scientific reports

OPEN

A promising deep learning-assistive algorithm for histopathological screening of colorectal cancer

Cowan Ho[1,6], Zitong Zhao[2,6], Xiu Fen Chen[2,6], Jan Sauer[3], Sahil Ajit Saraf[3], Rajasa Jialdasani[3], Kaveh Taghipour[3], Aneesh Sathe[3], Li-Yan Khor[2,4], Kiat-Hon Lim[2,4] & Wei-Qiang Leow[2,4,5✉]

Colorectal cancer is one of the most common cancers worldwide, accounting for an annual estimated 1.8 million incident cases. With the increasing number of colonoscopies being performed, colorectal biopsies make up a large proportion of any histopathology laboratory workload. We trained and validated a unique artificial intelligence (AI) deep learning model as an assistive tool to screen for colonic malignancies in colorectal specimens, in order to improve cancer detection and classification; enabling busy pathologists to focus on higher order decision-making tasks. The study cohort consists of Whole Slide Images (WSI) obtained from 294 colorectal specimens. Qritive's unique composite algorithm comprises both a deep learning model based on a Faster Region Based Convolutional Neural Network (Faster-RCNN) architecture for instance segmentation with a ResNet-101 feature extraction backbone that provides glandular segmentation, and a classical machine learning classifier. The initial training used pathologists' annotations on a cohort of 66,191 image tiles extracted from 39 WSIs. A subsequent application of a classical machine learning-based slide classifier sorted the WSIs into 'low risk' (benign, inflammation) and 'high risk' (dysplasia, malignancy) categories. We further trained the composite AI-model's performance on a larger cohort of 105 resections WSIs and then validated our findings on a cohort of 150 biopsies WSIs against the classifications of two independently blinded pathologists. We evaluated the area under the receiver-operator characteristic curve (AUC) and other performance metrics. The AI model achieved an AUC of 0.917 in the validation cohort, with excellent sensitivity (97.4%) in detection of high risk features of dysplasia and malignancy. We demonstrate an unique composite AI-model incorporating both a glandular segmentation deep learning model and a classical machine learning classifier, with excellent sensitivity in picking up high risk colorectal features. As such, AI plays a role as a potential screening tool in assisting busy pathologists by outlining the dysplastic and malignant glands.

Artificial intelligence (AI) is rapidly revolutionizing the field of pathology[1,2]. Founded on the specialty practice of interpreting expressed histomorphological changes in cellular or tissue structure caused by disease processes, pathology has maintained its clinical utility till today[3]. The objective evaluation of histological slides by highly-trained pathologists remains the gold standard for cancer diagnosis[4]. With ever increasing workloads on pathologists, this time-consuming and manpower intensive work has recently seen the advent of computational pathology[5] largely enabled by whole slide images (WSIs) which are digital counterparts of traditional glass slides and which have received selected FDA approval for primary clinical diagnosis[5,6]. Through application of medical image analysis, machine learning and deep convolutional neural networks (CNN), artificial intelligence have been used to inspect WSIs and produce computer-aided diagnosis (CAD) of cancers[1,5,7–9]. These CADs have demonstrated non-inferiority in the identification of malignancy compared to traditional means[8,10–12].

While human pathologists can outperform such AI systems, they are subject to fatigue, time-constraints and observer bias in clinical settings. As such CNN has the additional benefit of unimpaired accuracy only subject to the operational capacity of its processing hardware. With ever-increasing workloads, the integration of artificial intelligence into the field of computational pathology is a growing necessity[8]. With an annual estimated 1.8 million cancer cases and 900,000 cancer deaths globally it places significant strain on healthcare

[1]Yong Loo Lin School of Medicine, National University Singapore, Singapore, Singapore. [2]Department of Anatomical Pathology, Singapore General Hospital, 20 College Road, Singapore 169856, Singapore. [3]Qritive Pte. Ltd., Singapore, Singapore. [4]Duke-NUS Medical School, Singapore, Singapore. [5]School of Biological Sciences, Nanyang Technological University, Singapore, Singapore. [6]These authors contributed equally: Cowan Ho, Zitong Zhao and Xiu Fen Chen. ✉email: leow.wei.qiang@singhealth.com.sg

systems[13,14]. The number needed to diagnose (NND) is further increased as some studies showed that as few as 7.4% of colonoscopy biopsies had any positive findings at all[15]. This means that an AI, which is capable of reliably screening the colonoscopy biopsies, could considerably reduce the workload of a practicing pathologist[16,17]. Additionally, it is worth noting that the assistive effect of AI on pathology is not limited to the evaluation of slides but extends to the acquisition of the target tissue of interest during sampling in a clinical setting such as narrow-band imaging[18–24]. If combined with in-vivo endoscopic assessments, AI can effectively revolutionize and streamline current diagnostic workflows[1,25].

However, the development of CNN for colonic histopathology lags behind that for breast and prostate tissue[1,5,8]. We attempted to address these issues with our own CNN model by replacing outdated heatmaps and saliency maps with segmentation as the output[11,26,27]. Unlike other studies that typically rely purely on established datasets or open source segmentation architecture, we independently designed our own model which was trained and validated on separate training and validation sets[1,6,28].

