The aim of this study is to develop an automated and scalable PDF text extraction system using Python, incorporating Optical Character Recognition (OCR) and advanced parsing techniques to accurately process diverse document structures while ensuring efficiency, reliability, and usability.

Feature	**PyPDF2**	**PDFMiner**	**Pdfplumber**
Text Extraction	Basic	Advanced	Advanced
Layout Preservation	Minimal	Excellent	Good
Table Extraction	Limited	Moderate	Excellent
Manipulation (Split/Merge)	Excellent	Limited	Limited
Ease of Use	Easy	Complex	Moderate
Performance	Fast	Slower	Moderate
Use Case	Simple PDF manipulation and text extraction	Detailed text extraction from complex PDFs	Table extraction, detailed text extraction