"What You Need to Know" content

WHAT YOU NEED TO KNOW

- **Background**

Capsule endoscopy is a key non-invasive tool for assessing the small bowel in Crohn's disease (CD), but manual review is time-consuming and variable. Artificial intelligence offers the potential for automation and standardization.

- **Findings**

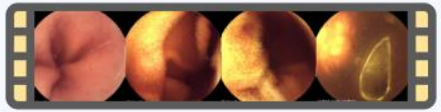
INTELCAPE achieved expert-level accuracy in detecting small-bowel lesions in CD. It decreased interpretation time by 65.7% and boosted clinician diagnostic accuracy by 18.2 percentage points, and was especially helpful for less-experienced readers.

- **Implications for patient care**

INTELCAPE provides rapid, accurate analysis of capsule endoscopy to make timely decisions and reduce workload, and could become a valuable tool in CD management.

Graphical Abstract

Step 1:
Intestine Segmentation Module

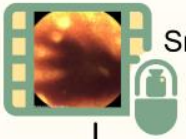


Capsule endoscopy Video

Label

Small intestine segment timespan

Step 2:
Small Intestine Lesion Detection Module



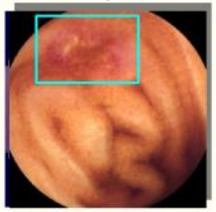
Small intestine video frame

Label

Annotate

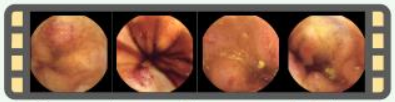
Congestion	Erosion
AU	Deep ulcer
Hemorrhage	Polypoid lesion
Parasitic infection	Others

Intestine lesion types



Lesion bounding box

Step 3:
CD Diagnosis Module



Capsule endoscopy Video

Label

CD/non-CD

AI-assisted Diagnosis

Comparison of 20 doctors of different levels before and after using the artificial intelligence algorithm

Diagnostic accuracy (76. 7%→94.8%)

Interpretation time (67.9 min→22.5 min)

Clinical Gastroenterology and Hepatology

INTELCAPE: A Deep Learning-Powered System for Automated, High-Accuracy Crohn's Disease Diagnosis via Capsule Endoscopy

Short title: AI system for CD via CE

Dejun FAN[1,2,†], Yize Mao[3,†], Feng Liang[4,†], Zheng Liu[5,†], Huayu Li[6,†], Jian Tang[7], Yanan Liu[1], Mingjie Wang[8], Yuting Qian[8], Jie Chen[9], Neng Wang[10], Tao Yang[1], Shuangyi Tan[11], Guanbin Li[12,*], Feng Gao[2,13,*], Jiancong Hu[1,*], Xiaojian Wu[2,14, #,*]

1. Department of Gastrointestinal Endoscopy, The Sixth Affiliated Hospital, Sun Yat-sen University

2. Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, The Sixth Affiliated Hospital, Sun Yat-sen University

3. Department of Pancreatobiliary Surgery, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou 510060, P. R. China.

4. Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Longgang, Shenzhen, China

5. School of Computer Science and Engineering, Sun Yat-sen University

6. State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou 510060, P. R. China.

7. Department of Gastroenterology, The Sixth Affiliated Hospital, Sun Yat-sen University

8. Department of Gastroenterology, Ruijin Hospital, Shanghai Jiao Tong University, School of Medicine

9. Department of Gastroenterology, The Third Xiangya Hospital of Central South University,

Changsha, China.

10. Endoscopy Center of Gastroenterology, Suining Central Hospital

11. School of Science and Engineer, The Chinese University of Hong Kong, Shenzhen

12. Computer Science and Engineering, Sun Yat-sen University

13. Biomedical Innovation Center, The Sixth Affiliated Hospital, Sun Yat-sen University

14. Department of General Surgery (Colorectal surgery), The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, 510080

[†]These authors contributed equally

[#]Lead contact

Correspondence:

Guanbin Li

Computer Science and Engineering, Sun Yat-sen University

E-mail: liguanbin@mail.sysu.edu.cn

Feng Gao

Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, The Sixth Affiliated Hospital, Sun Yat-sen University

Biomedical Innovation Center, The Sixth Affiliated Hospital, Sun Yat-sen University

E-mail: gaof57@mail.sysu.edu.cn

Jiancong Hu

Department of Gastrointestinal Endoscopy, The Sixth Affiliated Hospital, Sun Yat-sen University

E-mail: hujianc@mail.sysu.edu.cn

Xiaojian Wu

Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, The Sixth Affiliated Hospital, Sun Yat-sen University

E-mail: wuxjian@mail.sysu.edu.cn

Grant support:

This research was supported in part by Noncommunicable Chronic Diseases-National Science and Technology Major Project (2024ZD0520105) and The program of Guangdong Provincial Clinical Research Center for Digestive Diseases (2020B1111170004)

Disclosures:

No conflicts of interest.

Data transparency statement:

- For all data, please email all requests for academic use of raw and processed data to the lead contact (wuxjian@mail.sysu.edu.cn). All requests will be evaluated based on institutional and departmental policies to determine whether the data requested are subject to intellectual property or patient privacy obligations. Data can only be shared for non-commercial academic purposes and will require a formal materials transfer agreement.
- All original code has been deposited at https://github.com/liuzh385/INTELCAPE and are publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Abbreviations:

AI – artificial intelligence; AUC – area under the curve; CD – Crohn's disease; CE – capsule endoscopy; CI – confidence interval; IoU – intersection over union; MRMC – multi-reader, multi-case; OR – odds ratio; pp – percentage points; SAHSYSU – Sixth Affiliated Hospital of Sun Yat-sen University

ABSTRACT

Background and aims:

Capsule endoscopy (CE) is a non-invasive technique for diagnosing Crohn's disease (CD); however, manual interpretation of CE videos is time-consuming and error-prone. We developed an artificial intelligence system, INTELCAPE, to automate CE video analysis for accurate and efficient CD diagnosis.

Methods:

This retrospective, multi-center study used data from two Chinese hospitals. A multi-task deep learning framework segmented small-intestine regions, detected lesions, and diagnosed CD using CE videos from 757 (Cohort 1) and 115 (Cohort 2) patients. INTELCAPE integrated the ResNet, Transformer, and EfficientNet architectures for hierarchical processing. Performance was benchmarked against clinicians using three metrics. This study received Ethics Committee approval (2024ZSLYEC-040).

Results:

INTELCAPE achieved state-of-the-art performance across all tasks. For small-intestine segmentation, the model showed intersection over union (IoU) scores of 94.82% (Cohort 1, 95% confidence interval [CI] = 93.28%–96.36%) and 96.87% (Cohort 2, 95% CI = 94.63%–99.12%). For lesion detection, it achieved area under the curve (AUC) values of 0.993 (Cohort 1) and 0.980 (Cohort 2), with 99.33% classification accuracy, which was comparable to that of specialists (97.83%) but superior to that of residents (91.05%, $p < 0.001$). For CD diagnosis, INTELCAPE demonstrated robust generalizability, achieving AUCs of 0.982 (Cohort 1) and 0.984 (Cohort 2) with 90% diagnostic accuracy, comparable to that of specialists (93.33%) but 10-fold faster ($p < 0.001$). INTELCAPE improved doctors' diagnostic accuracy (76.7% to 94.8%, $p < 0.001$), while reducing their interpretation time (67.9 to 22.5 min, $p < 0.001$).

Conclusion:

INTELCAPE improved CD diagnosis by automating CE video analysis, thereby enhancing accuracy and efficiency, particularly for less-experienced clinicians.

Keywords: capsule endoscopy; artificial intelligence; Crohn’s disease; AI-assisted diagnosis

INTRODUCTION

Crohn's disease (CD) is a chronic, progressive, inflammatory intestinal disease that can involve any region of the gastrointestinal (GI) tract, and the small intestine is one of the most commonly affected areas.[1, 2] Accurate diagnosis and monitoring of CD remain challenging, particularly as its incidence and prevalence continue to increase globally.[3-5] Conventional approaches, including enteroscopy and radiographic imaging, show limitations in detecting early-stage disease, particularly in the small intestine, where lesions can be subtle or difficult to visualize.[6, 7] Although standard ileocolonoscopy is essential, it primarily visualizes the terminal ileum and colon, leaving most of the small bowel unexamined. This represents a major diagnostic gap, since up to one-third of patients with CD may have disease confined to the proximal small bowel.[8] While enteroscopy techniques can assess the small bowel, they are associated with several limitations.

Capsule endoscopy (CE) is an important non-invasive tool for evaluating the small bowel and is well established in the diagnosis and management of small-bowel CD. Its primary advantage is its non-invasiveness, which allows direct visualization of the entire small-intestine mucosa without the need for sedation or the risks associated with conventional endoscopy.[9] In comparison with enteroscopy, CE is less invasive and generally well-tolerated, and it enables visualization of mucosal abnormalities in otherwise inaccessible regions.[10] Consequently, CE is effective for detecting characteristic small-intestine lesions of CD, such as ulcers, erosions, and inflammatory changes, and can inform treatment decisions.[11, 12] Despite its notable merits, CE does have limitations. Notably, the large volume of image data generated by CE makes manual review time-consuming and susceptible to human error.[13, 14] Therefore, approaches that mitigate some of these limitations of CE while leveraging its advantages may improve its diagnostic utility.

Artificial intelligence (AI) offers a potential approach to address some of these limitations.[15] AI-driven algorithms, particularly deep learning models, have shown promise in automating medical image analysis, offering potential benefits for diagnostic accuracy and efficiency.[16-18] These algorithms can streamline image interpretation, reduce human error, and provide more consistent diagnoses. However, only a few studies have attempted AI analysis of CE videos and investigated how they may assist in CD diagnosis. Several preliminary studies have investigated AI algorithms for specific tasks in CE video analysis—such as detecting ulcers or stenosis—and for facilitating CD diagnosis and stratification.[19] Despite encouraging preliminary results, additional investigations are required to develop and validate integrated AI systems capable of comprehensively evaluating full CE videos in patients with CD. Therefore, in this multi-center study, we developed an AI-based system, INTELCAPE, to support CD diagnosis. This system automatically extracts small-intestine segments from full CE videos, identifies suspicious lesions, and provides diagnostic predictions, with the aim of addressing some of the limitations of conventional CE review.

METHODS

Study Design

This retrospective, multi-center study was conducted using data from The Sixth Affiliated Hospital of Sun Yat-sen University (SAHSYSU; Cohort 1, a gastroenterology specialty hospital) and Ruijin Hospital (Cohort 2, a tertiary care hospital) in China for model development and fine-tuning. A reader-assistance assessment was further performed as a simulated evaluation using retrospective data. These centers employed different CE devices, with video resolutions of 360 × 360 and 576 × 576 pixels. Supplementary Table 1 presents the summary of patients and videos.

The final diagnosis of CD was determined by multidisciplinary team consensus, incorporating clinical, laboratory, radiological, and endoscopic data. The CE findings were adjunctive, with diagnosis primarily relying on histopathological results and cross-sectional imaging. Cohort enrollment was based on CE availability rather than CE findings to minimize selection bias.

We sought to improve diagnostic accuracy and efficiency for CD across varying clinician experience levels. INTELCAPE, a multi-task deep learning system, was developed to perform small-intestine segmentation, lesion detection, and CD diagnosis from full CE videos (Figure 1A). The system demonstrated accurate diagnostic predictions and efficient computational performance, allowing complete diagnostic processing within clinically feasible timeframes (Figure 1B). Two expert gastroenterologists established ground truth labels. The Ethics Committee of SAHSYSU approved the study's protocol (approval number: 2024ZSLYEC-040), and informed consent was obtained.

