Development and validation of artificial intelligence-based prescreening of large-bowel biopsies taken in the UK and Portugal: a retrospective cohort study

Mohsin Bilal*, Yee Wah Tsang*, Mahmoud Ali, Simon Graham, Emily Hero, Noorul Wahab, Katherine Dodd, Harvir Sahota, Shaobin Wu, Wenqi Lu, Mostafa Jahanifar, Andrew Robinson, Ayesha Azam, Ksenija Benes, Mohammed Nimir, Katherine Hewitt, Abhir Bhalerao, Hesham Eldaly, Shan E Ahmed Raza, Kishore Gopalakrishnan, Fayyaz Minhas, David Snead†, Nasir Rajpoot†

Lancet Digit Health 2023; 5: e786–97

*Joint first authors

†Joint last authors

Tissue Image Analytics Centre, Department of Computer Science (M Bilal PhD, S Graham PhD, N Wahab PhD, W Lu PhD, M Jahanifar MSc, Prof A Bhalerao PhD, S E A Raza PhD, F Minhas PhD, Prof N Rajpoot PhD) **and Warwick Medical School** (Prof D Snead MBBS), **University of Warwick, Coventry, UK; Department of Pathology, University Hospitals Coventry and Warwickshire National Health Service Trust, Coventry, UK** (Y W Tsang MBBS, M Ali MBBS, E Hero MBBS, K Dodd MBBS, H Sahota MBBS, A Robinson MBBS, A Azam MBBS, M Nimir MBBS, K Hewitt MBBS, H Eldaly MBBS, K Gopalakrishnan MBBS, Prof D Snead, Prof N Rajpoot); **Histofy, Birmingham, UK** (S Graham, Prof D Snead, Prof N Rajpoot); **Department of Artificial Intelligence and Data Science, National University of Computer and Emerging Sciences, Islamabad, Pakistan** (M Bilal); **Department of Pathology, University Hospitals of Leicester National Health Service Trust, Leicester, UK** (E Hero); **Department of Pathology, East Suffolk and North Essex National Health Service Foundation Trust, Colchester, UK** (S Wu MBBS); **Department of Pathology, The Royal Wolverhampton National Health Service Trust, Wolverhampton, UK** (K Benes MBBS); **The Alan Turing Institute, London, UK** (Prof N Rajpoot)

Correspondence to:
Nasir Rajpoot, Tissue Image Analytics Centre, Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK
N.M.Rajpoot@warwick.ac.uk

Summary

Background Histopathological examination is a crucial step in the diagnosis and treatment of many major diseases. Aiming to facilitate diagnostic decision making and improve the workload of pathologists, we developed an artificial intelligence (AI)-based prescreening tool that analyses whole-slide images (WSIs) of large-bowel biopsies to identify typical, non-neoplastic, and neoplastic biopsies.

Methods This retrospective cohort study was conducted with an internal development cohort of slides acquired from a hospital in the UK and three external validation cohorts of WSIs acquired from two hospitals in the UK and one clinical laboratory in Portugal. To learn the differential histological patterns from digitised WSIs of large-bowel biopsy slides, our proposed weakly supervised deep-learning model (Colorectal AI Model for Abnormality Detection [CAIMAN]) used slide-level diagnostic labels and no detailed cell or region-level annotations. The method was developed with an internal development cohort of 5054 biopsy slides from 2080 patients that were labelled with corresponding diagnostic categories assigned by pathologists. The three external validation cohorts, with a total of 1536 slides, were used for independent validation of CAIMAN. Each WSI was classified into one of three classes (ie, typical, atypical non-neoplastic, and atypical neoplastic). Prediction scores of image tiles were aggregated into three prediction scores for the whole slide, one for its likelihood of being typical, one for its likelihood of being non-neoplastic, and one for its likelihood of being neoplastic. The assessment of the external validation cohorts was conducted by the trained and frozen CAIMAN model. To evaluate model performance, we calculated area under the convex hull of the receiver operating characteristic curve (AUROC), area under the precision-recall curve, and specificity compared with our previously published iterative draw and rank sampling (IDaRS) algorithm. We also generated heat maps and saliency maps to analyse and visualise the relationship between the WSI diagnostic labels and spatial features of the tissue microenvironment. The main outcome of this study was the ability of CAIMAN to accurately identify typical and atypical WSIs of colon biopsies, which could potentially facilitate automatic removing of typical biopsies from the diagnostic workload in clinics.

Findings A randomly selected subset of all large bowel biopsies was obtained between Jan 1, 2012, and Dec 31, 2017. The AI training, validation, and assessments were done between Jan 1, 2021, and Sept 30, 2022. WSIs with diagnostic labels were collected between Jan 1 and Sept 30, 2022. Our analysis showed no statistically significant differences across prediction scores from CAIMAN for typical and atypical classes based on anatomical sites of the biopsy. At 0·99 sensitivity, CAIMAN (specificity 0·5592) was more accurate than an IDaRS-based weakly supervised WSI-classification pipeline (0·4629) in identifying typical and atypical biopsies on cross-validation in the internal development cohort (p<0·0001). At 0·99 sensitivity, CAIMAN was also more accurate than IDaRS for two external validation cohorts (p<0·0001), but not for a third external validation cohort (p=0·10). CAIMAN provided higher specificity than IDaRS at some high-sensitivity thresholds (0·7763 *vs* 0·6222 for 0·95 sensitivity, 0·7126 *vs* 0·5407 for 0·97 sensitivity, and 0·5615 *vs* 0·3970 for 0·99 sensitivity on one of the external validation cohorts) and showed high classification performance in distinguishing between neoplastic biopsies (AUROC 0·9928, 95% CI 0·9927–0·9929), inflammatory biopsies (0·9658, 0·9655–0·9661), and atypical biopsies (0·9789, 0·9786–0·9792). On the three external validation cohorts, CAIMAN had AUROC values of 0·9431 (95% CI 0·9165–0·9697), 0·9576 (0·9568–0·9584), and 0·9636 (0·9615–0·9657) for the detection of atypical biopsies. Saliency maps supported the representation of disease heterogeneity in model predictions and its association with relevant histological features.

Interpretation CAIMAN, with its high sensitivity in detecting atypical large-bowel biopsies, might be a promising improvement in clinical workflow efficiency and diagnostic decision making in prescreening of typical colorectal biopsies.

Funding The Pathology Image Data Lake for Analytics, Knowledge and Education Centre of Excellence; the UK Government's Industrial Strategy Challenge Fund; and Innovate UK on behalf of UK Research and Innovation.

Introduction

Colorectal cancer is the fourth most common cancer in the UK and the second leading cause of cancer-related deaths in the UK and the USA.[1,2] More than 42 000 people are diagnosed with colorectal cancer every year in the UK[3] and more than 140 000 are diagnosed every year in the USA. Most colorectal cancers develop from polyps, a precancerous outgrowth of tissue from the lining of the colon. Colonoscopy has long been the reference standard for examining the rectum and colon for precancerous polyps, tumours, or other problems.[4,5] A tissue sample (biopsy) that is taken during a colonoscopy helps to make a definitive diagnosis of any colonic atypicalities. Microscopic examination of the histopathological features of colorectal biopsies is the standard way of providing a definitive diagnosis of any colonic pathology. A wide range of colonic atypicalities can be found, including carcinoma, polyps (eg, non-neoplastic or neoplastic), inflammation or ulceration caused by infection, medication-induced atypicalities, inflammatory bowel disease, and microscopic colitis.

Improvements in and increased roll-out of large-bowel cancer screening programmes, which are designed to detect colorectal cancer at the earliest stage, have contributed to an increase in the workload of pathologists. Furthermore, early-stage disease is often more difficult to detect than late-stage disease, and the requirement for precise diagnostic information to provide the best standard of care is contributing to the increased workload of pathologists and pathology laboratories worldwide.[6] The digitisation of cellular-pathology laboratories and an accelerated transition to digital pathology during the COVID-19 pandemic[7] offer an opportunity for automated prescreening of large-bowel biopsies to address the workload of pathologists and to improve clinical management.

