp. 521-525.

[14]. Maganti Venkatesh et al., "Application of Multilingual OCR Algorithm for Converting Text from Images and PDFs," Proceedings of the 5th International Conference on Data Science, Machine Learning and Applications; Volume 2. ICDSMLA 2023. Lecture Notes in Electrical Engineering, vol 1274. Springer, Singapore.