Participants and Dataset

INTELCAPE was trained and tested on 757 retrospective CE videos from SAHSYSU and 115 videos from Ruijin Hospital, which represented a distinct patient population and clinical setting (Figure 1B). Due to differences in device configuration between the two hospitals, which can decay model versatility, models of different stages were initially trained on the larger SAHSYSU dataset and fine-tuned using the smaller Ruijin Hospital dataset. Eligible patients underwent CE between January 2015 and September 2023 and met one of the following criteria: (i) no small-intestine lesions identified on CE (no lesion group), (ii) a diagnosis of CD confirmed by histopathological or clinical gold standard assessments (including initial diagnosis, follow-up, and remission periods), or (iii) small-intestine lesions observed on CE

with CD ultimately excluded after a minimum follow-up period of 12 months. These lesions included erosions and ulcers attributable to alternative etiologies (e.g., non-steroidal anti-inflammatory drug [NSAID]-related or non-specific inflammatory changes). This design ensured that the dataset reflected clinically relevant confounding lesions commonly encountered in routine practice. In CE studies, the capsule was required to traverse the ileocecal valve without retention, and the original images had to meet predefined quality standards for analysis. The exclusion criteria included prior intestinal resection, the need for endoscopic assistance for capsule insertion, incomplete data, or the absence of diagnostic confirmation. A total of 1005 videos were initially enrolled in this study. After the application of the inclusion criteria, 133 videos (accounting 13.2%) were excluded because they showed (1) incomplete small-intestine examination (n = 124); (2) presence of debris, bubbles, or poor illumination (n = 5); or (3) excessive blurring or rapid motion (n = 4). Ultimately, 872 videos were included for further analysis.

Overview of the INTELCAPE System

INTELCAPE employs a hierarchical three-stage processing pipeline, including modules for intestine segmentation, lesion detection, and CD diagnosis (Figure 1A). The intestine-segmentation module distinguishes the small-intestine section from CE videos, with the workflow depicted in Figure 2A. This module utilizes a ResNet-Transformer architecture (Figure 2B). The lesion-detection module identifies lesion image frames in the small-intestine section and localizes lesion regions using bounding boxes. Figure 3A depicts the workflow of the lesion-detection module. This module utilizes an efficient convolutional neural network-based model for per-image lesion probability prediction and a background-aware, weakly supervised localization mechanism incorporating B-CAM[20] for lesion region localization (Figure 3B). Furthermore, the CD diagnosis module conducts a diagnosis of CE videos based

on a sequence of lesion images filtered in the previous stage. The workflow is depicted in Figure 4A. It utilizes a Transformer-based model for video-level CD classification through spatiotemporal analysis of CE videos (Figure 4B). To assess clinical assistance ability, we adopted the multi-reader, multi-case (MRMC) framework and compared the diagnostic accuracy and interpretation time of 20 clinicians of varying experience levels (Figure 5A). These clinicians diagnosed the same set of 30 patients before and after INTELCAPE assistance with a washout period of 5 weeks. As shown in Figure 5C, CE videos were preprocessed to prepare data for training the three modules by verifying video quality and unifying the video format and resolution. Expert clinicians labeled the data for different modules. Details regarding data processing and labeling, the implementation of the three modules in the INTELCAPE pipeline, and clinician assistance are described in Supplementary Methods.

The accuracy and efficiency of INTELCAPE were evaluated for small-intestine segmentation, lesion detection, and CD diagnosis, and the results were compared to those obtained by three clinician groups (resident, general, and specialist clinicians) with varying years of experience. We further investigated the extent to which the model could enhance diagnostic accuracy and efficiency by assisting clinicians with different levels of experience, particularly less-experienced clinicians.

Statistical Analysis

Model performance was evaluated by measuring the intersection over union (IoU) and area under the curve (AUC) for segmentation and classification tasks, respectively. Intestine-segmentation performance was evaluated at the video level by comparing the predicted and reference small-intestine segments using IoU values. Lesion-detection performance was evaluated at the frame level by comparison with expert frame annotations. The diagnostic performance for Crohn's disease was evaluated at the video level, with one aggregated

prediction per CE study compared against the reference standard diagnosis. Clinician performance was compared using the Student's *t*-test in Python, with statistical significance set at $p < 0.05$. Diagnostic accuracy and efficiency in the clinical assistance assessment were analyzed using an MRMC framework. The accuracy of binary diagnosis results was evaluated using a generalized linear mixed model. Efficiency was assessed using a mixed-effects linear regression model on log-transformed reading time (Python statsmodels v0.14.4). All authors had access to the study data and reviewed and approved the final manuscript.

RESULTS

CE Segmentation

The performance of INTELCAPE in small-intestine segmentation was evaluated using the corresponding test dataset (Figure 2A). Figure 2C shows the schematic diagrams of the stomach, small intestine, and large intestine from the two capsule videos. The timeline of the CE video frames is shown in Figure 2D. Comparison of the model prediction of the small-intestine segment with the ground truth yielded a high IoU score of 94.63%. The ground truth of the small-intestine transit times was comparable between the two datasets (Cohort 1: 276.5 ± 108.8 min; Cohort 2: 272.0 ± 115.7 min; Figure 2E). Additionally, the model achieved overall AUCs of 0.928 (95% confidence interval [CI]: 0.92–0.93) and 0.952 (95% CI: 0.95–0.96) in the two datasets, demonstrating robust performance in GI tract localization. The AUCs were 0.929 (95% CI: 0.92–0.93) and 0.941 (95% CI: 0.93–0.96) for stomach segmentation and 0.928 (95% CI: 0.92–0.93) and 0.922 (95% CI: 0.90–0.94) for small-intestine segmentation across the two datasets. In the Cohort 1 test dataset, large-intestine segmentation achieved an AUC of 0.931 (95% CI: 0.92–0.94) (Figure 2F). When evaluated using the IoU metric, the model achieved values of 94.82% (95% CI: 93.28%–96.36%) and 96.87% (95% CI: 94.63%–99.12%)

for the Cohort 1 and 2 test datasets, respectively (Figure 2G). For cohorts 1 and 2, the model achieved sensitivities of 85.24% and 81.16%, specificities of 92.62% and 90.58%, positive predictive values of 85.24% and 81.16%, and negative predictive values of 92.62% and 90.58%, respectively (Supplementary Table 2). Subclass results are presented in Supplementary Tables 3–5. The model significantly outperformed clinical experts in terms of segmentation efficiency. A comparative analysis of 20 CE videos showed that the model processed each video in an average of 78.3 ± 19.0 s, substantially faster than specialist (217.2 ± 42.9 s), general (278.5 ± 55.3 s), and resident (354.2 ± 63.8 s) clinicians (Figure 2H).

Small-Intestine Lesion Detection

INTELCAPE's performance in small-intestine lesion classification was evaluated on the corresponding test dataset (Figure 3A). Figure 3C shows images illustrating lesion location and heat maps for the eight types of small-intestine lesions. The model achieved AUCs of 0.993 (95% CI: 0.99–0.99) and 0.980 (95% CI: 0.96–0.99) in the Cohort 1 and 2 test datasets, respectively (Figure 3D). To further validate the superiority of the model, a total of 1,800 images were selected. They were divided into three groups (600 images per group), and the model's performance was compared with that of three clinicians with varying levels of experience (Figure 3A). In terms of classification accuracy, INTELCAPE achieved a mean accuracy of 99.33% ± 0.3%, significantly outperforming specialist (97.83% ± 1.2%), general (94.94% ± 1.8%), and resident clinicians (91.05% ± 2.4%) (Figure 3E). Moreover, the model's average processing time of 20 ± 0.5 ms was faster than that of specialist (5.1 ± 0.5 s), general (6.2 ± 0.8 s), and resident (8.2 ± 1.4 s) clinicians (Figure 3F). The model achieved sensitivities of 99.35% and 91.98%, specificities of 98.74% and 93.64%, positive predictive values of 98.05% and 93.56%, and negative predictive values of 99.58% and 92.80% in cohorts 1 and 2, respectively (Supplementary Table 6).

CD Diagnosis

The CD diagnosis performance of INTELCAPE was evaluated on the corresponding test data (Figure 4A). Figure 4C shows the representative CD images from both centers, with three images selected from each center. For CD diagnosis, INTELCAPE achieved AUCs of 0.982 (95% CI: 0.95–1.00) and 0.984 (95% CI: 0.93–1.00) in the Cohort 1 and 2 test datasets, respectively (Figure 4D). To further assess clinical utility, a comparative analysis was conducted using 30 CE videos (10 CD-positive and 20 CD-negative cases) stratified into three balanced cohorts (Figure 4A). The diagnostic performance of INTELCAPE was benchmarked against those of three clinicians with varying levels of clinical expertise. It achieved a mean diagnostic accuracy of 90.00% (Figure 4E), comparable to that of specialist clinicians (93.33% ± 4.71%) and superior to those of general (86.67% ± 12.47%) and resident (83.33% ± 9.43%) clinicians. Notably, INTELCAPE demonstrated substantial efficiency gains in diagnostic processing time (Figure 4F), averaging 265 ± 107 s in analyzing 10 video cohorts, a 10-fold reduction in comparison with the processing time of the specialist clinicians (2632 ± 373 s). This efficiency advantage was more pronounced against general (2922 ± 515 s) and resident (4009 ± 596 s) clinicians. The model achieved sensitivities of 100% and 100%, specificities of 85.00% and 81.25%, positive predictive values of 77.78% and 72.73%, and negative predictive values of 100% and 100% in cohorts 1 and 2, respectively (Supplementary Table 7). Supplementary Figure 4 shows lesion subgroup confusion matrices for CD classification.

To assess the robustness of INTELCAPE's diagnostic performance in clinically challenging scenarios, we conducted a sensitivity analysis involving cases where the initial multidisciplinary team (MDT) diagnosis was "CD not excluded" (n = 89). These represented patients with equivocal findings for whom a definitive diagnosis could not be established at the time of CE. We evaluated the model's performance under two extreme scenarios: a worst case

for specificity, where all indeterminate cases were conservatively considered true negatives (non-CD), and a best case for sensitivity, where all were considered true positives (CD). Under the worst-case assumption, the model's specificity for CD diagnosis was 0.2022. Under the best-case assumption, the sensitivity was 0.8364. These analyses delineated the upper and lower bounds of performance for this complex subgroup and highlight that the system's output should be interpreted with particular caution in cases with inherently ambiguous clinical and endoscopic presentations.

Reader Assistance

In this phase, an MRMC analysis was used to evaluate the diagnostic accuracy and reading time of clinicians with different experience (Supplementary Table 8) before and after assistance from INTELCAPE. INTELCAPE significantly improved diagnostic accuracy and reading time across all readers and subgroups. As shown in Figure 5B, the overall accuracy increased from 76.7% pre-AI to 94.8% post-AI, yielding an absolute improvement of 18.2 percentage points (pp)~~% percentage points (pp)~~ (95% CI: 14.8–21.5 pp). Post-AI, the odds of correctly diagnosing a case were 2.32 (95% CI: 1.99–2.82) times higher than pre-AI. The overall prediction time decreased from 67.9 min pre-AI to 22.5 mins post-AI, showing a 65.7% reduction (95% CI: -66.6% to -64.8%; time ratio, 0.34 [95% CI: 0.33–0.35]). Notably, the benefit was greater for trainees. As shown in Table 1 and Supplementary Table 9, specialist clinicians, who had high pre-AI accuracy of 91.7% and reading time of 1,486 s, achieved an accuracy improvement of 6.8 pp (95% CI: 1.7–21.2 pp; odds ratio [OR], 5.36 [95% CI: 1.56–88.3]) and a reduction of 30.9% in reading time (95% CI: -37.6% to -23.6%; time ratio, 0.69 [95% CI: 0.62–0.76]) post-AI. Trainees showed the largest accuracy gain (+20.1 pp, 95% CI: 15.7–24.3 pp; OR, 5.47 [95% CI: 3.81–8.52]) and reduction in reading time (-69.2%, 95% CI: -70.0% to -68.4%; time ratio, 0.31 [95% CI: 0.30–0.32]). Confusion matrices illustrate detailed diagnostic data from 20

physicians evaluating 30 clinical cases with (Supplementary Figure 2) and without AI assistance (Supplementary Figure 3).