Approximately 30–40% of large-bowel endoscopic biopsies are reported as typical and contain no discernible pathology, meaning pathologists spend a substantial amount of time looking for pathology that is non-existent.[8] Prescreening colon biopsies with an artificial intelligence (AI) tool could save time, allowing pathologists to dedicate more time to atypical biopsies, for which their expertise is needed most. This approach is different to the current contribution of AI to pathology, in which screening is modelled to differentiate between typical biopsies and cancer biopsies[9–12] or between cancer biopsies and inflammatory biopsies.[13,14] Current AI approaches might have benefits in improving the speed of reporting, but do not filter out the large number of typical biopsies that still need to be reported by a human pathologist. AI-based diagnostic tools for large-bowel biopsy prescreening to improve pathology workflow is an unmet need of high clinical relevance. Developing high-sensitivity AI models will lead to an effective, reliable, and low-risk AI-based prescreening tool that is suitable for the clinical workflow. To the best of our knowledge, there is no existing AI tool that can prescreen colorectal biopsies for both non-neoplastic and neoplastic atypicalities. We believe that this tool is a current need as connecting modern digital

Research in context

Evidence before this study

We searched Google Scholar without language restrictions from Jan 1, 2018, to Aug 30, 2022, using the search terms "screening or pre-screening of (colorectal or colon or rectal) cancer or biopsies AND (machine OR deep) learning OR (artificial intelligence OR AI)" and analysed the first 50 scientific articles when ranked by relevance, citation metrics, author metrics, publication date, and journal metrics. Several studies have used deep learning or machine learning to identify colorectal cancer or polyps in typical or inflammatory cancer slides directly from regular, whole-slide images (WSIs) of tissue sections. However, previous studies did not attempt to prescreen colonic biopsies with high sensitivity to differentiate the two major types of atypical colonic biopsy slides (ie, inflammatory and neoplastic) from typical slides. Studies have also suggested that previous, weakly supervised approaches did not achieve the level of accuracy needed for clinical practice.

Added value of this study

Motivated by an unmet clinical need, we developed a weakly supervised deep-learning model for multiclass WSI classification (with confidence prediction and using only slide-level labels) for prescreening colon biopsies. Our model was more accurate at identifying typical and atypical colonic biopsies in one internal development cohort and three external validation cohorts than state-of-the-art, weakly supervised classification methods developed in 2021.

Implications of all the available evidence

At a sensitivity value of 0·99 in predicting atypical colon biopsies, our artificial-intelligence (AI) tool can potentially be used in clinical workflows to automatically report typical colonoscopies. This could reduce the workload of pathologists and improve diagnostic efficiency and patient management. After large-scale validation and enhanced domain generalisation with data from multiple cohorts, a trial of our CAIMAN-based model could be conducted for clinical practice. AI-based prediction of slide labels and spatial mapping of the tissue microenvironment also offer the potential to improve diagnostic practice for other types of cancer as CAIMAN can be used to train models for other cancer types, which will only require WSIs with clinical diagnostic reports. In the future, weakly supervised and interpretable AI could be combined for improved diagnostic practice.

technologies and human expertise can provide a timely diagnosis, decrease cancer fatalities, and reduce undertreatment or overtreatment.

Since 2020, an increasing number of weakly supervised deep-learning methods for whole-slide image (WSI) classification have been proposed for various histopathology problems.[15–21] An attractive feature of weakly supervised methods is their ability to enable automatic classification without the need for detailed pixel-level or region-level annotations. Weakly supervised deep learning is a type of machine learning that requires slide or case labels only, which is less manual labelling than traditional supervised learning. These methods can work efficiently on thousands of WSIs, often after dividing them into smaller parts as image tiles, resulting in millions of image tiles. In 2021, Oliveira and colleagues[22] identified the limitations of existing algorithms that are applied for histopathology and emphasised the need for more accurate methods in clinical practice. They highlighted the need for large datasets and appropriate learning methods to improve prediction accuracy. In our previous study,[23] we proposed an iterative draw and rank sampling (IDaRS)-based weakly supervised WSI-classification pipeline for the prediction of molecular pathways, including microsatellite instability and genetic mutations of *BRAF*, *KRAS*, and *TP53* in colorectal cancer.

In this Article, we present a customised deep-learning algorithm for prescreening of colorectal biopsies based on digitised WSIs of biopsy slides that aimed to distinguish between typical and atypical biopsies to assist in histopathological examination. This multiclass, weakly supervised learning algorithm aimed to distinguish between typical, inflammatory, and neoplastic colonic biopsies and to learn confidence in tile-level predictions. We aimed to show that the proposed Colorectal AI Model for Abnormality Detection (CAIMAN) model can be used for prescreening colon biopsies and to establish its accuracy via cross-validation in a large internal development cohort and external validation in three unseen external validation cohorts.

Methods

Study design

This retrospective cohort study was conducted with an internal development cohort of slides, which was acquired from University Hospitals Coventry and Warwickshire National Health Service (NHS) Foundation Trust, and three external validation cohorts of WSIs, which were acquired from two hospitals in the UK (ie, South Warwickshire NHS Foundation Trust and East Suffolk and North Essex NHS Foundation Trust) and one clinical laboratory in Portugal (ie, IMP Diagnostics).[22] A randomly selected subset of all large bowel biopsies was obtained between Jan 1, 2012, and Dec 31, 2017. The AI training, validation, and assessments were done between Jan 1, 2021, and Sept 30, 2022 WSIs with diagnostic labels were collected between Jan 1 and Sept 30, 2022. The internal development cohort of large-bowel biopsies was scanned at University Hospitals Coventry and Warwickshire. All external cohorts were also large-bowel biopsies, which were prepared and scanned at their local premises.

This study was conducted under Health Research Authority National Research Ethics approval (15/NW/0843; IRAS 189 095) and Pathology Image Data Lake for Analytics, Knowledge and Education (PathLAKE) research ethics committee approval (REC 19/SC/0363; IRAS project ID 257 932; South Central Oxford C Research Ethics Committee). Data collection and use of the IMP Diagnostics dataset was done in accordance with national legal and ethical standards.[22]

Data collection and preparation

The internal development cohort contained 5054 slides from 2080 patients with diagnostic clinical reports who were indicated for bowel-cancer screening. This cohort contained the randomly selected subset of all colonoscopic biopsies. We excluded small-bowel slides and slightly atypical pathologically insignificant slides, including some spirochaetoses and melanosis slides from all cohorts. We also excluded slides that were severely out of focus (appendix 1 p 3).

See Online for appendix 1

All slides in the internal development cohort were diagnostic standard, haematoxylin and eosin stained, formalin-fixed paraffin-embedded (FFPE) histology slides that were scanned at a magnification of 40 (0·275 microns per pixels [MPP]) with the GE Omnyx JP2 scanner at University Hospitals Coventry and Warwickshire. We used all slides from the 2080 patients. The slides were reviewed by two pathologists, with one additional pathologist providing consensus diagnoses when there was a discrepancy on slide-level diagnoses (one of YWT, KG, SW, EH, AR, AA, HE, KD, HS, MN, KH, KB, or DS). The three external validation cohorts were used for independent validation of the proposed CAIMAN tool for prescreening colon-tissue slides. These cohorts contained 148 patients from East Suffolk and North Essex NHS Foundation Trust, 257 patents from South Warwickshire NHS Foundation Trust, and 1131 patients from the IMP Diagnostics laboratory. The sample sizes were randomly chosen for the UK cohorts; the Portuguese cohort had been used in a previous study. All slides in the external cohorts were diagnostic standard, haematoxylin and eosin stained, and FFPE. The specimen sites and percentages of biopsies from University Hospitals Coventry and Warwickshire, as well as the atypical subconditions that were present in colorectal biopsies from all cohorts, are shown in appendix 1 (p 2). The data selection for both the internal development and external validation cohorts is also shown in appendix 1 (p 3).

