TEXT DETECTION AND EXTRACTION WITH TEXT TRANSLATION AND TEXT-TO-SPEECH USING DEEP LEARNING TECHNIQUES

A PROJECT REPORT

Submitted by

AATHIRAJU M

GOWRI SHANKAR S

KIRUSHIGAN R

RAJADURAI M

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

K.RAMAKRISHNAN COLLEGE OF TECHNOLOGY

(An Autonomous Institution, affiliated to Anna University Chennai and Approved by AICTE, New Delhi)

SAMAYAPURAM – 621 112

MAY, 2025

TEXT DETECTION AND EXTRACTION WITH TEXT TRANSLATION AND TEXT-TO-SPEECH USING DEEP LEARNING TECHNIQUES

A PROJECT REPORT

Submitted by

AATHIRAJU M (811721243001)

GOWRI SHANKAR S (811721243017)

KIRUSHIGAN R (811721243025)

RAJADURAI M (811721243303)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

K.RAMAKRISHNAN COLLEGE OF TECHNOLOGY

(An Autonomous Institution, affiliated to Anna University Chennai and Approved by AICTE, New Delhi)

SAMAYAPURAM – 621 112

MAY, 2025

K.RAMAKRISHNAN COLLEGE OF TECHNOLOGY

(AUTONOMOUS)

SAMAYAPURAM – 621 112

BONAFIDE CERTIFICATE

Certified that this project report titled **"TEXT DETECTION AND EXTRACTION WITH TEXT TRANSLATION AND TEXT-TO-SPEECH USING DEEP LEARNING TECHNIQUES"** is the bonafide work of **M.AATHIRAJU (811721243001), S.GOWRI SHANKAR(811721243017), R.KIRUSHIGAN (811721243025), M.RAJADURAI (81171243303)** who carried out the project under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE	**SIGNATURE**
Dr. T.Avudaiappan M.E., Ph.D.,	Mrs.S.Geetha M.E.,
HEAD OF THE DEPARTMENT	**SUPERVISOR**
Department of Artificial Intelligence K.Ramakrishnan College of Technology (Autonomous) Samayapuram – 621 112	ASSISTANT PROFESSOR Department of Artificial Intelligence K.Ramakrishnan College of Technology (Autonomous) Samayapuram – 621 112

Submitted for the viva-voce examination held on ………………

INTERNAL EXAMINER **EXTERNAL EXAMINER**

DECLARATION

We jointly declare that the project report on **"TEXT DETECTION AND EXTRACTION WITH TEXT TRANSLATION AND TEXT - TO-SPEECH USING DEEP LEARNING TECHNIQUES"** is the result of original work done by us and best of our knowledge, similar work has not been submitted to **"ANNA UNIVERSITY CHENNAI"** for the requirement of Degree of **BACHELOR OF TECHNOLOGY**. This project report is submitted on the partial fulfilment of the requirement of the award of Degree of **BACHELOR OF TECHNOLOGY**.

Signature

AATHIRAJU M

GOWRI SHANKAR S

KIRUSHIGAN R

RAJADURAI M

Place: Samayapuram

Date:

ACKNOWLEDGEMENT

It is with great pride that we express our gratitude and in-debt to our institution “**K.Ramakrishnan College of Technology (Autonomous)**”, for providing us with the opportunity to do this project.

We are glad to credit honourable chairman **Dr. K.RAMAKRISHNAN**, **B.E.,** for having provided for the facilities during the course of our study in college.

We would like to express our sincere thanks to our beloved Executive Director **Dr. S. KUPPUSAMY, MBA, Ph.D.,** for forwarding to our project and offering adequate duration in completing our project.

We would like to thank **Dr. N. VASUDEVAN, M.E., Ph.D.,** Principal, who gave opportunity to frame the project the full satisfaction.

We whole heartily thank to **Dr. T.AVUDAIAPPAN**, **M.E., Ph.D.,** Head of the department, **ARTIFICIAL INTELLIGENCE** for providing his encourage pursuing this project.

We express our deep and sincere gratitude to our project guide **Mr.S.GEETHA M.E.,** Department of **ARTIFICIAL INTELLIGENCE,** for her incalculable suggestions, creativity, assistance and patience which motivated us to carry out this project.

We render our sincere thanks to Course Coordinator and other staff members for providing valuable information during the course.

We wish to express our special thanks to the officials and Lab Technicians of our departments who rendered their help during the period of the work progress.

ABSTRACT

The rapid advancement of deep learning technologies has opened up new possibilities for text detection, translation, and text-to-speech synthesis, significantly enhancing accessibility and communication. Text detection and extraction with text translation and text to speech using deep learning techniques aims to develop an integrated framework that combines these capabilities into a single, user-friendly system. Utilizing advanced deep learning models such as EasyOCR, the system detects text in images, translates it into multiple languages, and converts it into natural- sounding speech. The system comprises three main components: text detection and extraction, translation engine, and text-to-speech synthesis. EasyOCR accurately identifies and extracts text from various image sources, ensuring high precision and robustness. The extracted text is processed by a multilingual translation engine, capable of translating text into multiple target languages, making the content accessible to a wider audience. The translated text is then converted into speech using advanced text- to-speech models, providing an auditory representation of the text. Text detection and extraction with text translation and text to speech using deep learning techniques demonstrates the practical applications of deep learning in enhancing accessibility and global communication. By integrating these features into a cohesive framework, this system represents a significant step forward in making written content more accessible and engaging for users worldwide.

TABLE OF CONTENTS

LIST OF TABLE

LIST OF FIGURES

LIST OF ABBREVIATIONS

CNN - Convolutional Neural Networks

OCR - Optical Character Recognition

RNN - Recurrent Neural Networks

TTS - Text-To-Speech

CHAPTER 1

INTRODUCTION

The text detection and extraction with text translation and text to speech using deep learning techniques aims to develop a comprehensive deep learning framework that integrates text detection, translation, and text-to-speech synthesis. By utilizing advanced models like EasyOCR, the system detects text from images, translates it into multiple languages, and converts it into natural-sounding speech. This innovative approach enhances accessibility and communication, making written content universally accessible and engaging for users with varying needs and language preferences, ultimately fostering a more inclusive digital environment.

1.1 BACKGROUND

The fast-paced developments in deep learning technologies have transformed the areas of text detection, translation, and text-to-speech (TTS) synthesis. The conventional techniques for text detection used handcrafted features and rule-based methods, which tended to struggle with different fonts, sizes, and intricate backgrounds. But the emergence of deep learning models, especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), has greatly enhanced the accuracy and resilience of text detection systems. EasyOCR is an advanced Optical Character Recognition (OCR) software that uses deep learning methods to identify and recognize text from images. Utilizing CNNs to extract features and RNNs to model sequences, EasyOCR easily works with various text styles and cluttered backgrounds, making it a very stable text detection solution. Simultaneously, machine translation has also experienced tremendous advancement with the advent of deep learning architectures, particularly Transformer networks. They have shown remarkable performance in text translation from and to multiple languages by identifying long-range dependencies and contextual cues more efficiently than before. The Transformer architecture based on self-attention has become the foundation of numerous contemporary translation systems, allowing translation between different languages with ease and precision. Text-to- speech synthesis has seen its transformation with the use of deep learning methods. Contemporary TTS models like Tacotron and WaveNet use deep neural networks to synthesize natural speech from text. Tacotron transcribes text into a mel-spectrogram, which is further converted into audio using a vocoder such as WaveNet. These models have drastically enhanced the quality and expressiveness of synthesized speech to a level that is almost identical to that of human speech. Through the incorporation of these advanced deep learning methods, your project seeks to build an integrated framework that promotes accessibility and communication. The intended system will scan text from images, translate it into several languages, and render it into natural-sounding voice. This new method can dramatically influence numerous applications, including helping visually impaired individuals, language learning, and real-time translation and vocalization for tourists.

1.2 PROBLEM STATEMENT

- Printed and handwritten text is not universally accessible, particularly for visually impaired individuals and those who speak multiple languages.
- Existing methods for text detection, translation, and text-to-speech synthesis lack the robustness and accuracy needed to handle diverse text styles, complex backgrounds, and multiple languages.
- Users currently rely on multiple tools to achieve text detection, translation, and text-to- speech synthesis, leading to inefficiency and inconvenience.
- There is a need for a unified, efficient, and accurate solution that integrates text detection, translation, and text-to-speech synthesis into a cohesive framework.

1.3 OBJECTIVES

The text detection and extraction with text translation and text to speech using deep learning techniques aims to develop an advanced deep learning framework that integrates text detection, translation, and text-to-speech synthesis into a seamless and user-friendly system. The framework will first identify and extract text from various sources, such as images or scanned documents, using computer vision techniques. Once the text is detected, it will be processed through a translation model, allowing it to be converted into different languages to enhance multilingual communication. Finally, the translated text will be converted into speech using text-to-speech synthesis, making the information accessible in an audible format.By combining these three key functionalities into a single system, the framework addresses multiple challenges related to accessibility and communication. It enables users to access and understand information across language barriers while also providing an audio output for individuals who may have visual impairments or prefer spoken content. This integrated approach enhances usability and ensures that information is more widely available, making it a valuable tool in various domains such as education, accessibility services, and cross-cultural communication.

- Text Detection and Extraction: The first objective is to accurately detect and extract printed or handwritten text from various image sources. By leveraging state-of-the- art deep learning models such as EasyOCR, the system will identify and extract text with high precision and robustness, even in diverse conditions such as varying fonts, sizes, and complex backgrounds. This component ensures that the text detection process is reliable and efficient.
- Translation Engine: The second objective is to translate the extracted text into multiple target languages. The project will utilize advanced machine translation models, such as Transformer networks, to achieve seamless and accurate translation.
- The translation engine will capture long-range dependencies and contextual information effectively, ensuring that the translated text is both accurate and contextually appropriate. This component enhances the accessibility of the content by making it available in multiple languages.
- Text-to-Speech Synthesis: The third objective is to convert the translated text into natural-sounding speech. The project will employ cutting-edge text-to-speech synthesis models, such as Tacotron and WaveNet, to generate high-quality and expressive speech from text inputs. This component provides an auditory representation of the text, making it accessible to visually impaired users and supporting language learning.

- User-Friendly Interface: The final objective is to develop a user-friendly interface that seamlessly integrates the text detection, translation, and text-to-speech synthesis components. The interface will be designed to be intuitive and easy to use, allowing users to perform all three tasks within a single platform. This cohesive framework ensures that users can efficiently and effectively interact with the system to achieve their desired outcomes.

CHAPTER 2

LITERATURE SURVEY

2.1 EXTRACTION ALGORITHM OF ENGLISH TEXT INFORMATION FROM COLOR IMAGES BASED ON RWT

Yaqin Wang, Lu Xun ,Qu Yuan

The text discusses advancements in extracting meaningful information from color images using techniques like radial wavelet transforms and complex Morlet wavelets. It highlights challenges such as weak signal extraction and directional sensitivity affecting image boundary detection. Innovative methods are proposed, including radial wavelet entropy for English text extraction from color images and a scene text recognition approach using D-SIFT features and spatio-temporal histograms. Furthermore, the integration of selective ensemble learning and model compression based on English text samples aims to enhance extractor efficiency and applicability across various domains, emphasizing improved performance and reduced computational requirements.

