

Dataset Name	Patch Size	Total Patch Count	WSI Count and Usage	Purpose of Use	Study Name	File Size
DigestPath2019	256 × 256	29,334 malignant, 28,419 benign	660 WSIs (250 malignant, 410 benign); patches were extracted from WSIs	To develop a segmentation model using noisy labels	<i>Hard Sample Aware Noise Robust Learning (Zhu et al., 2022)</i>	~ 3 GB
Chaoyang	512 × 512	1816 Normal, 1163 Serrated, 2244 Adenocarcinoma, 937 Adenoma, Total: 6160 (Train: 4021, Test: 2139)	Number of slides not specified; patches were extracted from colon slides and labeled by 3 expert pathologists	To develop a model robust to real-world label noise	<i>Hard Sample Aware Noise Robust Learning (Zhu et al., 2022)</i>	660 MB
NCT-CRC-HE-100K	224 × 224	100,000 patches; 9 tissue types: Adipose tissue, Background, Debris, Lymphocytes, Mucus, Smooth muscle, Normal colon mucosa, Cancer-associated stroma, Colorectal adenocarcinoma epithelium	Patches were extracted from 86 H&E-stained colorectal cancer tissue slides	Tissue classification	<i>Deep learning for colon cancer histopathological images analysis</i>	15.22 GB
WARWICK (GlaS)	160 × 140	165 patches (benign, malignant)	165 images derived from 16 H&E-stained histological sections of stage T3 or T4 colorectal adenocarcinoma	Tumor region segmentation	<i>Deep learning for colon cancer histopathological images analysis</i>	260 MB
CRC-5000	150 × 150	5000 patches, 8 tissue types: Adipose, Complex stroma, Debris, Empty, Lymphocytes, Mucosa, Stroma, Tumor	10 WSIs (5000×5000); patches were extracted	Classification using RESNET model and segmentation using SEGNET	<i>Deep learning for colon cancer histopathological images analysis</i>	1.23GB