DISCUSSION

In this study, we developed INTELCAPE, an AI-based system to improve CD diagnosis on the basis of CE videos. This system automatically extracts small-intestine segments from a full CE video, identifies lesions, and diagnoses CD, while reducing interpretation time and maintaining diagnostic accuracy.

Enteroscopy, whether single- or double-balloon, can be time-consuming and require specific expertise, and is typically done under sedation with its attendant risks.[21] The rate of complications is also higher, particularly in patients with severe disease or altered anatomy.[22-24] These factors limit its widespread use in routine clinical practice.[25] In contrast, CE is a non-invasive tool for evaluating the small intestine, providing visualization of the mucosal surface without the need for sedation. CE also has limitations. Notably, CE generates thousands of images per examination, imposing a substantial analytical burden on clinicians. Manual analysis is labor-intensive and prone to human error, increasing the risk of missed lesions, particularly those with subtle appearances.[9]

Deep learning-based video analysis techniques for CE have recently shown advancements in small-bowel lesion detection and CD diagnosis. Relevant studies have shown that AI models achieve lesion detection AUCs of ~0.99 on well-curated datasets and diagnostic accuracies comparable to those of experienced clinicians. However, most of the previous studies focused on single tasks (e.g., ulcer detection) and demonstrated limitations in multi-center device adaptability as well as in supporting primary care settings.[13, 16, 26-288] Therefore, we developed

the INTELCAPE system to address both the inherent limitations of CE and the unmet needs in existing AI research for CE-based classification suggestive of CD. Although CD can affect any region of the GI tract, approximately 80% of the patients with CD show small-intestine involvement, of whom 30% have exclusive small-intestine disease.[29, 3030] By focusing analysis on the small intestine, INTELCAPE can streamline the diagnostic process and reduce interpretation time. It can also automatically identify and present potential lesion frames, which may enhance diagnostic efficiency, particularly for less-experienced clinicians. In our evaluation, INTELCAPE was able to identify mucosal abnormalities suggestive of CD, such as ulcers and erosions, with performance comparable to that of specialist gastroenterologists. Non-specific erosions, such as those related to NSAID use, represent a common diagnostic confounder in CE. INTELCAPE may help address this challenge by leveraging global lesion patterns and distribution to discriminate Crohn's disease from non-Crohn's conditions, rather than relying on isolated lesion morphology. These capabilities could be valuable in settings where access to specialist expertise is limited or when clinicians need to review large volumes of CE studies. Our results showed that INTELCAPE effectively assisted clinicians and achieved 18.2 percentage point~~% pp~~ improvement in diagnostic accuracy and 65.7% reduction in reading time. These findings suggest that INTELCAPE could serve as a decision-support tool, potentially increasing clinician efficiency and accuracy.

The system integrates three core functions as follows: small-bowel segmentation, lesion detection, and CD diagnosis. The potential implications of this approach for clinical practice are as follows: (i) Diagnostic support: By providing standardized image analysis, systems such as INTELCAPE could help less-experienced clinicians or those in primary care settings achieve diagnostic accuracy closer to that of specialists in tertiary hospitals, potentially mitigating disparities in healthcare access. (ii) Intelligent decision support: Clinicians in geographically

dispersed locations can obtain real-time standardized diagnostic recommendations through cloud-based platforms, markedly enhancing CD management standardization in resource-constrained regions. (iii) Workflow reengineering: Automated analysis can substantially reduce the time clinicians spend reviewing CE videos. Overall, INTELCAPE could support the long-term management of CD by shortening diagnostic cycles and improving diagnostic consistency, thereby facilitating timely treatment decisions.

Furthermore, the development approach employed for INTELCAPE may inform the creation of explainable AI systems for medical video analysis. AI systems that generate diagnoses directly from raw video data typically lack transparency in their decision-making process. However, by leveraging the key characteristics of CD, INTELCAPE focused on the most relevant video segment, extracting key video frames that show various abnormal lesions. Thus, the final diagnosis was made on the basis of lesions that were interpretable by clinicians.

Our study has some limitations. First, the system's accuracy depends on the training data's quality and diversity. Its performance in more complex or atypical cases, which are not sufficiently represented in our datasets, requires further evaluation. Second, the system was not inherently generalizable across cohorts. Models trained on specific cohorts using specific devices require fine-tuning to new devices/cohorts. Thus, further validation in prospective cohort studies is also necessary. Third, the risk of capsule retention remains a concern in patients with intestinal strictures, potentially necessitating complementary diagnostic techniques such as magnetic resonance imaging or computed tomography enterography. Fourth, the CE-based analysis provided by INTELCAPE is intended only as an auxiliary diagnostic reference. Therefore, a definitive diagnosis of CD must be made by clinicians integrating multidimensional clinical information. Fifth, although this study was adequately powered to

detect changes in diagnostic accuracy, the sample size resulted in relatively wide CIs for some predictive values (e.g., positive and negative predictive values) in the reader-assessment study. The stability of this approach should be further confirmed in larger prospective studies. Sixth, the diagnostic accuracy reported for Stage 3 is based on AI-selected clips enriched for lesions. While this reflects the system's intended workflow, it may overestimate performance on unfiltered, full-length videos in real-world settings. Finally, because the system's performance was measured on CE studies of adequate quality after manual quality control, the reported metrics represent best-case conditions and not real-world "all-comers" CE. Therefore, the effectiveness of this approach in broader clinical practice requires prospective evaluation.

In conclusion, INTELCAPE represents a proof-of-concept AI system that, under curated conditions, demonstrates high performance in automating CE video analysis for CD and has potential to assist clinicians, particularly those with less experience. Nevertheless, its deployment in real-world clinical workflows will require prospective validation on "all-comers" populations, robust handling of poor-quality videos, and device-specific fine-tuning. With further refinement, such systems could become valuable tools in CD management, especially in settings with limited access to specialist care.

AUTHOR CONTRIBUTIONS

Conceptualization: Dejun Fan, Yize Mao, Feng Liang, Zheng Liu, Guanbin Li, Feng Gao, and Xiaojian Wu; Methodology: Dejun Fan, Feng Liang, Zheng Liu, Guanbin Li, Jie Chen, and Shuangyi Tan; Deep learning model development and data analysis: Zheng Liu, Huayu Li, Yanan Liu, Mingjie Wang, Yuting Qian, and Guanbin Li; Statistical analyses: Yize Mao, Jian Tang, Neng Wang, and Tao Yang; Data collection, preprocessing, and annotation: Dejun Fan, Huayu Li, Jiancong Hu, Jie Chen, Yize Mao, and Yuting Qian; Clinical experiment

implementation: Dejun Fan, Jian Tang, Yanan Liu, Mingjie Wang, Jie Chen, and Neng Wang; Quality control and validation: Feng Gao, Xiaojian Wu, Tao Yang, and Shuangyi Tan; Visualization: Zheng Liu, Yuting Qian, and Shuangyi Tan; Funding acquisition: Feng Gao, Xiaojian Wu, and Guanbin Li; Writing – original draft: Dejun Fan, Yize Mao, Feng Liang, and Zheng Liu; Writing – review & editing: All authors; Supervision and project coordination: Guanbin Li, Feng Gao, Jian Tang, and Xiaojian Wu

FIGURE LEGENDS

Figure 1. INTELCAPE System Overview

(A) Three-stage AI pipeline framework for intestine segmentation (n = 872), lesion detection (n = 46,344), and CD diagnosis (n = 363), covering eight lesion types. (B) Overall procedures, including data sourcing, labeling, training, and validation and comparison with doctors, and AI-assisted diagnosis.

CD, Crohn's disease; AI, artificial intelligence

Figure 2. Intestine Segmentation

(A) Data source and workflow. (B) Model: image frame encoding → multi-frame fusion → classification. (C) Example: images from three GI regions. (D) Example: frame sequence with prediction probabilities for intestine segmentation. (E) Result: model-processing time across cohorts. (F) Result: ROC curves for recognizing different intestine segments. (G) Result: IoU performance. (H) Result: per-video processing time in comparison with clinicians. (***$p < 0.001$)

AI, artificial intelligence; CD, Crohn's disease; GI, gastrointestinal; ROC, receiver operating characteristic; IoU, intersection over union

Figure 3. Lesion Detection

(A) Data source and workflow. (B) Model: EfficientNet with background suppression. (C) Example: lesion marked with bounding boxes and heatmaps. (D) Result: ROC curves. (E) Result: accuracy in comparison with clinician assessments. (F) Result: per-image processing time in comparison with clinicians. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$)

ROC, receiver operating characteristic

Figure 4. CD Diagnosis

(A) Data source and workflow. (B) Model: feature extraction → region selection → Transformer fusion → classification. (C) Example: CD and non-CD images. (D) Result: ROC curves. (E) Result: accuracy in comparison with clinicians. (F) Result: per-video processing time in comparison with clinicians. (***$p < 0.001$)

CD, Crohn's disease; ROC, receiver operating characteristic

Figure 5. INTELCAPE-Assisted Diagnosis

(A) Reader assistance by AI.

(B) Result: Statistically significant improvement in diagnostic accuracy (76.7% to 94.8%) and reduction in time (67.9 min to 22.5 min).

(C) End-to-end diagnostic workflow.

AI, artificial intelligence

Supplementary Figures

Supplementary Fig. 1. Model Architectures

(A) EfficientNet-B4 with MBConv and convolution layers. (B) Two-layer Transformer: segment-level and whole-video fusion for CD classification.

CD, Crohn's disease

Supplementary Fig. 2. Reader-Assistance Study Without AI

Performance of 15 clinicians in diagnosing CE videos without INTELCAPE assistance.

CE, capsule endoscopy; AI, artificial intelligence

Supplementary Fig. 3. Reader-Assistance Study With AI

Performance of the same clinicians assisted by INTELCAPE.

AI, artificial intelligence

Supplementary Fig. 4. Lesion Subgroup Confusion Matrices for CD Classification

CD, Crohn’s disease

Supplementary Fig. 5. SHAP Analysis for Interpreting Feature Importance of the Encoded CLS Token for CD Diagnosis Results

(A) Bee swarm summary plot of the top 20 important features for CD classes.

(B) Bee swarm summary plot of the top 20 important features for non-CD classes.

(C) Feature importance plot of the top 30 features ordered by total absolute SHAP values.

CD, Crohn’s disease; SHAP, Shapley additive exPlanations

Table

Table 1. Diagnostic accuracy before and after AI assistance (multi-reader, multi-case analysis, performing a two-way (reader-case) bootstrap with 2,000 replicates).

