華中科技大學

MDPBench: A Benchmark for Multilingual Document Parsing in Real-World Scenarios

Can OCR read documents from around the world?

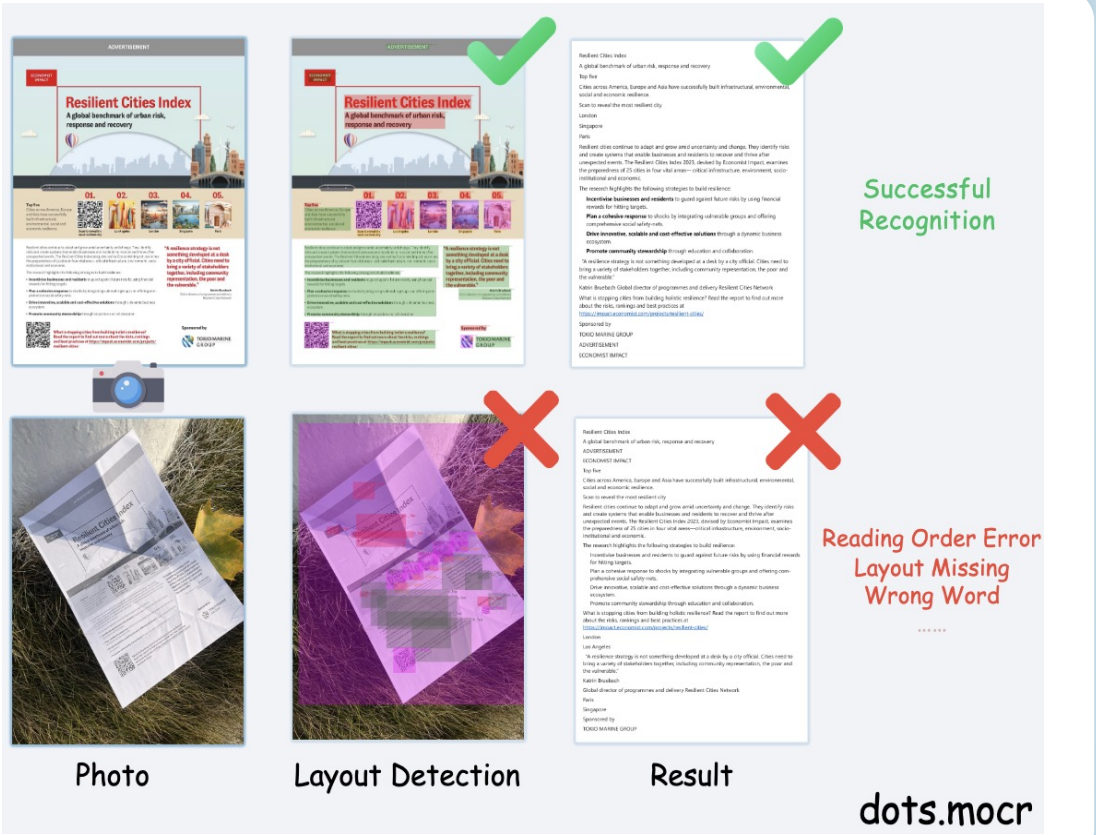
Paper: https://arxiv.org/pdf/2603.28130
GitHub: https://github.com/Yuliang-Liu/MultimodalOCR
Hugging Face: https://huggingface.co/datasets/Delores-Lin/MDPBench
ModelScope: https://modelscope.cn/datasets/DeloresLin/MDPBench

1. Motivation

Better multilingual parsing turns real-world documents into high-quality LLM pre-training data.



