Digital pathological slides from Hungarian colorectal cancer screening (Hungarian-Colorectal-Screening)

Redirection Notice

This page will redirect to https://doi.org/10.7937/TCIA.9CJF-0127 in about 10 seconds.

Summary

In this study, 200 digital whole-slide images are published which were collected via hematoxylin-eosin stained colorectal biopsy. This dataset contains the raw MIRAX (mrxs) formatted data. The samples were selected from the archives of the 2nd Department of Pathology of Semmelweis University, Budapest and were scanned with a 3DHistech Pannoramic 1000 Digital Slide Scanner at the highest available, 40x magnification. This is a single center dataset ensuring consequent and homogeneous data processing and patient handling. The related publication shows, how these data can be utilized for training an artificial neural network in order to detect pathological conditions.

Acknowledgements

The research was financed by the Thematic Excellence Programme (Tématerületi Kiválósági Program, 2020-4.1.1.-TKP2020) of the Ministry for Innovation and Technology in Hungary, within the framework of the DigitalBiomarker thematic programme of the Semmelweis University. This work was supported by the National Research, Development and Innovation Office of Hungary grants OTKA 128881 and K128780, the National Quantum Technologies Program and the Hungarian Artificial Intelligence National Laboratory.

Data Access

Data Access

Data Type	Download all or Query/Filter	License
Tissue Slide Images (MRXS, 392 GB)	*Download* *Search* (Download and apply the IBM-Aspera-Connect plugin to your browser to retrieve this faspex package)	CC BY 4.0
Clinical data (CSV, 4 kB)	*Download*	CC BY 4.0

Detailed Description

Detailed Description

Image Statistics	
Modalities	Pathology
Number of Patients	200
Number of Images	200
Images Size (GB)	392

Note about the data:

From the article, these data include hematoxylin- and eosin- (H&E) stained whole slide imaging (WSI), with resolution 0.1213 m/pixel, as acquired by 3DHistech Pannoramic 1000 Digital Slide Scanner.

ICD10 explanations

The supporting csv metadata contains ICD10 codes for each slide. Below are some helpful links about this standard and the differences you might see depending on if you use the international ICD10 codes (which are coded with 4 characters), the Hungarian ICD10 codes (which are coded with 5 characters), or the Institute of 2.Pathology and 1. Pathology at the Semmelweis University which use an extended version of ICD10 and has 6 characters.

Introduction to ICD10:

https://www.cdc.gov/nchs/data/dvs/icd10fct.pdf

https://datadictionary.nhs.uk/supporting_information/international_classification_of_diseases__icd_.html

https://icd.who.int/browse10/Content/statichtml/ICD10Volume2_en_2019.pdf

Code structure, restricted characters, and code length (note that the Hungarian variant does not use * and -):

https://datadictionary.nhs.uk/data_elements/icd-10_code.html

https://www.cdc.gov/nchs/icd/icd10cm_pcs_background.htm

Further notes:

Some ICD10 codes in the list are shorter than 6 characters. These are older samples from when the institute used only 5 characters.

Citations & Data Usage Policy

Citations & Data Usage Policy

Users must abide by the TCIA Data Usage Policy and Restrictions. Attribution should include references to the following citations:

Data Citation

Pataki, B. A., Olar, A., Ribli, D., Pesti, A., Kontsek, E., Gyongyosi, B., Bilecz, A., Kovács, T., Kovács, K. A., Zsofia, Kiss, A., Szócska, M., Pollner, P., & Csabai, I. (2021). *Digital pathological slides from Hungarian (Europe) colorectal cancer screening* (Version 2) [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/TCIA.9CJF-0127

Publication Citation

Pataki, B. Á., Olar, A., Ribli, D., Pesti, A., Kontsek, E., Gyöngyösi, B., Bilecz, Á., Kovács, T., Kovács, K. A., Kramer, Z., Kiss, A., Szócska, M., Pollner, P., & Csabai, I. (2022). **HunCRC: annotated pathological slides to enhance deep learning applications in colorectal cancer screening**. In Scientific Data (Vol. 9, Issue 1). https://doi.org/10.1038/s41597-022-01450-y

TCIA Citation

Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. **The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository**, Journal of Digital Imaging, Volume 26, Number 6, December, 2013, pp 1045-1057. DOI: 10.1007/s10278-013-9622-7

Other Publications Using This Data

TCIA maintains a list of publications which leverage TCIA data. If you have a manuscript you'd like to add please contact the TCIA Helpdesk.

Versions

Version 2 (Current): Updated 2022/09/20

Data Type	Download all or Query/Filter	License
Images (MRXS, 392 GB)	*Download* *Search* (Download and apply the IBM-Aspera-Connect plugin to your browser to retrieve this faspex package)	CC BY 4.0
Clinical Data (CSV)	*Download*	CC BY 4.0

Added missing Data0033.dat files from folders 094, 158, 170, & 186

Version 1: Updated 2022/06/03

Data Type	Download all or Query/Filter	License
Images (MRXS, 392 GB)	*Download* (Download and apply the IBM-Aspera-Connect plugin to your browser to retrieve this faspex package)	CC BY 4.0
Clinical Data (CSV)	*Download*	CC BY 4.0