BMC Research Notes

DATA NOTE Open Access

ERCPMP: an endoscopic image and video dataset for colorectal polyps morphology and pathology

Mojgan Forootan[1], Mohsen Rajabnia[2], Ahmad R. Mafi[1], Hamed Azhdari Tehrani[3], Erfan Ghadirzadeh[1], Mahziar Setayeshfar[4], Zahra Ghaffari[1], Mohammad Tashakoripour[1], Mohammad Reza Zali[1] and Hamidreza Bolhasani[5*]

Abstract

This dataset contains demographic, morphological and pathological data, endoscopic images and videos of 191 patients with colorectal polyps. Morphological data is included based on the latest international gastroenterology classification references such as Paris, Pit and JNET classification. Pathological data includes the diagnosis of the polyps including Tubular, Villous, Tubulovillous, Hyperplastic, Serrated, Inflammatory and Adenocarcinoma with Dysplasia Grade & Differentiation.

Objectives: Today the most important challenge of developing accurate algorithms for medical prediction, detection, diagnosis, treatment and prognosis is data. ERCPMP is an Endoscopic Image and Video Dataset for Recognition of Colorectal Polyps Morphology and Pathology. This dataset can be used for developing deep learning algorithms for polyps detection, classification, and segmentation.

Data description: Images were captured with Olympus colonoscope and are presented in RGB format, JPG type with the resolution of 368 * 256 pixels and 96 dpi. The name of each file (image or video) includes pathological diagnosis, grade and JNet classification of the related polyp.

Keywords Colorectal polyps, Dataset, Endoscopy, Colonoscopy, Morphology, Surface pattern, Pathology, Artificial intelligence

Objective

Colorectal cancer (CRC) is a significant cause of mortality worldwide, responsible for an estimated 1.9 million new cases and 935,000 deaths globally among 5.2 million diagnosed cases in 2020 [1]. It is the third most prevalent malignancy worldwide and the second major cause of cancer-related mortality [1]. Detecting CRC early through screening methods like colonoscopy, fecal occult blood tests, and sigmoidoscopy is crucial for improving patient outcomes, which can detect polyps and early-stage malignancies that can be excised before they progress [2, 3].

*Correspondence:
Hamidreza Bolhasani
hamidreza.bolhasani@srbiau.ac.ir
[1]Gastroenterology and Liver Disease Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[2]Alborz University of Medical Sciences, Alborz, Iran
[3]Department of Hematology and Medical Oncology, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[4]Iran University of Medical Sciences, Tehran, Iran
[5]Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)	Reference
Data File 1	ERCPMP_v5_Morphology_Pathological_Data.xlsx	MS Excel file (.xlsx)	Mendeley https://doi.org/10.17632/7grhw5tv7n.6	[9]
Data File 2	ERCPMP_v5_Images_Vidoes.rar	Image (.JPG) and Video (.MP4)	Mendeley https://doi.org/10.17632/7grhw5tv7n.6	[9]
Data File 3	ERCPMP_v5_Supplementary_Info.pdf	Document (PDF)	Mendeley https://doi.org/10.17632/7grhw5tv7n.6	[9]

Colorectal polyps are atypical growths found in the colon or rectum, often discovered during routine colonoscopy exams [4]. Most CRCs develop from precancerous adenomatous polyps [4, 5]. It has been demonstrated that early diagnosis and excision of precancerous colorectal polyps dramatically reduces the risk of colorectal cancer. The excision of such polyps during colonoscopy can prevent the development of cancer from these lesions [3].

In recent times, considerable endeavors have been undertaken to anticipate and identify various forms of cancer by utilizing artificial intelligence (AI) and its subfields, such as machine learning and deep learning [6–8]. The initial crucial phase towards accomplishing this objective involves obtaining an appropriate dataset. Consequently, this study sought to create a meticulously structured collection of images and videos encompassing demographic information, histopathological attributes (including grading, differentiation, and diagnosis), and morphological characteristics (such as size, circumference, Paris class, Pit pattern, JNET classification, and LST type) of colorectal polyps.

ERCPMP [9] is a histopathological and morphological image and video dataset of 191 patients diagnosed with colorectal polyps including 796 images and 21 videos in total. These numbers are related to the current version, but it is under development to bring more data in the next versions. For queries regarding the latest updates and more information about this dataset, please refer to: https://databiox.com

Data description

ERCPMP [9] is the name of the prepared image and video dataset of this research. This is a morphological and histopathological image and video dataset of 191 patients diagnosed with colorectal polyps including 796 images and 21 videos in total. These numbers are related to the current version, but it is under development to bring more data in the next versions. Images were captured with Olympus colonoscope and are presented in RGB format, JPG type with the resolution of 368 * 256 pixels and 96 dpi. Videos were captured with the same device and are presented in MP4 type. An overview of data files is presented in Table 1, and a summary of technical information of the dataset is introduced in the supplementary file. File names in the dataset are based on patient codes provided in the accompanied excel file. The excel file contains anonymized information on each patient's demographic data in addition to each polyp morphological and histopathological labeling.

