Journal of Imaging Informatics in Medicine
https://doi.org/10.1007/s10278-024-01309-1

ORIGINAL PAPER

Applying Deep-Learning Algorithm Interpreting Kidney, Ureter, and Bladder (KUB) X-Rays to Detect Colon Cancer

Ling Lee[1] · Chin Lin[1,2,3] · Chia-Jung Hsu[3,4] · Heng-Hsiu Lin[3,4] · Tzu-Chiao Lin[5] · Yu-Hong Liu[5] · Je-Ming Hu[1,5,6]

Received: 1 August 2024 / Revised: 26 September 2024 / Accepted: 14 October 2024

Abstract

Early screening is crucial in reducing the mortality of colorectal cancer (CRC). Current screening methods, including fecal occult blood tests (FOBT) and colonoscopy, are primarily limited by low patient compliance and the invasive nature of the procedures. Several advanced imaging techniques such as computed tomography (CT) and histological imaging have been integrated with artificial intelligence (AI) to enhance the detection of CRC. There are still limitations because of the challenges associated with image acquisition and the cost. Kidney, ureter, and bladder (KUB) radiograph which is inexpensive and widely used for abdominal assessments in emergency settings and shows potential for detecting CRC when enhanced using advanced techniques. This study aimed to develop a deep learning model (DLM) to detect CRC using KUB radiographs. This retrospective study was conducted using data from the Tri-Service General Hospital (TSGH) between January 2011 and December 2020, including patients with at least one KUB radiograph. Patients were divided into development ($n=28{,}055$), tuning ($n=11{,}234$), and internal validation ($n=16{,}875$) sets. An additional 15,876 patients were collected from a community hospital as the external validation set. A 121-layer DenseNet convolutional network was trained to classify KUB images for CRC detection. The model performance was evaluated using receiver operating characteristic curves, with sensitivity, specificity, and area under the curve (AUC) as metrics. The AUC, sensitivity, and specificity of the DLM in the internal and external validation sets achieved 0.738, 61.3%, and 74.4%, as well as 0.656, 47.7%, and 72.9%, respectively. The model performed better for high-grade CRC, with AUCs of 0.744 and 0.674 in the internal and external sets, respectively. Stratified analysis showed superior performance in females aged 55–64 with high-grade cancers. AI-positive predictions were associated with a higher long-term risk of all-cause mortality in both validation cohorts. AI-enhanced KUB X-ray analysis can enhance CRC screening coverage and effectiveness, providing a cost-effective alternative to traditional methods. Further prospective studies are necessary to validate these findings and fully integrate this technology into clinical practice.

Keywords Artificial intelligence · Deep learning model · Colorectal cancer · Kidney, ureter and bladder radiographs

Introduction

Colorectal cancer (CRC) is the third most diagnosed cancer and the fourth leading cause of cancer-related deaths globally, with prevalence increasing significantly in recent years [1]. CRC screening has been shown to be highly effective in reducing both the incidence and mortality of the disease [2]. Common screening methods include guaiac-based fecal occult blood tests (gFOBTs), immunochemical tests (FITs), and colonoscopy, among which FOBT is currently the most

✉ Je-Ming Hu
jeminghu@gmail.com

1 School of Medicine, National Defense Medical Center, Taipei, R.O.C, Taiwan

2 Military Digital Medical Center, Tri-Service General Hospital, National Defense Medical Center, Taipei, R.O.C, Taiwan

3 School of Public Health, National Defense Medical Center, Taipei, R.O.C, Taiwan

4 Medical Informatics Office, Tri-Service General Hospital, National Defense Medical Center, Taipei, R.O.C, Taiwan

5 Division of Colorectal Surgery, Department of Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei, R.O.C, Taiwan

6 Graduate Institute of Medical Sciences, National Defense Medical Center, No 325, Section 2, Cheng-Kung Road, Neihu 114, Taipei, R.O.C, Taiwan

Published online: 31 October 2024

used. However, owing to the inconvenience of sample collection, improper ways of sample collection by patients, and low patient compliance, the results of FOBT can be affected. All these methods often lead to colonoscopy, which allows polypectomy and ongoing surveillance or symptom evaluation [3]. However, colonoscopy suffers from low coverage in CRC screening owing to fears of its invasive nature and procedural risks, and its effectiveness is highly dependent on the operator [4]. Therefore, a high coverage method for identifying high-risk CRC is necessary.

Kidney, ureter, and bladder (KUB) radiography, a type of plain abdominal radiography, is primarily used to assess the urinary system, such as to detect kidney stones, and helps diagnose other conditions affecting the gastrointestinal (GI) system by analyzing gas patterns, obstructions, and calcifications of the bowel [5, 6]. There are some limitations to KUB radiography, including low sensitivity and specificity for diagnosing specific diseases, presentation of nonspecific findings, and potential for misinterpretation [7, 8]. Moreover, a study indicated particularly poor results for diagnosing conditions such as appendicitis, pyelonephritis, pancreatitis, and diverticulitis, as well as showed high interobserver variability in the assessment of colon obstruction [9]. Despite these challenges, KUB X-ray images are widely used and considered the first-line investigation in the emergency room (ER), especially in patients with abdominal issues [7]. Several conditions may lead to colonic obstruction, of which CRC is one of the major leading causes [10]. In addition, physicians have used KUB to evaluate early large bowel obstruction (LBO) with a sensitivity of 84% [11], suggesting the potential of using KUB X-ray images as a detection tool for CRC, particularly when enhanced with advanced techniques for recognizing specific patterns.

Recent advancements in artificial intelligence (AI), specifically deep learning models (DLMs), have shown promise in the field of colonic diseases for tasks such as identifying and classifying colonic polyps, as well as improving detection rates of early cancerous lesions [12]. One study demonstrated a convolutional neural network (CNN) identified polyps with an AUC of 0.991 [13]. DLMs have been extensively applied in computed tomography (CT), magnetic resonance imaging (MRI), and colonoscopy for detecting bowel lesions [14, 15]. While KUB radiography has a lower resolution than the previous mentioned radiographs, previous studies using DLMs with plain abdominal radiography have demonstrated significant potential, such as a high area under the curve (AUC) of 0.961, for identifying small bowel obstructions [16]. To the best of our knowledge, there remains a lack of DLM-enabled KUB X-rays for detecting CRC, which may be useful for identifying high-risk patients because of their high coverage rate.