Supplementary Table

Supplementary Table 1. Cohort Detailed Information

Supplementary Table 2 Small Intestine Segmentation Model Performance Evaluation

Supplementary Table 3 Stomach Recognition Model Performance Evaluation

Supplementary Table 4 Small Intestine Recognition Model Performance Evaluation

Supplementary Table 5 Large Intestine Recognition Model Performance Evaluation

Supplementary Table 6. Lesion Detection Performance

Supplementary Table 7 Crohn's Disease Diagnosis Model Performance Evaluation

Supplementary Table 8. Reader Experience

Supplementary Table 9. Diagnostic Time Before and After AI Assistance (multi-reader, multi-case analysis using the mixed-effects model)

REFERENCES

1. Dolinger M, Torres J, Vermeire S. Crohn's disease. Lancet 2024;403:1177-1191.
2. Rieder F, Mukherjee PK, Massey WJ, et al. Fibrosis in IBD: from pathogenesis to therapeutic targets. Gut 2024;73:854-866.
3. Olen O, Askling J, Sachs MC, et al. Mortality in adult-onset and elderly-onset IBD: a nationwide register-based cohort study 1964-2014. Gut 2020;69:453-461.
4. Dharni K, Singh A, Sharma S, et al. Trends of inflammatory bowel disease from the Global Burden of Disease Study (1990-2019). Indian J Gastroenterol 2024;43:188-198.
5. Zhu X, Yue M, Zhang X, et al. Global Disease Burden of Immune-Mediated Inflammatory Diseases (IMIDs), 1990–2021. Med Research 2025;1:285-296.
6. Gerson LB, Fidler JL, Cave DR, et al. ACG Clinical Guideline: Diagnosis and Management of Small Bowel Bleeding. Am J Gastroenterol 2015;110:1265-87; quiz 1288.
7. Lamb CA, Kennedy NA, Raine T, et al. British Society of Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. Gut 2019;68:s1-s106.
8. Le Berre C, Trang-Poisson C, Bourreille A. Small bowel capsule endoscopy and treat-to-target in Crohn's disease: A systematic review. World J Gastroenterol 2019;25:4534-4554.
9. Kwack WG, Lim YJ. Current Status and Research into Overcoming Limitations of Capsule Endoscopy. Clin Endosc 2016;49:8-15.
10. Goran L, Negreanu AM, Stemate A, et al. Capsule endoscopy: Current status and role in Crohn's disease. World J Gastrointest Endosc 2018;10:184-192.
11. Ben-Horin S, Lahat A, Ungar B, et al. Capsule Endoscopy-Guided Proactive Treat-to-Target Versus Continued Standard Care in Patients With Quiescent Crohn's Disease: A Randomized Controlled Trial. Gastroenterology 2025;169:85-93 e3.
12. Levartovsky A, Eliakim R. Video Capsule Endoscopy Plays an Important Role in the Management of Crohn's Disease. Diagnostics (Basel) 2023;13.
13. Ukashi O, Soffer S, Klang E, et al. Capsule Endoscopy in Inflammatory Bowel Disease: Panenteric Capsule Endoscopy and Application of Artificial Intelligence. Gut Liver 2023;17:516-528.
14. Pennazio M, Rondonotti E, Despott EJ, et al. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2022. Endoscopy 2023;55:58-95.
15. Dhali A, Kipkorir V, Maity R, et al. Artificial Intelligence-Assisted Capsule Endoscopy Versus Conventional Capsule Endoscopy for Detection of Small Bowel Lesions: A Systematic Review and Meta-Analysis. J Gastroenterol Hepatol 2025.

16. Brodersen JB, Jensen MD, Leenhardt R, et al. Artificial Intelligence-assisted Analysis of Pan-enteric Capsule Endoscopy in Patients with Suspected Crohn's Disease: A Study on Diagnostic Performance. J Crohns Colitis 2024;18:75-81.
17. Messmann H, Bisschops R, Antonelli G, et al. Expected value of artificial intelligence in gastrointestinal endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. Endoscopy 2022;54:1211-1231.
18. Chahal D, Byrne MF. A primer on artificial intelligence and its application to endoscopy. Gastrointest Endosc 2020;92:813-820 e4.
19. Yamada A, Niikura R, Otani K, et al. Automatic detection of colorectal neoplasia in wireless colon capsule endoscopic images using a deep convolutional neural network. Endoscopy 2021;53:832-836.
20. Zhu L, She Q, Chen Q, et al. Background-aware classification activation map for weakly supervised object localization. IEEE Transactions on Pattern Analysis and Machine Intelligence 2023;45:14175-14191.
21. Elena RM, Riccardo U, Rossella C, et al. Current status of device-assisted enteroscopy: Technical matters, indication, limits and complications. World J Gastrointest Endosc 2012;4:453-61.
22. Heine GD, Hadithi M, Groenen MJ, et al. Double-balloon enteroscopy: indications, diagnostic yield, and complications in a series of 275 patients with suspected small-bowel disease. Endoscopy 2006;38:42-8.
23. May A, Nachbar L, Ell C. Double-balloon enteroscopy (push-and-pull enteroscopy) of the small bowel: feasibility and diagnostic and therapeutic yield in patients with suspected small bowel disease. Gastrointest Endosc 2005;62:62-70.
24. Waddingham W, Kamran U, Kumar B, et al. Complications of diagnostic upper Gastrointestinal endoscopy: common and rare - recognition, assessment and management. BMJ Open Gastroenterol 2022;9.
25. Bourreille A, Ignjatovic A, Aabakken L, et al. Role of small-bowel endoscopy in the management of patients with inflammatory bowel disease: an international OMED-ECCO consensus. Endoscopy 2009;41:618-37.
26. Dhali A, Kipkorir V, Maity R, et al. Artificial Intelligence-Assisted Capsule Endoscopy Versus Conventional Capsule Endoscopy for Detection of Small Bowel Lesions: A Systematic Review and Meta-Analysis. J Gastroenterol Hepatol 2025;40:1105-1118.
27. Andrade P, Mascarenhas M, Mendes F, et al. AI-Assisted Capsule Endoscopy for Detection of Ulcers and Erosions in Crohn's Disease: A Multicenter Validation Study. Clin Gastroenterol

Hepatol 2025.
28. Klang E, Barash Y, Margalit RY, et al. Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. Gastrointest Endosc 2020;91:606-613 e2.
29. Torres J, Mehandru S, Colombel JF, et al. Crohn's disease. Lancet 2017;389:1741-1755.
30. Carretero C, Bojorquez A, Eliakim R, et al. Updates in the diagnosis and management of small-bowel Crohn's disease. Best Pract Res Clin Gastroenterol 2023;64-65:101855.

KEY RESOURCES TABLE

REAGENT or RESOURCE	**SOURCE**	**IDENTIFIER**
Deposited data		
Center 1 Capsule Endoscopy Data	This paper	
Center 2 Capsule Endoscopy Data	This paper	
Software and algorithms		
ResNet, EfficientNet + Transformer	This paper	
Python version 3.8	Python Software Foundation	https://www.python.org/
Torch version 2.4	Pytorch platform	https://pytorch.org/
Opencv-python 4.8	Opencv toolkit	https://opencv.org/
Pandas version 2.2	Pandas toolkit	https://pandas.pydata.org/
Numpy version 1.26	Numpy toolkit	https://numpy.org/
Statsmodels version 0.14.4	Statsmodels toolkit	https://www.statsmodels.org/