This study was conducted with retrospective data from histopathology archives relating to samples taken during clinical care, and for which consent for research had not been taken. Gathering consent retrospectively was not feasible and was deemed not necessary by the research ethics committee.

Data preprocessing

Tissue regions were segmented from each WSI to extract tiles that corresponded to the tissue areas only (figure 1A). Square tiles of 256×256 pixels were extracted with a stride of 128 pixels (50% overlap) from a downscaled version of a WSI at the objective magnification of 5 from the segmented tissue regions, corresponding to 2·2 MPP. A tile was kept for subsequent processing if it contained more than 50% tissue specimen. That tiles on the border of the tissue content included the white image background was normal. WSIs with fewer than four tiles were excluded. We conducted an image-quality check to identify and remove blurry artifacts and pen marks using Canny edge detection in Python OpenCV version 4.5.5.64.[20]

To evaluate the robustness of the algorithm to staining variations, we downsampled images from the external validation cohorts to match the resolution of the internal development cohort. Large tiles that were obtained from the external validation cohorts were downscaled with bilinear interpolation to the required pixel resolution at an objective magnification of 5 with the resize() function of Python image library version 6.2.0.

For **ImageNet** see https://www.image-net.org/

More than 50 heatmaps of correctly predicted biopsies were reviewed by a pathologist (YWT) and an arbitrary selection of those biopsies were also verified by the entire team of 13 pathologists (YWT, KG, SW, EH, AR, AA, HE, KD, HS, MN, KH, KB, and DS). The false-positive and false-negative samples were reviewed by two pathologists (YWT and KG) and, if needed, were discussed by the pathology-review board of five pathologists (YWT, KG, AR, HE, and DS).

Weakly supervised deep-learning model CAIMAN

We proposed colorectal biopsy prescreening as a three-class classification problem, in which each WSI was classified into one of three classes (ie, typical, atypical non-neoplastic, and atypical neoplastic). The proposed CAIMAN model was developed by training a deep convolutional neural network, an advanced computer programme that can see and comprehend images by automatically detecting and learning distinguishing features (including subvisual cues and patterns within the image), for three-class classification on image tiles from training slides of the internal development cohort (figure 1). Prediction scores of image tiles were then aggregated into three prediction scores for the whole slide, one for its likelihood of being typical, one for its likelihood of being non-neoplastic, and one for its likelihood of being neoplastic. The assessment of the external validation cohorts was conducted by the trained and frozen CAIMAN model.

As training labels were only available at the WSI level, and not all tiles in a WSI were predictive of the diagnostic label of the WSI, the CAIMAN model used a weakly supervised classification pipeline that was based on our previously published IDaRS algorithm.[23] Similar to IDaRS, CAIMAN works by conducting iterative draw and rank sampling of tiles from a set of training WSIs. However, not similar to IDaRS, CAIMAN addresses colon biopsy prescreening as a multiclass, weakly supervised learning problem and learns to predict confidence scores.

Instead of using all tiles, we chose three subsets of image tiles from each slide for training (figure 1B). For the first two iterations, we split the whole training set into two sets by randomly dividing the tiles of each slide into two halves, one for the first training iteration and one for the second training iteration. We obtained the top 50% of tiles from the first two iterations and combined them into an initial fixed set for the rest of the training iterations. For the next and subsequent training iterations, we had an initial fixed subset of top tiles, a subset of a randomly taken proportion of tiles (d), and a subset of the top-ranked proportion of the training tiles (k). The parameters for CAIMAN were empirically chosen as d=50% and k=50%, with a maximum of ten training iterations and a batch size of 1024 tiles. We refined the backbone ResNet34 network,[24] which was originally pretrained on the ImageNet dataset, on the internal development cohort.

For weakly supervised training in an end-to-end way, we added another output branch of a single neuron in the backbone neural network that learned to estimate confidence for the prediction of each tile (figure 1B).[25] This approach incentivised the neural network to produce confidence estimates that accurately reflected its ability to make correct predictions for any input tile, as per its weakly supervised label, in exchange for a reduction in loss. We modified the batch training loss by combining the tile and confidence losses with the slide loss for atypical slides only. Slide loss for each atypical slide was calculated separately by finding the mean of the losses of all tiles of the atypical slide and the losses of all typical tiles. A mathematical description of loss function is provided in appendix 1 (p 6). During the training process, we iteratively used CAIMAN to produce confidence-weighted prediction scores for each tile and selected the top-ranking k tiles of each slide (ie, tiles with high likelihood of being atypical).

The CAIMAN model assigned three probability values and a confidence score to each tile at the time of inference (figure 1C). The scale of probability was between 1=max/100% probability and 0=min/0% probability. The probabilities corresponded to the likelihood of a tile belonging to a typical, atypical non-neoplastic, and atypical neoplastic histology via multiclass classification. At the time of inference, we obtained confidence-weighted prediction scores by multiplying the confidence prediction by each of the three probability values separately. Confidence-weighted scores of all tiles in a