In this study, based on previous research highlighting the potential of integrating DLM with medical imaging and considering that KUB imaging is widely used for its non-invasive, cost-effective, and high coverage nature in first-line evaluating bowel obstructions compared to CT, MRI, and colonoscopy, additionally, the low compliance rate of FOBT further motivates the exploration of alternative screening methods. Therefore, we hypothesized that DLMs can effectively classify KUB images for the detection of CRC as a screening method. Although AI-enabled KUB analysis may identify some false-positive cases such as gFOBTs and FITs, a previous study has demonstrated a higher risk of mortality in AI-predicted false-positive cases compared to true-negative cases [17]. Therefore, we proposed exploring the additional potential of the AI model for predicting prediction from gFOBTs and FITs.

This study used a large dataset from two hospitals to develop a robust AI model for CRC detection using KUB X-rays. We presented a novel AI-enabled approach for early CRC detection using KUB x-ray images, offering a promising alternative to traditional screening methods with high-grade tumor detection potential. This cost-effective, accessible, and less invasive method could positively impact clinical practices and improve colorectal cancer tumor detection rate.

Related Works

The application of computer vision in the diagnosis of colorectal cancer can be traced back to a study in 1990 [18], which utilized a machine vision analysis system to differentiate between colonic adenoma, adenocarcinoma, and normal colonic tissue. In recent years, with the rapid advancement of AI, its use in medicine has become increasingly prominent. Among AI techniques, deep learning models, particularly those based on CNN, have emerged as a primary tool for medical imaging tasks. Research in AI-assisted colorectal cancer diagnosis encompasses various modalities such as CT, MRI, colonoscopy, and histopathological imaging, and addresses tasks including detection, classification, segmentation, and survival prediction [15]. In the following paragraph, recent studies with objectives similar to this research will be discussed.

In 2024, Marcello Di Giammarco et al. focuses on utilized deep learning for colon cancer diagnosis through histological image analysis [19]. Using 5000 images per class (benign tissue and adenocarcinoma) via data augmentation, the study tested several models, including ResNet50, DenseNet, VGG19, Inception-V3, EfficientNet, and MobileNet. MobileNet achieved the best performance, with near 99% accuracy. A key feature of the study is the use of explainable AI, employing Class Activation Mapping (CAM) techniques such as Grad-CAM to provide visual explanations, improving model reliability. The results highlight model ability to effectively classify colon cancer.

In 2022, Muthu Subash Kavitha et al. primarily used a ResNet-18 CNN model with transfer learning as the primary approach for detecting colorectal cancer and analyzing histopathological images [20]. The primary data used in the model consists of medical imaging data, particularly endoscopic and whole-slide images, which are processed using CNN architectures. The research emphasizes the development of end-to-end and transfer learning techniques to automate feature extraction, reduce manual intervention, and improve the accuracy of cancer detection. The main result of the study indicates that deep learning models, particularly CNN-based architectures, show high diagnostic accuracy in predicting invasive cancer from medical images.

In 2021, Masud et al. focus on developing a deep learning framework to classify lung and colon cancer using histopathological images [21]. The dataset used in this study is the LC25000 dataset, which contains 25,000 color images of five different tissue types, including both benign and malignant variations of lung and colon cancer tissues. They employed a CNN model for the classification task, extracting features using image processing techniques such as 2D Discrete Fourier Transform (DFT) and 2D Discrete Wavelet Transform (DWT). The proposed model achieved a maximum classification accuracy of 96.33%, demonstrating its high reliability for identifying various types of lung and colon cancer tissues. The study concludes that the method can significantly aid in cancer diagnosis, offering a highly accurate, automated system that reduces the effort and time required for manual diagnosis by medical professionals.

Paper Organization

The paper is organized as follows: the method section covers data sources, model architecture, and statistical analysis; the results present dataset characteristics, model performance, and survival analysis; the discussion compares our findings with state-of-the-art approaches and acknowledges study limitations; finally, conclusions and suggestions for future research are provided.

Methods

Data Source

This study was approved by the Institutional Review Board of Tri-Service General Hospital (TSGH), Taipei, Taiwan (IRB no. C202305019). We performed a retrospective 2-site study of the TSGH system between January 2011 and December 2020. The inclusion criteria of this study included patients who underwent at least one KUB X-ray at an academic medical center during the study period. The exclusion criteria for this study involved utilizing international classification of diseases (ICD) code to exclude patients with a prior diagnosis of colorectal or other cancer history before undergoing KUB X-ray imaging during the study period. A total of 135,704 patients who underwent at least one KUB X-ray at an academic medical center during the study period were included. Figure 1 illustrates the generation of each dataset. There were 90,343 patients without a history of CRC before the KUB X-ray examination, and they were divided into 50%

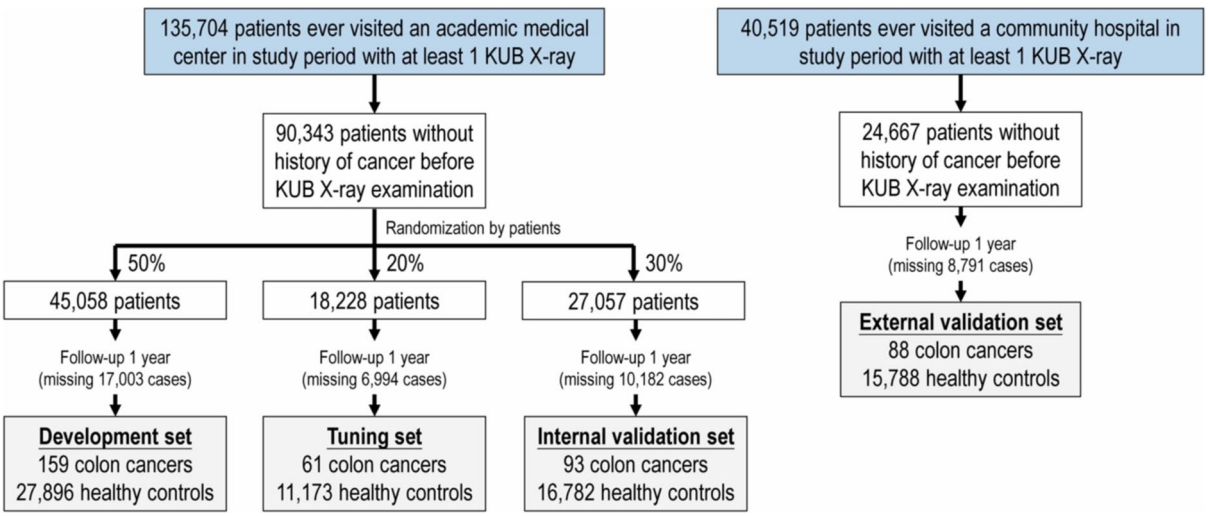

Fig. 1 Schematic of datasets generation. This chart was devised to assure a robust and reliable data set for training, validating, and testing of the network. Once a patient's data were placed in one of the datasets, that individual's data were used only in that set, avoiding "cross-contamination" among the training, validation, and test data sets. We selected only one KUB X-ray for each patient at random and followed up 1 year. The details of the flow chart and how each of the datasets was used are described in the "Methods" section

(45,058 patients), 20% (18,228 patients), and 30% (27,057 patients) groups through randomization. We selected only one KUB radiograph for each patient randomly and followed up for one year without censored cases to establish the development, tuning, and internal validation sets. The numbers of patients with colon cancer and healthy controls in the development set for training the DLM, tuning set to guide the training process, and internal validation set for conducting an accuracy test of the DLM were 159 and 27,896, 61 and 11,173, as well as 93 and 16,782, respectively. Additionally, 88 patients with CRC and 15,788 healthy controls at a community hospital meeting the same criteria were included in the external validation set. Similarly, we selected only one KUB X-ray image per patient in both validation sets.