This successful application of a highly functional CNN to colonic biopsy WSI demonstrates the capability of AI in detecting epithelial tumours from a non-neoplastic background. This further shows that AI is very relevant in the specific field of colonic histopathology[5]. Expanding on these research developments, we describe our experience in this single-centre pilot study.

Materials and methods

Our AI algorithm consisted of two distinct models. The first is a gland segmentation model that identifies potentially high risk regions on WSIs. The second involves a slide classification model that classifies WSIs as either 'high risk' or 'low risk'. Using two separate models ensures robustness of the results and gives the operating pathologist more insight into the reasoning behind the slide classification. This allows for easier resolution of incongruent diagnoses and closer surveillance of the output data at this early stage of validation.

To produce these WSIs, our laboratory processed colonic specimens with haematoxylin and eosin (H&E) and scanned them with the Philips digital pathology whole-slide scanner, which satisfies current best practice recommendations[29].

Inclusion and exclusion criteria. We included samples from colon biopsies (May to June 2019) and resections (June 2019 to October 2019) obtained from Singapore General Hospital's pathology archives and The Cancer Genome Atlas (TCGA). All slides from our institution were anonymized and de-identified H&E slides. We excluded all slides that did not contain mucosa, were poor in image quality (e.g., image being blurred, tissue being folded) or contained malignancies other than colonic adenocarcinoma. We also excluded slides suggestive of mucinous adenocarcinoma and signet ring carcinoma as these entities were poorly represented in our training data. This ensured proper robustness of our AI model in achieving our objective of detecting classic adenocarcinoma histology.

Slide classifier. The gland segmentation model is a deep neural network that is built according to a Faster Region Based Convolutional Neural Network (Faster-RCNN) architecture for instance segmentation with a ResNet-101 feature extraction backbone. This is a standard architecture for segmentation tasks and many state-of-the-art instance segmentation models utilize this architecture or a variant thereof. The cost of this high accuracy complex model is a longer runtime. This is deemed as an acceptable trade-off for this device due to the relatively high importance of accuracy in the context of cancer detection. The slide classification model is a gradient-boosted decision tree classifier that uses the outputs of the gland segmentation model to classify slides as either high or low risk.

The training data for our gland segmentation model was extracted from WSIs. Ten WSIs were of colon resections obtained from TCGA, while 21 WSIs were of colon resections and 8 WSIs were of colon biopsies obtained from Singapore General Hospital.

These slides underwent ground truth annotation by an expert pathologist studying the dataset. Slides were classified as either "high risk" or "low risk". The WSIs were further annotated with labels from the following 7 categories: (1) benign glands, (2) glands that are either characteristic for adenocarcinoma or high-grade dysplasia, (3) low-grade dysplasia, (4) blood vessels, (5) necrosis, (6) mucin or (7) inflammation (Fig. 1).

Training algorithm and protocols. To train our gland segmentation model, heuristic algorithms were used to identify tissue regions on WSIs. These were then divided into tiles of 775×522 pixels each. Tiles that did not contain any tissue (as determined by a tissue masking algorithm) were discarded. Likewise, ground truth annotations were separated into corresponding tiles so that each pixel was clearly labeled as either background or one of the categories described above. This resulted in a total of 73,546 image and ground truth tiles. Of these, 66,191 tiles (corresponding to 90% of all tiles) were used to train the neural network itself. We used the standard training paradigm for neural networks of backpropagation with stochastic gradient descent to adjust the parameters of the network. This is an iterative training procedure that gradually adjusts parameters to minimize a loss function. Five percent of all tiles (3678 tiles) were used for model selection. The neural network was trained with a constant learning rate of 0.016 over 15 epochs. We also performed data augmentation to artificially increase the training set and make the resulting model robust against predictable variations in data, such as imaging artifacts or variations in staining intensity. We have detailed the augmentations in Table 1. At regular intervals throughout the training, the model was evaluated on this development set. At the end of the training, the model parameters which resulted in the best performance on this development set were selected as the final model parameters. These yielded the final 3677 tiles (corresponding to 5% of all tiles) which were used to prevent overfitting.