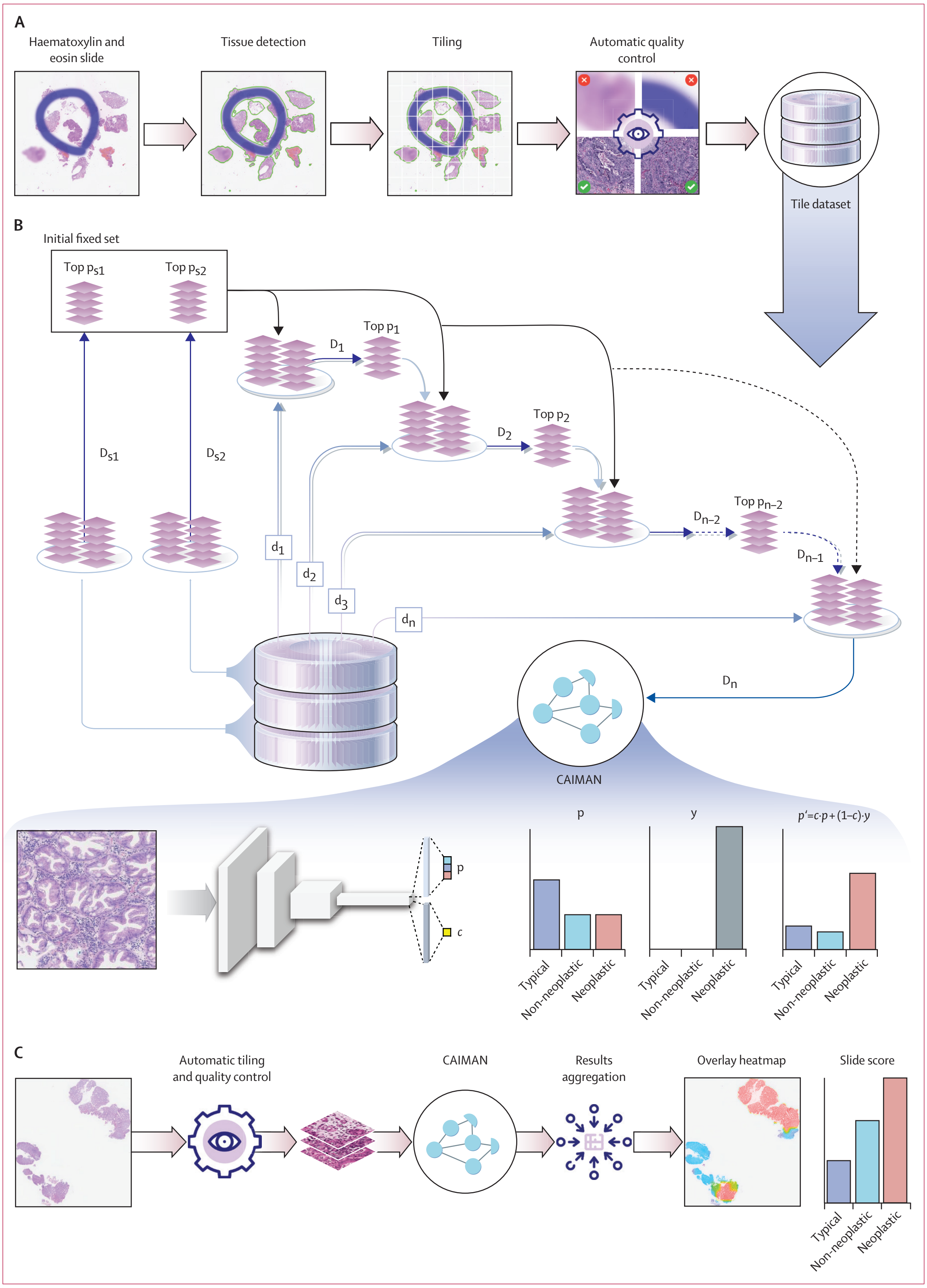

***Figure 1:* AI-based colon-biopsy prescreening with CAIMAN**
(A) Preprocessing of whole-slide images. (B) CAIMAN for training data per iteration. (C) CAIMAN inference. AI=artificial intelligence. *c*=the confidence of a tile on its prediction produced by CAIMAN. CAIMAN=Colorectal AI Model for Abnormality Detection. D_i=CAIMAN deep-learning model at the ith training iteration. d_i=draw subset of image tiles. D_{s1}=CAIMAN deep-learning model at the first training iteration. p=the probability of a tile belonging to diagnostic categories produced by CAIMAN. *p′*=confidence-weighted probability. Top p_1=top 50% of tiles at each training iteration. Top p_{s1}=first half of the training set. Top p_{s2}=remaining half of the training set.

WSI were considered for average aggregation into a single score, which represented the inference scores of the WSI being typical, atypical non-neoplastic, and neoplastic. We then mapped the score of each tile to colours to generate overlay heatmaps; blue showed the likelihood of a tile being typical, green showed the likelihood of a tile being non-neoplastic, and red showed the likelihood of a tile being neoplastic. For the generation of saliency maps we used the internal development cohort.

For **PyTorch** see https://pytorch.org/

We used PyTorch for the implementation of our CAIMAN deep learning model. A set of data augmentations—including random rotation with angles at 0°, 90°, 180°, and 270°; random horizontal and vertical flip transformations; colour jitter with brightness, contrast, saturation of 0·3, and hue of 0·05; randomly adjusted sharpness with factor 2; and random auto contrast—were applied on-the-fly on all training tiles. The training sets in all batches were carefully curated to ensure that no images were repeated. All tests were conducted on an Nvidia DGX-2 Deep Learning System with 16 × 32GB Tesla V100 Volta graphics processing units (GPUs) in a shared environment. The deep-learning model was built on two parallel GPUs with ten worker threads, with each GPU having a dedicated random-access memory of 32GB. The pseudocode of the CAIMAN algorithm from a theoretical point of view is provided in appendix 1 (pp 6–7).

The main outcome of this study was the ability of CAIMAN to accurately identify typical and atypical WSIs of colon biopsies, which could potentially facilitate automatic removing of typical biopsies from the diagnostic workload in clinics. The secondary outcome was the assignment of diagnostic labels (ie, neoplastic or inflammatory) to atypical slides.

Statistical analysis

We conducted three-fold internal cross-validation with case-controlled stratification for the performance evaluation of CAIMAN using the internal development cohort. In this evaluation protocol, the dataset was divided into three subsets on the basis of case identifiers (ie, unique identifiers for each patient) to ensure that all WSIs from one patient were in the same fold. For each fold of the cross-validation, two subsets were used for training, with WSIs in the training set randomly split into training and validation sets, whereas the third was an unseen internal test set. The model with the best performance, in terms of area under the convex hull of the receiver operating characteristic curve (AUROC) from the validation sets, was chosen as the model with the best trained performance to obtain predictions from the remaining unseen test set.

For the performance evaluation, we used AUROC, the average precision of the area under the precision-recall curve (AUPRC), and specificity values at three chosen thresholds of sensitivity. The means of all evaluation metrics were found for multiple folds of each experiment. We reported the mean and SD of AUROC and AUPRC. Furthermore, we reported mean specificity values at sensitivity values of 0·95, 0·97, and 0·99 as points of reference to evaluate the effectiveness and robustness of the CAIMAN tool in a real-world clinical setting. We compared CAIMAN results with results obtained from IDaRS,[23] with the same training and testing splits for the internal development and external validation cohorts.

All performance metrics were obtained from WSI scores. To generate the slide-level label, CAIMAN used a so-called average aggregation scheme, which found the mean of tile-level prediction scores in a slide for both non-neoplastic and neoplastic classes. We evaluated the classification performance of CAIMAN for non-neoplastic and neoplastic classes separately and for the atypical class by adding non-neoplastic and neoplastic scores of each slide into a combined atypicality score. When a slide contained both inflammatory and neoplastic regions, CAIMAN calculated both atypicality scores and the scores were combined to provide a final atypicality score. This approach was consistent with our aim of identifying and removing typical biopsies, even in the presence of mixed atypical features.

Statistical significance for the difference of model prediction scores on anatomical sites of the biopsy, which would show that the prediction performance of the model was not biased towards any anatomical site, was assessed via Mann-Whitney U test, as was the comparison between CAIMAN and IDaRS. p>0·05 was used to define a difference that was not statistically significant.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, writing of the report, or decision to submit for publication.

Results

Participant characteristics for the internal development cohort are provided in appendix 1 (p 3). Demographic data were not collected for the external cohorts due to their ethics approvals.

The results of CAIMAN assessment and IDaRS for cross-validation and independent evaluation on the internal development cohort and external validation cohorts are shown in figure 2. Our analysis showed no statistically significant differences across prediction scores from CAIMAN for typical and atypical classes based on anatomical sites of the biopsy (appendix 1 p 13). CAIMAN produced specificities of 0·8382 (95% CI 0·8316–0·8448) against a sensitivity value of 0·99, 0·9682 (0·9669–0·9695) against a sensitivity value of 0·97, and 0·9832 (0·9819–0·9846) against a sensitivity value of 0·95 for typical versus neoplastic slides; specificities of 0·4793 (0·4715–0·4871) against a sensitivity value of 0·99, 0·7114 (0·7068–0·7161) against a sensitivity value of 0·97, and 0·8003 (0·7959–0·8047) against a sensitivity value of 0·95 for typical versus non-neoplastic slides; and specificities of 0·5592 (0·5544–0·5641) against a