Implementation of the DLM

The KUB X-ray images were labeled as a binary classification according to the electronic medical record (EMR), which recorded patients with or without CRC. The model architecture and training details were revised from those of a previous study that used a 121-layer DenseNet convolutional network [22]. This model was trained using data from the development set, and the hyperparameters were tuned based on model performance in the tuning set. The prediction output of the model was the probability of each label with the presence or absence of CRC. The technical details were the same as those used in our previous study [22]. Detailed information of the DLM architecture and training process were provided in Supplemental Appendix S1.

Interpreting Model Predictions

To gain a better understanding of the prediction by our model, we created a heat map to identify the locations in the KUB X-ray image that contributed the most to the network classification using a class activation map (CAM). The most significant features used by our model in its predictions were visualized on an overlaid heatmap image as in Fig. 5.

Baseline Information and Data Collection

The baseline information was obtained from the EMR of each hospital. Disease histories were based on new diagnoses according to the corresponding International Classification of Diseases (ICD), Ninth and Tenth Revisions, or laboratory tests. CRC histories were collected from the Taiwan Cancer Registry Coding Manual, which includes new diagnoses at our hospitals. Furthermore, we excluded ICD codes for cancers to avoid duplicate cases and ensure the feasibility of our data resources.

Outcome

Several outcomes of interest in both the internal and external validation cohorts were monitored. The primary outcome was the ability of the DLM to detect individual patients with CRC at the time of screening. The follow-up period was calculated from the date of the randomly selected KUB radiograph for each patient. The secondary outcome was subsequent all-cause mortality in patients without CRC who had a positive prediction by the DLM. Data for at-risk patients were censored at the last known hospital encounter to limit bias from incomplete records.

Statistical Analysis

Characteristics of the different datasets were presented as mean and standard deviation, number of patients, or percentage, as appropriate. Comparisons were made using either analysis of variance or chi-square tests where suitable. The performance of our model in detecting CRC was assessed using receiver operating characteristic (ROC) curve analysis, with the AUC, sensitivity, and specificity used to demonstrate the performance. The operating point was selected based on the maximum Youden index in the tuning set and applied to both validation sets using the same value. We also stratified patients using baseline information to explore the performance of the model across different populations.

Additionally, we performed a Kaplan–Meier survival analysis using the available follow-up data stratified by AI-positive predictions for the outcome of all-cause mortality. A Cox proportional hazards model was used to calculate the hazard ratios (HRs) with 95% confidence intervals (95% CIs) for all data. To account for potential competing risks with all-cause mortality, the R package "cmprsk" was used to calculate cumulative incidence.

Results

Baseline Characteristics

Characteristics of patients in the development, tuning, as well as internal and external validation cohorts, are shown in Table 1. There were 159 (0.6%), 93 (0.6%), and 88 (0.6%) patients with colon cancer in the development, internal validation, and external validation cohorts, respectively. In the development cohort, patients were younger, the proportion of males was lower, and the disease history was shorter than in the external validation cohort. The comparisons between patients with colon cancer and controls are shown in Table 2. There were 52 (55.9%) and 31 (35.2%) patients with tumor stages 3–4 (high grade) in the internal and external validation cohorts, respectively. In both validation cohorts, the

Table 1 Baseline characteristics in each dataset

	Development	Tuning	Internal validation	External validation
Colon cancer				
With	159 (0.6%)	61 (0.5%)	93 (0.6%)	88 (0.6%)
Without	27,896 (99.4%)	11,173 (99.5%)	16,782 (99.4%)	15,788 (99.4%)
KUB X-ray source				
ED	9285 (33.1%)	3740 (33.3%)	5566 (33.0%)	5491 (34.6%)
IPD	2526 (9.0%)	982 (8.7%)	1553 (9.2%)	1291 (8.1%)
OPD	16,244 (57.9%)	6512 (58.0%)	9756 (57.8%)	9094 (57.3%)
Demography				
Gender (male)	15,106 (57.0%)	6108 (57.8%)	9170 (57.5%)	9303 (58.6%)
Age (years)	46.9 ± 22.2	47.0 ± 22.2	46.7 ± 22.2	49.9 ± 21.6
BMI (kg/m^2)	23.8 ± 4.6	23.9 ± 4.6	23.8 ± 4.6	24.2 ± 4.4
Disease history				
DM	2900 (10.9%)	1171 (11.1%)	1682 (10.6%)	2509 (15.8%)
HTN	407 (1.5%)	157 (1.5%)	263 (1.6%)	393 (2.5%)
HLP	3747 (14.1%)	1463 (13.8%)	2186 (13.7%)	3699 (23.3%)
CKD	1257 (4.7%)	464 (4.4%)	733 (4.6%)	932 (5.9%)
CAD	2034 (7.7%)	853 (8.1%)	1212 (7.6%)	1895 (11.9%)
HF	649 (2.4%)	265 (2.5%)	374 (2.3%)	662 (4.2%)
COPD	1812 (6.8%)	732 (6.9%)	1113 (7.0%)	1864 (11.7%)

ED, emergency department; *IPD*, inpatient department; *OPD*, outpatient department; *BMI*, body mass index; *DM*, diabetes mellitus; *HTN*, hypertension; *HLP*, hyperlipidemia; *CKD*, chronic kidney disease; *CAD*, coronary artery disease; *HF*, heart failure; *COPD*, chronic obstructive pulmonary disease

