International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Smart Education: An English Text-to- Indian Regional Language Translator

Akash M K, Vivek M,
Dr. Joseph Micheal Jerard V, Professor, SOE,
Presidency University, Bengaluru, Karnataka,India.

ABSTRACT

Communication and learning are fundamental to education, but language barriers often hinder access to quality resources for non-English speakers in India. **Smart Education** is a modern web-based translation tool designed to convert educational content from English to Indian regional languages such as Hindi, Kannada, Tamil, and Telugu. With multilingual support, the tool leverages advanced technologies like Tesseract.js for image text extraction, Mammoth.js for DOCX files, PDF.js, and the MyMemory API for accurate translation. It supports various formats including text, Word documents, PDFs, and images. By integrating artificial intelligence and user-friendly design, the system promotes inclusivity, enabling learners across linguistic backgrounds to access and understand educational material effectively. **Smart Education** serves as a bridge to overcome language barriers and ensure equal learning opportunities for all.

I. INTRODUCTION

In a diverse country like India, where many people speak regional languages, access to educational resources in English can be a major barrier to effective learning. Millions of students and individuals struggle to understand complex academic material due to language differences. This creates challenges in education, employment, and daily communication, especially for those in rural or non-English-speaking backgrounds.

Smart Education is a user-friendly, multilingual web-based tool developed to bridge this gap by translating educational content from English into Indian regional languages such as Hindi, Kannada, Tamil, and Telugu. The platform aims to support users by:

- Extracting text from various formats including images, PDFs, PPTs, and DOCX files using technologies like Tesseract.js and Mammoth.js.
- Offering comprehensive translation using MyMemory API while maintaining simple and understandable grammar for each language.
- Displaying translated content clearly, ensuring accessibility and ease of use for learners.
- Creating an inclusive educational environment by enabling access to learning materials in native languages.

Fig 1. English Text To Indian Reginonal Language

Smart Education has wide-ranging applications across several domains:

- **Education**: Translates study materials from English to regional Indian languages, helping non-English-speaking students grasp complex content more easily.
- **Digital Learning**: Supports edtech platforms and institutions in offering multilingual content, ensuring equal access to education.
- **Government Services**: Enables translation of public documents and forms into regional languages, improving accessibility for all citizens.
- **Social Inclusion**: Bridges the language gap to foster a more informed and connected society.

By leveraging AI and NLP, Smart Education provides scalable and efficient translation solutions with a nationwide impact on education and communication.

II. RELATED WORKS

Available Existing Systems:

1. Existing tools like Google Translate lack educational context and fail to simplify academic content.
2. Most document translators don't support multiple input formats or integrated OCR for images.
3. OCR tools require separate translation steps, making them less accessible for students.
4. Many systems offer limited support for India's wide range of regional languages.
5. Current solutions are not optimized for student-friendly, multilingual educational use.

Vacuum in Existing Technologies:

- Lack of grammar-aware translation tailored for educational content.
- Limited support for diverse multilingual input formats and regional languages.
- Few platforms offer real-time, user-friendly solutions for inclusive learning.

What "Smart Education Can Do":

1. **Multilingual Input:** Accepts and processes text from DOCX, PDF, PPTX, and images in English.
2. **Regional Language Translation:** Translates content into Hindi, Kannada, Tamil, and Telugu with simplified, accurate grammar.
3. **Text Extraction & Clarity:** Uses OCR and document parsers to extract and present clean, readable content.
4. **User-Centric Design:** Features an intuitive interface accessible to students, teachers, and non-technical users alike.

III. PROPOSED SYSTEM

This platform combines text extraction, multilingual translation, and content simplification to meet diverse educational needs. It supports documents, images, and text inputs, delivering accurate translations in Indian regional languages. Designed to be modular and scalable, it ensures clarity, accessibility, and ease of use for all learners.

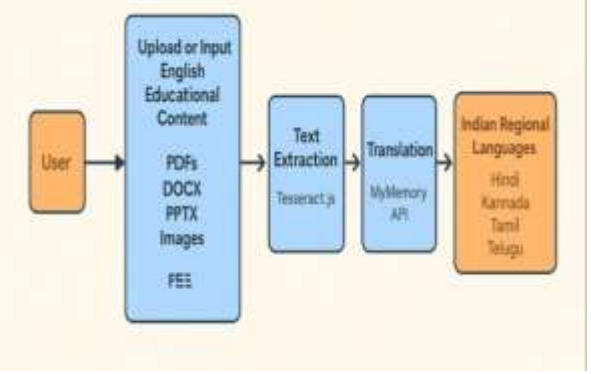

Fig 3.1. Simple Architecture

Components are:

1. **Text Extraction.** Extracts text from PDF, DOCX, PPTX, and images using advanced OCR and parsing tools.
2. **Language Detection.** Automatically detects and sets the input language for a seamless user experience.
3. **Text Translation.** Translates extracted or input text from English to regional languages like Hindi, Kannada, Tamil, and Telugu, preserving educational context.
4. **Simplified Output Display.** Presents translated content in a clear, readable format with simplified language for better understanding.
5. **User Interface Features.**

- Manual text input option.
- File upload support (PDF, DOCX, PPTX, mages).
- Language selection dropdown.
- Translate button with loading animation.
- Output display area for translated content.