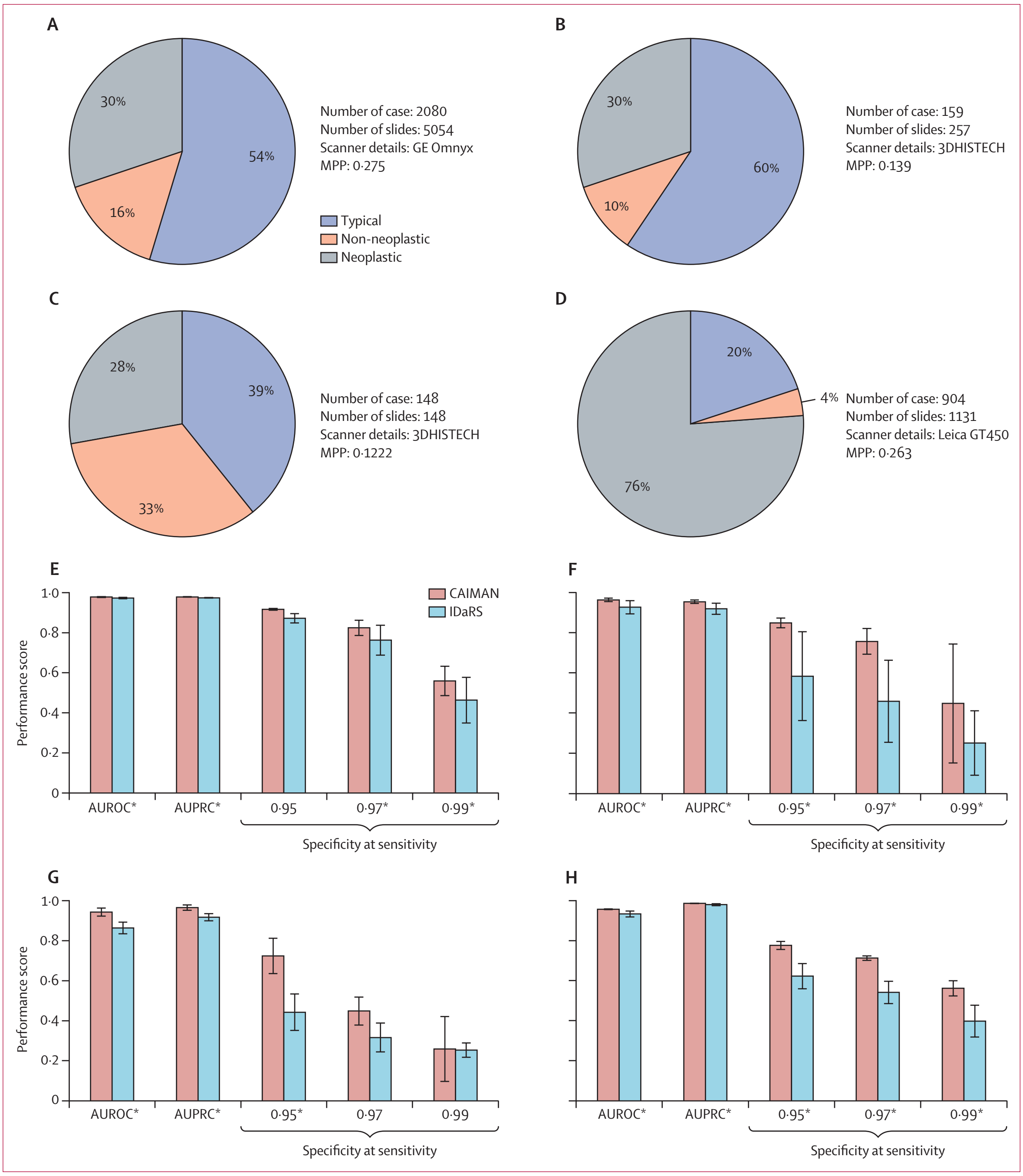

***Figure 2:* Patient cohorts, internal cross-validation and external validation results of CAIMAN and IDaRS**

(A) University Hospitals Coventry and Warwickshire NHS Foundation Trust cohort. (B) South Warwickshire NHS Foundation Trust cohort. (C) East Suffolk and North Essex NHS Foundation Trust cohort. (D) IMP Diagnostics cohort. (E) University Hospitals Coventry and Warwickshire NHS Foundation Trust cohort (internal cross-validation). (F) South Warwickshire NHS Foundation Trust cohort (independent external validation). (G) East Suffolk and North Essex NHS Foundation Trust cohort (independent external validation). (H) IMP Diagnostics cohort (independent external validation). Error bars show SD. AI=artificial intelligence. AUROC=area under the convex hull of the receiver operating characteristic curve. AUPRC=average precision of the area under the precision-recall curve. CAIMAN=Colorectal AI Model for Abnormality Detection. IDaRS=iterative draw and rank sampling. MPP=microns per pixels. NHS=National Health Service. *Statistically significant.

sensitivity value of 0·99, 0·8246 (0·8209–0·8283) against a sensitivity value of 0·97, and 0·9175 (0·9169–0·9182) against a sensitivity value of 0·95 for typical versus atypical slides.

However, at 0·99 sensitivity, CAIMAN (specificity 0·5592) was more accurate than an IDaRS-based weakly supervised WSI-classification pipeline (0·4629) in identifying typical and atypical biopsies on

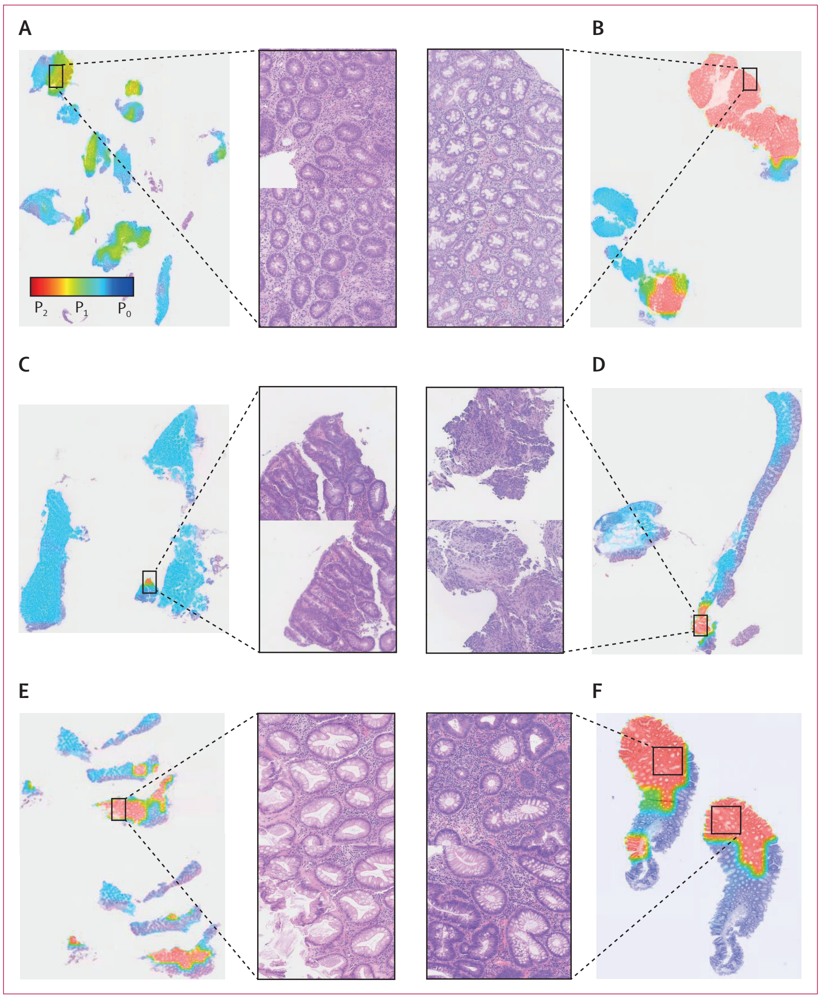

Figure 3: **CAIMAN scores at the tile level overlaid on original WSIs with corresponding strongly predictive tiles**
(A) CAIMAN scores of a non-neoplastic slide at the tile level, overlaid on original WSIs. (B) CAIMAN scores of a neoplastic slide at the tile level, overlaid on original WSIs. (C) CAIMAN scores of a neoplastic, small-fragment, low-grade dysplasia slide at the tile level, overlaid on original WSIs. (D) CAIMAN scores of a neoplastic hyperchromasia slide at the tile level, overlaid on original WSIs. (E) Diagnostic category error of the neoplastic slide. (F) Diagnostic category error of the neoplastic slide. Overlay heatmaps of slides labelled with diagnostic errors but that were found to be predicted correctly by CAIMAN are shown. AI=artificial intelligence. CAIMAN=Colorectal AI Model for Abnormality Detection. P_0=normal. P_1=non-neoplastic. P_2=neoplastic. WSI=whole-slide image.

cross-validation in the internal development cohort (p<0·0001; figure 2E; appendix p 13). At 0·99 sensitivity, CAIMAN was also more accurate than IDaRS for two external validation cohorts (both p<0·0001), but not for a third external validation cohort (p=0·10; figure 2F–H; appendix p 13).