| High Risk Tumor Tissue | Low Risk Tumor Tissue | Dysplasia | Inflammation | Blood Vessel |
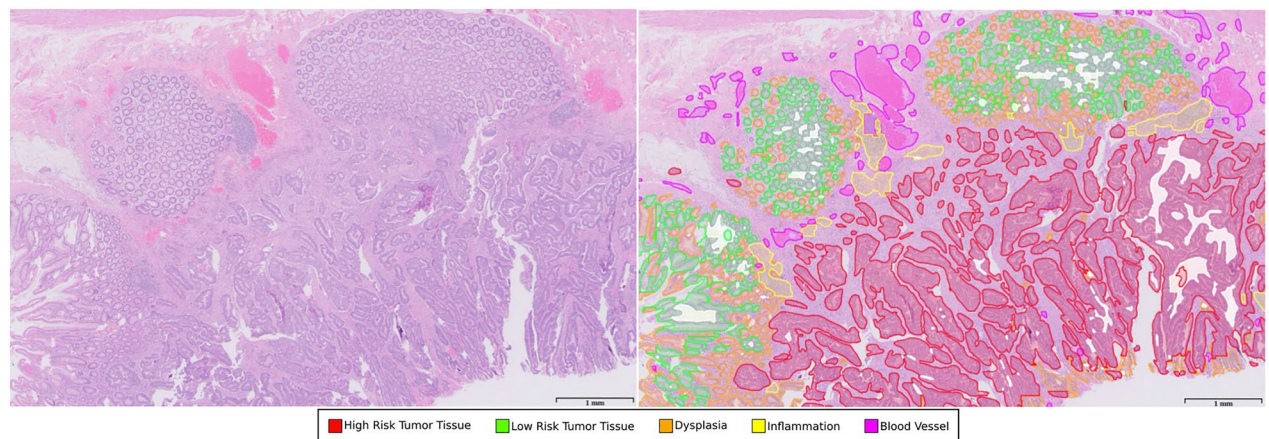
Figure 1. A sample analysis on a whole slide image (WSI). Left: original image. Right: the segmentation model highlighted regions of the WSI as (1) likely benign or normal (green), (2) likely dysplastic (orange), and (3) likely malignant (red). The AI model also segmented blood vessels (pink) and inflammation (yellow), and these segmentations were taken into account for slide labeling.

| Augmentation method	Details
Rotation	Images were randomly rotated by 0°, 90°, 180°, or 270° with equal probability
Mirroring	Images were randomly mirrored along either the horizontal axis, the vertical axis, or neither axis with equal probability
Contrast	Increases or decreases the contrast of an image by shifting pixel values either away from or towards the mean intensity of the image, respectively, by a uniformly sampled random multiplicative factor between 0.7 and 1.3
Brightness	Increases or decreases the brightness of an image by increasing or decreasing the pixel values, respectively, by a uniformly sampled random multiplicative factor between 0.7 and 1.3
H&E color augmentation	The image is transformed from the RGB color space into the H&E color space so the first image channel now corresponds to the intensity of the hematoxylin and the second to the eosin stains. These two channels are now modulated by adding or subtracting uniformly sampled random numbers between -0.05 and 0.05 to each channel. The image is then converted back to the RGB color space
	This augmentation is meant to simulate the effects of variability in the staining intensity
Gaussian noise	Every pixel is additively modulated by a normally distributed number (sigma = 0.1). This is meant to simulate various digital image artifacts that can occur during the scanning process
Gaussian blur	A Gaussian blur effect is applied to the entire image (radius = 0.1) to simulate microscope images that may be slightly out of focus

Table 1. Data augmentations applied during training.

The slide classification model was trained with data comprising 105 WSIs of colon resection slides from Singapore General Hospital. As with the segmentation model, resection slides were intentionally chosen to train our model to ensure that the final device was able to generalize on new datasets and no bias was introduced into the feature selection at this stage. As tissue on colon resection slides are typically much larger and have different composition compared to colon biopsy slides, a model that performs equally well on both types of slides can be assumed to generalize well over new data.

The ground truth annotations used in the training dataset for the slide classification model was created by two pathologists analyzing the WSIs. Slides were classified as either "high risk" (contains sign of adenocarcinoma or dysplasia) or "low risk" (contains normal histology/inflammation/reactive changes without signs of adenocarcinoma or dysplasia). Disagreements between pathologists were solved through discussion. Upon agreement, this final diagnosis was utilized as the ground truth (Fig. 2).

Validation protocols. The fully trained segmentation model was applied to all 105 WSIs. The segmentation results were aggregated to obtain numeric features for each slide. Features were selected from an ensemble of potential features based on their relative impact on the slide classification. These were: (1) the total area classified as either adenocarcinoma or dysplasia by the segmentation model with a prediction certainty of greater than 70% (this area is expressed as a percentage of the total tissue), (2) the average prediction certainty for adenocarcinoma or dysplasia objects weighted by their areas, (3) an additional Boolean flag that was set to 1 if the slide had adenocarcinoma or dysplasia predictions with a prediction certainty of greater than 85% that cover area of at least 0.1 mm^2, and (4) the bottom 1-percentile of adenocarcinoma or dysplasia predictions weighted by their areas. Once these numeric features were obtained, fivefold cross validation was applied and an ensemble of models was trained, tuned, and compared. The model and hyperparameter combination with the optimal cross
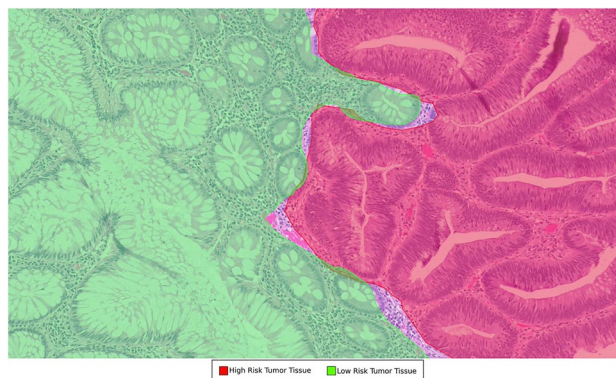
Figure 2. Labelling of "high-risk" (red) versus "low-risk" (green) regions based on the ground truth. A tile is labeled high-risk if there is overlap with any amount of high-risk ground truth annotations. Otherwise, the tile is labeled as low-risk. In this example, the highlighted tile would be labeled high-risk. The same rule was also applied to AI annotations. If a tile contains any amount of high-risk annotation given by AI, the tile is labeled as high-risk by AI, and low-risk otherwise.