Table 2 Comparison between colon cancer cases and controls

	Internal validation			External validation		
	Case	Control	*p*-value	Case	Control	*p*-value
Grade						
3–4	52 (55.9%)			31 (35.2%)		
1–2	36 (38.7%)			46 (52.3%)		
Unknown	5 (5.4%)			11 (12.5%)		
KUB X-ray source			< 0.001			< 0.001
ED	21 (22.6%)	5545 (33.0%)		21 (23.9%)	5470 (34.6%)	
IPD	23 (24.7%)	1530 (9.1%)		19 (21.6%)	1272 (8.1%)	
OPD	49 (52.7%)	9707 (57.8%)		48 (54.5%)	9046 (57.3%)	
Demography						
Gender (male)	53 (57.6%)	9117 (57.5%)	0.986	50 (56.8%)	9253 (58.6%)	0.734
Age (years)	71.9 ± 14.8	46.6 ± 22.1	< 0.001	70.4 ± 15.0	49.8 ± 21.6	< 0.001
BMI (kg/m^2)	23.0 ± 3.7	23.8 ± 4.6	0.148	23.4 ± 4.3	24.2 ± 4.4	0.134
Disease history						
DM	15 (16.3%)	1667 (10.5%)	0.072	18 (20.5%)	2491 (15.8%)	0.230
HTN	0 (0.0%)	263 (1.7%)	0.410	4 (4.5%)	389 (2.5%)	0.174
HLP	13 (14.1%)	2173 (13.7%)	0.907	26 (29.5%)	3673 (23.3%)	0.165
CKD	9 (9.8%)	724 (4.6%)	0.038	11 (12.5%)	921 (5.8%)	0.008
CAD	8 (8.7%)	1204 (7.6%)	0.692	16 (18.2%)	1879 (11.9%)	0.070
HF	6 (6.5%)	368 (2.3%)	0.021	8 (9.1%)	654 (4.1%)	0.030
COPD	12 (13.0%)	1101 (6.9%)	0.022	16 (18.2%)	1848 (11.7%)	0.060

ED, emergency department; *IPD*, inpatient department; *OPD*, outpatient department; *BMI*, body mass index; *DM*, diabetes mellitus; *HTN*, hypertension; *HLP*, hyperlipidemia; *CKD*, chronic kidney disease; *CAD*, coronary artery disease; *HF*, heart failure; *COPD*, chronic obstructive pulmonary disease

number of KUB radiographs from the inpatient department (IPD) was significantly higher in patients with cancer than in those without cancer. In addition, patients with cancer were significantly older than those without cancer in both validation cohorts.

Performance of KUB X-Ray to Detect CRC

The AUC of the performance of DLM in detecting CRC using KUB was 0.738, with corresponding sensitivity and specificity of 61.3% and of 74.4%, respectively, in the internal validation cohort, while in the external validation cohort, the AUC was 0.656, with corresponding sensitivity and specificity of 47.7% and 72.9%, respectively, as shown in Fig. 2. Furthermore, in internal and external validation cohorts, the algorithm performed better in predicting high-grade CRC, with AUCs of 0.744 and 0.674, respectively. The model was further evaluated with F1-score, accuracy, and confusion matrix as shown in Supplementary Table 1 and Fig. S1. In addition, we performed a stratified analysis to compare AUCs in detecting CRC using our DLM algorithm. The results are shown in Fig. 3. In the internal validation cohort, the model had an outstanding prediction performance (AUC = 0.977 95%CI 0.937–1.000) for female patients who were in the group of 55–64 years old and were diagnosed with high-grade colon cancer. In the external validation cohort, the model performed well for females under 55 years of age and those aged 55–64 with high-grade cancer, with AUCs of 0.913 (95% CI 0.832–0.995) and 0.914 (95% CI 0.754–1.000), respectively. Furthermore, in this analysis, the AUC showed a decreasing trend for females who were diagnosed with high-grade cancer after 74 years of age in both the internal and external validation cohorts. No other significant differences were observed between the subgroups.

