Data: Single−column text in image scans of Arabic Wikipedia pages with noise added artificially (n = 8800; 100 per engine and noise type).
Noise codes: 'col'=colour, 'bin'=binary, 'blur'=blur, 'weak'=weak ink, 'snp'=salt&pepper, 'wm'=watermark, 'scrib'=scribbles, 'ink'=ink stains.