validation performance, a gradient boosting classifier, was selected. Inputs to these models were the features described in the previous step and the output was a binary classification into high risk or low risk. The highest performing model with greatest accuracy was selected to be trained and finalized on the entire dataset consisting of 105 WSIs.

Declarations. We confirm that all methods were carried out in accordance with relevant guidelines and regulations. All experimental protocols were approved by the SingHealth Centralized Institutional Review Board. Informed consent was obtained from all subjects and/or their legal guardian(s) prior to having their data used in this study.

Results

The composite AI model, consisting of the gland segmentation and slide classification, was fully trained on the 105 resection WSIs and validated on the separate dataset of 150 biopsy WSIs. The biopsy WSIs were classified into high risk and low risk, similar to the resection slides used for training the slide classifier. After obtaining the output labels produced by the AI algorithm from the validation set, we compared them to the ground truth labels produced by our expert pathologists. We calculated the number of true positives, true negatives, false positives, and false negatives. Using these statistics, we then calculated several performance metrics, such as the sensitivity, specificity, and area under the receiver-operator characteristic curve (AUC).

Our AI model classified 119 of the 150 biopsies correctly. There were 31 errors consisting of 2 false negatives and 29 false positives. With these, the AI model achieved high sensitivity of 97.4% and lower specificity of of 60.3 with an AUC of 91.7 (Fig. 3). These performance metrics and trade-off between sensitivity and specificity is demonstrated in Fig. 4. By changing the prediction threshold to values above 0.7, we could obtain both sensitivity and specificity values above 80%. This prediction threshold refers to the probability cut-off in which the AI classifies a slide as either high risk or low risk. We deliberately set this threshold of 0.7 to suit the operation needs of our institution. Such a threshold allows the model to achieve greater sensitivity in detection of malignancy and functioning as a triage system. The prediction threshold can be adjusted easily to suit other operational requirements as the user deems appropriate.

Discussion

Upon evaluation of the output data from the validation set, its high AUC of 91.7 demonstrates a good concordance between the AI model and the expert pathologist. With a sensitivity of 97.4, it validates our AI model to function as a screening tool that minimizes false negatives extensively. This favour of sensitivity over specificity shows its usability in assistive workflow and has added practicality into the application of a clinical workflow. Since our AI model is designed solely as an assistive tool, the final diagnosis during reporting remains with the pathologist. This high sensitivity ensures with greater certainty that all lesions suspicious for malignancy are highlighted for further pathological review. This triage system is part of a diagnostic workflow we recommend the AI model be implemented in. We also propose four other possible ways in which AI model could be incorporated in pathology workflow; it could act as first reader, second reader, triage, and pre-screening (Fig. 5)[10,30,31].

Although multiple studies have reported the role of AI in colorectal polyp detection and characterization, our study is unique in the usage of segmentation as the output. The prevailing trends in the literature are based mainly on heat maps or saliency maps[9,26,32]. We believe that segmentation offers significant advantages over heatmaps. In segmentation, WSIs are split apart into small patches which are analysed by an AI algorithm. These segmentation-based annotations are individually more detailed, have more intuitive presentation of results to pathologists and enable pathologists to derive more insightful quantitative information from the WSIs segment[11,27]. There is also

Biopsies		Ground Truth	
		High Risk	Low Risk
Prediction	High Risk	TP: 75	FP: 29
	Low Risk	FN: 2	TN: 44

| Statistic	Biopsies
Sensitivity	97.4
Specificity	60.3
Positive Likelihood Ratio	2.5
Negative Likelihood Ratio	0.04
Accuracy	79.3
AUC	91.7
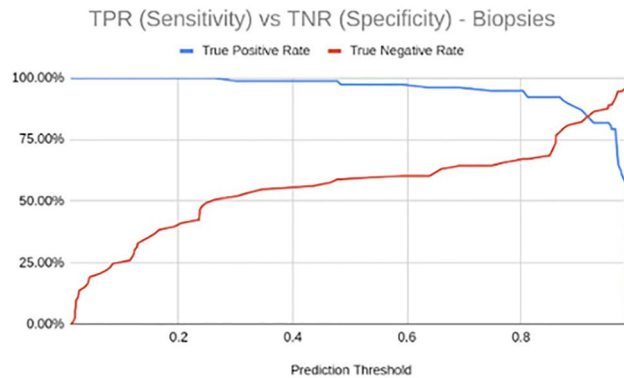
Figure 3. Performance data from applying the AI model on the validation set of 150 WSIs.
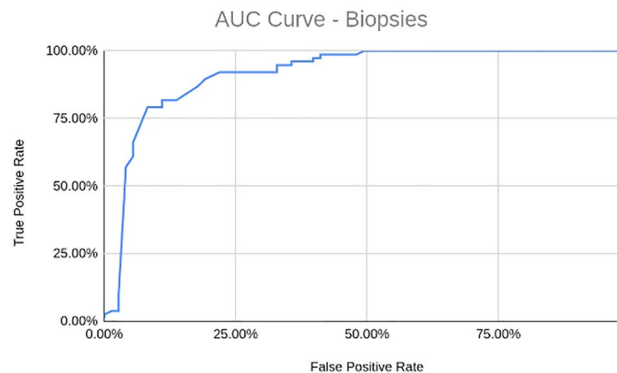
Figure 4. AUC curve from applying the AI model on the validation set of 150 WSIs. The system achieves an AUC of 91.7%.