Languages preserve diverse knowledge and cultures



Some archives and notes exist only as photographs

2. Existing Benchmarks

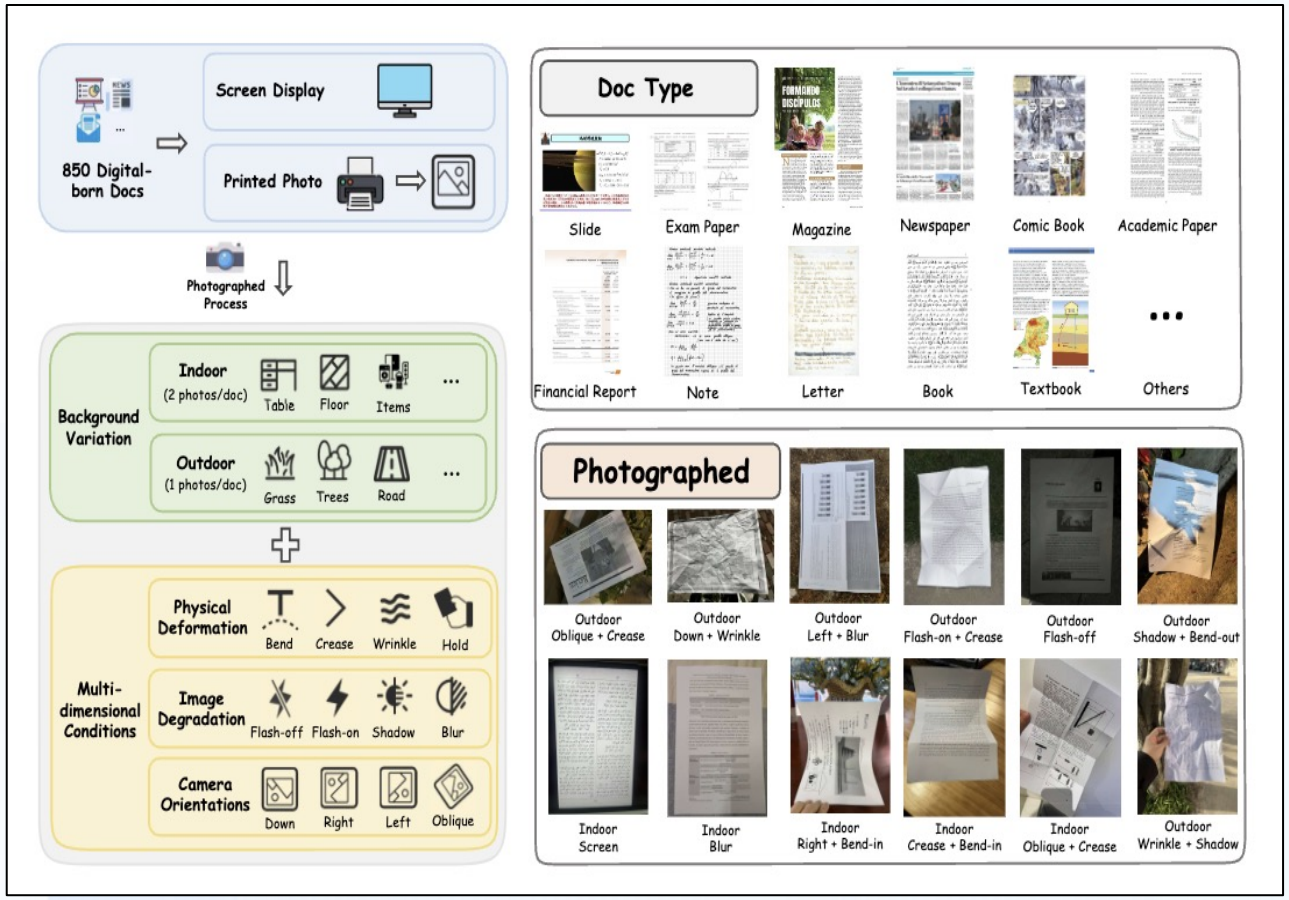
Prior benchmarks mainly cover English and Chinese. MDPBench covers 17 languages.

Table 1: Comparison of document parsing benchmarks.**DB**: Digital-Born; **Photo.**: Photographed; **SR**: Screen Re-photograph; **PD**: Physical Deformation; **ID**: Image Degradation; **CO**: Camera Orientations; **BV**: Background Variation.