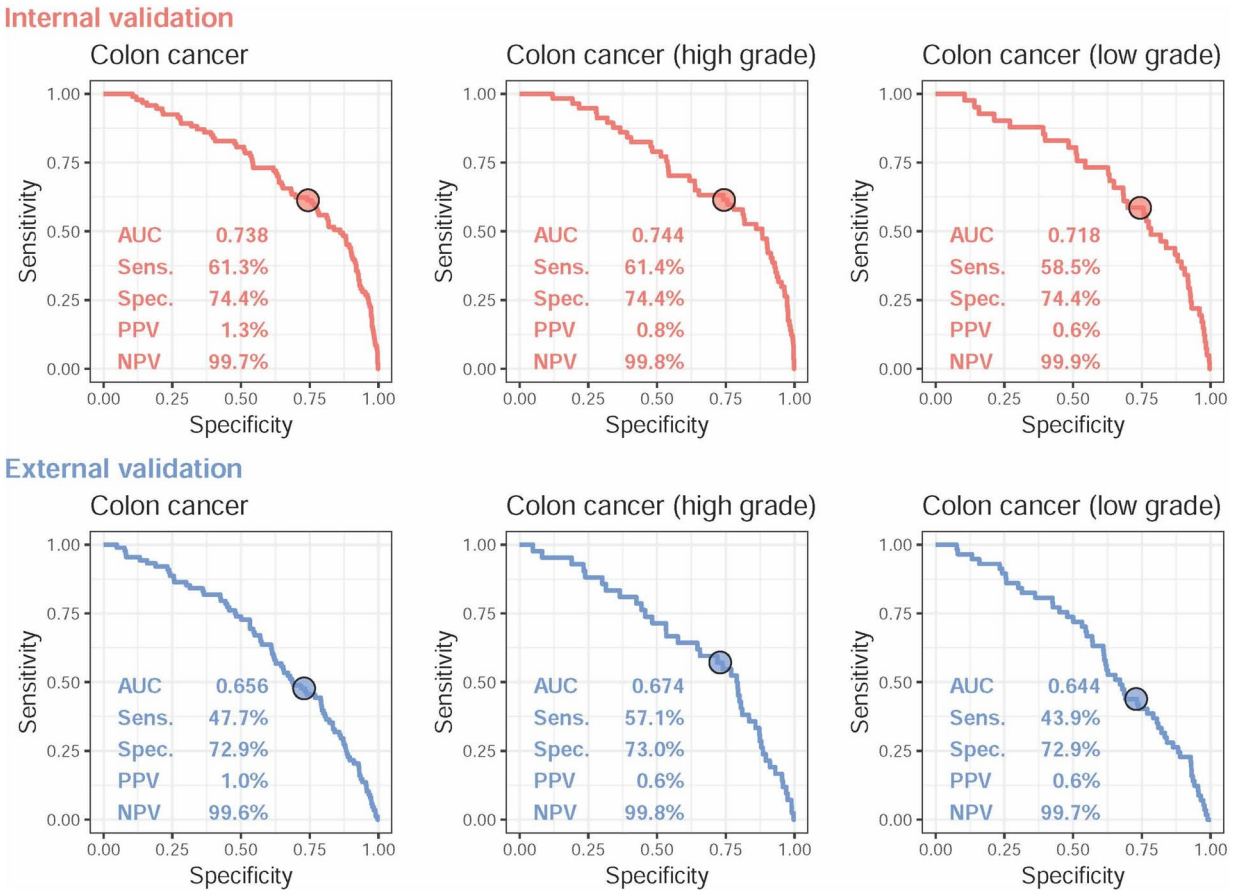

Fig. 2 The ROC curve of DLM predictions based on KUB X-ray to detect colon cancer. The cutoff point was selected based on the maximum of Youden's index in tuning set and presented using a circle mark, and the area under ROC curve (AUC), sensitivity (Sens.), specificity (Spec.), positive predictive value (PPV), and negative predictive value (NPV) were calculated based on it. The high grade was the pathology or clinical tumor stage of ≥ 3, and the low grade was the tumor stage of ≤ 2. Patients without pathological or image report were excluded in these grade-stratified analyses

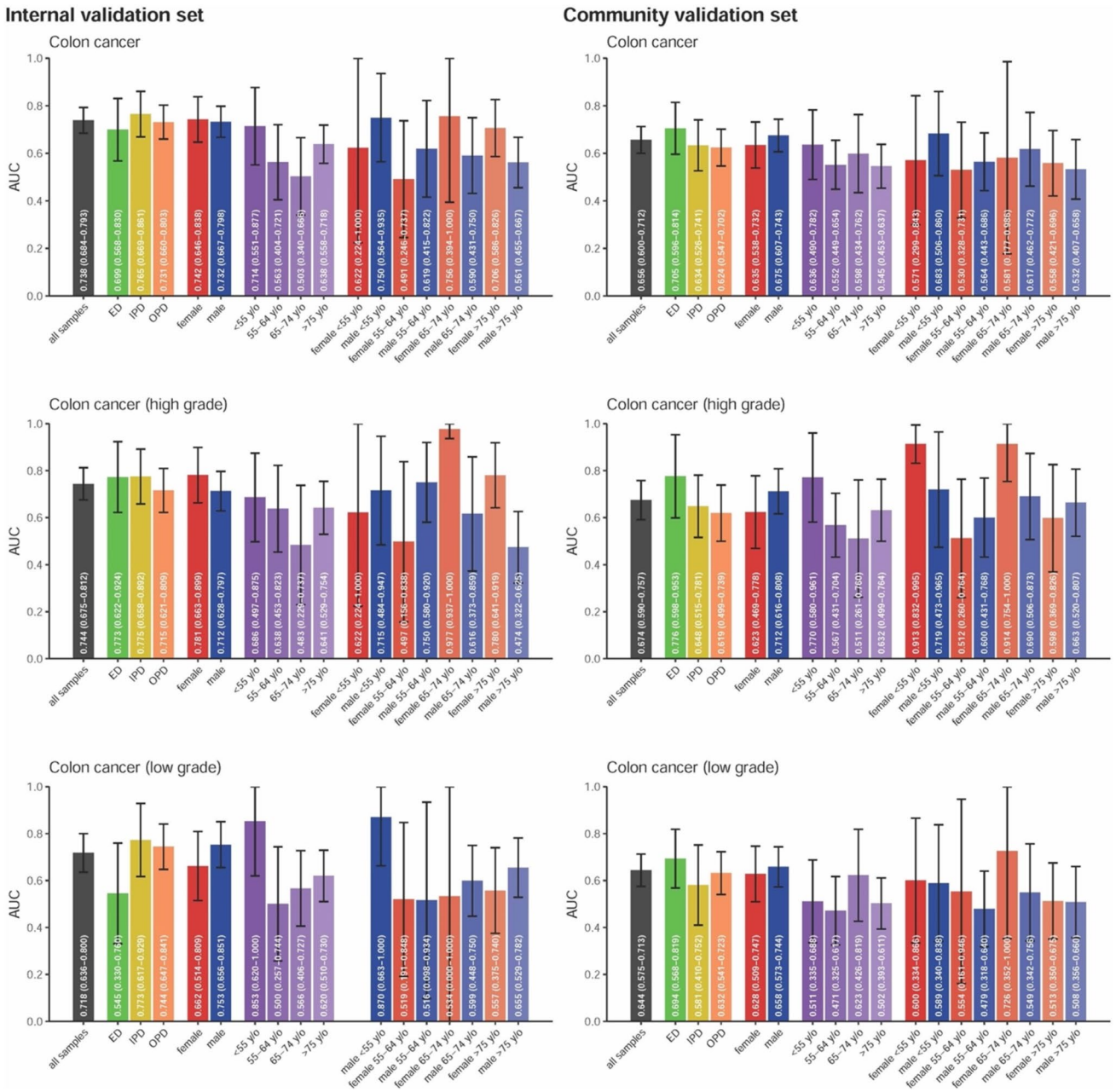

Fig. 3 Stratified analysis for the area under ROC curve (AUC) comparison in detecting colon cancer using DLM predictions based on KUB X-ray. The analyses are stratified by the source of KUB X-ray and demographic data

Prediction of Long-Term Incidence of All-Cause Mortality

The predicted incidence of all-cause mortality in patients without colon cancer stratified by the AI-predicted cancer after adjustment for age and sex is shown in Fig. 4. In the internal validation cohort, the cumulative all-cause mortality was 3.4% at 2.5 years and 5.3% at 5 years for AI-positive patients compared with 0.4% and 0.9% for AI-negative patients, respectively. In the external validation cohort, the rates were 2.7% and 5.5% for AI-positive patients, as well as 0.4% and 1% for AI-negative patients. This indicated a higher risk of future death when patients without cancer were detected positive for cancer than those detected negative for cancer by our DLM (internal and external validation cohorts HR [95% CI]: 2.13 [1.60–2.83] and 1.64 [1.24–2.16]).