Figure 1

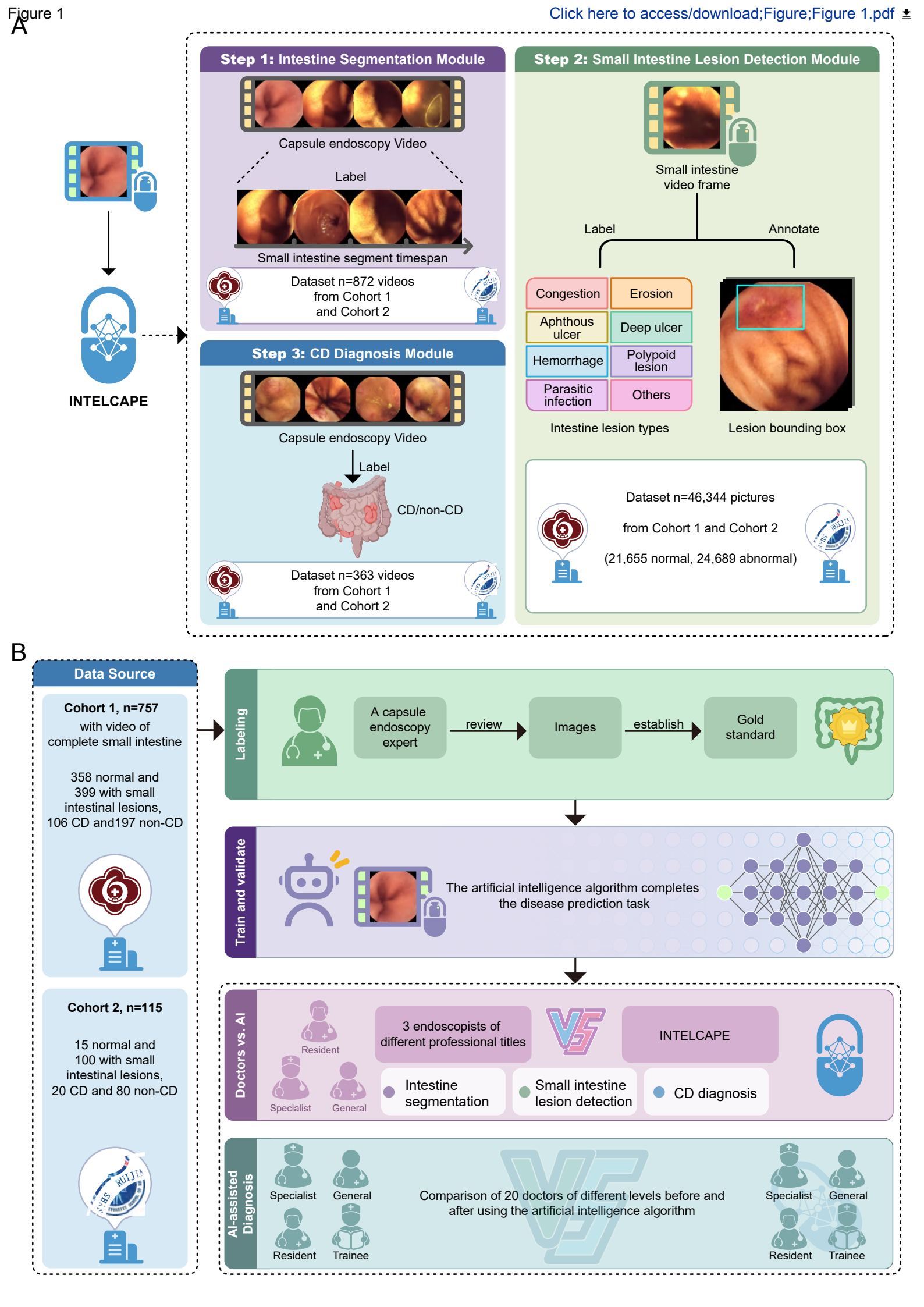


Figure 2

Click here to access/download;Figure;Figure 2-reviesed.pdf

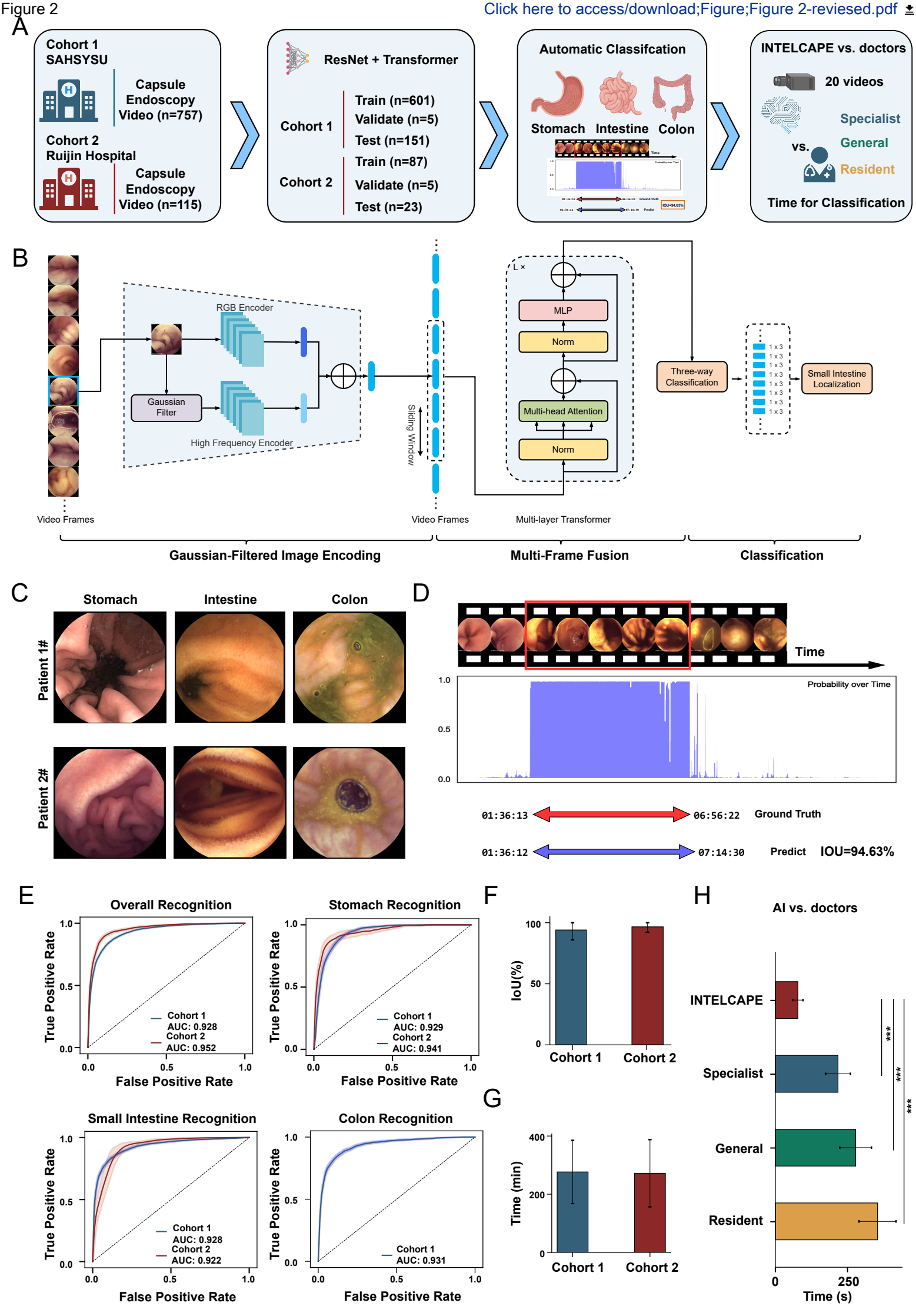


Figure 3

Click here to access/download;Figure;Figure 3-revised.pdf