Merits

- Better text extraction
- More efficient processing
- Advanced image analysis

Demerits

- Struggles with weak signals
- Issues with boundary detection
- Complex to implement

2.2 EFFICIENT TEXT BOUNDING BOX IDENTFICATION USING MASK R-CNN

Phanthakan , Dittaya Wanvarie, Nagul Harojananone

This study focuses on text detection for OCR applications, addressing challenges in existing models like CRAFT, PMTD, and SPCNET, which struggle with post-processing and multiline text errors. CRAFT, for instance, merges bounding boxes from consecutive lines, causing inaccuracies. To overcome these issues, the study proposes using Mask R-CNN, an instance segmentation model that treats each text element as an independent object. This approach eliminates post- processing inefficiencies and effectively handles multiline text. The model is tested on various document types, including bankbooks, ID cards, invoices, and passports, demonstrating superior accuracy, adaptability, and efficiency. By removing post- processing dependencies, the model reduces inference time and resource consumption, making it highly suitable for real-world OCR applications.

Merits

- Eliminates post-processing inefficiencies
- Handles multiline text more effectively
- Improves accuracy and adaptability

Demerits

- Complex model implementation
- Requires high computational resources
- May need further refinement for varied document types

2.3 CURSIVE TEXT RECOGNITION IN NATURAL SCENE IMAGES USING DEEP C-RNN

Mark R, Md.Asikuzzaman, Asghar Ali, Mehwish Leghari

The paper addresses the challenges of recognizing Urdu text in natural scene images, which is more complex than standard OCR due to variations in font styles, orientations, occlusions, and lighting conditions. A segmentation-free deep convolutional recurrent neural network (CRNN) is proposed, where whole word images are processed without pre-segmenting characters. The model consists of a deep convolutional neural network (CNN) with shortcut connections for feature extraction, a recurrent neural network (RNN) for decoding, and connectionist temporal classification (CTC) for mapping sequences to labels. To enhance accuracy, architectures like VGG-16, VGG-19, ResNet-18, and ResNet-34 are explored. A large-scale dataset of Urdu text images is developed, and experiments show that the proposed CRNN model outperforms other architectures.

Merits

- Handles complex Urdu text variations
- No need for character segmentation
- Outperforms other architectures

Demerits

- High computational requirements
- Dependent on large datasets
- May struggle with extreme occlusions

2.4 CAMERA KEYBOARD: A NOVEL INTERACTION TECHNIQUE FOR TEXT ENTRY THROUGH SMARTPHONE CAMERA

Alessio Bellino, Valeria Herskovic

Camera Keyboard is a text entry technique that enables smartphones to extract and digitize text from physical sources like business cards and IDs using their cameras. After capturing an image, Optical Character Recognition (OCR) via Google Cloud Vision API processes the text, preserving its original layout. Users can then select words or lines by tapping on them. Integrated into the standard mobile QWERTY keyboard, Camera Keyboard provides a seamless way to digitize text across any application (e.g., Gmail, WhatsApp). A user study with 18 participants compared it to Google Keyboard and a physical desktop keyboard, showing that Camera Keyboard was faster in most cases, making text entry more efficient and convenient.

Merits

- Digitizes text from physical sources quickly
- Preserves original layout for accuracy
- Seamlessly integrates with mobile keyboards

Demerits

- Depends on OCR accuracy
- May struggle with poor image quality
- Limited by supported languages and formats

2.5 SMART TEXT SCANNER: AN ENHANCED CAMERA-BASED INPUT METHOD FOR MULTILINGUAL TEXT EXTRACTION AND CONVERSION

Priya Natarajan, Rahul Mehta

Smart Text Scanner introduces an innovative method for capturing and processing text from physical documents such as forms, signage, and ID cards using a smartphone camera. Once an image is captured, advanced Optical Character Recognition (OCR) techniques powered by deep learning, including EasyOCR and translation services, are applied to accurately extract and translate text while preserving the layout. This system enables users to interactively select and convert text into speech in real-time through gTTS, providing accessibility and convenience across multiple languages. Integrated with mobile interfaces, this method supports seamless communication and accessibility across apps like messaging platforms, emails, and browsers. A usability study conducted with 18 participants demonstrated that Smart Text Scanner significantly improved the speed and accuracy of multilingual text extraction and comprehension compared to traditional input methods.

Merits

- Extracts and digitizes text efficiently from physical sources
- Supports multilingual translation and speech output
- Enhances accessibility for visually impaired users
- Integrates with mobile applications for real-time use

Demerits

- OCR performance depends on image clarity and lighting
- Limited offline support for translation and TTS
- May require internet access for optimal language support

CHAPTER 3
SYSTEM ANALYSIS

3.1 EXSITING SYSTEM

The existing systems for text detection, translation, and text-to-speech synthesis typically function as separate entities, each relying on its own algorithms to carry out its specific task. While these technologies are effective individually, they are not inherently connected, which means users often have to switch between multiple tools to complete a single workflow. For instance, a person may need to use one application to detect and extract text from an image, another to translate the extracted text into a different language, and yet another to convert the translated text into speech. This fragmentation creates inefficiencies, making the process cumbersome and less accessible, especially for those who require a seamless and intuitive solution for communication and accessibility.

3.1.1 Algorithm Used

Text Detection and Extraction

Identifying and extracting text from images using OCR techniques for further processing.

- **Convolutional Recurrent Neural Networks (CRNN)**

  CRNNs combine Convolutional Neural Networks (CNNs) for feature extraction with Recurrent Neural Networks (RNNs) for sequence modeling.

Translation

Converting text from one language to another while retaining its original meaning.

Transformer Networks

Transformer networks use a self-attention mechanism to capture long- range dependencies and contextual information, making them highly effective for translation tasks.

Text-to-Speech (TTS) Synthesis

Transforming written text into natural-sounding spoken words using deep learning models.

- **Tacotron**

  Tacotron is an end-to-end TTS system that converts text to a mel- spectrogram, which is then transformed into audio using a vocoder.

- **WaveNet**

  WaveNet is a generative model that produces high-quality, natural- sounding speech.

3.1.2 Drawbacks

- Computational Intensity: Deep learning models require substantial computational resources and large datasets for training, making them less accessible for smaller projects or organizations.
- Accuracy and Robustness: Traditional OCR and translation methods often lack the accuracy and robustness needed to handle diverse text styles, complex backgrounds, and multiple languages.
- Emotional Nuance: Text-to-Speech (TTS) systems, even advanced ones, may struggle to convey emotional nuance, leading to less engaging and natural-sounding speech.
- Integration: Existing systems typically operate in isolation, requiring users to rely on multiple tools for text detection, translation, and text-to-speech synthesis, leading to inefficiency and inconvenience.
- Resource Limitations: The requirement for extensive data and computational power limits the widespread adoption and practical deployment of these models.
- User Experience: Separate tools can complicate the user experience, making it cumbersome to achieve seamless text detection, translation, and speech synthesis.

3.2 PROPOSED SYSTEM

The proposed system aims to integrate text detection, translation, and text-to- speech synthesis into a single, user-friendly platform. Utilizing advanced deep learning models like EasyOCR for text detection, the system accurately extracts text from images. The extracted text is then translated into multiple languages using Transformer networks, ensuring seamless and accurate translation. Finally, the translated text is converted into natural-sounding speech using Google Text-to-Speech (gTTS), which provides high-quality speech synthesis. This integrated framework enhances accessibility by providing a unified solution for text detection, translation, and vocalization, making written content more accessible and engaging for users with varying needs and language preferences. The user-friendly interface ensures efficient and convenient interaction with the system.

3.2.1 Algorithms Used

- **EasyOCR**

EasyOCR is used for Text Detection and Extraction, the EasyOCR utilizes Convolutional Recurrent Neural Networks (CRNN). CRNN combines Convolutional Neural Networks (CNNs) for feature extraction with Recurrent Neural Networks (RNNs) for sequence modeling. CNNs detect and extract features from images, while RNNs recognize the sequential nature of text, enabling accurate text detection and extraction from various image sources.

- **Transformer Networks**

Transformer Networks is used for Text Translation. Transformers use a self- attention mechanism to capture long-range dependencies and contextual information. They consist of an encoder-decoder architecture that processes input text and generates translated text in the target language. This architecture ensures seamless and accurate translation across multiple languages.

- **Google Text-to-Speech (gTTS)**

  Google Text-to-Speech (gTTS) is used for Text-to-Speech Synthesis, gTTS leverages advanced text-to-speech synthesis models to convert written text into natural-sounding speech. It supports multiple languages and provides high-quality speech synthesis, enhancing accessibility for visually impaired users and facilitating language learning.

- **Seq2Seq (Sequence-to-Sequence) Model**

  Seq2Seq (Sequence-to-Sequence) Model is used for Translation, Seq2Seq models use an encoder-decoder architecture with RNNs or LSTMs (Long Short-Term Memory) to translate text sequences from one language to another. They are particularly effective for handling variable-length input and output sequences.

- **WaveRNN**

  WaveRNN is used for Text-to-Speech Synthesis. WaveRNN is a compact and efficient neural network that generates high-quality speech waveforms from mel- spectrograms. It is a lighter alternative to WaveNet, making it suitable for deployment on resource-constrained devices.

3.2.2 Advantages

- High Accuracy
- Cost-Effective
- Educational Aid
- Multilingual Support
- Seamless Integration
- Real-Time Processing
- Enhanced Accessibility
- Improved Decision-Making

CHAPTER 4

SYSTEM SPECIFICATIONS

4.1 HARDWARE SPECIFICATIONS

- GPU: NVIDIA GeForce RTX 3050/3090 or NVIDIA Tesla V100 with a minimum of 8 GB VRAM (for deep learning model training)
- CPU: A multi-core CPU, preferably with at least 8 cores e.g., Intel i7/i9 or AMD Ryzen 7/9
- MEMORY: 16GB or more (to handle large datasets efficiently)
- STORAGE: 1TB SSD (for fast data access)

4.2 SOFTWARE SPECIFICATIONS

- Programming Language: Python (Python 3.7 or later)
- Libraries:
- Deep Learning: TensorFlow or PyTorch (for model development)
- EasyOCR
- gTTS
- Googletransfor text translation)
- NLTK
- Scikit-learn
- Environment: VS Code
- Database: PostgreSQL or MongoDB
- Version Control: Git
- Operating System: Linux or Windows 10/11

4.3 SOFTWARE DESCRIPTION

The text detection and extraction with text translation and text to speech using deep learning techniques leverages advanced deep learning techniques to create an integrated platform for text detection, translation, and text-to-speech synthesis. Built using Python, it utilizes TensorFlow and PyTorch for deep learning, EasyOCR for text detection, and Hugging Face Transformers for translation. Google Text-to-Speech (gTTS) provides natural-sounding speech synthesis. Running on an Ubuntu 20.04 LTS operating system with high-performance CPU and GPU, the system ensures efficient data processing and model training. The modular architecture allows for easy updates, while the user-friendly interface enables seamless interaction. This unified approach enhances accessibility and communication, making written content universally accessible and engaging for users with varying needs.