more explicability behind prediction made by the system. Segmentation also allows for calculation of tumor area and higher potential for biomarker discovery while allowing pre-annotation of whole-slide images using AI[33]. Finally, it is also more versatile and is able to segregate more elements on the slide, which means we can view multiple parameters on a single image[1]. For example, in addition to being able to segment the colonic tissue into three main categories: likely benign or normal (green), likely dysplastic (orange), and likely malignant (red), our AI system also segments additional structures such as blood vessels (pink) and inflammation (yellow) (Fig. 1). Additional parameters can be easily programmed into our algorithm, allowing for more detailed information.

By having separate training and validation sets, we have avoided overfitting and reduced bias in our study. We have also compared our F1 scores over existing segmentation algorithms with similar convolutional neural network architecture (Table 2). This composite metric is based on precision and recall sameness which directly correlates to the product of sensitivity and specificity in the AI model[25,27]. The vast majority of current AI studies utilize fully-supervised learning based on established datasets which are available as open-source data[1,25,27]. There are few similar studies that have presented an AI algorithms based on an architecture consisting of two independently trained models, though an attempt has been made to separately segment glands from the background to identify gland edges[1,7]. There has also been the development of a neural network architecture that is similar to but less efficient than the RCNN architecture which is more commonly used in instance segmentation
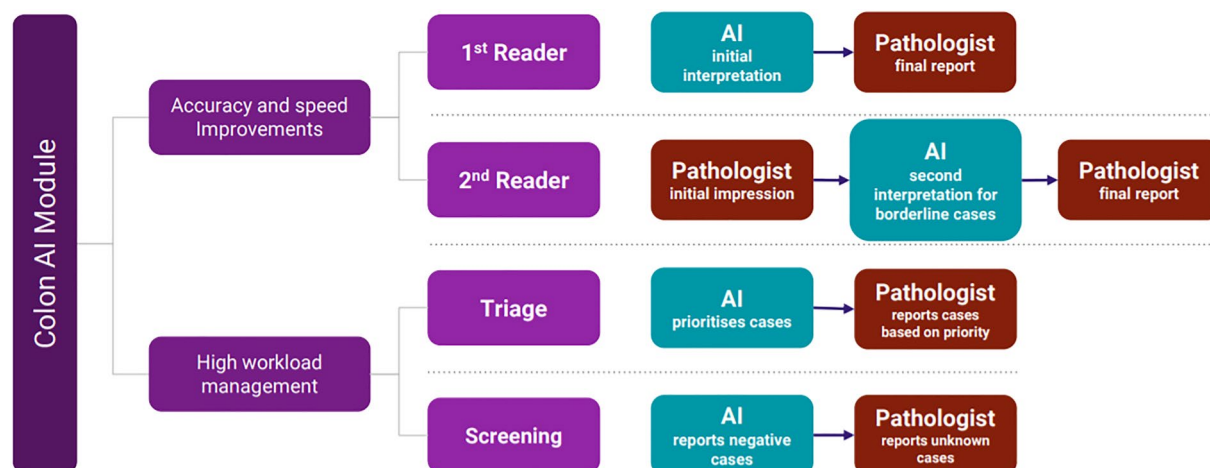
Figure 5. Proposed workflow for the use of AI in pathology.

Study	Objectives	Dataset	F1 score
Xu et al. (2017)	Segmentation of colon glands	GLAS challenge (165 images)	0.893–0.843
Xu et al. (2016)	Nuclei segmentation	537 images from Case Western Reserve University	0.858–0.771
Korbar et al. (2017)	Deep Neural Network Visualization to Interpret WSI Analysis Outcomes for Colorectal Polyps	176 WSIs from Dartmouth-Hitchcock Medical Center	0.925–0.841
MIMO—Net[35]	Various studies	Various studies	0.913–0.724
DeepLab v3+[36]	Various studies	Various studies	0.862–0.764
SegNet[37]	Various studies	Various studies	0.858–0.783
FCN—8[37]	Various studies	Various studies	0.783–0.692
Qritive Colon AI (current study)	Glandular segmentation deep learning model to detect high risk colorectal polyps	WSIs produced from 294 colorectal specimens from Singapore General Hospital	0.974–0.856

Table 2. Comparison of studies with similar convolutional neural network architecture. *GLAS* Gland Segmentation in Colon Histology Images Challenge Contest, *WSI* whole slide images.

tasks[34]. This architecture consisted of three arms—one to separate glands from the background, one to identify gland edges, and a third one to perform instance bounding box detection.