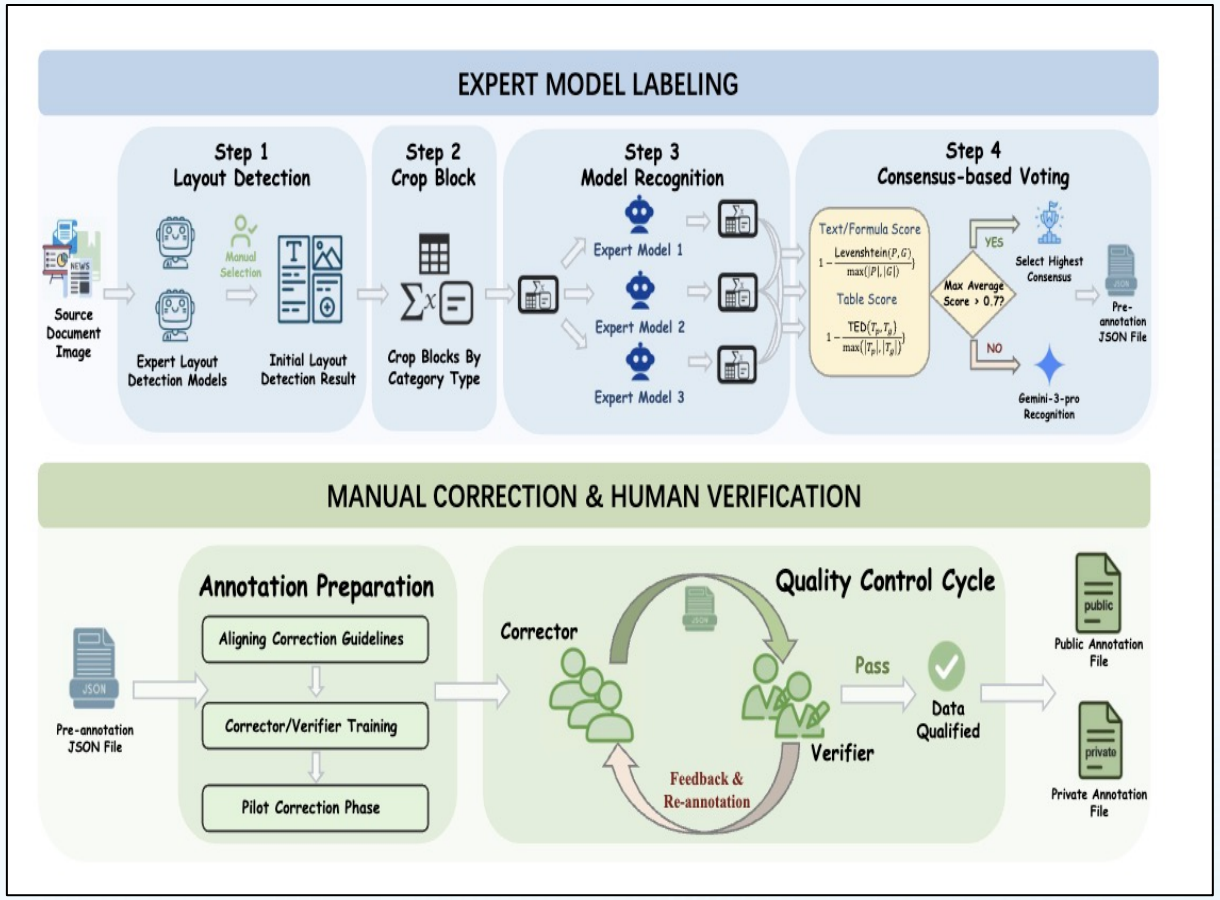
Benchmark	Languages								Type	Image Count	Photograph Conditions				
	Num	ZH	EN	FR	ES	RU	AR	Others			SR	PD	ID	CO	BV
FoxPage [24]	2	✓	✓					—	DB	212					
olmOCR-Bench [34]	1		✓					—	DB	1402					
OmniDocBench-v1.5 [31]	2	✓	✓					—	DB	1355					
DocPTBench [11]	2	✓	✓					—	DB / Photo.	2362		✓	✓	✓	✓
Real5-OmniDocBench [55]	2	✓	✓					—	DB / Photo.	6775	✓	✓	✓		
MDPBench	17	✓	✓	✓	✓	✓	✓	DE, HI, ID, IT, JP, KO, NL, PT, TH, VI, ZH-T	DB / Photo.	3400	✓	✓	✓	✓	✓

3. Building MDPBench

MDPBench uses model labeling, manual correction and independent human verification.



Multilingual document collection



Document parsing annotation

4. Overall Results

Gemini-3-Pro leads at 86.4%. The strongest open-source model, dots.mocr, scores 80.5%.

Table 2: Performance of general VLMs, specialized VLMs, and pipeline tools on MDP-Bench.