CHAPTER 5

SYSTEM DESIGN

The architectural framework guiding a text extraction and processing system using OCR technology. A client application on web, mobile, or desktop allows users to upload images or capture them via a live camera. The image undergoes preprocessing using OpenCV before being processed by an OCR engine (EasyOCR) to extract text. The extracted text is displayed on the client device and can be enhanced with additional features. These include text translation, which converts the text into different languages, and text-to-speech functionality, which reads the text aloud. The system optimizes data extraction efficiency, enabling real- time text digitization from images. This integration streamlines text recognition for various applications, making it a versatile solution for digital text processing.

5.1 SYSTEM ARCHITECTURE

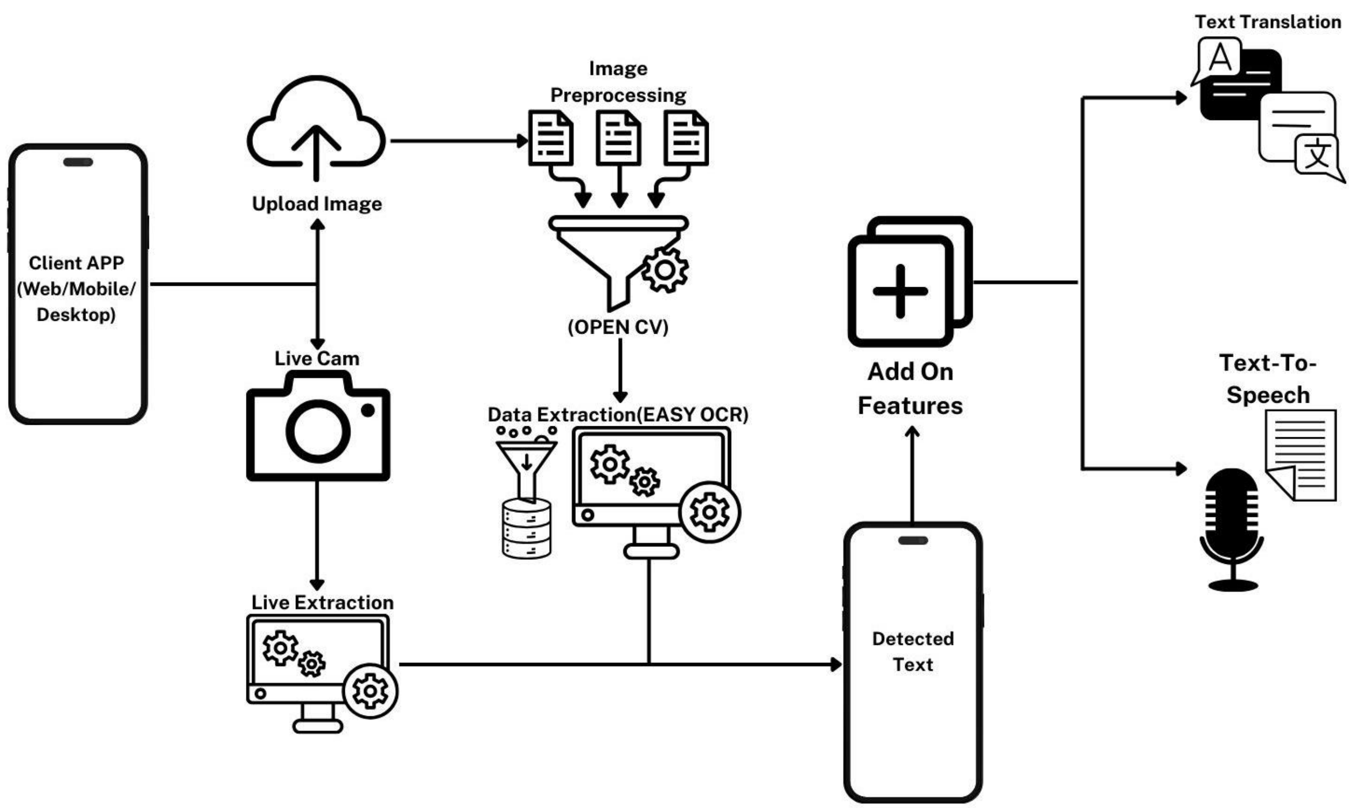

Fig. 5.1 System Architecture

5.2 DATA FLOW DIAGRAM

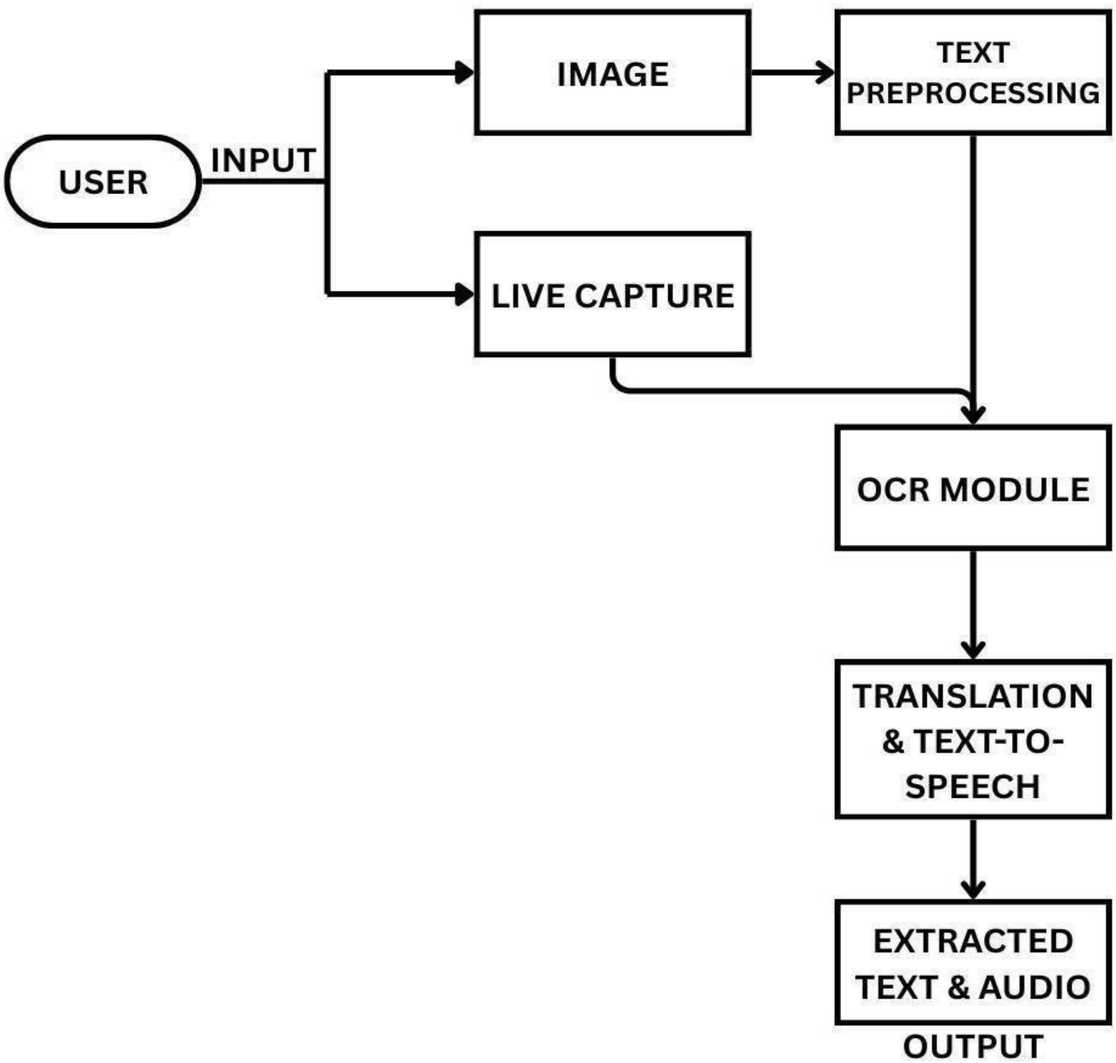

Fig. 5.2 Data Flow Diagram

5.3 USE CASE DIAGRAM

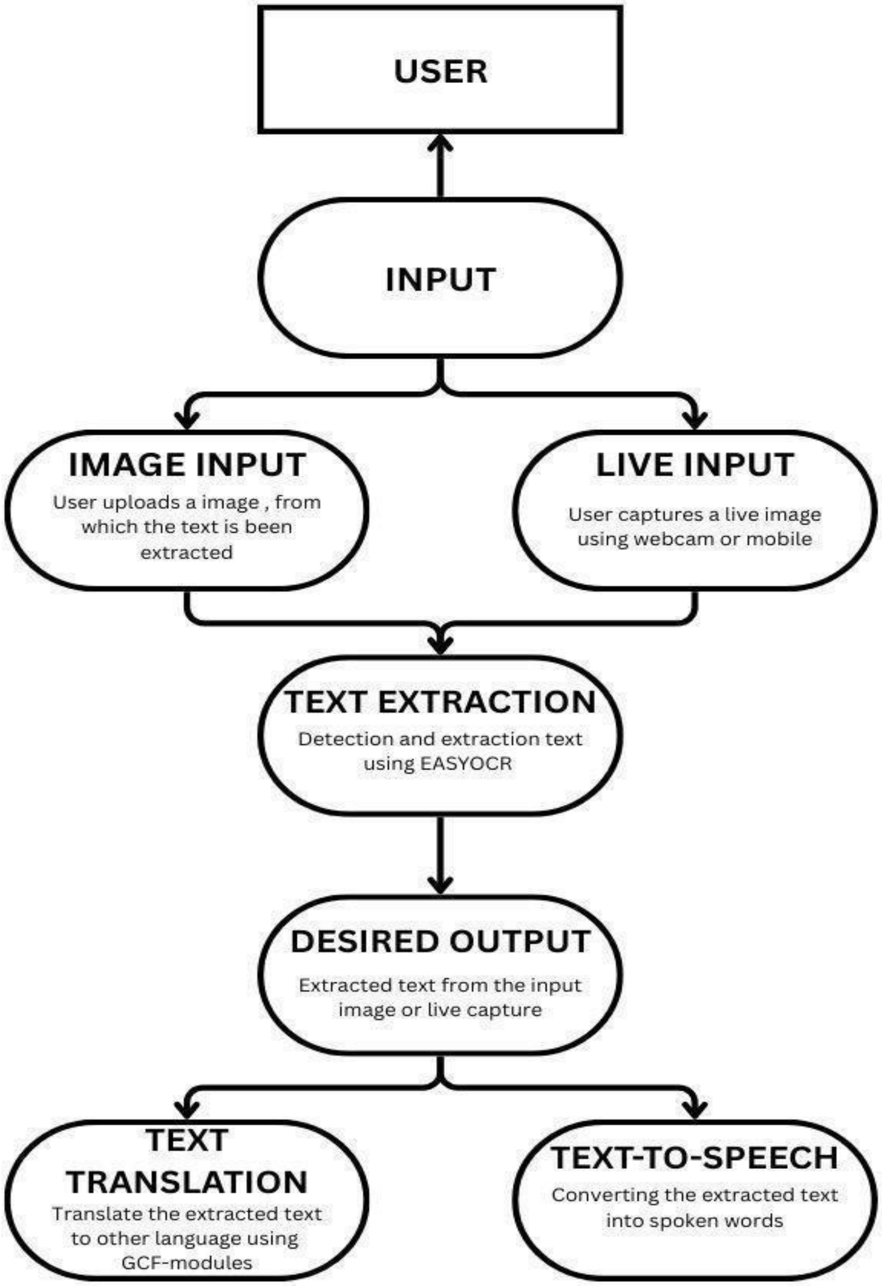

Fig. 5.3 Use Case Diagram

5.4 ACTIVITY DIAGRAM

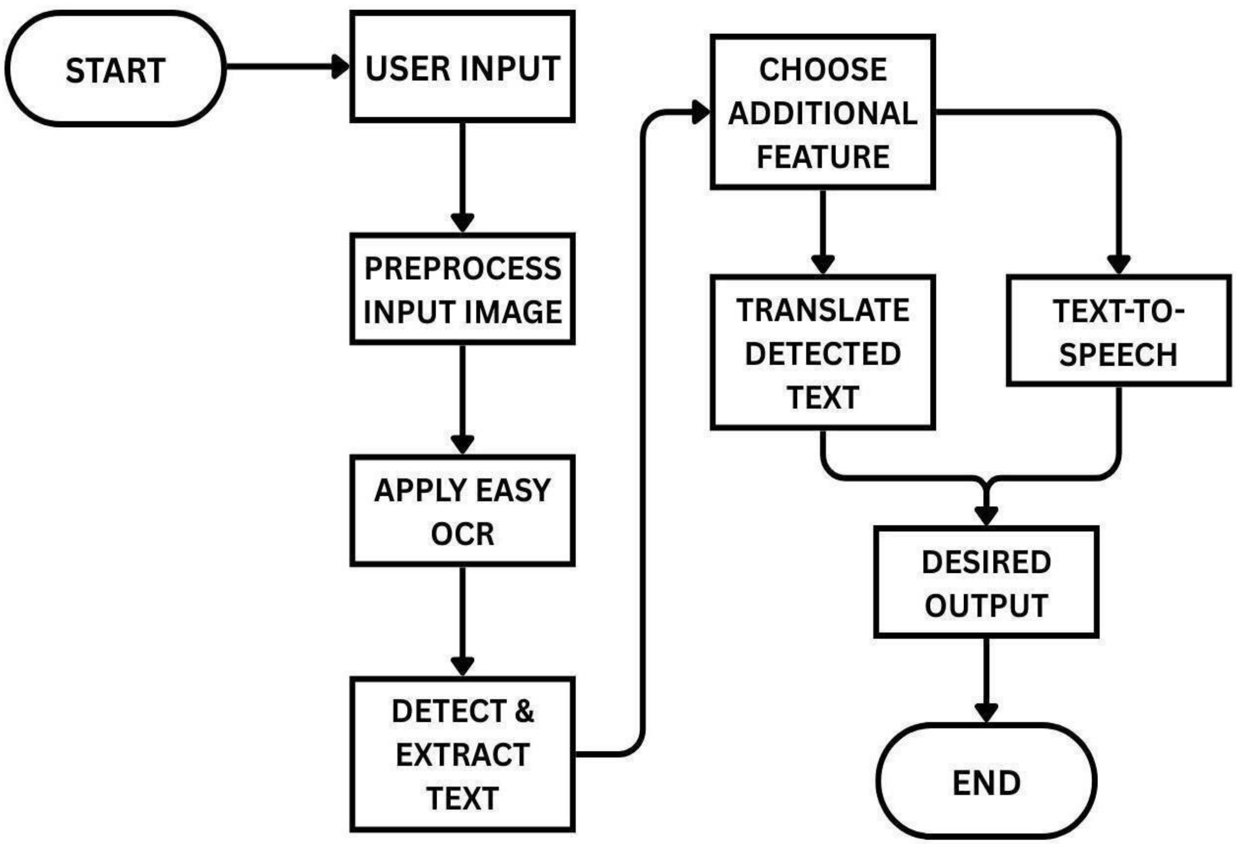

Fig. 5.4 Activity Diagram

5.5 SEQUENTIAL DIAGRAM

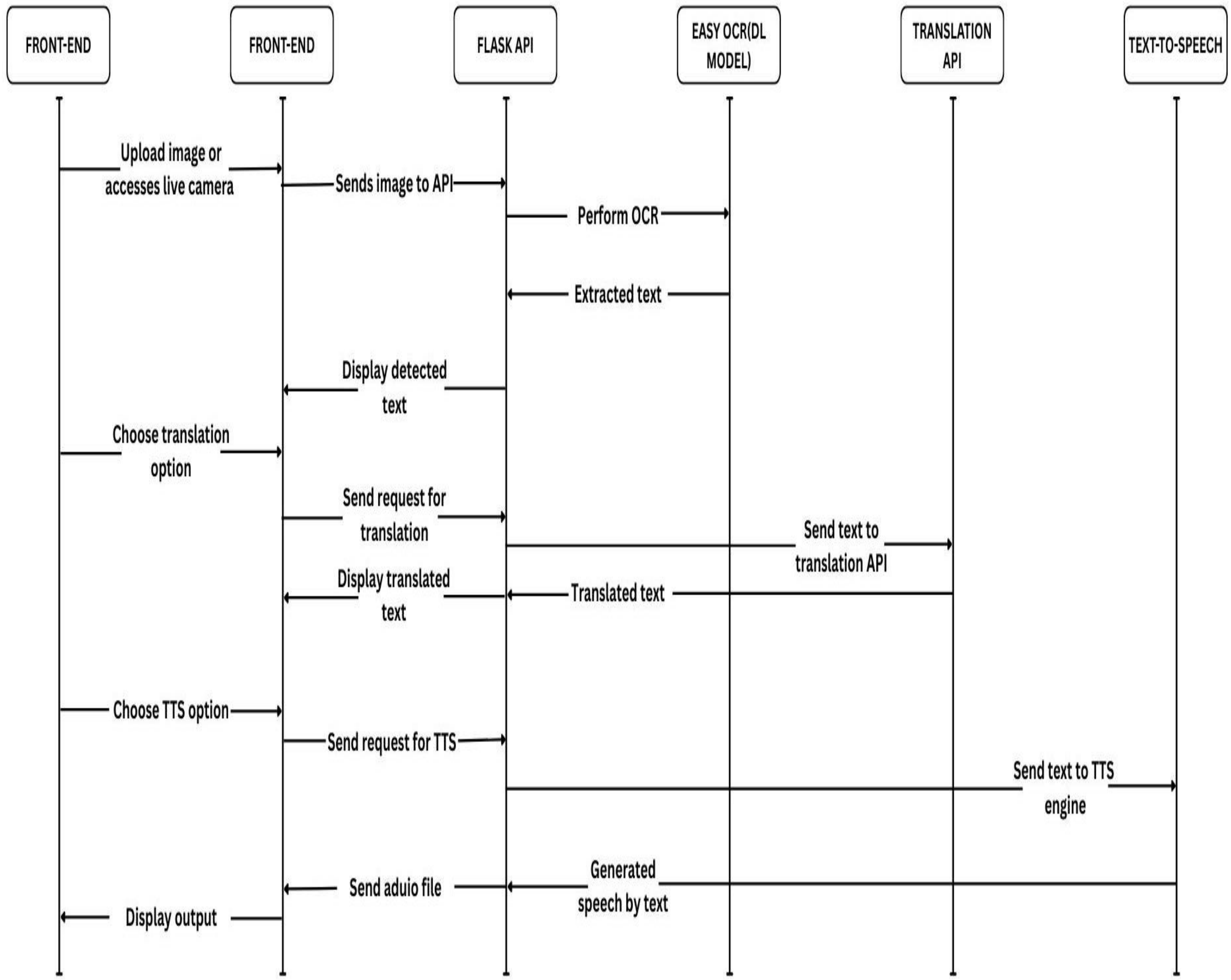

Fig. 5.5 Sequential Diagram

CHAPTER 6

MODULES DESCRIPTION

6.1 LIST OF MODULES

- Text Detection and Extraction Module
- Translation Module
- Text-to-Speech Synthesis Module
- User Interface Module
- Data Preprocessing Module
- Post-Processing Module

6.2 TEXT DETECTION AND EXTRACTION MODULE

The Text Detection and Extraction Module serves as the foundational step in processing text from images or documents. This module employs advanced optical character recognition (OCR) techniques to identify textual regions within an image. Initially, preprocessing methods such as contrast enhancement and edge detection are used to highlight possible text areas, ensuring accurate recognition even in challenging environments with varied lighting conditions and complex backgrounds. By leveraging deep learning-based approaches like convolutional neural networks (CNNs), the module efficiently detects and segments textual components, irrespective of font style, orientation, or language. This robust detection mechanism is essential for applications dealing with natural scene images, handwritten text, or document processing.