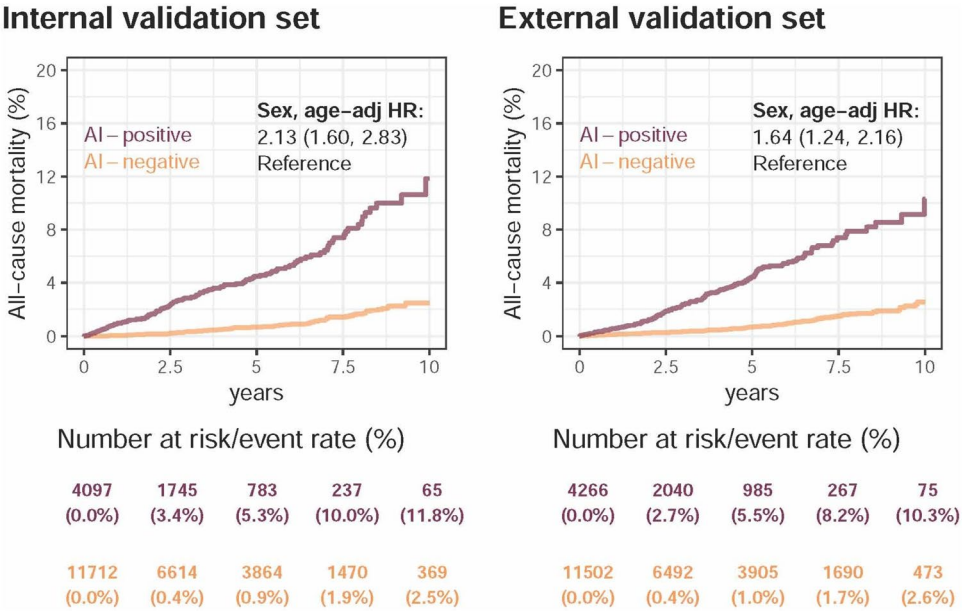

Fig. 4 Long-term incidence of developing all-cause mortality in patients without colon cancer. The analyses are conducted both in internal and external validation sets. The table shows the at-risk population and cumulative risk for the given time intervals in each risk stratification

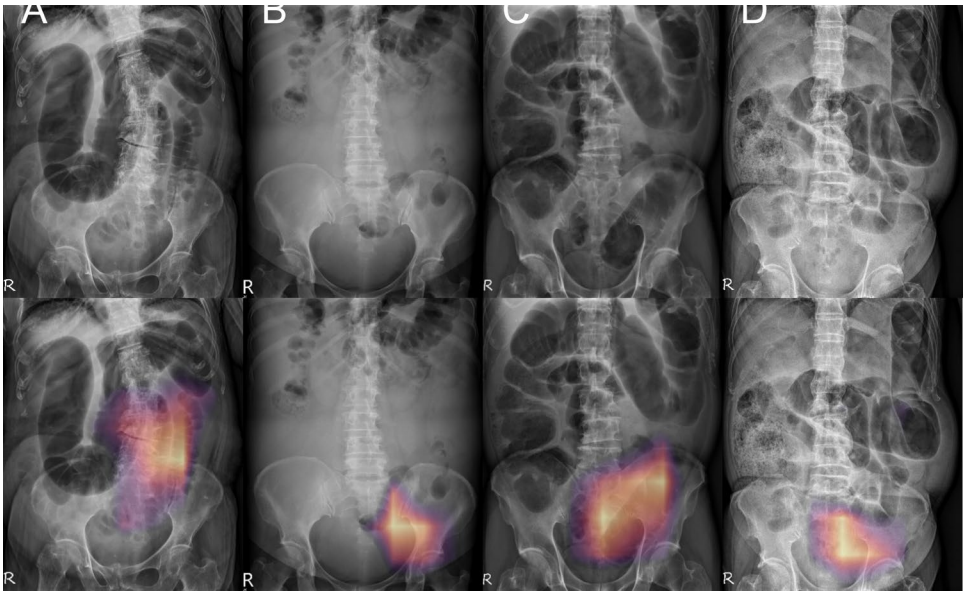

Fig. 5 Class activation map (CAM) of selected patients in validation sets. Images demonstrate how artificial intelligence (AI) using kidney, ureter, bladder (KUB), and X-ray makes decisions. Up: Original KUB X-ray in posterior-anterior view. Down: Class activation maps generated by overlays colored probability maps onto the original image. Observations indicate that AI-enabled KUB X-ray primarily uses information from the large bowel region (peripheral), while excluding regions overlapping with small bowel region (central). This phenomenon may be attributed to the anatomical structure of large bowel that commonly lies in the peripheral region of abdomen, which provides AI-enabled KUB X-ray with a clearer perspective for assessing colonic lesion or manifestations such as large bowel obstruction. The AI-enabled KUB X-ray mainly focused on the large bowel located in the lower part of the abdomen around pelvic region: **A** a 99-year-old woman with hypertension and senile dementia and with the diagnosis of adenocarcinoma of descending colon; **B** a 65-year-old woman with abdominal distention and bloody stool and with diagnosis of adenocarcinoma of sigmoid colon; **C** a 61-year-old man with epigastric pain and constipation and a diagnosis of adenocarcinoma of sigmoid colon; and **D** an 79-year-old man with DM and HCVD and with a diagnosis of adenocarcinoma of sigmoid-rectal junction

Lesion Localization Through Heat-map

Our study provided an additional feature using a heat map for highlighting potential cancer regions in KUB X-rays. In Fig. 5, four KUBs that were mainly distal CRC cases are presented, which are highlighted in the predicted heatmap. The true cancerous region highly overlapped with the areas highlighted by the AI, which was verified by experienced physicians in the EMR.

Discussion

Effective and broadly used CRC screening is essential to increase the chances of successful treatment and reduce CRC-related mortality [23]. To the best of our knowledge, this is the first study to explore a noninvasive and accessible CRC screening method using AI-KUB. In this study, we developed a DLM to detect CRC using KUB X-ray images, achieving an AUC of 0.738, with sensitivity and specificity of 61.3% and 74.4%, respectively, in the internal validation cohort. Furthermore, the DLM demonstrated enhanced performance in detecting high-grade CRC, with an AUC of 0.744. Additionally, our AI-KUB algorithm may serve as a noninvasive and broadly applicable screening tool for assessing the long-term risk of all-cause mortality in patients without CRC.

There are several manifestations of CRC, including LBO, bowel dilation, and bowel wall thickening [24]. LBO is primarily caused by colonic adenocarcinoma, accounting for over 60% of all cases. KUB is often the first imaging study performed in patients suspected of having LBO [11]. However, the accuracy of human-read KUB in diagnosing LBO ranges from 50 to 80%, with moderate specificity [25]. In our study, our DLM achieved an AUC of 0.74, indicating relatively good performance in detecting CRC. This difference may be because the KUB is a 2-dimensional image, which physicians often investigate merely by identifying the bowel diameter and air–liquid level, resulting in limited capability [11]. In contrast, compared with the accuracy of human-read KUB, our results suggest that AI may have the ability to identify additional findings to detect CRC. In summary, AI-KUB has the potential and reasonable capabilities to identify patterns, such as LBO and dilated large bowels, for CRC detection.

