

Abstract

Optical Character Recognition (OCR) plays a critical role in digitizing academic literature, enabling search, analysis, and knowledge extraction from scanned documents.

Introduction

However, academic papers present unique challenges to OCR systems due to complex layouts, multi-column formats, mathematical symbols, and varied font styles.

Conclusion

This study evaluates the performance of a modern OCR pipeline on single-page academic documents.