Once the text regions are identified, the extraction process involves breaking down these segments into recognizable characters or words. Traditional OCR approaches often require predefined character segmentation, which may lead to inaccuracies in complex texts. However, modern techniques, particularly those based on deep neural networks, adopt segmentation-free methods that recognize entire words or lines at once, improving efficiency and accuracy. The extracted text undergoes further refinement, such as spell-checking and contextual corrections, ensuring reliability before passing it on for translation or speech synthesis. This module is crucial for streamlining data processing, reducing errors, and preparing textual information for subsequent operations.

6.3 TRANSLATION MODULE

The Translation Module is responsible for converting extracted text into different languages while maintaining the original meaning and contextual integrity. Machine translation systems use sophisticated transformer-based architectures to analyze linguistic patterns and structures. These models employ tokenization, which breaks down text into smaller components for efficient processing. An attention mechanism then evaluates word relationships within sentences to ensure translations capture nuances and maintain coherence. This method significantly improves translation accuracy compared to conventional phrase-based models, making the system more effective in handling complex sentence structures.

To refine translations, semantic analysis and grammatical correction processes are integrated into the module. These enhancements help ensure output remains natural and meaningful. Additionally, specialized terminology databases assist in maintaining accuracy when dealing with technical, legal, or medical content. This module also incorporates user feedback to continuously improve translation quality, adapting to linguistic variations over time. Through ongoing refinement and optimization, it enables seamless multilingual communication, supporting applications in business, education, accessibility, and cross-cultural interactions.