High-grade CRC tumors, which show over 50% histological gland formation, indicate poorer differentiation than low-grade tumors and generally have a worse prognosis and lower survival rates [26, 27]. Therefore, we assumed that the current screening method, FOBT, would have better screening performance for high-grade CRC tumors. However, few previous studies have demonstrated this finding. One possible explanation may be that most of the screening cases are asymptomatic and are more correlated with low-grade CRC tumors, while high-grade CRC tumors usually develop more symptomatically, which may directly lead to a diagnostic colonoscopy; therefore, the screening performance of high-grade CRC may be less discussed [28]. In addition, a CRC epidemiology report showed that high-grade CRC is associated with a poor 5-year survival rate of only 12% [29]. In our study, our model achieved relatively good performance in detecting high-grade CRC, with AUC of 0.744 and 0.674 in both validation cohorts. This result indicates that our algorithm has the potential to serve as an effective opportunistic screening tool for patients with high-grade CRC who are asymptomatic to have early awareness before performing a diagnostic colonoscopy, enabling early detection, and improving survival rates.

The manifestations of CRC, such as a dilated large bowel, increased gas retention, and colonic obstruction, may consequently lead to death [24]. Despite the potential of AI to identify these characteristics and increase the diagnosis rate, there remains a concern among physicians about false positives that may result in psychological distress in patients who are alerted [30]. Because of the limited research on using AI-enabled abdominal X-rays, we found that some previous studies using AI-enabled chest X-ray (CXR) to diagnose disease also encountered false-positive issues [31, 32]. However, these studies consistently found a correlation between false-positive results and preexisting abnormalities or adverse disease outcomes. In addition to effectively identifying patients with CRC, AI-positive results may predict long-term risk of mortality [33]. In our study, we also validated that the long-term incidence of all-cause mortality in patients without a CRC diagnosis, stratified by AI-positive results, was significantly increased by 1.5- to twofold compared with those stratified by AI-negative results in both validation cohorts. These results suggest that AI can identify not only the abnormalities of the targeted disease but also subtle findings that are linked to severe medical crises, which explains the findings of increased all-cause mortality in the AI-positive group. However, further prospective studies are needed to validate this predictive model.

Statistically, left-sided CRC accounts for most CRC cases, a distribution that was also observed in our study [34]. Compared to right-sided CRC, left-sided tumors exhibit different molecular characteristics and histology, such as a higher prevalence of polypoid-like tumors [34]. Our DLM demonstrated a reasonable ability to detect CRC lesions using KUB radiography, particularly in the distal colon as shown in Fig. 5. This result aligns with the findings from FOBT screening and may be attributed to the etiology of CRC, where tumors in the distal colon typically present as polyps, unlike the flat morphology observed in the proximal colon [35, 36]. However, further investigation is required.

We also compared our method to some state-of-the-art approaches. Most of the studies use CNN-based architecture to accomplish the image classification tasks. Marcello Di Giammarco et al. proposed a study testing different DLMs using histological images and achieved over AUC of 0.99 on MobileNet [19]. This outstanding performance may because of the histological image has a more microscopic view to the pathological aspects of the disease. However, in the same study, they also try DenseNet but only reach AUC of 0.696, while our DLM achieved AUC of 0.74, indicating that our approach of using KUB image may have sufficient potential in training model for detecting colon cancer.

AI-enabled KUB radiographs have several clinical applications. First, it may serve as an improved tool for CRC screening owing to its extensive coverage. Compared with the current CRC screening method, the gFOBT, which has an AUC of 0.77, our DLM demonstrated an AUC of 0.74 [37]. Although the AUC was slightly lower than that of gFOBT, AI-KUB may be more effective owing to issues of adherence and limited coverage associated with gFOBT [38]. Second, integrating AI with KUB enables not only the detection of urinary issues and early assessments in abdominal emergencies but also opportunistic CRC screening. Third, this algorithm can be incorporated into primary healthcare to identify patients at long-term risk of mortality.

This study had several limitations. First, being a retrospective study, further prospective studies are necessary to validate the performance and clinical applications of our AI-KUB algorithm. Second, cancer registration files may miss diagnoses from other hospitals, but the influence is likely minimal. Third, the limited number of CRC cases in our study could affect the model's robustness and generalizability, necessitating further evaluation. Additionally, there was an issue of the data imbalance that the tumor location of the cases in our study were mostly in the distal colon, the further cases of proximal colon lesion included may be necessary for improving our algorithm. Fourthly, the relatively low performance of our DLM in the external validation cohort may be due to the inadequate CRC cases and the patients from the community hospital may not be representative enough to the population. Lastly, despite heat map analyses confirming their rationality, the explanations for how AI recognizes images to detect lesions remain unclear. Further evaluation and validation are still required for clinical use.

Conclusion

In conclusion, this study introduced a novel approach for early CRC detection using AI-enabled KUB X-ray images, marking an advancement in non-invasive cancer screening methods. The application of the DLM to KUB X-rays offers a promising alternative to traditional screening techniques and is effective for high-grade tumor detection. This study not only demonstrated the potential for improving early detection but also indicated a higher risk of long-term all-cause mortality in patients without CRC. By providing a cost-effective, accessible, and less invasive option, this AI-enabled method may potentially impact current clinical practices and patient outcomes in CRC care.

There are several contributions of this study: (i) It has a potential to use as a real-time opportunistic screening tool. (ii) It can provide an additional screening method for younger population. (iii) It is a more non-invasive and cost-effective screening method.

In this study, we demonstrated that KUB imaging, one of the most used modalities, has the potential to serve as a valuable image type for training DLMs in the detection of colorectal cancer. In future work, we planned to explore other state-of-the-art models trained on KUB X-ray images, while also addressing dataset imbalances, with the goal of developing a non-invasive alternative screening method for colorectal cancer.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10278-024-01309-1.

Authors' Contributions All authors contributed to the study conception and design. Material preparation and data collection were done by Chia-Jung Hsu and Heng-Hsiu Lin. Data analysis were performed by Chin Lin. The first draft of the manuscript was written by Ling Lee, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability The data supporting the findings of this study are available from the corresponding author upon reasonable request. Due to privacy and ethical considerations, the data are not publicly accessible.

Declarations

Ethics Approval This study was approved by the Institutional Review Board of Tri-Service General Hospital (TSGH), Taipei, Taiwan (IRB no. C202305019).

Consent to Participate This retrospective study was conducted using anonymized patient data from the hospital database.

Consent to Publish This retrospective study was conducted using anonymized patient data from the hospital database.

Competing Interest The authors declare no competing interests.