Our study was not without limitations. Our small sample size is susceptible to heterogenicity of data and that could hence diminish it's applicability to clinical practice in screening a general population. Further studies on a larger dataset are underway. Additionally, while the performance data from the validation had a high AUC of 91.7, it still contained 2 false negatives but 29 false positives. This was largely attributable to the manipulation of the prediction threshold values. In a clinical institution with AI model being applied to patient care in the context of cancer diagnosis, the system was designed to favour sensitivity over specificity to ensure usability in assistive workflow. This allows the AI model to categorize all high grade dysplasia and adenocarcinoma as 'high risk'. This carry-over effect, when applied onto the validation set, resulted in higher false positives as the prediction certainty of the various features were forced into a binary classification, leading to images with relatively low prediction certainty of 70% being highlighted as 'high risk'.

The issue of false positives can be addressed by further training and validation of the composite AI model. It is estimated that at least 10,000 WSIs are required to train a weakly supervised AI model for histopathology even before accounting for the lack of the heterogeneity encountered in WSIs with regard to staining, anomalies, tissue textural variation and polymorphism[12,33,38]. Our dataset will need further expansion to a much larger sample size and conduct a large-scale multi-site clinical validation to improve the quality of our AI model to reduce misclassification rate and achieve better results. Ideally, we intend to eventually remove the machine learning classifier and transition entirely into a gland segmentation model. The machine classifier was added to avoid overfitting as we expand on our current dataset. Currently, two separate models provide more insight into the reasoning behind the slide classification. This allows for easier deconflicting of incongruent diagnoses and closer surveillance of the output data as we further refine our AI algorithm.

Additionally, current segmentation output models utilise the most frequent tile classification output as the resulting predicted slide label and aggregated to produce a slide-level output but this provides limited contextual knowledge with a narrow field of view, while grossly increasing computational complexity[39–43]. Such patch-based models are not consistent with the manner pathologists analyse slides under microscopes and fail to take into consideration the characteristics of the surrounding structures and the overall morphology of the WSI[1]. While approaches to alleviate this issue such as designing multi-magnification networks exist, they have largely been

limited[44]. Some authors have proposed a variant of a fully conventional network (FCN) which consists of dense scanning, anchor layers, or combining FCNs with CNNs[41,42].

Furthermore, more work can be done in regard to pre-processing of slides for increased standardisation, including, but not limited to, stain normalization, augmentation and stain transfer[1]. While our laboratory's process of H&E staining and scanning with the Philips digital pathology whole-slide scanner satisfies current best practice recommendations, it is worth noting that earlier works in digital pathology had previously assumed staining attenuated optical density and applied uniformity assumptions[29]. However, this has fallen out of favour with chemical staining and tissue morphology now being considered in the generation of stain transfer matrices[45–48].

Conclusion

In summary, we demonstrated that our unique composite AI model incorporating a glandular segmentation deep learning model and a machine learning classifier has promising ability in picking up high risk colorectal features. The high sensitivity highlighted its role as a potential screening tool to create initial interpretations and assist pathologists to streamline the workflow, thereby effectively reducing the diagnostic burden on pathologists. Ongoing calibration and training of our composite AI model will improve its accuracy in risk classification of colorectal specimens.

