This document was scanned from a physical original.

The text layer must be recovered via OCR.

Section 1: Background

Optical Character Recognition (OCR) transforms raster

images of text into machine-readable strings. Modern

OCR engines such as Tesseract and Surya achieve high

accuracy on clean scans with standard typefaces.

Section 2: Challenges

Skew, noise, and low resolution degrade OCR quality.

Confidence scoring helps identify unreliable regions.