References

1. Xi Y, Xu P: Global colorectal cancer burden in 2020 and projections to 2040. Transl Oncol 14:101174, 2021
2. Meester RG, et al.: Colorectal cancer deaths attributable to nonuse of screening in the United States. Ann Epidemiol 25:208-213 e201, 2015

3. Ladabaum U, Dominitz JA, Kahi C, Schoen RE: Strategies for Colorectal Cancer Screening. Gastroenterology 158:418-432, 2020
4. Young PE, Womeldorph CM: Colonoscopy for colorectal cancer screening. J Cancer 4:217-226, 2013
5. James B, Kelly B: The abdominal radiograph. Ulster Med J 82:179-187, 2013
6. Brisbane W, Bailey MR, Sorensen MD: An overview of kidney stone imaging techniques. Nat Rev Urol 13:654-662, 2016
7. Hampson F, Shaw AS. Assessment of the acute abdomen: role of the plain abdominal radiograph. Reports in Medical Imaging 3:93-105, 2010. https://doi.org/10.2147/RMI.S13837
8. Artigas Martin JM, Marti de Gracia M, Rodriguez Torres C, Marquina Martinez D, Parrilla Herranz P: Routine abdominal X-rays in the emergency department: a thing of the past? Radiologia 57:380-390, 2015
9. Markus JB, Somers S, Franic SE, Moola C, Stevenson GW: Interobserver variation in the interpretation of abdominal radiographs. Radiology 171:69-71, 1989
10. Frago R, Ramirez E, Millan M, Kreisler E, del Valle E, Biondo S: Current management of acute malignant large bowel obstruction: a systematic review. Am J Surg 207:127-138, 2014
11. Jaffe T, Thompson WM: Large-Bowel Obstruction in the Adult: Classic Radiographic and CT Findings, Etiology, and Mimics. Radiology 275:651-663, 2015
12. Azer SA: Challenges Facing the Detection of Colonic Polyps: What Can Deep Learning Do? Medicina (Kaunas) 55, 2019. https://doi.org/10.3390/medicina55080473
13. Urban G, et al.: Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. Gastroenterology 155:1069-1078.e1068, 2018
14. Ozawa T, Ishihara S, Fujishiro M, Kumagai Y, Shichijo S, Tada T: Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks. Therap Adv Gastroenterol 13:1756284820910659, 2020
15. Pacal I, Karaboga D, Basturk A, Akay B, Nalbantoglu U: A comprehensive review of deep learning in colon cancer. Comput Biol Med 126:104003, 2020
16. Kim DH, et al.: An artificial intelligence deep learning model for identification of small bowel obstruction on plain abdominal radiographs. Br J Radiol 94:20201407, 2021
17. Huang YC, et al.: Artificial intelligence-enabled electrocardiographic screening for left ventricular systolic dysfunction and mortality risk prediction. Front Cardiovasc Med 10:1070641, 2023
18. Graham AR, Paplanus SH, Bartels PH: A diagnostic expert system for colonic lesions. Am J Clin Pathol 94:S15-18, 1990
19. Di Giammarco M, Martinelli F, Santone A, Cesarelli M, Mercaldo F: Colon cancer diagnosis by means of explainable deep learning. Scientific Reports 14:15334, 2024
20. Kavitha MS, Gangadaran P, Jackson A, Venmathi Maran BA, Kurita T, Ahn B-C: Deep Neural Network Models for Colon Cancer Screening. Cancers 14:3707, 2022
21. Masud M, Sikder N, Nahid A-A, Bairagi AK, Alzain MA: A Machine Learning Approach to Diagnosing Lung and Colon Cancer Using a Deep Learning-Based Classification Framework. Sensors 21:748, 2021
22. Rajpurkar P, et al.: Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med 15:e1002686, 2018
23. Zheng S, Schrijvers JJA, Greuter MJW, Kats-Ugurlu G, Lu W, de Bock GH: Effectiveness of Colorectal Cancer (CRC) Screening on All-Cause and CRC-Specific Mortality Reduction: A Systematic Review and Meta-Analysis. Cancers (Basel) 15, 2023. https://doi.org/10.3390/cancers15071948
24. Cappell MS, Batke M: Mechanical obstruction of the small bowel and colon. Med Clin North Am 92:575-597, viii, 2008
25. Nelms DW, Kann BR: Imaging Modalities for Evaluation of Intestinal Obstruction. Clin Colon Rectal Surg 34:205-218, 2021
26. Lee CH, Cheng SC, Tung HY, Chang SC, Ching CY, Wu SF: The Risk Factors Affecting Survival in Colorectal Cancer in Taiwan. Iran J Public Health 47:519-530, 2018
27. Chen K, Collins G, Wang H, Toh JWT: Pathological Features and Prognostication in Colorectal Cancer. Curr Oncol 28:5356-5383, 2021
28. Lee MH, Hinshaw JL, Kim DH, Pickhardt PJ: Symptomatic Versus Asymptomatic Colorectal Cancer: Predictive Features at CT Colonography. Acad Radiol 23:712-717, 2016
29. Rawla P, Sunkara T, Barsouk A: Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Prz Gastroenterol 14:89-103, 2019
30. Mori Y, Bretthauer M: Addressing false-positive findings with artificial intelligence for polyp detection. Endoscopy 53:941-942, 2021
31. Wehbe RM, et al.: DeepCOVID-XR: An Artificial Intelligence Algorithm to Detect COVID-19 on Chest Radiographs Trained and Tested on a Large U.S. Clinical Data Set. Radiology 299:E167-E176, 2021
32. Tsai DJ, Lin C, Lin CS, Lee CC, Wang CH, Fang WH: Artificial Intelligence-enabled Chest X-ray Classifies Osteoporosis and Identifies Mortality Risk. J Med Syst 48:12, 2024
33. Hsiang CW, et al.: Detection of Left Ventricular Systolic Dysfunction Using an Artificial Intelligence-Enabled Chest X-Ray. Can J Cardiol 38:763-773, 2022
34. De Renzi G, Gaballo G, Gazzaniga P, Nicolazzo C: Molecular Biomarkers according to Primary Tumor Location in Colorectal Cancer: Current Standard and New Insights. Oncology 99:135-143, 2021
35. Allison JE, et al.: Screening for colorectal neoplasms with new fecal occult blood tests: update on performance characteristics. J Natl Cancer Inst 99:1462-1470, 2007
36. Nawa T, et al.: Differences between right- and left-sided colon cancer in patient characteristics, cancer morphology and histology. J Gastroenterol Hepatol 23:418-423, 2008
37. Meklin J, SyrjAnen K, Eskelinen M: Fecal Occult Blood Tests in Colorectal Cancer Screening: Systematic Review and Meta-analysis of Traditional and New-generation Fecal Immunochemical Tests. Anticancer Res 40:3591-3604, 2020
38. Miller Wilson LA, et al.: Opportunities and Challenges in Screening for Colorectal Cancer. Popul Health Manag 26:246-253, 2023