Images were captured with Olympus colonoscope and are presented in RGB format, JPG type with the resolution of 368 * 256 pixels and 96 dpi. The name of each file (image or video) includes pathological diagnosis, grade and JNet classification of the related polyp.

The ERCPMP dataset [9] is distinct from similar datasets due to the presence of both morphological and histopathological characteristics of colorectal polyps in addition to including more polyp samples with various features. Generally, six necessary steps were implemented to arrange this dataset as listed:

1. Patients with colorectal polyps were diagnosed and their demographics were recorded.
2. Polyp anatomical features, morphology features, and surface pattern were assessed and classified.
3. Polyp samples were referred for histopathologic assessment.
4. Histologic diagnosis and grading were recorded.
5. A written informed consent was obtained from patients to include their clinical details.
6. The dataset was organized.

Limitations

The limitations of the study were the lack of access to reports, photos, and especially videos, and pathology reports of some patients that had been done in the past. Another limitation was the unpredictability of the pathological type of polyps, which caused asymmetry in the number of different pathological types of polyps.

Abbreviations

CRC	Colorectal Cancer
EMR	Endoscopic Mucosal Resection
ESD	Endoscopic Submucosal Dissection
FTRD	Full-Thickness Resection Device
JNet	Japanese Narrow Band Imaging Expert Team
LST	Laterally Spreading Tumor
LST-G-H	Laterally Spreading Tumor, Granular-Homogenous
LST-G-NM	Laterally Spreading Tumor, Granular-Nodular Mixed
LST-NG	Laterally Spreading Tumor, Non-Granular
LST-NG-FE	Laterally Spreading Tumor, Non-Granular-Flat Elevated
LST-NG-PD	Laterally Spreading Tumor, Non-Granular-Pseudo Depressed
LST-G	Laterally Spreading Tumor, Granular

NBI Narrow-Band Imaging

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13104-024-07062-6.

Supplementary Material 1

Author contributions

M.F. is principal investigators and owner of data.M.Z. is project administrator.E.G. and M.R. wrote the main manuscripts.H.A., A.M., Z.G. revised the paper and helped to gather data.M.F., M.R., and M.T labeled the data. M.R., M.S., Z.G. H.A. prepared the data and labeled them.H.B. organized and created the dataset, methodology and wrote the dataset description.

Funding

Not Applicable.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

This study has been approved by the ethics committee of the Gastroenterology and Liver Disease Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences. According to ethical principles, the datasets completely anonymous. Informed consent was obtained from all subjects and/or their legal guardian(s)" in the ethical approval and consent to participate sub-section.

Competing interests

The authors declare no competing interests.

Received: 27 February 2024 / Accepted: 19 December 2024

Published online: 28 December 2024

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin. 2021;71(3):209–49.
2. Ladabaum U, Dominitz JA, Kahi C, Schoen RE. Strategies for colorectal cancer screening. Gastroenterology. 2020;158(2):418–32.
3. Petimar J, Smith-Warner SA, Rosner B, Chan AT, Giovannucci EL, Tabung FK. Adherence to the World Cancer Research Fund/American Institute for Cancer Research 2018 recommendations for Cancer Prevention and Risk of Colorectal CancerWCRF/AICR recommendations and Colorectal Cancer Risk. Cancer Epidemiol Biomarkers Prev. 2019;28(9):1469–79.
4. Sawicki T, Ruszkowska M, Danielewicz A, Niedźwiedzka E, Arłukowicz T, Przybyłowicz KE. A review of colorectal cancer in terms of epidemiology, risk factors, development, symptoms and diagnosis. Cancers. 2021;13(9):2025.
5. Poudineh M, Poudineh S, Hosseini Z, Pouramini S, Fard SS, Fadavian H, et al. Risk factors for the development of cancers. Kindle. 2023;3(1):1–118.
6. Viscaino M, Bustos JT, Muñoz P, Cheein CA, Cheein FA. Artificial intelligence for the early detection of colorectal cancer: a comprehensive review of its advantages and misconceptions. World J Gastroenterol. 2021;27(38):6399.
7. Wang P, Xiao X, Glissen Brown JR, Berzin TM, Tu M, Xiong F, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nat Biomedical Eng. 2018;2(10):741–8.
8. Hassan C, Wallace MB, Sharma P, Maselli R, Craviotto V, Spadaccini M, et al. New artificial intelligence system: first validation study versus experienced endoscopists for colorectal polyp detection. Gut. 2020;69(5):799–800.
9. Forootan M, et al. ERCPMP-v5: an endoscopic image and video dataset for Recognition of Colorectal Polyps Morphology and Pathology. Mendeley Data. 2024. https://doi.org/10.17632/7grhw5tv7n.6.

Publisher's note