

Arabic OCR Leaderboard (2025–2026)

Rank	OCR Tool / Model	Key Notes on Arabic Accuracy & Usability	Size / OS	Output
1	DeepSeekOCR	✓ Excellent character/word recognition, preserves semantic flow, minimal errors, readable structure. Great for educational texts, exam passages, and structured content.	6.22 GB, Heavy	Outputs Markdown
2	QARI OCR	✓ Strong for Arabic script words sometimes misread. Needs post-processing for clean text.	4.12 GB, Vision-Language (Qwen2-VL-2B-Instruct), fine-tuned on Arabic, Lightweight	Output sometimes fragmented with HTML noise
3	PaddleOCR VL	✓ Very good for Arabic, especially on structured documents and tables. Slightly behind DeepSeek/QARI in character-level accuracy, but excellent for layout preservation.	1.79 GB, Lightweight, Requires WSL/Linux	Outputs Structured Text
4	PaddleOCR v5	✗ Good Arabic recognition on printed text; tested on Windows but does not show top performance.	Lightweight, Windows compatible	
5	Tesseract OCR	✗ Lowest accuracy (74–80% on Arabic). Struggles with cursive, diacritics, and word ligatures. Only suitable for very simple printed text.	Lightweight, Cross-platform	Outputs Plain Text