Model	Overall			Latin										Non-Latin								Private
	All	Digit.	Photo.	Avg.	DE	EN	ES	FR	ID	IT	NL	PT	VI	Avg.	AR	HI	JP	KO	RU	TH	ZH	ZH-T	All
General VLMs																							
Gemini-3-pro-preview [14]	**86.4**	90.4	**85.1**	**88.4**	**91.2**	**90.6**	**83.4**	**82.7**	**91.5**	**91.6**	**87.7**	**91.4**	85.9	**84.1**	**89.4**	**90.4**	74.8	85.5	**84.9**	**80.6**	**85.1**	**82.1**	**89.8**
kimi-K2.5 [38]	77.5	85.0	75.0	81.6	85.9	86.2	72.7	71.0	80.6	86.6	77.4	87.6	**86.2**	72.9	75.8	74.5	72.5	70.9	61.8	67.0	81.7	78.6	81.2
Doubao-2.0-pro [4]	74.2	78.9	72.8	75.7	82.8	74.4	69.0	70.0	73.3	82.0	69.9	83.4	76.5	72.5	81.3	75.7	65.8	74.7	63.3	71.9	71.9	75.2	79.5
Claude-Sonnet-4.6 [1]	73.1	85.0	69.3	79.2	79.8	80.6	72.8	66.5	82.3	83.3	76.7	88.0	83.1	66.2	67.8	71.7	63.4	64.3	70.8	65.2	61.3	65.1	77.6
ChatGPT-5.2-2025-12-11 [30]	68.6	85.6	63.0	75.2	70.8	79.4	71.4	60.0	77.7	78.5	71.6	85.0	82.1	61.1	64.9	63.4	55.8	65.4	60.7	63.8	56.3	58.7	74.0
Qwen3-VL-Instruct-8b [2]	68.3	78.4	65.0	73.6	73.7	71.4	69.3	66.2	68.5	79.1	78.3	82.2	73.4	62.5	63.1	58.4	59.9	61.9	57.9	62.0	62.6	73.8	70.8
Qwen3.5-Instruct-9B [36]	65.7	74.8	62.7	72.5	72.8	72.0	72.0	64.4	66.2	77.6	74.5	79.1	74.0	58.2	53.4	56.2	55.7	60.3	54.7	56.7	60.8	67.5	68.9
InternVL-3.5-8B [44]	42.7	59.7	37.0	53.4	39.8	64.2	47.5	42.7	53.8	60.6	52.2	63.2	57.0	30.6	8.2	9.0	45.6	30.3	26.1	10.8	55.3	59.3	45.3
Specialized VLMs																							
dots.mocr [52]	80.5	**90.5**	77.2	81.7	82.6	87.4	71.3	70.1	84.5	89.3	83.2	86.8	79.9	79.2	83.3	83.6	**75.0**	78.7	71.2	77.9	84.6	79.6	82.8
PaddleOCR-VL-1.5 [8]	78.3	87.4	75.2	81.2	84.8	83.0	75.7	78.1	83.9	85.2	80.6	80.2	78.9	74.9	71.3	67.7	69.5	**86.0**	76.0	68.4	84.8	75.7	80.7
dots.ocr [20]	76.5	88.8	72.3	79.1	79.7	81.2	69.2	67.1	82.5	87.8	78.8	86.9	79.1	73.5	75.9	77.3	70.6	68.5	66.8	73.3	79.1	76.2	79.7
olmOCR2 [35]	70.4	79.9	67.2	76.7	75.7	77.3	72.5	68.9	70.6	81.0	72.0	88.0	84.0	63.3	59.0	60.8	59.4	70.6	65.8	59.2	68.6	63.4	76.1
PaddleOCR-VL [7]	69.6	87.6	63.6	72.1	78.2	79.3	62.9	66.0	77.4	78.4	67.9	72.0	66.6	66.7	65.8	68.4	59.9	77.8	56.9	57.8	78.2	68.5	70.9
HunyuanOCR [37]	68.3	80.2	64.3	72.4	75.0	73.1	63.0	66.1	69.9	80.3	61.4	81.9	80.6	63.7	68.3	73.1	55.6	68.9	52.2	60.7	66.8	64.2	68.6
GLM-OCR [12]	67.3	77.9	63.7	78.7	82.7	84.5	75.8	76.2	79.7	82.8	80.2	77.4	69.2	54.3	21.7	39.6	65.5	61.2	64.2	27.4	78.5	76.7	68.8
MonkeyOCRv1.5 [50]	65.0	84.3	58.6	67.4	70.8	74.9	55.6	60.3	73.8	75.9	66.3	67.2	61.4	62.4	60.1	56.8	57.0	78.9	51.7	55.6	74.8	64.1	69.0
Nanonets-ocr2-3B [28]	64.2	79.2	59.3	71.4	76.7	76.4	61.8	66.1	68.4	78.5	74.1	74.2	66.0	56.2	60.2	59.2	52.1	54.7	45.5	44.6	68.3	65.1	67.6
Nanonets-OCR-s [27]	63.7	78.8	58.7	71.3	75.1	78.5	61.2	62.5	70.3	81.0	69.6	75.9	67.5	55.0	59.5	61.8	55.9	51.2	43.5	39.5	67.4	61.5	66.6
MonkeyOCR-pro-3B [21]	52.2	68.0	47.0	65.1	71.7	77.9	55.9	62.1	66.2	74.5	66.3	71.1	40.2	37.6	4.6	4.2	55.2	60.5	42.6	9.1	72.2	52.4	53.6
DeepSeek-OCR [47]	51.8	80.7	42.2	54.5	55.0	58.3	44.1	43.2	60.9	69.3	52.4	53.0	54.1	48.9	56.9	52.2	49.1	28.2	36.2	49.4	59.7	59.2	54.5
MinerU-2.5-VLM [29]	46.3	61.9	40.8	63.0	68.8	78.4	54.7	57.3	67.5	75.2	60.4	58.8	46.0	27.4	1.3	9.0	39.1	14.7	8.6	11.3	72.9	62.2	48.7
Pipeline Tools																							
PP-StructureV3 [9]	45.4	56.2	41.7	59.8	60.4	68.7	54.4	49.8	69.6	68.9	55.5	58.4	52.7	28.9	1.0	7.7	56.2	15.4	7.5	11.9	72.2	59.1	49.6
MinerU-2.5-pipeline [29]	33.5	57.6	25.4	46.5	54.3	58.3	38.4	43.6	51.9	56.5	43.9	44.0	27.6	18.7	1.2	5.3	24.5	6.8	4.2	6.4	53.9	47.2	36.2

