

An OCR-Based Solution for Digitizing Handwritten Old Documents with Regional Language Translation

Abstract:

The preservation and accessibility of handwritten old documents, particularly those in regional languages, pose significant challenges due to the diversity in handwriting styles, material degradation over time, and the complexities of translating text into regional languages. These challenges limit public access to valuable documents, emphasizing the need for efficient digitization solutions.

This idea presents an Optical Character Recognition (OCR)-based solution aimed at digitizing handwritten, old registered documents and translating them into regional languages. The proposed system utilizes advanced OCR technology to convert scanned images of handwritten records into machine-readable text. Furthermore, a regional language translation model is integrated to ensure the translation of the digitized content into local languages, thereby enhancing public accessibility.

Preliminary results indicate that the OCR system efficiently handles various handwriting styles, while the translation model ensures the generation of region-specific language outputs. This approach offers a promising solution for preserving and making old documents more accessible to a broader audience.