With the rise of AI-driven translation solutions, there has also been a focus on personalization and customization. Users can adjust the style, tone, or dialect preferences within the module, ensuring that translations align with specific audience needs. For instance, a business document might require a more formal and precise translation, while a social media post might benefit from a casual and conversational tone. This flexibility enhances the effectiveness of the module across various domains, making it an indispensable tool for individuals, businesses, and academic institutions alike. By refining translations to fit specific contexts, the module adapts to the diverse requirements of modern communication.

At its core, the Translation Module is an essential innovation that breaks linguistic boundaries and fosters global connectivity. Its ability to transform text while preserving meaning and readability highlights the immense progress made in computational linguistics.

The continued development of more intelligent, context-aware models ensures that translations become increasingly seamless, making information more accessible to speakers of different languages. By refining and expanding its capabilities, the module plays a crucial role in promoting cross-cultural understanding, paving the way for more inclusive and interconnected interactions across the world.

6.4 TEXT-TO-SPEECH SYNTHESIS MODULE

The Text-to-Speech (TTS) Module is a sophisticated technology designed to convert written text into human-like speech, offering an auditory experience that feels natural and engaging. This module relies on deep learning techniques and advanced speech synthesis algorithms to generate expressive and intelligible voices, making digital interactions more accessible and immersive. Unlike earlier robotic-sounding speech systems, modern TTS models incorporate nuanced intonations, rhythm adjustments, and emotional expressions that closely resemble human speech patterns. As a result, the synthesized voice appears more fluid and engaging rather than flat and mechanical, enhancing communication in accessibility tools, virtual assistants, and automated response systems.

With the rise of multilingual communication, modern TTS modules support a wide range of languages, ensuring accessibility for diverse populations. By incorporating linguistic databases, these systems can synthesize speech in multiple languages while maintaining accurate pronunciation and grammatical flow. This feature is particularly beneficial for global applications such as multilingual customer support, language translation platforms, and international broadcasting services. Furthermore, TTS technology facilitates seamless cross-lingual interactions, allowing individuals to consume information in their preferred language while preserving clarity and coherence in spoken communication. As globalization drives the need for more inclusive AI solutions, the expansion of multilingual speech synthesis remains a crucial area of development.

Another key function of the TTS module is its application in assistive technologies, empowering individuals with disabilities to access digital content through auditory means. For visually impaired users, screen readers utilizing TTS technology enable easy navigation of websites, documents, and applications without requiring visual input. Similarly, individuals with speech impairments benefit from AI-driven voice synthesis tools that provide an expressive outlet for communication. By bridging accessibility gaps, the TTS module enhances inclusivity and enables equal participation in digital environments. As AI-driven voice synthesis continues to evolve, efforts to refine expressiveness, emotional intonation, and personalized voice settings contribute to a more human-like interaction experience.

The integration of TTS modules into modern smart devices and AI-powered applications has transformed the way individuals engage with technology. From voice assistants guiding users through daily tasks to audiobook narrators delivering engaging content, TTS technology has become an indispensable part of interactive digital experiences. Future advancements aim to refine contextual awareness, enabling AI-generated voices to adapt dynamically based on conversational cues and audience preferences. As research progresses, developers explore the potential of emotionally intelligent speech synthesis, where AI can replicate subtle vocal emotions to create more relatable and responsive interactions. These innovations pave the way for a future where AI-driven voices seamlessly blend into everyday life, enhancing communication and accessibility on a global scale.

6.5 USER INTERFACE MODULE

The User Interface (UI) Module plays a crucial role in creating seamless and intuitive interactions between users and digital systems. It serves as the bridge between human intent and technological execution, ensuring that users can engage effortlessly with software, applications, and devices. This module is designed to enhance accessibility, responsiveness, and overall user experience by incorporating well-structured graphical layouts, interactive elements, and adaptive functionalities. By prioritizing ease of use and efficiency, the UI module eliminates complexities and streamlines interactions, making technology more approachable and inclusive for diverse user groups.

A well-developed UI module dynamically adjusts based on user preferences, device compatibility, and accessibility needs. Whether interacting through touch screens, voice commands, or gesture-based inputs, it tailors itself to the most intuitive mode of engagement. This responsiveness is essential in modern applications, where users expect instant feedback and fluid transitions between actions. Additionally, the module integrates personalization features that allow users to customize layouts, themes, and navigation structures according to their preferences. By offering a tailored experience, it fosters greater engagement, ensuring that interactions feel natural and user-friendly across various platforms.

Security and reliability are also integral to the UI module, providing users with confidence in their interactions. Secure authentication methods, data encryption, and intuitive error-handling mechanisms safeguard user information while ensuring uninterrupted functionality. The module incorporates feedback loops that notify users of errors or required actions in a clear and concise manner, preventing confusion and enhancing system reliability. With real-time updates and improvements, UI designs continuously evolve to meet user expectations while maintaining system stability and security.

The User Interface Module ultimately transforms digital interactions into engaging, efficient, and accessible experiences. By combining aesthetics, usability, adaptability, and inclusivity, it ensures that users can navigate applications effortlessly and interact with technology seamlessly. As advancements continue, AI-driven UI systems will further refine personalization, predictive interfaces, and intelligent interactions, shaping the future of digital connectivity.

6.6 DATA PREPROCESSING MODULE

The Data Preprocessing Module is a fundamental component that ensures images are in the best possible condition for text detection. When raw images are received, they often contain imperfections such as poor lighting, blurriness, unwanted noise, or distortions that can interfere with accurate text extraction. To address these issues, the module applies a series of enhancements to optimize image quality.

First, resizing ensures that all images are uniform in size, allowing the detection system to function consistently without being affected by varying dimensions. Normalization helps balance contrast and brightness levels, so that text remains visible regardless of lighting inconsistencies. Additionally, noise reduction techniques filter out unnecessary disturbances in an image, such as grainy textures or background clutter, which could confuse the text detection algorithm. Together, these processes refine the image, making text clearer and easier to detect.

Beyond basic enhancements, the preprocessing module also corrects distortions that might arise from skewed camera angles or warped text in photographed documents. If a piece of text appears tilted or stretched, the module applies corrective transformations to realign it, ensuring the detection system can interpret characters correctly. By meticulously preparing images through these procedures, the system can handle inputs of varying quality and maintain accuracy across different conditions.

6.7 POST-PROCESSING MODULE

The Post-Processing Module, on the other hand, is responsible for refining the extracted content, ensuring the final output is polished and coherent. Once text is detected, raw extraction can sometimes contain imperfections—such as minor recognition errors, inconsistent formatting, or gaps in clarity. To enhance accuracy, the module implements correction techniques that review the identified text for potential errors, such as misspellings or misplaced characters, and adjust accordingly.

When translation is involved, the module helps refine linguistic aspects, ensuring the translated text maintains proper grammatical structure and context. Certain phrases may require reordering or slight modifications for readability, and this module helps fine-tune the output so that the final version sounds natural. For users relying on synthesized speech, adjustments are made to optimize voice modulation, pacing, and clarity, so the speech sounds smooth and comprehensible rather than robotic or fragmented.

Additionally, the post-processing module plays a crucial role in ensuring uniform formatting. Whether text needs to be structured in paragraphs, lists, or specific styles, this component ensures everything is presented in a well-organized manner. This level of refinement helps eliminate inconsistencies and makes the final output more user-friendly. By enhancing text readability, translation precision, and audio clarity, this module significantly improves the overall experience for users, making the system reliable and effective across different applications.

CHAPTER 7

RESULTS AND PERFORMANCE COMPARISON

7.1 RESULTS

Table 7.1 Performance Comparison

Aspect	EasyOCR + Translation API + TTS Engine *(used)*	Tesseract OCR + Google Translate + gTTS	Amazon Textract + AWS Translate + Amazon Polly
OCR Accuracy	Works well with printed and handwritten text, though complex layouts can be tricky.	Reliable for printed text but struggles with handwriting and detailed formats	Highly accurate for both printed and handwritten text, even in structured documents.
Translation Quality	Provides good accuracy but may have trouble with nuanced phrases.	Decent translations, though some complex text may lose its meaning.	Produces high-quality translations, especially for professional and technical content.
Speech Quality	Generates fairly natural speech but may sound robotic at times.	Understandable speech but lacks expressiveness	Delivers lifelike speech with customizable tones and styles.
Speed & Performance	Fast processing with minimal hardware demands.	Works at a reasonable pace but may slow down with complex tasks.	Optimized for speed and large-scale applications.
Ease to use	Integrates well but requires some setup.	Simple for basic OCR and translation but lacks advanced features.	Powerful but requires familiarity with AWS services.

Using EasyOCR, a Translation API, and a TTS Engine is a great way to get accurate, quick, and user-friendly results for this project. It does a solid job recognizing both printed and handwritten text, making it reliable for different types of content. The translation API ensures that language conversions are smooth and meaningful, even for complex phrases. Plus, the TTS engine delivers clear and natural speech, making the output more engaging and easy to understand. Since this approach is fast, lightweight, and simple to integrate, it's a practical choice for projects that need real-time processing without requiring heavy computing power. By using this method, the project can run efficiently, adapt easily, and provide seamless communication, making it the ideal solution for reaching project goals successfully.