References

1. Srinidhi, C. L., Ciga, O. & Martel, A. L. Deep neural network models for computational histopathology: A survey. *Med Image Anal.* **67**, 101813. https://doi.org/10.1016/j.media.2020.101813 (2021).
2. Pinckaers, H. & Litjens, G. Neural ordinary differential equations for semantic segmentation of individual colon glands. arXiv: 1910.10470 (2019) **(NeurIPS)**.
3. Pathology MeSH Descriptor Data 2021. D010336.
4. Kather, J. N. *et al.* Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS Med.* **16**(1), 1–22. https://doi.org/10.1371/journal.pmed.1002730 (2019).
5. Iizuka, O. *et al.* Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Sci. Rep.* **10**(1), 1–11. https://doi.org/10.1038/s41598-020-58467-9 (2020).
6. Mukhopadhyay, S. *et al.* Whole slide imaging versus microscopy for primary diagnosis in surgical pathology. *Am. J. Surg. Pathol.* **42**, 39–52 (2018).
7. Kainz, P., Pfeiffer, M. & Urschler, M. Segmentation and classification of colon glands with deep convolutional neural networks and total variation regularization. *PeerJ* **2017**(10), 1–28. https://doi.org/10.7717/peerj.3874 (2017).
8. Malik, J. *et al.* Colorectal cancer diagnosis from histology images: A comparative study. arXiv. Published online March 26, 2019:1–12. Accessed 28 February 2021. arXiv:1903.11210.
9. Lui, T. K. L., Guo, C. G. & Leung, W. K. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: A systematic review and meta-analysis. *Gastrointest. Endosc.* **92**(1), 11-22.e6. https://doi.org/10.1016/j.gie.2020.02.033 (2020).
10. Dembrower, K. *et al.* Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: A retrospective simulation study. *Lancet Digit. Health* **2**(9), e468–e474. https://doi.org/10.1016/S2589-7500(20)30185-0 (2020).
11. Barisoni, L., Lafata, K. J., Hewitt, S. M., Madabhushi, A. & Balis, U. G. J. Digital pathology and computational image analysis in nephropathology. *Nat. Rev. Nephrol.* **16**(11), 669–685. https://doi.org/10.1038/s41581-020-0321-6 (2020).
12. Campanella, G. *et al.* Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **25**(8), 1301–1309. https://doi.org/10.1038/s41591-019-0508-1 (2019).
13. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**(6), 394–424. https://doi.org/10.3322/caac.21492 (2018).
14. Safiri, S. *et al.* The global, regional, and national burden of colorectal cancer and its attributable risk factors in 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol. Hepatol.* **4**(12), 913–933. https://doi.org/10.1016/S2468-1253(19)30345-0 (2019).
15. Morarasu, S., Haroon, M., Morarasu, B. C., Lal, K. & Eguare, E. Colon biopsies: Benefit or burden?. *J. Med. Life* **12**(2), 156–159. https://doi.org/10.25122/jml-2019-0009 (2019).
16. Rakha, E. A. *et al.* Current and future applications of artificial intelligence in pathology: A clinical perspective. *J. Clin. Pathol.* **74**(figure 1), 1–6. https://doi.org/10.1136/jclinpath-2020-206908 (2020).
17. Cui, M. & Zhang, D. Y. Artificial intelligence and computational pathology. *Lab. Investig.* https://doi.org/10.1038/s41374-020-00514-0 (2021).
18. Sharma, S. & Seth, U. Artificial intelligence in colonscopy. *J. Pract. Cardiovasc. Sci.* **3**(3), 158. https://doi.org/10.4103/jpcs.jpcs_2_18 (2017).
19. Luo, Y. *et al.* Artificial intelligence-assisted colonoscopy for detection of colon polyps: A prospective, randomized cohort study. *J. Gastrointest. Surg.* https://doi.org/10.1007/s11605-020-04802-4 (2020).
20. Attardo, S. *et al.* Artificial intelligence technologies for the detection of colorectal lesions: The future is now. *World J. Gastroenterol.* **26**(37), 5606–5616. https://doi.org/10.3748/wjg.v26.i37.5606 (2020).
21. Hassan, C. *et al.* Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: A systematic review and meta-analysis. *Gastrointest. Endosc.* **93**(1), 77–85.e6. https://doi.org/10.1016/j.gie.2020.06.059 (2021).
22. Attardo, S. *et al.* World Journal of Gastroenterology Artificial intelligence technologies for the detection of colorectal lesions: The future is now. *World J. Gastroenterol.* **26**(37), 5606–5616. https://doi.org/10.3748/wjg.v26.i37.5606 (2020).
23. Hassan, C., Pickhardt, P. J. & Rex, D. K. A resect and discard strategy would improve cost-effectiveness of colorectal cancer screening. *Clin. Gastroenterol. Hepatol.* **8**(10), 865–869.e3. https://doi.org/10.1016/j.cgh.2010.05.018 (2010).
24. East, J. E. *et al.* Advanced endoscopic imaging: European Society of Gastrointestinal Endoscopy (ESGE) Technology Review. *Endoscopy* **48**(11), 1029–1045. https://doi.org/10.1055/s-0042-118087 (2016).
25. Zhao, Z. Q., Zheng, P., Xu, S. T. & Wu, X. Object detection with deep learning: A review. *IEEE Trans. Neural Netw. Learn Syst.* **30**(11), 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865 (2019).
26. Huang, Y. & Chung, A. C. S. Evidence localization for pathology images using weakly supervised learning. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) LNCS*, Vol. 11764, 613–621. https://doi.org/10.1007/978-3-030-32239-7_68 (2019).

27. Madabhushi, A. & Lee, G. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med. Image Anal.* **33**, 170–175. https://doi.org/10.1016/j.media.2016.06.037 (2016).

28. Arvaniti, E. *et al.* Automated Gleason grading of prostate cancer tissue microarrays via deep learning. *Sci. Rep.* **8**(1), 12054. https://doi.org/10.1038/s41598-018-30535-1 (2018).

29. Cross, S., Furness, P., Igali, L., Snead, D. & Treanor, D. Best practice recommendations for implementing digital pathology. *R. Coll. Pathol.* Published online 3–5. https://www.rcpath.org/uploads/assets/f465d1b3-797b-4297-b7fedc00b4d77e51/Best-practice-recommendations-for-implementing-digital-pathology.pdf (2018).