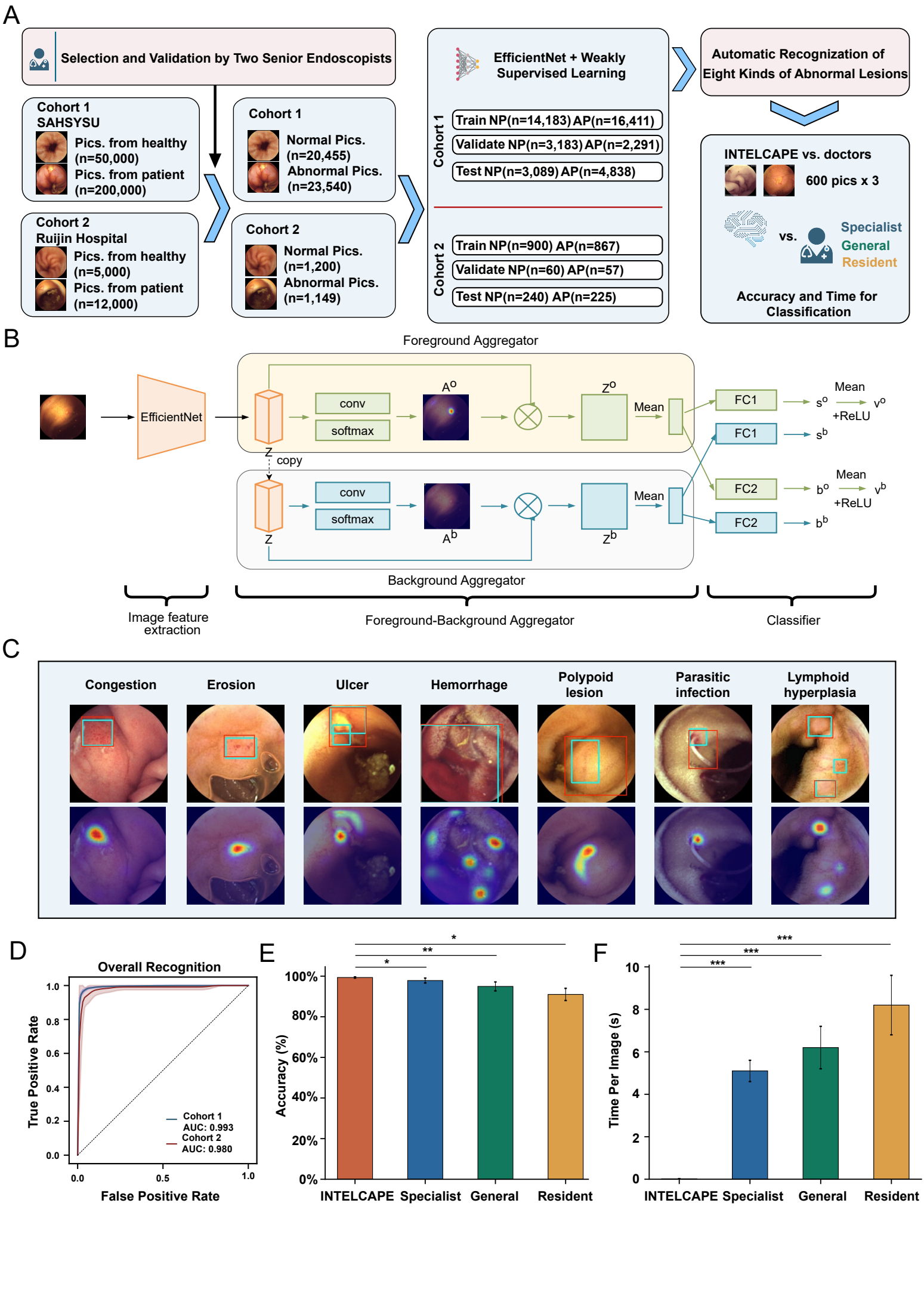


Figure 4

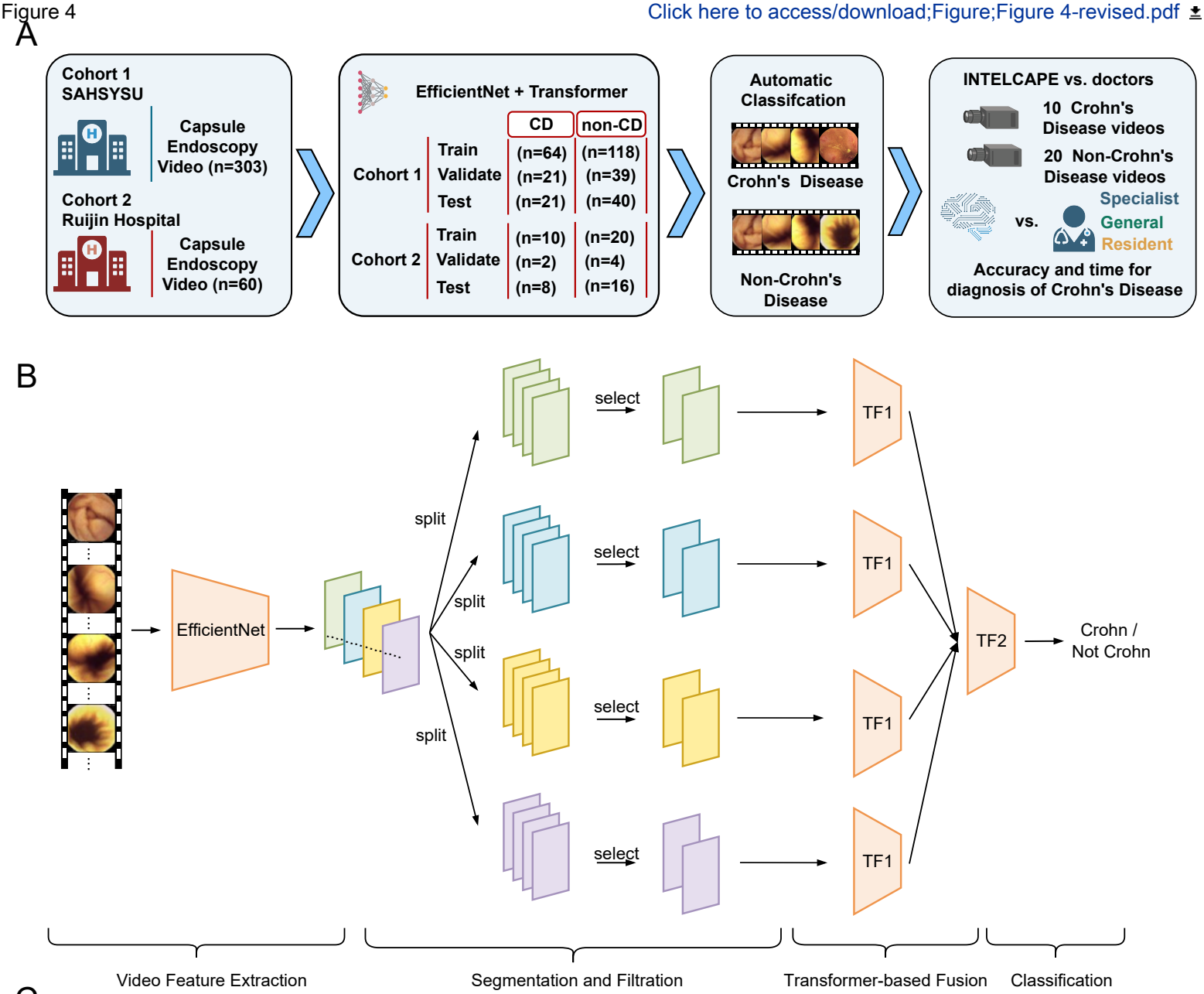
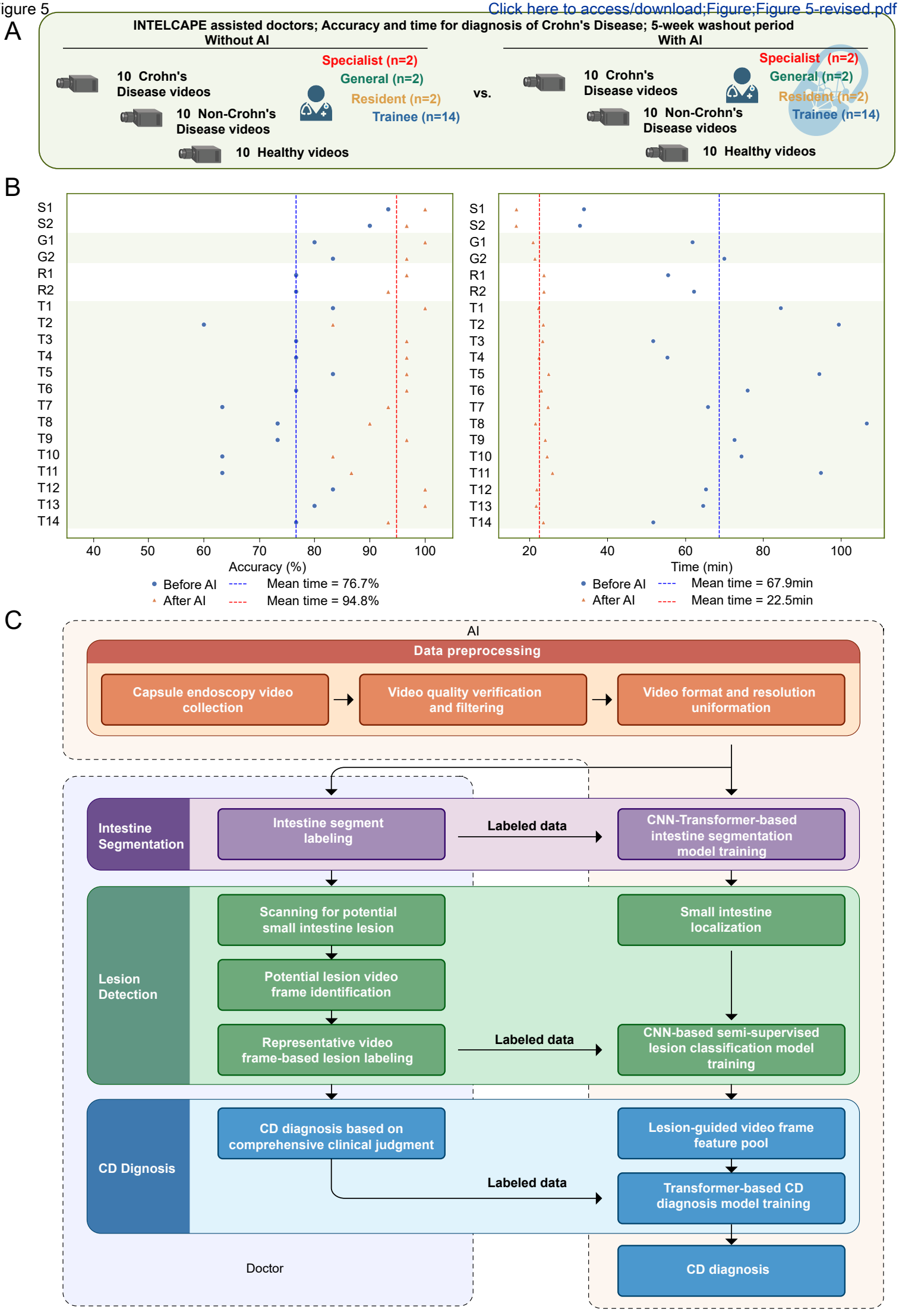
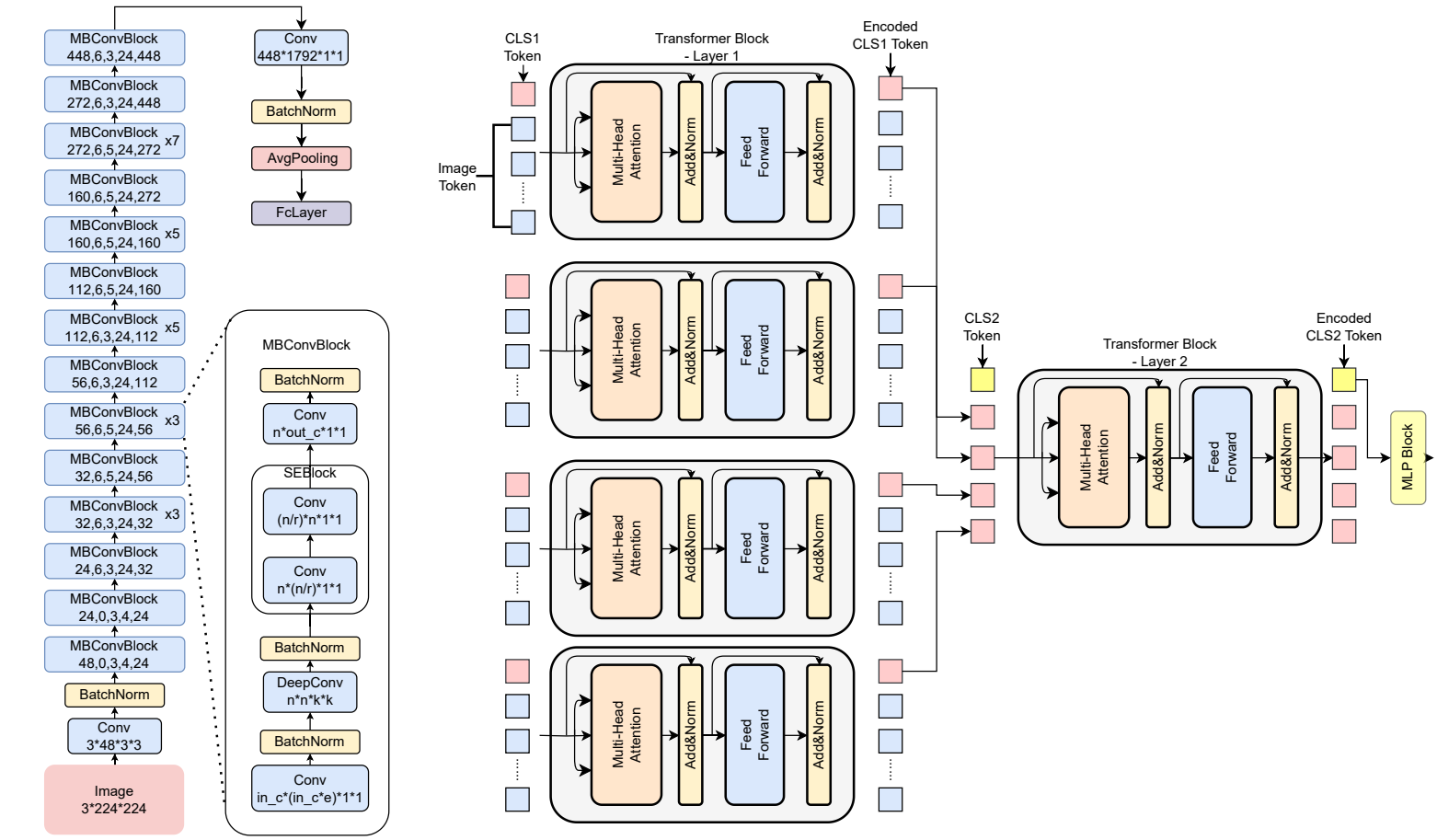