5. Challenges in Photographed Documents

Photographed documents reduce model scores by 17.8 points on average.



Fig. 4: Visualization of document parsing results of Gemini-3-Pro on photographed documents.

6. Performance Across Scripts

Scores on non-Latin scripts are 14.0 points lower on average than on Latin scripts.

Table 2: Performance of general VLMs, specialized VLMs, and pipeline tools on MDP-Bench.

Model	Overall			Latin									Non-Latin								Private
	All	Digit.	Photo	Avg.	DE	EN	ES	FR	ID	IT	NL	PT	VI	Avg.	AR	HI	JP	KO	RU	TH	ZH	ZH-T	All
General VLMs																							
Gemini-3-pro-preview [14]	**86.4**	90.4	**85.1**	**88.4**	**91.2**	**90.6**	**83.4**	**82.7**	**91.5**	**91.6**	**87.7**	**91.4**	85.9	**84.1**	**89.4**	**90.4**	74.8	85.5	**84.9**	**80.6**	**85.1**	**82.1**	**89.8**
kimi-K2.5 [38]	77.5	85.0	75.0	81.6	85.9	86.2	72.7	71.0	80.6	86.6	77.4	87.6	**86.2**	72.9	75.8	74.5	72.5	70.9	61.8	67.0	81.7	78.6	81.2
Doubao-2.0-pro [4]	74.2	78.9	72.8	75.7	82.8	74.4	69.0	70.0	73.3	82.0	69.9	83.4	76.5	72.5	81.3	75.7	65.8	74.7	63.3	71.9	71.9	75.2	79.5
Claude-Sonnet-4.6 [1]	73.1	85.0	69.3	79.2	79.8	80.6	72.8	66.5	82.3	83.3	76.7	88.0	83.1	66.2	67.8	71.7	63.4	64.3	70.8	65.2	61.3	65.1	77.6
ChatGPT-5.2-2025-12-11 [30]	68.6	85.6	63.0	75.2	70.8	79.4	71.4	60.0	77.7	78.5	71.6	85.0	82.1	61.1	64.9	63.4	55.8	65.4	60.7	63.8	56.3	58.7	74.0
Qwen3-VL-Instruct-8b [2]	68.3	78.4	65.0	73.6	73.7	71.4	69.3	66.2	68.5	79.1	78.3	82.2	73.4	62.5	63.1	58.4	59.9	61.9	57.9	62.0	62.6	73.8	70.8
Qwen3.5-Instruct-9B [36]	65.7	74.8	62.7	72.5	72.8	72.0	72.0	64.4	66.2	77.6	74.5	79.1	74.0	58.2	53.4	56.2	55.7	60.3	54.7	56.7	60.8	67.5	68.9
InternVL-3.5-8B [44]	42.7	59.7	37.0	53.4	39.8	64.2	47.5	42.7	53.8	60.6	52.2	63.2	57.0	30.6	8.2	9.0	45.6	30.3	26.1	10.8	55.3	59.3	45.3
Specialized VLMs																							
dots.mocr [52]	80.5	**90.5**	77.2	81.7	82.6	87.4	71.3	70.1	84.5	89.3	83.2	86.8	79.9	79.2	83.3	83.6	**75.0**	78.7	71.2	77.9	84.6	79.6	82.8
PaddleOCR-VL-1.5 [8]	78.3	87.4	75.2	81.2	84.8	83.0	75.7	78.1	83.9	85.2	80.6	80.2	78.9	74.9	71.3	67.7	69.5	**86.0**	76.0	68.4	84.8	75.7	80.7