CAIMAN provided higher specificity than IDaRS at some high-sensitivity thresholds for typical versus atypical slides (0·7763 *vs* 0·6222 for 0·95 sensitivity, 0·7126 *vs* 0·5407 for 0·97 sensitivity, and 0·5615 *vs* 0·3970 for 0·99 sensitivity for the IMP Diagnostics cohort; 0·8487 *vs* 0·5833 for 0·95 sensitivity, 0·7566 *vs* 0·4583 for 0·97 sensitivity, and 0·4474 *vs* 0·2500 for 0·99 sensitivity for the South Warwickshire NHS Foundation Trust cohort; 0·7241 *vs* 0·4425 for 0·95 sensitivity, 0·4483 *vs* 0·3161 for 0·97 sensitivity, and 0·2586 *vs* 0·2529 for 0·99 sensitivity for the East Suffolk and North Essex NHS Foundation Trust cohort; and 0·9175 *vs* 0·8721 for 0·95 sensitivity, 0·8246 *vs* 0·7628 for 0·97 sensitivity, and 0·5592 *vs* 0·4629 for 0·99 sensitivity for the internal development cohort). CAIMAN showed high classification performance in distinguishing between neoplastic biopsies (AUROC 0·9928, 95% CI 0·9927–0·9929), inflammatory biopsies (0·9658, 0·9655–0·9661), and atypical biopsies (0·9789, 0·9786–0·9792). On the three external validation cohorts, CAIMAN had AUROC values of 0·9431 (95% CI 0·9165–0·9697), 0·9576 (0·9568–0·9584), and 0·9636 (0·9615–0·9657) for the detection of atypical biopsies.

Analysis of errors in identifying atypical slides at a sensitivity value of 0·99 showed that CAIMAN did not correctly identify 22 (1%) of 2294 atypical slides, 6 (>1%) of 1500 neoplastic slides, or 16 (2%) of 794 non-neoplastic slides. During the analysis of these slides, we found one (5%) cancerous slide of 22 atypical neoplastic slides with specific diagnosis of atypical cells, nuclear pleomorphism, hyperchromasia; the atypical region was detected by AI, but the mean atypicality score was lower because only two tiles were atypical. Another slide (5%) included one possible tubular adenoma and one hyperplastic polyp. In the clinical reports, five slides of hyperplastic polyp and low-grade dysplasia were also labelled as typical, which indicates that the reference standard might be noisy (ie, inaccurately diagnosed). After further analysis by three pathologists (KG, YWT, and DS), three (50%) of six slides that were originally reported as neoplastic and 13 (81%) of 16 slides that were originally reported as non-neoplastic were found to be typical slides on re-review. We also examined typical slides that were predicted with high neoplastic or non-neoplastic scores and found that six (75%) of eight slides were misreported as typical. Prediction errors in non-neoplastic slides included a single focus of cryptitis and a single crypt abscess (eg, mild focal cryptitis, focal granuloma, or focus of active inflammation). In the typical class, we found some typical slides with quiescent ulcerative colitis and some slides with no substantial inflammation.

Algorithmic contributions of CAIMAN included a novel batch training loss function and modifications to the iterative draw and rank sampling process, which led to substantial improvements in the performance of CAIMAN, achieving a specificity of 56% compared with 46% in standard IDaRS at a sensitivity value of 0·99.[23] Algorithmic innovations are explained in appendix 1 (p 9). The performance of CAIMAN with another state-of-the-art, weakly supervised model on the internal development cohort, with multiple aggregation schemes, is also shown in appendix 1 (p 8).[15] The algorithm can also identify small and sparse features of atypicality (figure 3C, D).

Although quality control improved the specificity of the model, the frequency of image artifacts varied across different datasets. For example, the internal development cohort did not have any pen marks, whereas all external validation cohorts had pen marks.

An overlay heatmap of tiles in a non-neoplastic slide and strongly predictive non-neoplastic tiles is shown in figure 3. Overlay heatmaps of two neoplastic slides in which CAIMAN detected small neoplastic regions show one slide with a small fragment displaying low-grade dysplasia and another slide with a small fragment displaying hyperchromatic dysplasia (figure 3C, D). Overlay heatmaps of two slides that were predicted by CAIMAN to be atypical were suggested to be typical in clinical reports (figure 3E, F).

Saliency maps of strongly predictive atypical non-neoplastic tiles and atypical neoplastic tiles are shown in figure 4. Saliency maps highlight all pixels contributing to the prediction of atypical labels for a specific image tile. Image tiles corresponding to various atypicality sub-conditions (eg, hyperplastic polyps, poorly differentiated adenocarcinoma, dysplasia or adenocarcinoma, high-grade invasive carcinoma, active inflammation, chronic inflammation, collagenous colitis, ulceration, and inflammatory bowel disease) were predicted correctly by the CAIMAN model (figure 4B). Additional saliency maps for tile-level false positives and false negatives are in appendix 1 (pp 9–12).

Discussion

CAIMAN, a weakly supervised AI model for the prescreening of large bowel biopsies, can classify multiple diagnoses. Our method aimed to assist in histopathological examination by offering improved efficiency and reliability compared with current standards. By choosing a sensitivity value of 0·99, the workload of pathologists could be reduced by automated reporting of typical diagnoses, allowing more time for complex or atypical diagnoses. CAIMAN was more accurate than IDaRS at identifying typical and atypical biopsies in specificity values at all three high-sensitivity thresholds, which shows robustness and generalisability, for an internal development cohort and three external validation cohorts.

The CAIMAN model has been designed to be robust to staining variations by using colour augmentation during training, enabling the model to learn stain-invariant features. To evaluate the robustness of the algorithm to staining variations, we downsampled images from the external validation cohorts to match the resolution of the internal development cohort. Our results show that the algorithm could perform well for different cohorts, even when images are acquired with different scanners and have different MPP. We used a relatively large internal development cohort and an image-quality check before model training and evaluation to obtain diverse and representative information in terms of proportions of patients in each category (ie, typical, atypical non-neoplastic, and atypical neoplastic). Although quality control improved the specificity of the model, the frequency of image artifacts varied across different datasets. These findings highlight the importance of designing models that are robust to staining variations in digital pathology to improve their clinical use.

To implement an automated colon-biopsies screening system in clinical practice, Iizuka and colleagues[26]