Figure 5



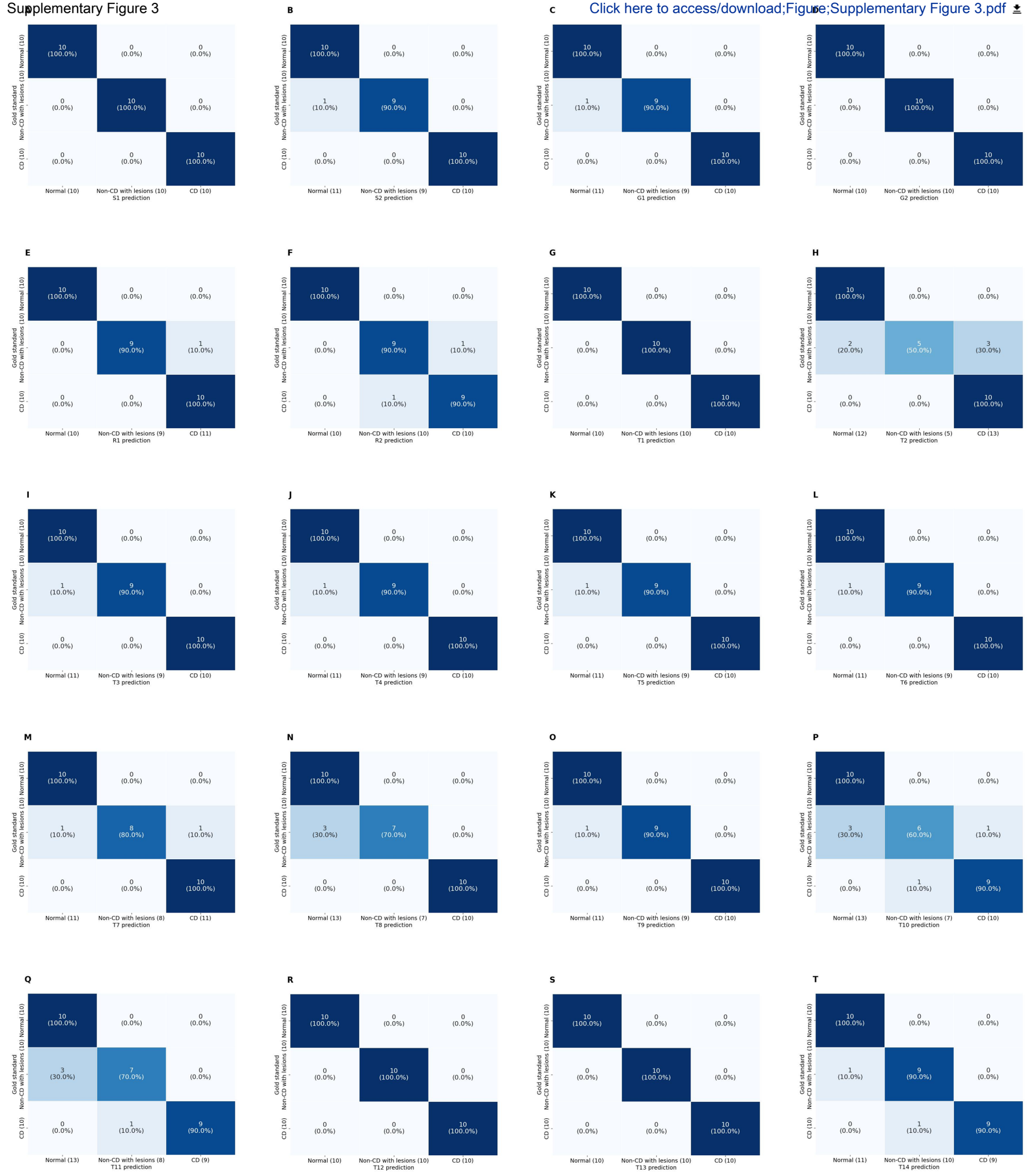
Supplementary Figure 1

Click here to access/download;Figure;Supplementary Figure 1.pdf



Supplementary Figure 2

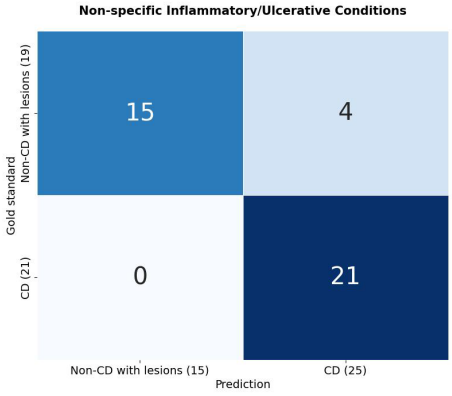
Supplementary Figure 3



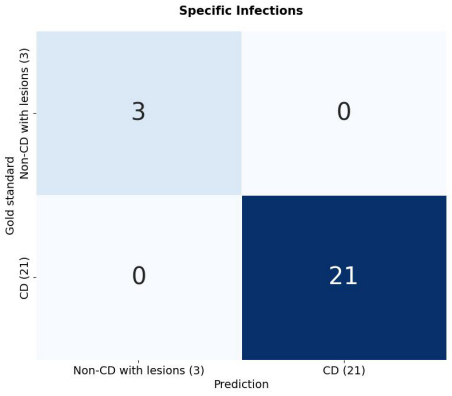
Supplementary Figure 4

Click here to access/download;Figure;Supplementary Figure 4.pdf

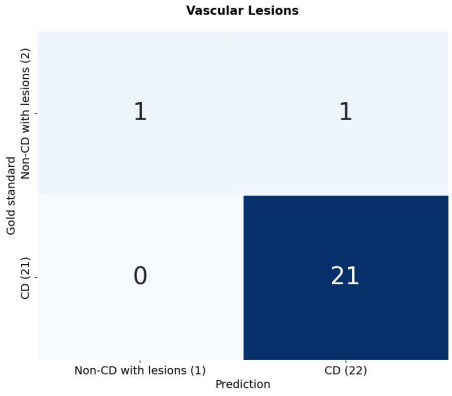
A



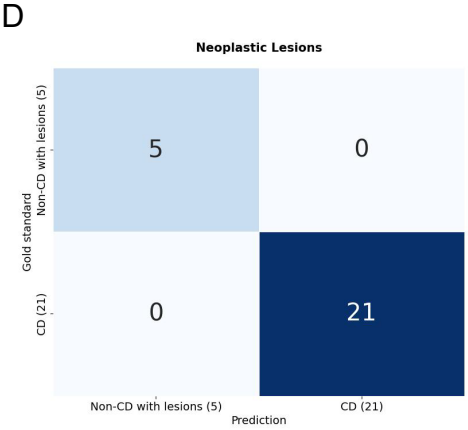
B



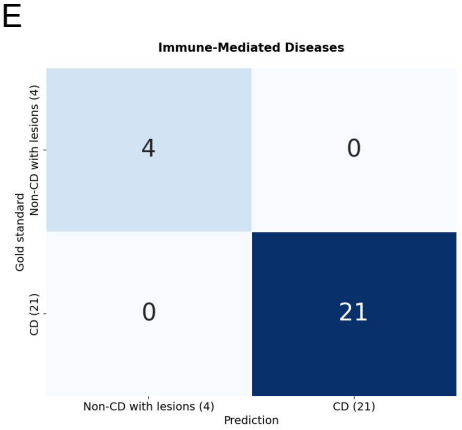
C



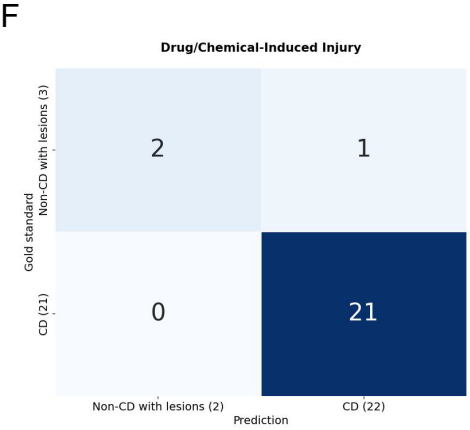
D



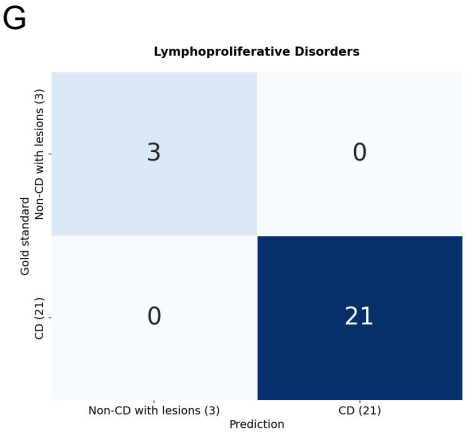
E



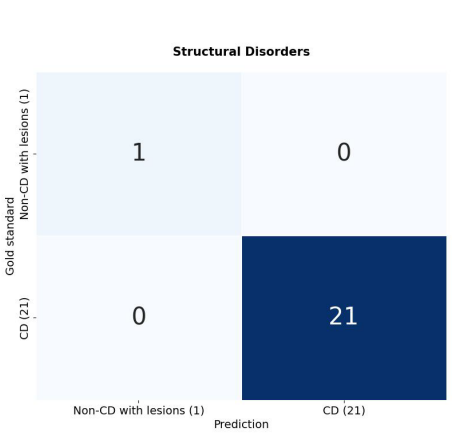
F



G



H



Supplementary Figure 5

Click here to access/download;Figure;Supplementary Figure 5.pdf

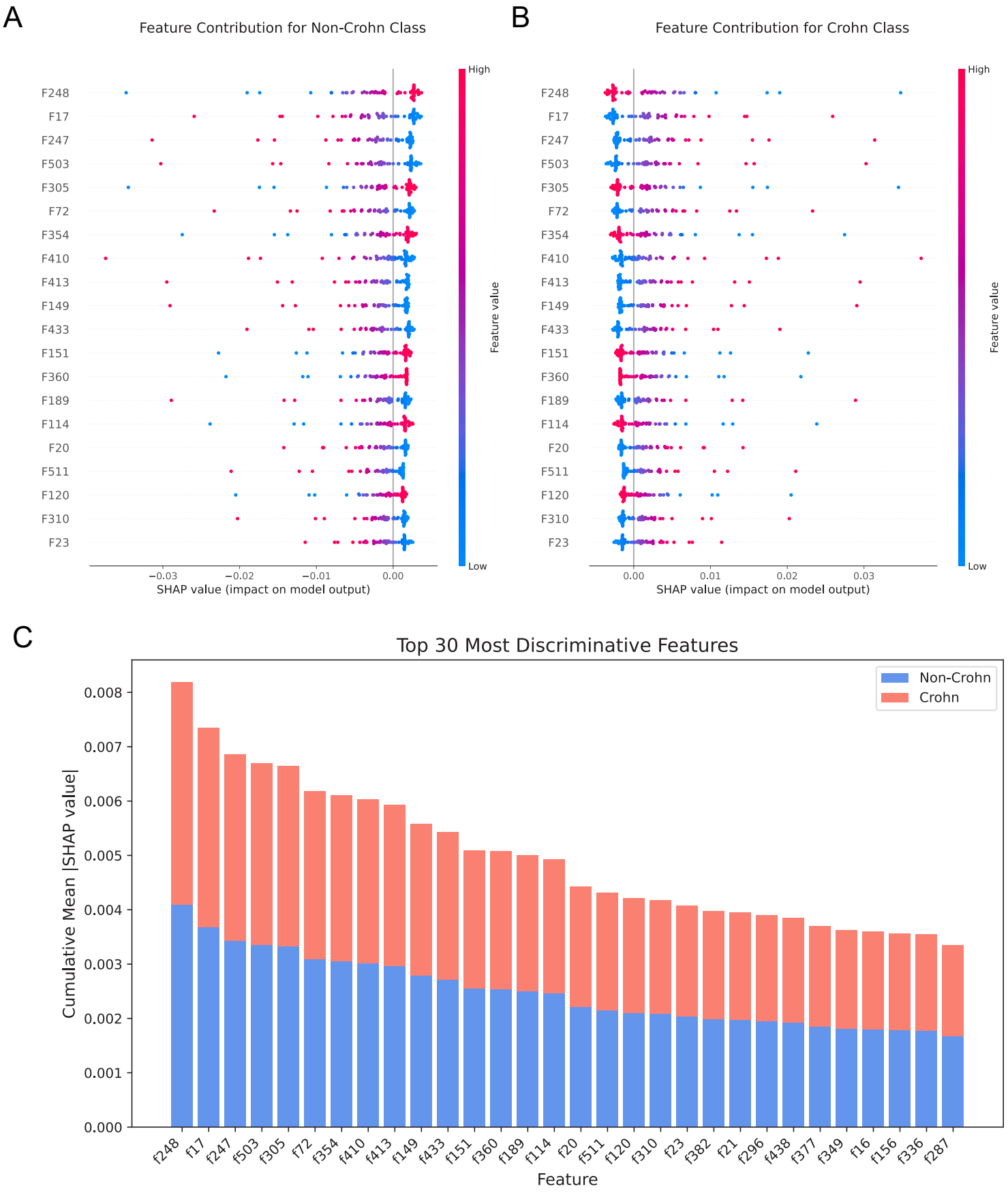


Table 1. Diagnostic accuracy before and after AI assistance (multi-reader, multi-case analysis, performing a two-way (reader-case) bootstrap with 2,000 replicates).

Subset	Pre-AI Accuracy	Post-AI Accuracy	Δ Accuracy (pp)	OR (95% CI)
Overall	76.7%	94.8%	+18.2 (14.8–21.5)	2.32 (1.99–2.82)
Specialist Group	91.7%	98.3%	+6.8 (1.7–13.3)	1.59 (1.16–2.25)
General Group	81.7%	98.3%	+16.6 (6.7–26.7)	2.33 (1.48–3.55)
Resident Group	76.7%	95.0%	+18.3 (8.3–28.4)	2.05 (1.39 – 3.15)
Trainee Group	73.8%	93.8%	+20.1 (15.7–24.3)	2.80 (1.92 – 2.79)

Supplementary Table 1 Cohort Detailed Information

Baseline characteristics	Internal Center SAHSYSU (n=757)	External Center Ruijin Hospital (n=115)
Age(years)		
Median	36	49
Sex~~Gender~~		
Female	217	41
Male	540	74
Smoke		
Yes	92	19
No	483	41
Unknown	182	55
Lesion Condition		
Congestion	264	
Erosion	56	
Ulcer	189	
Hemorrhage	21	
Polypoid Lesion	12	
Parasitic Infection	3	
Lymphoid hyperplasia	13	
Video Composition		
Normal	358	15
Lesion non-CD	201	80
CD	109	20
CD Not Excluded	89	0

Abbreviates

CD: Crohn’s disease

Supplementary Table 2 Small Intestine Segmentation Model Performance Evaluation

Type	Accuracy	AUC	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	F1
Cohort 1	90.16%	0.928 (95%CI:0.92-0.93)	85.24%	92.62%	85.24%	92.62%	85.24%
Cohort 2	87.44%	0.952 (95%CI: 0.95-0.96)	81.16%	90.58%	81.16%	90.58%	81.16%

Supplementary Table 3 Stomach Recognition Model Performance Evaluation

Type	Accuracy	AUC	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	F1
Cohort 1	94.28%	0.929 (95%CI:0.92-0.93)	86.85%	96.27%	86.22%	96.46%	86.54%
Cohort 2	88.61%	0.941 (95%CI: 0.93-0.96)	67.34%	94.36%	76.38%	91.44%	71.59%

Supplementary Table 4 Small Intestine Recognition Model Performance Evaluation

Type	Accuracy	AUC	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	F1
Cohort 1	85.25%	0.928 (95%CI:0.92-0.93)	83.92%	86.98%	89.24%	80.77%	86.50%
Cohort 2	91.07%	0.922 (95%CI: 0.90-0.94)	60.62%	93.68%	45.12%	96.53%	51.73%

Supplementary Table 5 Large Intestine Recognition Model Performance Evaluation

Type	Accuracy	AUC	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	F1
Cohort 1	90.96%	0.931 (95%CI:0.92-0.94)	87.05%	92.09%	76.21%	96.07%	81.27%
Cohort 2	/	/	/	/	/	/	/

Supplementary Table 6 Small Intestine Lesion Recognition Model Performance Evaluation

Type	Accuracy	AUC	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	F1
Cohort 1	98.70%	0.993 (95%CI:0.99-0.99)	99.35%	98.74%	98.05%	99.58%	98.70%
Cohort 2	92.81%	0.980 (95%CI: 0.96-0.99)	91.98%	93.64%	93.56%	92.80%	92.77%

Supplementary Table 7 Crohn's Disease Diagnosis Model Performance Evaluation

Type	Accuracy	AUC	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	F1
Cohort 1	90.16%	0.982 (95%CI:0.95-1.00)	100%	85.00%	77.78%	100%	87.50%
Cohort 2	87.50%	0.984 (95%CI: 0.93-1.00)	100%	81.25%	72.73%	100%	84.21%

Supplementary Table 8. Reader experience

Reader ID	Experience (yr)	Gastroenteroscopy operations per year	Capsule endoscopy interpretations per year	years of experience specializing in CD	Expertise/Training
Specialist 1 (S1)	15	3500	70	8	Endoscopist specializing in CD
Specialist 2 (S2)	9	4000	60	6	Endoscopist specializing in CD
General 1 (G1)	9	5000	35	3	Attending endoscopist
General 2 (G2)	5	4500	25	3	Attending endoscopist
Resident 1 (R1)	5	5000	80	1	Resident endoscopist
Resident 2 (R2)	3	5500	30	1	Resident endoscopist
Trainee 1 (T1)	10	500	0	0	GI endoscopy trainee
Trainee 2 (T2)	12	500	0	0	GI endoscopy trainee
Trainee 3 (T3)	7	700	0	0	GI endoscopy trainee
Trainee 4 (T4)	6	1500	0	0	GI endoscopy trainee
Trainee 5 (T5)	10	600	0	0	GI endoscopy trainee
Trainee 6 (T6)	13	500	0	0	GI endoscopy trainee
Trainee 7 (T7)	9	1000	0	0	GI endoscopy trainee
Trainee 8 (T8)	8	1500	0	0	GI endoscopy trainee
Trainee 9 (T9)	7	800	0	0	GI endoscopy trainee
Trainee 10 (T10)	10	300	0	0	GI endoscopy trainee
Trainee 11 (T11)	6	800	0	0	GI endoscopy trainee
Trainee 12 (T12)	4	1500	0	0	GI endoscopy trainee
Trainee 13 (T13)	7	1000	0	0	GI endoscopy trainee
Trainee 14 (T14)	7	700	0	0	GI endoscopy trainee

1.Specialist 1 and Specialist 2 were endoscopists specializing in gastroenterology, each with ≥5 years of experience in managing Crohn's disease.