dots.ocr [20]	76.5	88.8	72.3	79.1	79.7	81.2	69.2	67.1	82.5	87.8	78.8	86.9	79.1	73.5	75.9	77.3	70.6	68.5	66.8	73.3	79.1	76.2	79.7
olmOCR2 [35]	70.4	79.9	67.2	76.7	75.7	77.3	72.5	68.9	70.6	81.0	72.0	88.0	84.0	63.3	59.0	60.8	59.4	70.6	65.8	59.2	68.6	63.4	76.1
PaddleOCR-VL [7]	69.6	87.6	63.6	72.1	78.2	79.3	62.9	66.0	77.4	78.4	67.9	72.0	66.6	66.7	65.8	68.4	59.9	77.8	56.9	57.8	78.2	68.5	70.9
HunyuanOCR [37]	68.3	80.2	64.3	72.4	75.0	73.1	63.0	66.1	69.9	80.3	61.4	81.9	80.6	63.7	68.3	73.1	55.6	68.9	52.2	60.7	66.8	64.2	68.6
GLM-OCR [12]	67.3	77.9	63.7	78.7	82.7	84.5	75.8	76.2	79.7	82.8	80.2	77.4	69.2	54.3	21.7	39.6	65.5	61.2	64.2	27.4	78.5	76.7	68.8
MonkeyOCRv1.5 [50]	65.0	84.3	58.6	67.4	70.8	74.9	55.6	60.3	73.8	75.9	66.3	67.2	61.4	62.4	60.1	56.8	57.0	78.9	51.7	55.6	74.8	64.1	69.0
Nanonets-ocr2-3B [28]	64.2	79.2	59.3	71.4	76.7	76.4	61.8	66.1	68.4	78.5	74.1	74.2	66.0	56.2	60.2	59.2	52.1	54.7	45.5	44.6	68.3	65.1	67.6
Nanonets-OCR-s [27]	63.7	78.8	58.7	71.3	75.1	78.5	61.2	62.5	70.3	81.0	69.6	75.9	67.5	55.0	59.5	61.8	55.9	51.2	43.5	39.5	67.4	61.5	66.6
MonkeyOCR-pro-3B [21]	52.2	68.0	47.0	65.1	71.7	77.9	55.9	62.1	66.2	74.5	66.3	71.1	40.2	37.6	4.6	4.2	55.2	60.5	42.6	9.1	72.2	52.4	53.6
DeepSeek-OCR [47]	51.8	80.7	42.2	54.5	55.0	58.3	44.1	43.2	60.9	69.3	52.4	53.0	54.1	48.9	56.9	52.2	49.1	28.2	36.2	49.4	59.7	59.2	54.5
MinerU-2.5-VLM [29]	46.3	61.9	40.8	63.0	68.8	78.4	54.7	57.3	67.5	75.2	60.4	58.8	46.0	27.4	1.3	9.0	39.1	14.7	8.6	11.3	72.9	62.2	48.7
Pipeline Tools																							
PP-StructureV3 [9]	45.4	56.2	41.7	59.8	60.4	68.7	54.4	49.8	69.6	68.9	55.5	58.4	52.7	28.9	1.0	7.7	56.2	15.4	7.5	11.9	72.2	59.1	49.6
MinerU-2.5-pipeline [29]	33.5	57.6	25.4	46.5	54.3	58.3	38.4	43.6	51.9	56.5	43.9	44.0	27.6	18.7	1.2	5.3	24.5	6.8	4.2	6.4	53.9	47.2	36.2

7. Language-Specific Errors

Hindi: crucial diacritics are omitted (Arvind is misread as Aravid).

Russian: Cyrillic letters are confused with visually similar Latin letters.

Arabic: models can reverse the right-to-left reading order.