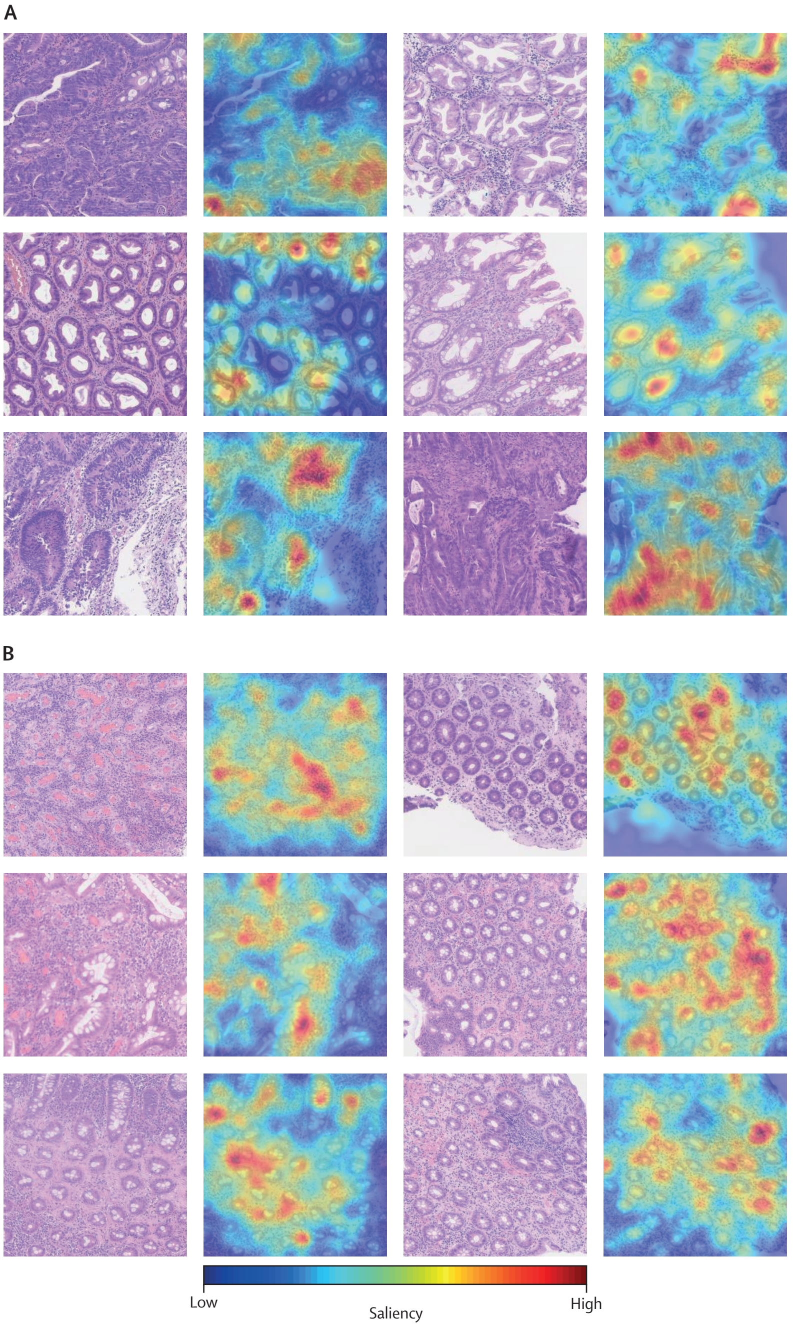

Figure 4: **Saliency maps of example image tiles that were accurately predicted by CAIMAN**
(A) Neoplastic biopsies. (B) Non-neoplastic biopsies. Regions contributing towards accurate predictions are highlighted. Gaussian smoothing has been applied on the saliency maps for improved visualisation for the purpose of the figure only. AI=artificial intelligence. CAIMAN=Colorectal AI Model for Abnormality Detection.

modelled the classification of epithelial tumours (ie, adenocarcinoma and adenoma) with a large development cohort of 4036 patients and two relatively small validation cohorts of 500 patients and 547 patients (ie, The Cancer Genome Atlas—Colon Adenocarcinoma cohort). Ho and colleagues[27] used gland segmentation and classification to model the categorisation of so-called low risk (eg, benign or inflammation) and high risk (eg, dysplasia or malignancy) slides with a small cohort of 294 patients. However, these models cannot identify diagnosis of precancerous lesions or non-neoplastic atypicalities and do not have high sensitivity. CAIMAN can differentiate between neoplastic, non-neoplastic, and typical large-bowel biopsies and identify hyperplastic polyps and dysplastic lesions. These abilities could provide a reliable and efficient prescreening stage of the regular clinical workflow and reduce the workload of pathologists.

CAIMAN showed high classification performance in distinguishing between neoplastic biopsies, inflammatory biopsies, and atypical biopsies, which was more accurate than reported accuracy figures of supervised and semi-supervised AI tools for colorectal-cancer diagnosis of neoplastic atypicalities proposed in 2021, considering multiple aspects.[9,28] For example, Wang and colleagues[9] proposed a fully supervised AI tool for cancer prediction versus typical prediction, whereas Yu and colleagues[28] proposed a semi-supervised AI tool that matches the performance of supervised AI when differentiating between cancer slides and typical slides. These two methods required both slide-level and laborious tile-level annotations to achieve a sensitivity of 0·982, a non-significant improvement compared with the mean performance of six pathologists (sensitivity of 0·975) on the same test set.[9] However, an AI pre-screening tool that yields high specificity but has reduced sensitivity is not desirable as an increase number of prescreening errors could be accepted and an increase number of neoplastic slides could be reported as typical.

CAIMAN improves sensitivity up to a value of 0·99 but has reduced specificity. Furthermore, it identifies non-neoplastic, neoplastic, and inflammatory atypicalities (appendix 1 pp 4–6). CAIMAN does not require any regional-level or cell-level annotations for model development, thereby saving the time and financial costs of laborious manual annotations. CAIMAN is also able to identify errors in the slide-level ground truth labels. The algorithm can also identify small and sparse features of atypicality, which might be overlooked under the microscope due to the vast amount of tissue content that has to be analysed manually.

One limitation of this study is a class imbalance across all cohorts (ie, a relatively high number of typical samples and a relatively low number of atypical inflammatory and neoplastic samples), which often occurs in medical datasets and might arguably overestimate AUROC values. To analyse the effect of the class imbalance, we reported an AUPRC measure as well. Rigorously reviewing all cohorts used for the development and evaluation of health-care-related AI before deployment is crucial.

A second limitation is that as the model was trained on a relatively small number of non-neoplastic slides with several subconditions, the classification results of CAIMAN were worse on non-neoplastic slides. The reduction in generalisation performance in the external validation cohorts might indicate the effects of differences in staining and scanners and of the varying composition of each cohort (ie, number of samples per class), which are different to the internal development cohort. However, the internal development cohort percentages of samples per class somewhat reflect the number of patients visiting UK hospitals.

Other limitations include the small sizes of the external validation cohorts and the diagnostic labels of our training cohort originating from a single laboratory, which might reduce clinical heterogeneity. However, our results for external validation cohorts show the generalisability of the proposed tool via the performance metrics.

The overall classification and generalisation performance on external validation cohorts can be enhanced if the amount of training data is increased with a multicentric cohort from which the model can learn from and through model-domain adaptation from data from multiple different scanners and centres. The diagnostic repertoire of the algorithm is a limitation (appendix 1 p 3), but we believe that continued improvement in training data and methods will enhance its performance on non-neoplastic diagnoses and other rare diagnoses before its consideration for clinical implementation. CAIMAN's training strategy allows for efficient retraining of the model on a new cohort, by contrast to models that require detailed cellular and regional annotations that are laborious and time-consuming. We also envisage that, after careful revision of the ground truth labels, the sensitivity and the specificity of the model could be further improved (having 100% correct ground truth labels are likely to improve the performance of CAIMAN). In future research, training and validation with large and diverse multicentric cohorts should be done. This could lead to improved model performance and generalisability before clinical deployment.

Although CAIMAN has shown promising results, a weakly supervised model with sensitivity greater than 99% and a higher level of specificity than that of this study is achievable with training on larger, multicentric cohorts and by ensuring that slide-level diagnostic labels are correct and consistent across the multiple centres. Such an outcome is likely to assist with digital pre-screening of colorectal biopsies for multiple atypicalities in a clinical setting. Despite little digital pathology in many pathology departments worldwide,[29,30] our results support the potential benefits of companion AI in improving diagnostic accuracy and reducing workload in histopathology. We will continue to explore solutions for implementation challenges and anticipate further advancements in this area of research.

A digital prescreening tool, such as CAIMAN, can increase efficiency in time and cost, increase reliability, and increase the accuracy of regular diagnostic screening by identifying and removing typical large-bowel biopsies from the workload of pathologists and by highlighting the neoplastic and non-neoplastic regions on the slide. We believe that AI-based prescreening for large-bowel biopsies is a current need, connecting modern digital technologies and human expertise to provide a timely diagnosis, decrease cancer fatalities, and reduce under-treatment or overtreatment.

Contributors

MB, DS, and NR designed the study with support from all co-authors. MB and NR developed the methods. MB wrote the code and conducted all tests. YWT, KG, MA, EH, AR, AA, HE, KD, HS, MN, KH, KB, SW, and DS provided diagnostic annotations of colonic biopsy slides. YWT, MB, FM, KG, MA, DS, and NR analysed and interpreted the results. MJ and MB made the data visualisations. FM, SEAR, AB, SG, MJ, NW, and WL provided technical and material support. MB, YWT, FM, DS, and NR drafted the manuscript. DS and NR supervised the study. MB, SG, FM, and NR directly accessed and verified the underlying data. All authors had full access to all the data in the study as and when needed and had final responsibility for the decision to submit for publication.