2.General 1 and General 2 were attending endoscopists from the Department of Gastroenterological Endoscopy at the SAHSYSU.

3.Resident 1 and Resident 2 were resident endoscopists receiving subspecialty training in Crohn's disease diagnosis.

4.Trainees 1 through 14 were physicians from major local hospitals, including departments of gastroenterology and gastrointestinal surgery, all enrolled in a gastrointestinal endoscopy training program at the SAHSYSU during the reader study. None had previous experience in Crohn's disease management.

Supplementary Table 9. Diagnostic time before and after AI assistance (multi-reader, multi-case analysis using the mixed-effects model).

Subset	Pre-AI Mean Time (s)	Post-AI Mean Time (s)	Post-AI/Pre-AI Time Ratio (95% CI)	% Change (95% CI)
Overall	4,072	1,350	0.34 (0.33–0.35)	–65.7 (–66.6 to –64.8)
Specialist Group	1,486	995.8	0.69 (0.62–0.76)	–30.9 (–37.6 to –23.6)
General Group	3,955	1,268	0.31 (0.30–0.33)	–68.6 (–70.4 to –66.7)
Resident Group	3,534	1,421	0.39 (0.36–0.42)	–60.9 (–63.6 to –57.9)
Trainee Group	4,335	1,403	0.31 (0.30–0.32)	–69.2 (–70.0 to –68.4)

INTELCAPE: A Deep Learning-Powered System for Automated, High-Accuracy Crohn's Disease Diagnosis via Capsule Endoscopy

Short title: AI system for CD via CE

Dejun FAN[1,2,†], Yize Mao[3,†], Feng Liang[4,†], Zheng Liu[5,†], Huayu Li[6,†], Jian Tang[7], Yanan Liu[1], Mingjie Wang[8], Yuting Qian[8], Jie Chen[9], Neng Wang[10], Tao Yang[1], Shuangyi Tan[11], Guanbin Li[12,*], Feng Gao[2,13,*], Jiancong Hu[1,*], Xiaojian Wu[2,14,#,*]

1. Department of Gastrointestinal Endoscopy, The Sixth Affiliated Hospital, Sun Yat-sen University
2. Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, The Sixth Affiliated Hospital, Sun Yat-sen University
3. Department of Pancreatobiliary Surgery, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou 510060, P. R. China.
4. Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Longgang, Shenzhen, China
5. School of Computer Science and Engineering, Sun Yat-sen University
6. State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou 510060, P. R. China.
7. Department of Gastroenterology, The Sixth Affiliated Hospital, Sun Yat-sen University
8. Department of Gastroenterology, Ruijin Hospital, Shanghai Jiao Tong University, School of Medicine
9. Department of Gastroenterology, The Third Xiangya Hospital of Central South University, Changsha, China.
10. Endoscopy Center of Gastroenterology, Suining Central Hospital
11. School of Science and Engineer, The Chinese University of Hong Kong, Shenzhen
12. Computer Science and Engineering, Sun Yat-sen University
13. Biomedical Innovation Center, The Sixth Affiliated Hospital, Sun Yat-sen University
14. Department of General Surgery (Colorectal surgery), The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, 510080

[†]These authors contributed equally

[#]Lead contact

SUPPLEMENTARY METHODS and RESULTS

Data Preprocessing and Labeling

Figure 5C illustrates the artificial intelligence (AI)-based diagnostic system data preprocessing and labeling workflow for Crohn's disease (CD) using capsule endoscopy (CE). We verified their quality and uniformly preprocessed them after collecting videos from CE. First, unqualified videos were filtered out, cases including 1) the presence of debris, bubbles, or dark, 2) excessive blur or rapid motion, and 3) incomplete small intestine examination. Approximately 13% of the originally collected videos were excluded based on the above quality control criteria before the annotation process began. Second, the picture frames were extracted from the videos. Board-certified gastroenterologists labeled the start and end of the small intestine in each video for intestinal segmentation. For lesion detection, gastroenterologists scanned videos of the small intestine segment, identified representative video frames with lesions, and used an auxiliary tool to mark the lesion locate with a bounding box and indicate the lesion type. Furthermore, the gastroenterologists provided CD diagnosis results for each video based on a comprehensive clinical judgment. A feature pool was constructed using only lesion-containing frames, reducing the computation burden during training. All datasets underwent rigorous de-identification before annotation. All annotation results were verified using dual-expert consensus to ensure annotation reliability.

Intestine Segmentation

The intestine segmentation module used a hybrid ResNet-Transformer architecture (Figure 2B), designed to classify frames by region (stomach, small intestine, or colon) and localize transitions. It captures long-range dependencies and attention mechanisms for automated video segmentation. Confidence-based boundary detection was applied to identify anatomical

transitions. Subsequently, the model was trained on 601 and 87 videos from Cohorts 1 and 2, respectively, with separate validation and test sets for both cohorts (Figure 2A).

Small Intestine Lesion Detection

Figure 3A shows the workflow of small intestine lesion detection, which is the second stage of the overall pipeline. A dual-path framework for lesion classification (a binary problem) and localization was implemented. For lesion frame identification, we employed EfficientNet (Supplementary Fig. 1A)—an efficient convolutional neural network (CNN) architecture—to process small intestine frames through deep feature extraction, followed by fully connected layers for lesion probability prediction (Figure 3B). This enables comprehensive screening of potential lesion frames throughout the CE video sequence. Additionally, a novel background-aware, weakly supervised localization mechanism was developed—incorporating B-CAM with background suppression loss. The architecture processes frame features using parallel aggregators to generate distinct foreground and background representations. These discriminative features are subsequently fed into a classification network, with the aggregators optimized via backpropagation. Final lesion localization was achieved through Grad-CAM-based activation mapping, providing precise spatial identification of the lesion regions. The dataset from Cohort 1 included 23,540 abnormal and 20,455 normal frames, and fine-tuning used 2,349 labeled frames from Cohort 2. Separate sets were defined for training, validation, and testing (Figure 3A). On a test set of 1,800 images divided evenly into three groups, we compared the performance of INTELCAPE with that of three clinicians with varying levels of experience.

CD Diagnosis

The final stage (Figure 4) used a Transformer-based model for video-level CD classification

through temporal-spatial analysis of small-intestine CE videos (Figure 4B). The system does not assign lesion type or etiologic labels to individual lesions. Rather, it infers Crohn's disease probability from global lesion patterns, distribution, and burden, analogous to expert clinical reasoning. Specifically, lesion frames were grouped into four anatomical regions, and the top 500 high-confidence frames per region were selected. A two-layer Transformer architecture (Supplementary Fig. 1B) was used to aggregate features, followed by a multilayer perceptron for classification, enabling effective integration of long-range dependencies and local features for the final classification. As shown in Figure 4A, the dataset from Cohort 1 included 182 training videos (64 CD-positive), along with 60 validation and 61 test videos. Fine-tuning used 30 videos (10 CD-positive) from Cohort 2, with 6 and 24 for validation and testing, respectively. To interpret feature importance for CD classification, we used Shapley Additive exPlanations (SHAP) on the CLS token (Supplementary Fig. 5). A significant SHAP value highlights the important influence of each feature in terms of the effectiveness of classification prediction, enabling researchers and clinicians to better understand the model's behavior and make more informed decisions. The most impactful latent features can act as a biomarker for reliable CD diagnosis.

Reader Assistance

To assess clinical utility, we compared the diagnostic accuracy and reading time of 20 clinicians with varying experience levels, diagnosing before and after INTELCAPE assistance following the multi-reader multi-case (MRMC) framework (Figure 5A). They include specialists (n=2), general clinicians (n=2), residential clinicians (n=2), and trainees (n=14). The null hypothesis assumed a diagnostic accuracy of 0.65, while the alternative hypothesis set an accuracy of 0.95. With a desired statistical power of 80% and a two-sided McNemar Test (significance level = 0.05), the minimum required sample size was calculated to be 28 videos. At last, the same set

of 30 test videos, including 10 cases for healthy, lesion, but non-CD, and CD, respectively, was used for clinician assessments with and without AI assistance, ensuring consistency in the evaluation material. The 20 clinicians were randomly divided into two groups. In the first assessment, the first group of clinicians read videos and diagnosed without INTELCAPE assistance, while the second group did the same things without assistance. In the second assessment, the first group of clinicians read videos and diagnosed with INTELCAPE assistance, while the second group did it without NTELCAPE assistance. Between these two assessments, a 5-week washout period was implemented for each clinician to minimize recall bias, preventing prior familiarity with cases from influencing post-AI diagnostic outcomes. Before AI assistance, clinicians reviewed the basic clinical information, scanned and interpreted the full CE video, and composed the diagnostic report. After AI assistance, they reviewed the basic clinical information, interpreted the 50 AI-selected video clips, each 3 s long at three frames per second, and composed the diagnostic report. The key difference was that the AI assistance replaced the labor-intensive, manual review of the full video with a targeted review of curated, lesion-focused clips. However, the reporting component remained constant. Diagnostic accuracy and reading time of each clinician for each case were recorded. The measured reading time encompassed the full duration from when the clinician opened the case to when they finalized and submitted their diagnostic report.

Interpreting CD Diagnosis Model

Providing interpretable biomarker for CD diagnosis is very helpful when clinicians use INTELCAPE's CD classification result as a diagnosis suggestion. Our model's decision is based on an aggregation of features from multiple, high-confidence frames. While an individual frame might be borderline, the collective evidence from the entire sequence allows the model to make a robust prediction. We use the SHAP analysis to study how the model synthesizes information

from various latent features to reach a diagnosis, rather than relying on a single ambiguous finding. The SHAP analysis results are shown in Supplementary Fig. 5. The bee swarm plot is designed to display an information-dense summary illustrating how the top features in a dataset affect the output of a model. Each observation in the data is represented by a single dot on each feature row. The vertical axis represents the features, sorted from top to bottom according to their importance as predictors. The position of a dot on a feature row is determined by the SHAP value of the corresponding feature, and the accumulation of dots on each feature row illustrates its density. The feature value determines the color of the dots, with red indicating large SHAP values and blue indicating small SHAP values.

We can infer from Supplementary Figs. 5(A) and 5(B) the most impactful features for classifying non-CD and CD patients. For example, feature 248, 17, and 247 are the top-3 significant dimensions that help distinguish CD. Note that the SHAP values of each feature are complementary for binary classification. The most impactful latent features can act as biomarker for reliable CD diagnosis. Supplementary Figs. 5(C) orders these impactful features by total absolute SHAP values.