Appendix 5

All Study ID's are referring to Appendix 8

| Dataset | Source | Type of Data | Size (No. of Samples used for Experiment) | Classes | Studies that reported the use |
|---|---|---|---|---|---|
| **Public Dataset** | lizard dataset from the CoNIC Challenge | Histopathology Images | 4981 | (1) neutrophil, (2) epithelial, (3) lymphocyte, (4) plasma, (5) eosinophil, (6) connective tissue | [28] [29] |
| | Kather, J. N. et al. Multi-class texture analysis in colorectal cancer histology. Sci. Rep. 6, 27988; | Histopathology Images | 5000 | (1) Debris, (2) Lymphocyte, (3) Mucus, (4) Muscle, (5) Normal, (6) Stroma, (7) Tumor | [39] |
| | | | | (1)ADIPOSE, (2)DEBRIS, (3) STROMA , (4)EMPTY , (5) MUCOSA, (6)TUMOR, (7) LYMPHO, (8) COMPLEX | [10] |
| | | | | (1) Tumor, (2)Stroma, (3) Complex, (4) Lympho, (5) Debris, (6)Mucosa, (7)Adipose, (8) Empty | [42] |
| | | | | (1) Tumour, (2) Stroma, (3) Complex, (4)Lympho, (5) Debris, (6) Mucosa, (7)Adipose, (8)Empty | [12] |
| | | | | (1) Tumor, (2) Stroma, (3) Complex , (4) Lympho, (5) Debris, (6) Mucosa, (7) Adipose , (8) Empty | [31] |
| | TCGA colorectal cancer cohort | Histopathology Images | 700 | (1) MSI, (2) MSS | [41] |
| | Cancer Genome Atlas (TCGA) | Histopathology Images | N/A | (1) high risk (2) low risk [4] | [33] |
| | | | | (1) Strong relevance, (2) Significant, (3) moderate, (4) weak [13] | [1] |
| | | | 1036 | (1) Loose non-tumor, (2) Dense non-tumor tissue, (3) Gastrointestinal Cancer (1) Metastatic, (2) Non-metastatic | [6] |
| | | Text files | 300 | (1) Metastatic, (2) Non-metastatic | [44] |
| | Kvasir V2 | Endoscopy Images | 4000 | (1) Left Cable, (2) Right Cable, (3) Foreceps Head, (4) Snare Head, (5) Non-instrument | [16] |
| | HyperKvasir dataset | Endoscopy Images | 16470 | (1) anatomical landmark, (2) pathological finding | [15] |
| | TCGA-COAD database | Histopathology Images + Text files | 879 | (1) Tumor, (2) Normal + 7 gene mutations can be identified: (1) CHD4, (2) MTOR, (3) MYH9, (4) NIN, (5) PIK3CA, (6) TP53, (7) ROS1 | [2] |
| | Li Kaidong, Mohammad I. Fathan, Krushi Patel, Tianxiao Zhang, Cuncong Zhong, Ajay Bansal, et al., "Colonoscopy Polyp Detection and Classification: Dataset Creation and Comparative Evaluations", arXiv preprint arXiv:2104.10824, 2021. | Endoscopy Images | 4, 719 | (1) Hyperplastic, (2) Adenomatous | [37] |
| | | | 28773 | (1) adenoma, (2) hyperplastic | [30] |