Declaration of interests

NR receives research funding from AstraZeneca and GlaxoSmithKline; is a GlaxoSmithKline Chair of Computational Pathology at the University of Warwick; and is a co-founder, director, and chief scientific officer of Histofy. SG is a co-founder of Histofy. DS is a director and a shareholder of Histofy; CAIMAN is in the process of being licensed to Histofy. FM receives research funding from GlaxoSmithKline. SW is a co-founder of C-Prio. All other authors declare no competing interests.

Data sharing

Original whole-slide images from the University Hospitals Coventry and Warwickshire National Health Service (NHS) Trust, the East Suffolk and North Essex NHS Foundation Trust, and the South Warwickshire NHS Foundation Trust are being added to the Pathology Image Data Lake for Analytics, Knowledge and Education (PathLAKE) data repository (https://www.pathlake.org/pathlake-data/) and will be made available upon request to the corresponding author. Information on obtaining data from the IMP Diagnostics cohort is provided in the original publication.[22] Scores generated by our algorithm are provided in appendix 2. Source code can be made available, subject to intellectual property constraints, by contacting Mohsin Bilal (mohsinbil@gmail.com) or Nasir Rajpoot (n.m.rajpoot@warwick.ac.uk).

See **Online** for appendix 2

Acknowledgments

This study was supported by the Pathology Image Data Lake for Analytics, Knowledge and Education (PathLAKE) Centre of Excellence for digital pathology and artificial intelligence, which is funded by the Data to Early Diagnosis and Precision Medicine area of investment and support of the UK Government's Industrial Strategy Challenge Fund and is managed and delivered by Innovate UK on behalf of UK Research and Innovation (file ref 104 689; application number 18 181). The views expressed in this Article are those of the authors and not necessarily those of the PathLAKE Consortium members, the National Health Service, Innovate UK, or UK Research and Innovation. Scott A Sanders and Naresh Chachlani assisted in providing slides of cases from South Warwickshire National Health Service Foundation Trust. Sean James assisted in scanning slides for the external cohorts. We thank Johnathan Breddy, Katie Martin, Salinder Tandi, Young Park, Giorgos Hadjigeorghiou, John Euinton, Chien-Hui Liao, and Thomas Leech for their assistance with administrative matters and for information-technology support for software and systems that were used in this study.

References

1 National Institute for Health and Care Excellence. Colorectal cancer. 2020. https://www.nice.org.uk/guidance/ng151 (accessed Dec 9, 2022).

2 American Cancer Society. Key statistics for colorectal cancer. 2023. https://www.cancer.org/cancer/types/colon-rectal-cancer/about/key-statistics.html (accessed Sept 1, 2023).

3 Bowel Cancer UK. Bowel cancer. 2023. https://www.bowelcanceruk.org.uk/about-bowel-cancer/bowel-cancer/ (accessed Sept 1, 2023).

4 Shmerling RH. Understanding the results of your colonoscopy. 2020. https://www.health.harvard.edu/staying-healthy/understanding-the-results-of-your-colonoscopy (accessed Aug 9, 2021).

5 Cancer.Net. Colorectal cancer: diagnosis. 2022. https://www.cancer.net/cancer-types/colorectal-cancer/diagnosis (accessed Aug 9, 2021).

6 Bainbridge S, Cake R, Meredith M, Furness P, Gordon B. Testing times to come? An evaluation of pathology capacity across the UK. 2016. https://www.cancerresearchuk.org/sites/default/files/testing_times_to_come_nov_16_cruk.pdf (accessed Sept 3, 2022).

7 Browning L, Colling R, Rakha E, et al. Digital pathology and artificial intelligence will be key to supporting clinical and academic cellular pathology through COVID-19 and future crises: the PathLAKE consortium perspective. *J Clin Pathol* 2021; **74:** 443–47.

8 Talbot IC, Price AB, Salto-Tellez M. Biopsy pathology in colorectal disease, 2nd edition. London: Hodder Arnold, 2006.

9 Wang KS, Yu G, Xu C, et al. Accurate diagnosis of colorectal cancer based on histopathology images using artificial intelligence. *BMC Med* 2021; **19:** 76.

10 Song Z, Yu C, Zou S, et al. Automatic deep learning-based colorectal adenoma detection system and its similarities with pathologists. *BMJ Open* 2020; **10:** e036423.

11 Zhou C, Jin Y, Chen Y, et al. Histopathology classification and localization of colorectal cancer using global labels by weakly supervised deep learning. *Comput Med Imaging Graph* 2021; **88:** 101861.

12 Yoshida H, Yamashita Y, Shimazu T, et al. Automated histological classification of whole slide images of colorectal biopsy specimens. *Oncotarget* 2017; **8:** 90719–29.

13 Moore M, Feakins RM, Lauwers GY. Non-neoplastic colorectal disease biopsies: evaluation and differential diagnosis. *J Clin Pathol* 2020; **73:** 783–92.

14 Lang-Schwarz C, Agaimy A, Atreya R, et al. Maximizing the diagnostic information from biopsies in chronic inflammatory bowel diseases: recommendations from the Erlangen International Consensus Conference on Inflammatory Bowel Diseases and presentation of the IBD-DCA score as a proposal for a new index for histologic activity assessment in ulcerative colitis and Crohn's disease. *Virchows Arch* 2021; **478:** 581–94.

15 Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng* 2021; **5:** 555–70.

16 Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 2019; **25:** 1301–09.

17 Kather JN, Pearson AT, Halama N, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med* 2019; **25:** 1054–56.

18 Bulten W, Kartasalo K, Chen PHC, et al. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. *Nat Med* 2022; **28:** 154–63.

19 Echle A, Grabsch HI, Quirke P, et al. Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. *Gastroenterology* 2020; **159:** 1406–16.

20 Schrammen PL, Ghaffari Laleh N, Echle A, et al. Weakly supervised annotation-free cancer detection and prediction of genotype in routine histopathology. *J Pathol* 2022; **256:** 50–60.

21 Bilal M, Nimir M, Snead D, Taylor GS, Rajpoot N. Role of AI and digital pathology for colorectal immuno-oncology. *Br J Cancer* 2022; **128:** 3–11.

22 Oliveira SP, Neto PC, Fraga J, et al. CAD systems for colorectal cancer from WSI are still not ready for clinical acceptance. *Sci Rep* 2021; **11:** 14358.

23 Bilal M, Raza SEA, Azam A, et al. Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study. *Lancet Digit Health* 2021; **3:** e763–72.

24 He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016. https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf (accessed Sept 1, 2023).

25 DeVries T, Taylor GW. Learning confidence for out-of-distribution detection in neural networks. *aRxiv* 2018; published online Feb 13. https://doi.org/10.48550/arXiv.1802.04865 (preprint).

26 Iizuka O, Kanavati F, Kato K, Rambeau M, Arihiro K, Tsuneki M. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Sci Rep* 2020; **10:** 1504.

27 Ho C, Zhao Z, Chen XF, et al. A promising deep learning-assistive algorithm for histopathological screening of colorectal cancer. *Sci Rep* 2022; **12:** 2222.

28 Yu G, Sun K, Xu C, et al. Accurate recognition of colorectal cancer with semi-supervised deep learning on pathological images. *Nat Commun* 2021; **12:** 6311.

29 Jahn SW, Plass M, Moinfar F. Digital pathology: advantages, limitations and emerging perspectives. *J Clin Med* 2020; **18:** 3697.

30 Koelzer VH, Grobholz R, Zlobec I, Janowczyk A, Swiss Digital Pathology Consortium. Update on the current opinion, status and future development of digital pathology in Switzerland in light of COVID-19. *J Clin Pathol* 2021; **75:** 687–89.