

## **Abstract**

The project, **"Image-to-Text and Voice-to-Text using EasyOCR and Speech Recognition,"** presents a comprehensive solution for extracting and transcribing text from both image and audio sources into editable digital format. Leveraging modern deep learning and signal processing technologies, the application offers dual functionality through a user-friendly graphical user interface (GUI) built with Tkinter.

For **image-based text extraction**, the system employs **EasyOCR**, a powerful and flexible OCR engine capable of recognizing both printed and handwritten text from a variety of image formats including PNG, JPG, and JPEG. Users can upload images directly via the GUI, with real-time rendering, resizing, and zooming enhancements powered by Pillow to optimize visual interaction and usability.

The **voice-to-text component** utilizes Python's **SpeechRecognition** library to accurately transcribe spoken words from audio files or live recordings. The application supports common audio formats and incorporates preprocessing techniques to improve recognition accuracy. Additionally, audio playback, recording with waveform visualization, and language selection options are integrated to enrich the user experience.

### **Keywords:**

Image-to-Text, Voice-to-Text, EasyOCR, Speech Recognition, Optical Character Recognition, Audio Transcription, Tkinter, GUI, Deep Learning, Signal Processing, Pillow, Python, OCR Engine, Audio Processing, Language Selection, Real-Time Rendering, Waveform Visualization, Editable Text Extraction, Handwritten Text Recognition, Live Audio Recording