| | | | | | |
|---|---|---|---|---|---|
| | Computer-Aided Classification of Gastrointestinal Lesions in Regular Colonoscopy :http://www.depeca.uah.es/colonoscopy_dataset/ | Endoscopy images | 440<br><br>47 | (1) Adenoma, (2) Hyperplastic, (3) Serrated<br><br>(1) malignant, (2) benign | [11]<br><br>[7] |
| | Warwick-QU Dataset | Histopathology Images | 4602 | (1) malignant, (2) benign | [4] |
| | National Cancer Center of Heidelberg | Histopathology Images | 9200 | (1) Normal Colonic Mucosa, (2) Colorectal Adenocarcinoma Mucosa | [5] |
| | Darthmouth-Hitchcock Medical Center | Histopathology Images | 176 | (1) Hyperplastic polyp, (2) Sessile Serated Polpy ,(3) Traditional Serated Adenoma ,(4) Tubular adenoma ,(5) Tubulovillous | [18] |
| | CRCHistoPhenotypes | Histopathology Images | 100 | (1) Epithelial, (2) Fibroblast, (3) Inflammatory, (4) Miscellaneous | [13] |
| | A. A. Borkowski, M. M. Bui, L. B. Thomas, C. P. Wilson, L. A. DeLand, and S. M. Mastorides, "Lung and colon cancer histopathological image dataset (lc25000) | Histopathology Images | 10000 | (1) benign, (2) adenocarcinoma | [36] |
| | Mayo Polyp Dataset | Endoscopy Images | 95 | (1) Adenoma, (2) Hyperplastic, (3) Serrated | [19] |
| | UniTOPatho | Histopathology Images | 292 | (1) Normal ,(2) Hyperplastic, (3) Tubular Adenoma(high-grade), (4) Tubular Adenoma(Low-grade), (5) Tubulo-Villous Adenoma(High-Grade dysplasia), (6) Tubulo-Villous Adenoma(Low-grade) | [38] |
| | | | | | |
| **Private Dataset** | **Source** | **Type of Data** | **Size (No. of Samples used for Experiment)** | **Classes** | **ID of Studies that reported the use** |
| | Chiba University of School of Medicine | Endoscopy Images | 190 | (1) Protruding, (2) flat, (3) recessed | [25] |
| | The Department of Pathology and Laboratory Medicine at One Medical Center Drive, HB 7261, Lebanon | Histopathology Images | 697 | (1)hyperplastic, (2) sessile serrated, (3) traditional serrated, (4) tubular, (5) tubulovillous/villous | [43] |
| | Gangnam-Real-Time Optical Diagnosis (READI) program / Seoul National University Hospital, Healthcare System Gangnam Center (training set) | Endoscopy Images | 2450 | (1) adenomatous, (2) hyperplastic | [21] |
| | Singapore General Hospital's pathology archives | Histopathology Images | ( check later) | (1) high risk (contains sign of adenocarcinoma or dysplasia) , (2) low risk (contains normal histology/inflammation/reactive changes without signs of adenocarcinoma or dysplasia) | [33] |
| | Department of Pathology, Sahloul University Hospital, Sousse (Tunisia). | Histopathology Images | 391 | (1) MSI, (2) MSS | [40] |
| | The Netherlands Cancer Institute, Antoni van Leeuwenhoek (Amsterdam, the Netherlands). | Hyperspectral imaging | 145 | (1) malignant, (2) benign | [34] |
| | EMIS-I dataset | Endoscopy Images | 14000 | (1) Left Cable, (2) Right Cable, (3) Forceps Head, (4) Snare Head, (5) Non-instrument | [16] |
| | database generated from the contribution of Urduliz and Biodonostia Hospitals | Endoscopy Images | 785 | (1) Not Dangerous [ IP, IS, IIa, IIb] , (2) Dangerous [IIc, when associated with Ip, Is, IIa or IIb], (3) Cancer [III] ( all according to Paris Classification) | [35] |
| | Department of Gastroenterology at the Hospital of Zhengzhou University, Henan Province from July 2017 to August 2018. | Endoscopy Images | 2082 | (1) Polyp, (2) Adenoma | [3] |
| | Sardjto Public Hospital Yogyakarta-Indonesia | Endoscopy Images | 314 | (1) Minimal, (2) Moderate, (3) Progressive | [32] |
| | Department for Internal Medicine (St. Elisabeth Hospital, Vienna). | Endoscopy images | 800 | (1) Healthy, (2) Abnormal | [45] |
| | custom dataset from Ming Chi University of Technology (mentioned from a hospital) | Endoscopy images | 2449 | (1) adenoma, (2) hyperplastic | [30] |
| | Aichi medical university hospital, Japan under the supervision of Dr. Kunio Kas | Endoscopy Images | 4304 | (1) Adenoma, (2) Hyperplastic, (3) Serrated | [11] |
| | Custom dataset (University of New York) | Colonography or Virtual colonoscopy images | 63 | (1) malignant, (2) benign | [17] |

| | | | | | |
|---|---|---|---|---|---|
| | Korea University Medical Center (KUMC), Seoul, Korea | Endoscopy Images + Histology report | 3000 | (1) normal, (2) tubular adenoma with low-grade dysplasia (TALGD), (3) tubular adenoma with high-grade dysplasia (TAHGD), (4) adenocarcinoma | [22] |
| | Research Institute for Nanodevices, Hiroshima University | Endoscopy Images | 6161 | (1) Type 1, (2) Type 2, (3) Type 3 ; NICE classification | [26] |
| | Department of Computer Science and Engineering<br>Sejong University<br>Seoul, Korea | Histopathology Images | 2854 | (1) Well Differentiated, (2)Moderately differentiated (3) poorly differentiated | [23] |
| | Graduate School of Electronic and Electrical Engineering, Kyungpook National University, Daegu 41566, South Korea | Histopathology Images | 7746 | (1) MSI-H , (2) non-MSI-H | [24] |
| | JR Hiroshima Hospital and Hiroshima University Hospital | Endoscopy images | 4056 | (1) Type 1, (2) Type 2, (3) Type 3 ; NICE classification | [27] |
| | An Efficient Method of Histological Cell Image Detection Based on Spatial Information Convolution Neural Network - Proceedings of the 3rd International Conference on Video and Image Processing<br>Custom dataset | Histopathology Images | 20000 | (1) Epithelial (2) Inflammatory (3) Fibroblasts, (4) Heteronuclear | [8] |
| | | | | | |
| | | | | | |
| | | | | | |