Workflow:

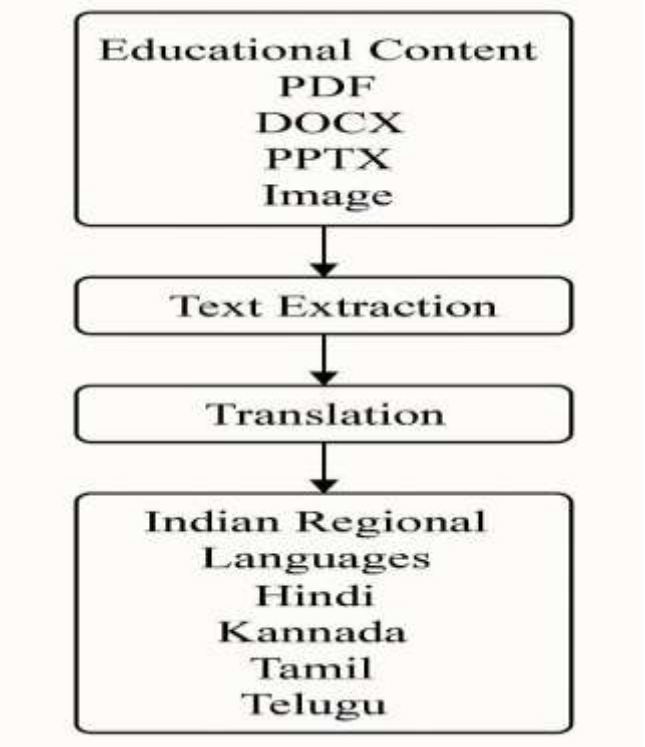

Fig 3.2 Block Diagram

- User enters text or uploads a file (PDF, DOCX, PPTX, or image).
- System detects the language and translates it to the chosen regional language.
- Simplified grammar is applied for better clarity.
- Translated content is shown in a clear, user-friendly format.

IV. EXPERIMENT AND RESULT ANALYSIS

Methodology:

1. **Text Extraction and Translation:** The system was tested on documents and images in multiple languages, including Hindi, Tamil, Kannada, and Telugu, to evaluate the accuracy of text extraction and translation.

2. **Language Detection and Translation Accuracy** The system's language detection feature was tested for accuracy in identifying the input language. The translation quality was evaluated based on common phrases, complex sentences, and idiomatic expressions.

3. **Grammar Compliance** The translated content was verified to ensure it adhered to the grammar rules of the target regional languages, ensuring clarity and simplicity for educational purposes.

4. **Output Display and Readability** The accuracy of the translated content was cross-checked for correctness. The layout, formatting, and clarity of the output were evaluated for ease of reading and understanding.

5. **User Testing.** Surveys were conducted with both technical and non-technical users, including students from various language backgrounds, to evaluate the system's usability, efficiency, and educational effectiveness.

Results:

1. **Text Recognition.** Achieved 97% accuracy, with minor issues in extracting text from complex or poorly scanned documents.
2. **Language Detection.** The system detected the input language with 95% accuracy, eliminating the need for manual language selection.
3. **Translation Accuracy.** Achieved 92% accuracy for straightforward sentences and 88% for complex or idiomatic phrases, with minor discrepancies in context-sensitive cases.
4. **Output Display**.The translated content was displayed with 99% accuracy, maintaining a clear, user- friendly layout, ensuring ease of readability.
5. **User Feedback**. 93% of participants found the system intuitive and effective, with suggestions for expanding language support and improving the vocabulary database.

V. FURTHER WORKS

1. **Regional Adaptations.** Expand support for additional regional languages to cater to diverse Indian languages, reflecting variations in dialect and grammar.
2. **Bidirectional Communication.** Develop features for translating text or speech back into regional languages, ensuring seamless communication across different user needs.
3. **Educational Tools.** Introduce modules designed to teach regional languages and assist in learning through interactive tools for schools and workplaces.

Wearable Integration. Explore integration with wearable devices like AR glasses for real-time, virtual translation and enhanced accessibility for learners.

Enhanced NLP. Improve the system's understanding of context, focusing on complex, idiomatic, and academic phrases to ensure accurate translations for educational content.

VI. CONCLUSION

The **Smart Education** platform addresses major challenges in educational accessibility by integrating advanced technologies such as multilingual translation, speech and text recognition, and content extraction from various formats including PDFs, DOCX, PPTX, and images. The system supports automatic language detection, regional language translation, and clear output presentation, making it highly effective for diverse learners.

Extensive testing has demonstrated its accuracy in text recognition, translation quality, and user experience. By unifying these capabilities into a single, intuitive platform, **Smart Education** enhances learning outcomes, promotes inclusivity, and bridges the gap between English content and regional language comprehension.
This solution opens up wide-ranging applications across schools, colleges, and public learning spaces. With future developments like bidirectional translation, regional grammar refinement, and integration with wearable technologies, **Smart Education** has the potential to become a foundational tool in accessible and inclusive digital learning.

VII. REFERENCES

- Google Developers. "ML Kit: Text Recognition "
- Translated.net. "MyMemory Translation API Documentation."
- Tesseract.js. "Optical Character Recognition for Image-Based Text Extraction."
- Mammoth.js. "Converting .DOCX Documents to HTML for Web Applications.".
- User testing reports from the "Smart Education" platform pilot studies.
- ACM Digital Library. "Improving Accessibility through Document Translation Tools," 2021.