30. Steiner, D. F. *et al.* Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. *Am. J. Surg. Pathol.* **42**(12), 1636–1646. https://doi.org/10.1097/PAS.0000000000001151 (2018).

31. Stenzinger, A. *et al.* Artificial intelligence and pathology: From principles to practice and future applications in histomorphology and molecular profiling. *Semin. Cancer Biol.* https://doi.org/10.1016/j.semcancer.2021.02.011 (2021).

32. Wang, S. *et al.* Comprehensive analysis of lung cancer pathology images to discover tumor shape and boundary features that predict survival outcome. *Sci. Rep.* **8**(1), 1–9. https://doi.org/10.1038/s41598-018-27707-4 (2018).

33. Gurcan, M. N. *et al.* Histopathological image analysis: A review. *IEEE Rev. Biomed. Eng.* **2**, 147–171. https://doi.org/10.1109/RBME.2009.2034865 (2009).

34. Xu, Y. *et al.* Gland instance segmentation using deep multichannel neural networks. *IEEE Trans. Biomed. Eng.* **64**(12), 2901–2912. https://doi.org/10.1109/TBME.2017.2686418 (2017).

35. Raza, S. E. A. *et al.* MIMONet: Gland segmentation using multi-input–multi-output convolutional neural network. In *Communications in Computer and Information Science*, Vol. 723, 698–706 (Springer, 2017). https://doi.org/10.1007/978-3-319-60964-5_61.

36. Xiao, W. T., Chang, L. J. & Liu, W. M. Semantic segmentation of colorectal polyps with DeepLab and LSTM networks. In *2018 IEEE International Conference on Consumer Electronics-Taiwan, ICCE-TW 2018* (Institute of Electrical and Electronics Engineers Inc., 2018). https://doi.org/10.1109/ICCE-China.2018.8448568.

37. Tang, J., Li, J. & Xu, X. Segnet-based gland segmentation from colon cancer histology images. In *Proceedings—2018 33rd Youth Academic Annual Conference of Chinese Association of Automation, YAC 2018* 1078–1082 (Institute of Electrical and Electronics Engineers Inc., 2018). https://doi.org/10.1109/YAC.2018.8406531.

38. Louis, D. N. *et al.* The 2016 World Health Organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol.* **3**, 803–820. https://doi.org/10.1007/s00401-016-1545-1 (2016).

39. Korbar, B. *et al.* Looking under the hood: Deep neural network visualization to interpret whole-slide image analysis outcomes for colorectal polyps. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Work* 2017-July, 821–827. https://doi.org/10.1109/CVPRW.2017.114 (2017).

40. Lin, H. *et al.* Fast ScanNet: Fast and dense analysis of multi-gigapixel whole-slide images for cancer metastasis detection. *IEEE Trans. Med. Imaging* **38**(8), 1948–1958. https://doi.org/10.1109/TMI.2019.2891305 (2019).

41. Lin, H. *et al.* ScanNet: A fast and dense scanning framework for metastatic breast cancer detection from whole-slide image. In *Proceedings—2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018.* Vol. 2018-January, 539–546 (Institute of Electrical and Electronics Engineers Inc., 2018). https://doi.org/10.1109/WACV.2018.00065.

42. Guo, Z. *et al.* A fast and refined cancer regions segmentation framework in whole-slide breast pathological images. *Sci. Rep.* **9**(1), 1–10. https://doi.org/10.1038/s41598-018-37492-9 (2019).

43. Chen, P. J. *et al.* Accurate classification of diminutive colorectal polyps using computer-aided analysis. *Gastroenterology* **154**(3), 568–575. https://doi.org/10.1053/j.gastro.2017.10.010 (2018).

44. Ho, D. J. *et al.* Deep multi-magnification networks for multi-class breast cancer image segmentation. *Comput. Med. Imaging Graph.* **88**, 1–35. https://doi.org/10.1016/j.compmedimag.2021.101866 (2021).

45. Macenko, M. *et al.* A method for normalizing histology slides for quantitative analysis. In *Proceedings—2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009* 1107–1110. https://doi.org/10.1109/ISBI.2009.5193250 (2009).

46. Vahadane, A. *et al.* Structure-preserving color normalization and sparse stain separation for histological images. *IEEE Trans. Med. Imaging* **35**(8), 1962–1971. https://doi.org/10.1109/TMI.2016.2529665 (2016).

47. Bejnordi, B. E. *et al.* Stain specific standardization of whole-slide histopathological images. *IEEE Trans. Med. Imaging* **35**(2), 404–415. https://doi.org/10.1109/TMI.2015.2476509 (2016).

48. Cho, H., Lim, S., Choi, G. & Min, H. Neural stain-style transfer learning using GAN for histopathological images (2003) 1–10. arXiv:1710.08543 (2017).

Author contributions
C.H. wrote the main manuscript with the support from X.F.C., Z.Z. and S.A.S. J.S., S.A.S., R.J., K.T. and A.S. designed the AI model and analysed the data. L.-Y.K. and K.-H.L. supervised and finalised the project. W.-Q.L. initiated the research initiative and supervised the project. All authors discussed the results and contributed to the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to W.-Q.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.