7.2 PERFORMANCE COMPARISON

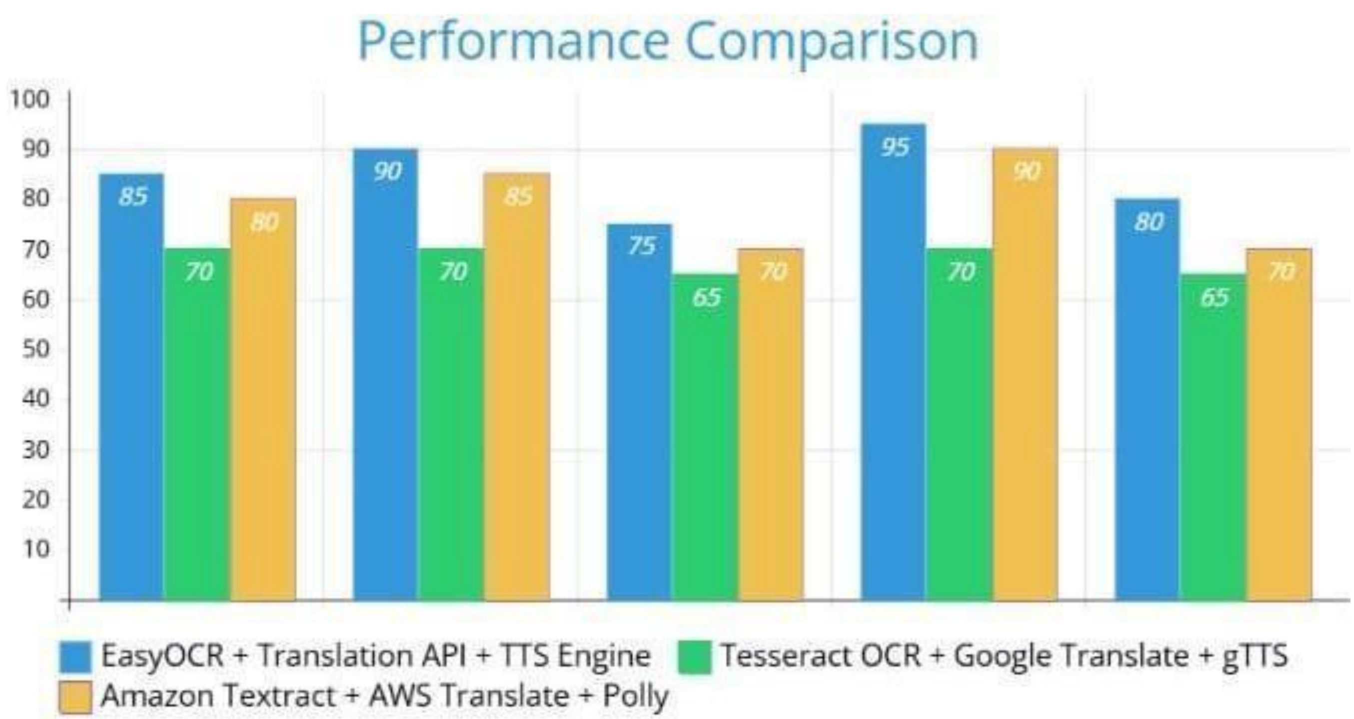

Fig. 7.1 Performance Comparison

CHAPTER 8

CONCLUSION AND FUTURE ENHANCEMENT

8.1 CONCLUSION

In conclusion, the text detection and extraction with text translation and text to speech using deep learning techniques system under consideration employs state-of- the-art deep learning technologies to create an integrated platform for text detection, translation, and text-to-speech synthesis. Through the use of advanced models like EasyOCR for text reading, Transformer networks for accurate translation, and Google Text-to-Speech (gTTS) for synthetic speech of high quality, the system realizes superior accuracy, reliability, and natural-sounding audio output. This converged solution enhances accessibility by translating written content into other languages and formats to serve users with varying language and accessibility requirements. A seamless user experience, supported by enhanced data preprocessing and post- processing modules, provides a fluid and efficient process. Unlike fragmented conventional solutions, this single system does away with the necessity of using multiple tools, providing a streamlined and effortless process. Through breaking the limitations of current approaches, the project offers a very powerful solution that facilitates better communication, rendering written information available to all and interesting for people from everywhere.

8.2 FUTURE ENHANCEMENT

- Future enhancements for the text detection and extraction with text translation and text to speech using deep learning techniques could involve several advanced features to further improve accuracy, usability, and adaptability. Here are a few possible directions:

- Expanded Language Support: To cater to a broader audience, future versions could incorporate support for additional languages, including dialects and low-resource languages. This would make the system more inclusive and versatile.
- Emotion and Tone in Speech Synthesis: Enhancing the text-to-speech synthesis module to convey different emotions and tones could make the synthesized speech more engaging and natural. By incorporating prosody modeling and emotion recognition, the system could generate speech that better reflects human emotions, improving user experience.
- Real-Time Processing and On-Device Deployment: Optimizing the system for real- time processing could significantly enhance usability, particularly in scenarios requiring immediate feedback, such as live translation. Techniques like model compression and efficient hardware utilization could enable on-device deployment, making the system more portable and accessible.
- Interactive and Adaptive Interface: Developing an adaptive user interface that learns from user interactions and preferences could make the system more personalized and intuitive. Features like voice commands, real-time adjustments based on user feedback, and customizable settings could enhance the overall user experience.
- Context-Aware Translation: Incorporating context-aware translation models that consider the broader context of documents or conversations could improve translation accuracy and coherence. Techniques like document-level translation, discourse analysis, and contextual embeddings could be integrated to achieve this.
- Accessibility Features: Adding features such as Braille output, haptic feedback, and customizable font sizes and colors could further enhance accessibility for users with visual impairments or other disabilities. Integrating these features would make the system more inclusive and user-friendly.
- Scalability and Cloud Integration: Developing a scalable architecture that can handle increased demand and integrating with cloud services would ensure the system's reliability and performance. Cloud integration would enable seamless updates, distributed processing, and storage, making the system more efficient and responsive to user needs.

- Integration with Assistive Technologies: Collaborating with developers of existing assistive technologies, such as screen readers and magnifiers, could lead to a more holistic and integrated solution for users with accessibility needs. By combining strengths, the system could provide a more comprehensive and seamless experience.
- Automatic Language Detection: Implementing automatic language detection for input text could streamline the translation process, making it more user-friendly and efficient. This feature would allow the system to automatically detect the language of the input text and translate it without requiring user intervention.
- Customizable Voice Options: Offering a variety of voice options in the text-to-speech module, including different accents, genders, and speech styles, could enhance user satisfaction and engagement. Users could choose the voice that best suits their preferences and needs.

APPENDIX A

SOURCE CODE

FRONTEND:

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>OCR Flask App</title>
  <link rel="stylesheet" href="{{ url_for('static', filename='styles.css') }}">
  <script src="{{ url_for('static', filename='script.js') }}" defer></script>
</head>
<body>

<div class="container">
  <h1 class="title">OCR Flask API</h1>
  <div class="options">
    <label for="imageUpload" class="button">Upload Image</label>
    <input type="file" id="imageUpload" accept="image/*" hidden>

    <label for="videoUpload" class="button">.    Upload Video</label>
    <input type="file" id="videoUpload" accept="video/*" hidden>

    <button id="startLive" class="button">🔴 Start Live OCR</button>
  </div>

  <div class="output">
    <textarea id="outputText" placeholder="Extracted text will appear here..."
```

```
readonly></textarea>
    </div>
</div>

</body>
</html>
```

CSS:

```
body {
    background: linear-gradient(45deg, #141E30, #243B55);
    font-family: Arial, sans-serif;
    color: white;
    text-align: center;
}

.container {
    margin-top: 50px;
    padding: 20px;
}

.title {
    font-size: 32px;
    font-weight: bold;
    margin-bottom: 20px;
}


.options {
    display: flex;
    flex-direction: column;
    gap: 15px;
```

```
    align-items: center;
}

.button {
    background: #007acc;
    padding: 15px 25px;
    font-size: 18px;
    font-weight: bold;
    border: none;
    border-radius: 8px;
    color: white;
    cursor: pointer;
    transition: 0.3s;
}

.button:hover {
    background: #005f99;
    transform: scale(1.05);
}
.output textarea {
    width: 80%;
    height: 200px;
    background: #2f2f2f;
    color: white;
    font-size: 16px;
    border: none;
    padding: 10px;
    border-radius: 8px;
    margin-top: 20px;
}
```

Java Script:

```
document.getElementById("imageUpload").addEventListener("change", function() {
    let formData = new FormData();
    formData.append("file", this.files[0]);

    fetch("/upload_image", { method: "POST", body: formData })
        .then(response => response.json())
        .then(data => document.getElementById("outputText").value = data.text);
});

document.getElementById("videoUpload").addEventListener("change", function() {
    let formData = new FormData();
    formData.append("file", this.files[0]);

    fetch("/upload_video", { method: "POST", body: formData })
        .then(response => response.json())
        .then(data => document.getElementById("outputText").value = data.text);
});

document.getElementById("startLive").addEventListener("click", function() {
    let socket = io.connect("http://" + document.domain + ":" + location.port);
    socket.on("live_text", function(data) {
        document.getElementById("outputText").value = data.text;
    });

    socket.emit("start_live_ocr");
});
```

Python:

```
import os
import cv2
import pytesseract
import easyocr
import numpy as np
from flask import Flask, render_template, request, jsonify
from flask_socketio import SocketIO, emit
from werkzeug.utils import secure_filename
from PIL import Image

# Initialize Flask app
app = Flask(_name_)
socketio = SocketIO(app, async_mode="eventlet")

# Set upload folder
UPLOAD_FOLDER = "uploads"
if not os.path.exists(UPLOAD_FOLDER):
    os.makedirs(UPLOAD_FOLDER)
app.config["UPLOAD_FOLDER"] = UPLOAD_FOLDER

# Configure Tesseract path
pytesseract.pytesseract.tesseract_cmd = r'C:\Program Files\Tesseract-OCR\tesseract.exe'

# Initialize EasyOCR
reader = easyocr.Reader(["en"])

# Route: Home Page
@app.route("/")
```

```python
def index():
    return render_template("index.html")

# Route: Image OCR
@app.route("/upload_image", methods=["POST"])
def upload_image():
    if "file" not in request.files:
        return jsonify({"error": "No file uploaded"})

    file = request.files["file"]
    if file.filename == "":
        return jsonify({"error": "No selected file"})
    filepath = os.path.join(app.config["UPLOAD_FOLDER"], secure_filename(file.filename))
    file.save(filepath)

    # Read and process image
    img = cv2.imread(filepath)
    extracted_text = pytesseract.image_to_string(img)
    return jsonify({"text": extracted_text})
# Route: Video OCR
@app.route("/upload_video", methods=["POST"])
def upload_video():
    if "file" not in request.files:
        return jsonify({"error": "No file uploaded"})

    file = request.files["file"]
    if file.filename == "":
        return jsonify({"error": "No selected file"})

    filepath = os.path.join(app.config["UPLOAD_FOLDER"], secure_filename(file.filename))
```

```
    file.save(filepath)

    # Process video
    cap = cv2.VideoCapture(filepath)
    extracted_text = ""

    while cap.isOpened():
        ret, frame = cap.read()
        if not ret:
            break

        text = pytesseract.image_to_string(frame)
        extracted_text += text + "\n"

    cap.release()
    return jsonify({"text": extracted_text})

