

Dataset	Strategy	Training Set (Pages)			Test Set (Pages)	
		Original	Processed (256 ²)	Processed (512 ²)	Original	Processed (512 ²)
Benchmark ¹	Following [18, 20]	143	120,174	804	102	582
Nabuco	Two-Fold	15	32,038	90	20	120
	Cross-Validation	20	48,400	120	15	90
CMATERdb	Five-Fold	4	5,308	24	1	6
		4	5,242	24	1	6
	Cross-Validation	4	4,942	24	1	6
		4	4,592	24	1	6
		4	2,140	24	1	6
		4	2,140	24	1	6

¹ "Benchmark" refers to training sets containing DIBCO 2009, H-DIBCO 2010, H-DIBCO 2012, BD, PHIBD, and SMADI; and test sets containing DIBCO 2011, DIBCO 2013, H-DIBCO 2014, H-DIBCO 2016, DIBCO 2017, H-DIBCO 2018, and DIBCO 2019.

