📊 Scoring Criteria

Each paper was scored on the following 8 metrics. All metrics are evaluated on a **0–5 scale**, defined as follows:

Metric	0	1	2	3	4	5
Paper Availability	Not available	Paywalled	Partially open	Open, lacks DOI	Open w/ DOI	Fully open (e.g. arXiv/Zenodo)
Code & Software Availability	None	Link broken	Incomplete or undocumented	Complete but limited docs	Good docs, installable	Versioned, reproducible, containerized
Dataset Availability	None	Proprietary	Partial	Public but unclear	Licensed and usable	Fully open, well-described
Computer Requirements	Not mentioned	Vague	Basic specs	Moderate specs + OS	OS + RAM + tool versions	Full specs + env (e.g. Conda, Docker)
GPU Requirements	Not mentioned	Assumed only	GPU required, unspecified	GPU mentioned	GPU + RAM specs provided	All specs + setup scripts provided
Documentation Quality	None	Minimal README	Incomplete or unclear	Functional but sparse	Clear and complete	Rich docs + tutorials or examples
Ease of Setup	Not reproducible	Needs manual editing	Works with effort	Mostly works	Setup script available	One-click setup (Docker, VM)
Result Reproducibility	No results	Unverifiable	Partially match	Mostly match	Matches with effort	Fully reproduced

📄 Scorecard Table

Paper ID	Paper	Code	Data	Comp. Req	GPU Req	Docs	Setup	Results	Avg Score	Notes
119	5	4	1	2	1	3	2	3	2.63	Missing datasets/hardware
49	5	4	3	3	3	3	2	4	3.38	High-quality artifacts, no container
39	5	5	4	4	2	4	3	4	3.88	Strongest overall, container missing
100	5	4	4	4	4	3	2	3	3.63	Code missing, but detailed results
191	5	4	4	3	4	4	2	2	3.50	No container, Joern parsing effort
179	5	1	5	4	1	4	3	3	3.25	Great data, weak automation
23	5	5	5	3	1	4	3	4	3.75	Needs Docker to improve
99	5	5	5	3	0	4	3	4	3.63	Mostly reproducible
13	5	5	5	4	4	4	3	4	4.25	Very strong, Docker would perfect it
63	5	5	5	4	4	4	2	4	4.13	Minor env tweaks needed