# Live OCR with WebSocket
@socketio.on("start_live_ocr")
def start_live_ocr():
    cap = cv2.VideoCapture(0)
    while cap.isOpened():
        ret, frame = cap.read()
        if not ret:
            break
        text = pytesseract.image_to_string(frame)
        emit("live_text", {"text": text})
    cap.release()
# Run Flask app
if _name___== "_main_":
    socketio.run(app, debug=True)
```

APPENDIX B

SCREENSHOTS

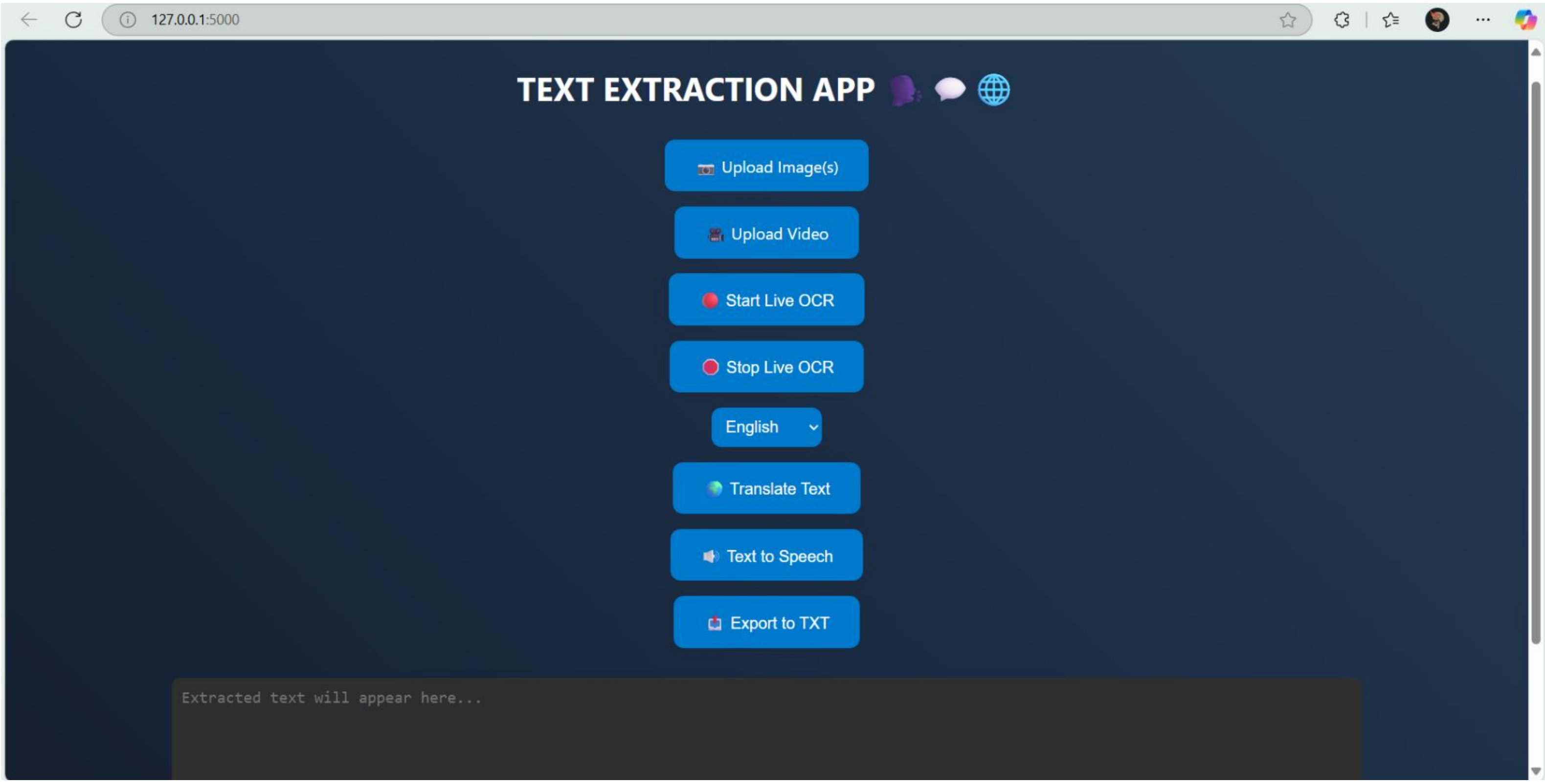

Fig. B.1 Interface

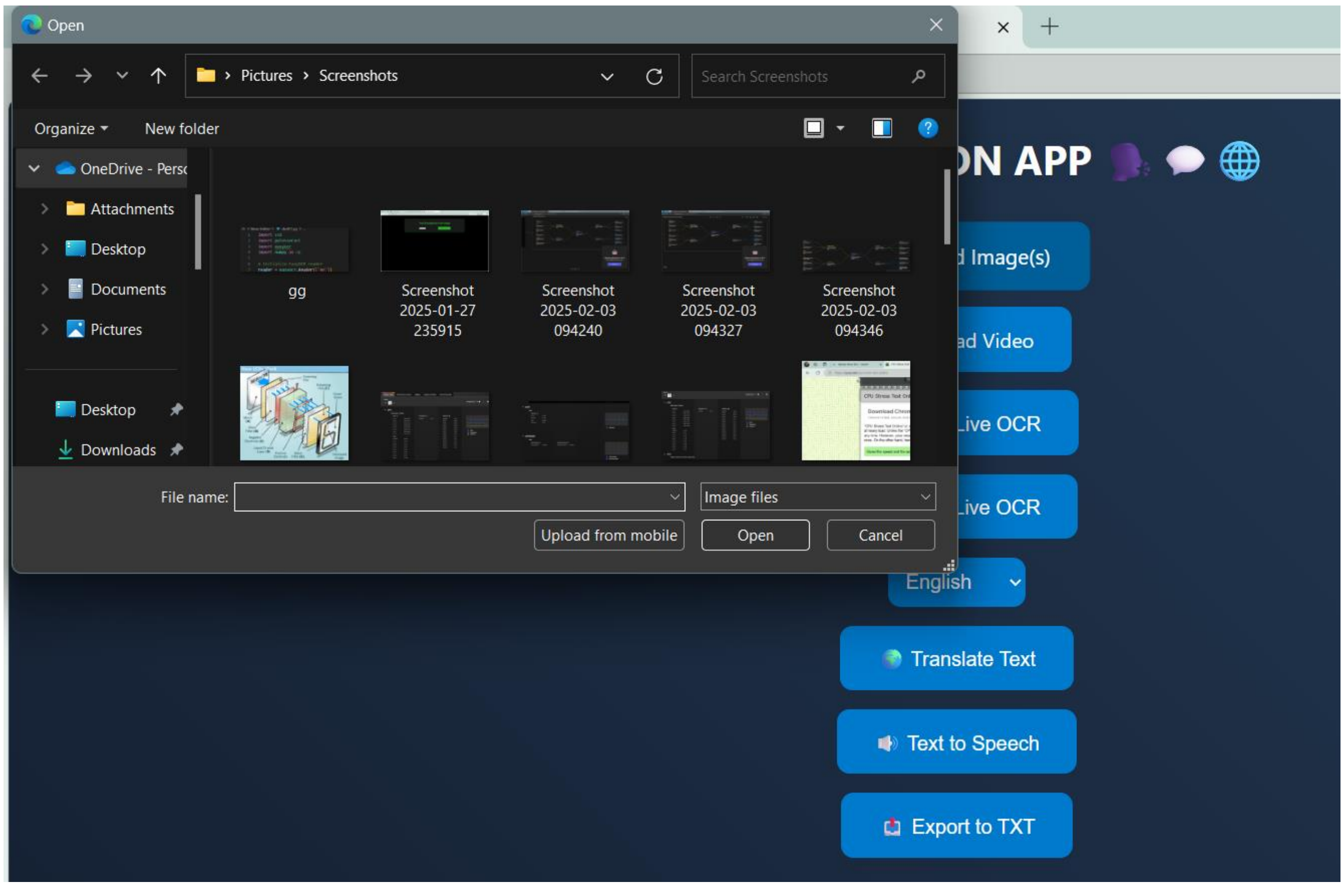

Fig. B.2 Uploading Image

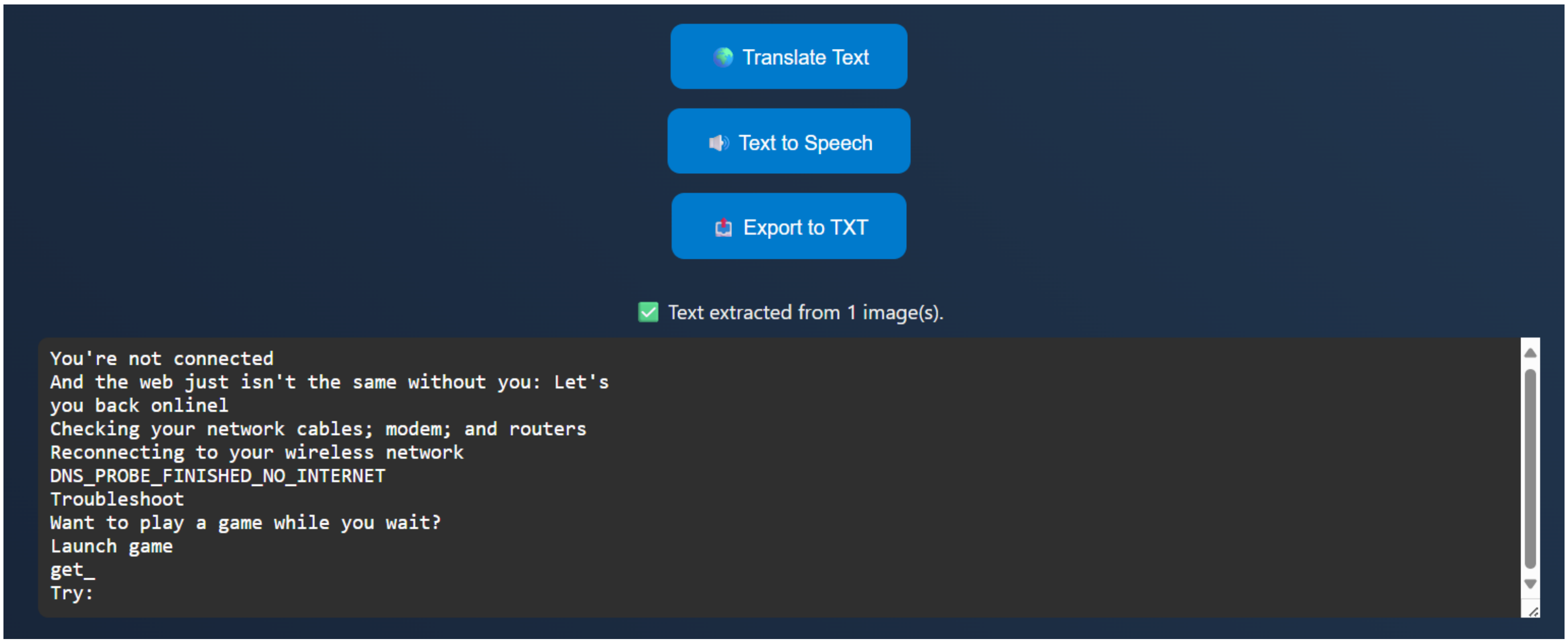

Fig. B.3 Text Extraction

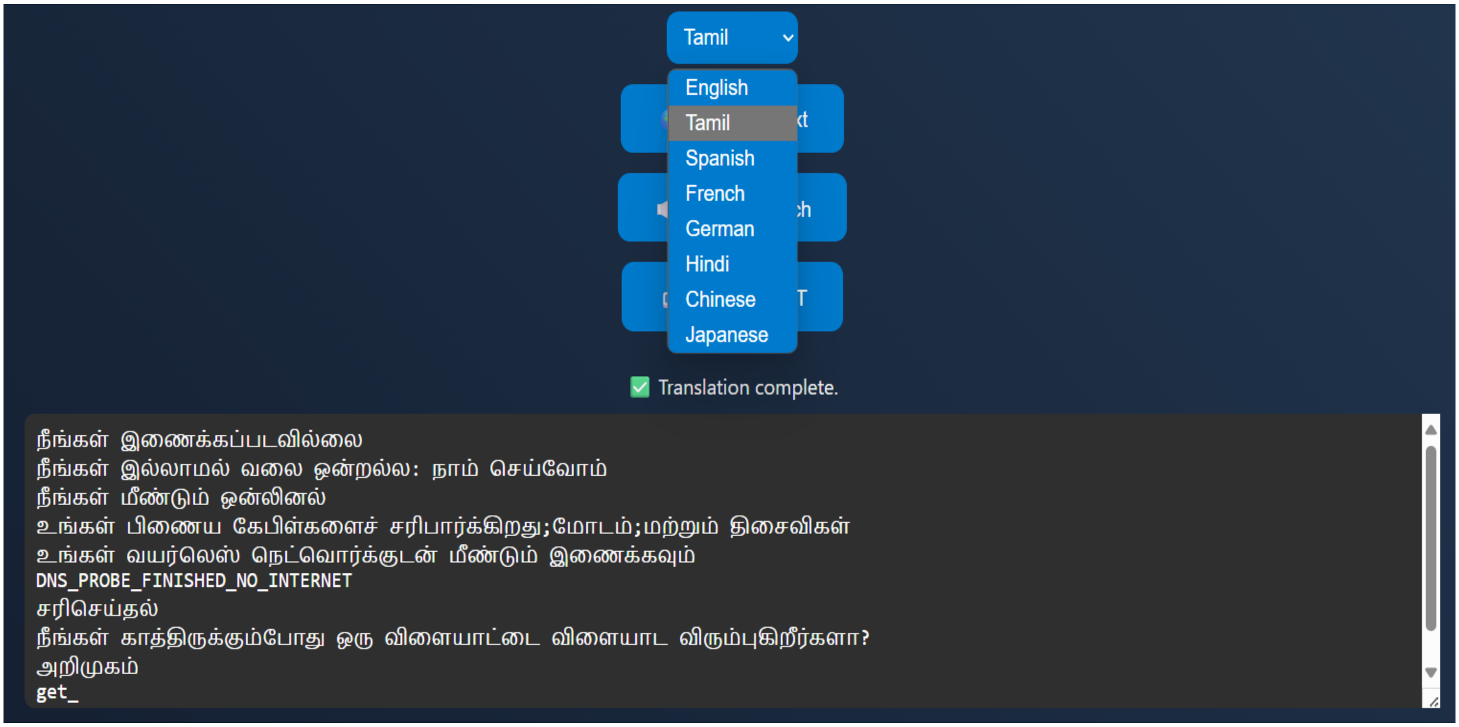

Fig. B.4 Text Translation

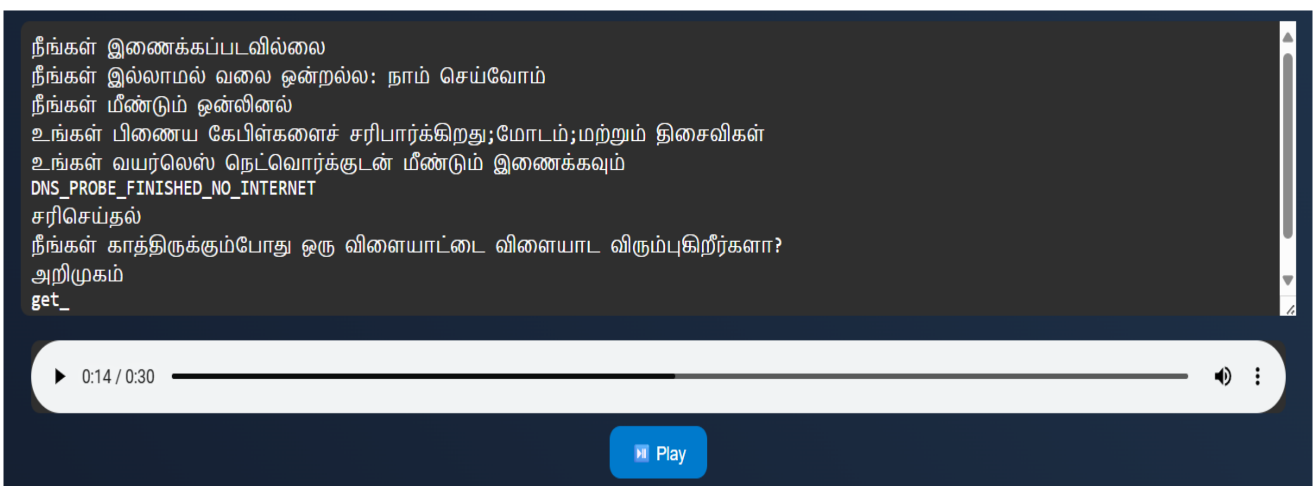

Fig. B.5 Output

REFERENCES

1. Sai Teja Krithik Putcha, Yelagandula Sai Venkata Rajam, K. Sugamya, Sushank Gopala, "Text Extraction and Translation Through Lip Reading using Deep Learning", Thesai Journal, 2025
2. Wang, Jiaan; Meng, Fandong; Zhou, Jie, arXiv: Deep Reasoning Translation via Reinforcement Learning, arXiv-01, 2025
3. Lux, Florian; Meyer, Sarina; Behringer, Lyonel; Zalkow, Frank; Do, Phat; Coler, Matt; Habets, Emanuël A. P.; Vu, Ngoc Thang, arXiv: Meta Learning Text-to-Speech Synthesis in over 7000 Languages, arXiv-02, 2024
4. Wang, Siyang; Szekely, Eva, Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), LREC-COLING-06, 2024
5. Tao, Ye; Lu, Chaofeng; Liu, Meng; Xu, Kai; Liu, Tianyu; Tian, Yunlong; Du, Yongjie, Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), LREC-COLING-07, 2024
6. Baek, Y., Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR-03, 2019
7. Zhou, X., IEEE Transactions on Pattern Analysis and Machine Intelligence, TPAMI-02, 2017
8. Busta, M., International Conference on Document Analysis and Recognition (ICDAR), ICDAR-05, 2017
9. Jaderberg, M., International Journal of Computer Vision, IJCV-01, 2016
10. Shi, B., IEEE Transactions on Image Processing, TIP-04, 2016
11. Luong, M.T., Conference on Empirical Methods in Natural Language Processing (EMNLP), EMNLP-10, 2015

P.S.R. ENGINEERING COLLEGE

An Autonomous Institution, Approved by AICTE & Affiliated to Anna University, Chennai.
Accredited by NBA & NAAC with A+ Grade and listed Under 12 (B) of the UGC Act 1956.
Sivakasi-626-140, Virudhunagar District, Tamil Nadu, India.

INTERNATIONAL CONFERENCE

ON

ADVANCES IN BIOTECHNOLOGY: LEVERAGING AI TOOLS FOR FUTURE INNOVATIONS

BLAIT - 2025

Organized by

DEPARTMENT OF BIOTECHNOLOGY

Certificate of Participation

This is to certify that Mr./Ms./Dr. M. AATHIRAJU
has presented/participated a paper entitled TEXT DETECTION & EXTRACTION USING DEEP LEARNING TECHNIQUES WITH ADD ON FEATURES AS TRASLATION & TEXT TO SPEECH
in the International Conference on Advances in Biotechnology: Leveraging AI Tools for Future Innovations organized by the Department of Biotechnology on March 13/03/2025 & 14/03/2025 at P.S.R. Engineering College, Sivakasi, Virudhunagar District, Tamil Nadu, India.

Coordinator | Convenor | Principal

P.S.R. ENGINEERING COLLEGE

An Autonomous Institution, Approved by AICTE & Affiliated to Anna University, Chennai.
Accredited by NBA & NAAC with A+ Grade and listed Under 12 (B) of the UGC Act 1956.
Sivakasi-626-140, Virudhunagar District, Tamil Nadu, India.

INTERNATIONAL CONFERENCE

ON

ADVANCES IN BIOTECHNOLOGY: LEVERAGING AI TOOLS FOR FUTURE INNOVATIONS

BLAIT - 2025

Organized by

DEPARTMENT OF BIOTECHNOLOGY

Certificate of Participation

This is to certify that Mr./Ms./Dr. S. GOWRI SHANKAR has presented/participated a paper entitled TEXT DETECTION & EXTRCTION USING DEEP LEARNING TECHNIQUES WITH ADD ON FEATURES AS TRASLATION & TEXT TO SPEECH in the International Conference on Advances in Biotechnology: Leveraging AI Tools for Future Innovations organized by the Department of Biotechnology on March 13/03/2025 & 14/03/2025 at P.S.R. Engineering College, Sivakasi, Virudhunagar District, Tamil Nadu, India.

Coordinator	Convenor	Principal

P.S.R. ENGINEERING COLLEGE

An Autonomous Institution, Approved by AICTE & Affiliated to Anna University, Chennai.
Accredited by NBA & NAAC with A+ Grade and listed Under 12 (B) of the UGC Act 1956.
Sivakasi-626-140, Virudhunagar District, Tamil Nadu, India.

INTERNATIONAL CONFERENCE

ON

ADVANCES IN BIOTECHNOLOGY: LEVERAGING AI TOOLS FOR FUTURE INNOVATIONS

BLAIT - 2025

Organized by

DEPARTMENT OF BIOTECHNOLOGY

Certificate of Participation

This is to certify that Mr./Ms./Dr. R. KIRUSHIGAN
has presented/participated a paper entitled TEXT DETECTION & EXTRACTION USING DEEP LEARNING TECHNIQUES WITH ADD ON FEATURES AS TRANSLATION & TEXT TO SPEECH
in the International Conference on Advances in Biotechnology: Leveraging AI Tools for Future Innovations organized by the Department of Biotechnology on March 13/03/2025 & 14/03/2025 at P.S.R. Engineering College, Sivakasi, Virudhunagar District, Tamil Nadu, India.

Coordinator | Convenor | Principal

P.S.R. ENGINEERING COLLEGE

An Autonomous Institution, Approved by AICTE & Affiliated to Anna University, Chennai.
Accredited by NBA & NAAC with A+ Grade and listed Under 12 (B) of the UGC Act 1956.
Sivakasi-626-140, Virudhunagar District, Tamil Nadu, India.

INTERNATIONAL CONFERENCE

ON

ADVANCES IN BIOTECHNOLOGY: LEVERAGING AI TOOLS FOR FUTURE INNOVATIONS

BLAIT - 2025

Organized by

DEPARTMENT OF BIOTECHNOLOGY

Certificate of Participation

This is to certify that Mr./Ms./Dr. M. RAJADURAI
has presented/participated a paper entitled TEXT DETECTION & EXTRACTION USING DEEP LEARNING TECHNIQUES WITH ADD ON FEATURES AS TRANSLATION & TEXT TO SPEECH
in the International Conference on Advances in Biotechnology: Leveraging AI Tools for Future Innovations organized by the Department of Biotechnology on March 13/03/2025 & 14/03/2025 at P.S.R. Engineering College, Sivakasi, Virudhunagar District, Tamil Nadu, India.

Coordinator

Convenor

Principal