Original Paper

Automating Colon Polyp Classification in Digital Pathology by Evaluation of a "Machine Learning as a Service" AI Model: Algorithm Development and Validation Study

David Beyer[1], BSc, MD; Evan Delancey[2], BSc, MSc; Logan McLeod[3], BSc, MSc

[1]Department of Lab Medicine and Pathology, University of Alberta, Edmonton, AB, Canada

[2]NGIS (Australia), Victoria, Australia

[3]Deptartment of Environmental Studies, University of Victoria, Victoria, BC, Canada

Corresponding Author:
David Beyer, BSc, MD
Department of Lab Medicine and Pathology
University of Alberta
5-411 Edmonton Clinic Health Academy
Edmonton, AB, T6G 1C9
Canada
Phone: 1 (780) 492 3111
Email: dbbeyer@ualberta.ca

Abstract

Background: Artificial intelligence (AI) models are increasingly being developed to improve the efficiency of pathological diagnoses. Rapid technological advancements are leading to more widespread availability of AI models that can be used by domain-specific experts (ie, pathologists and medical imaging professionals). This study presents an innovative AI model for the classification of colon polyps, developed using AutoML algorithms that are readily available from cloud-based machine learning platforms. Our aim was to explore if such AutoML algorithms could generate robust machine learning models that are directly applicable to the field of digital pathology.

Objective: The objective of this study was to evaluate the effectiveness of AutoML algorithms in generating robust machine learning models for the classification of colon polyps and to assess their potential applicability in digital pathology.

Methods: Whole-slide images from both public and institutional databases were used to develop a training set for 3 classifications of common entities found in colon polyps: hyperplastic polyps, tubular adenomas, and normal colon. The AI model was developed using an AutoML algorithm from Google's VertexAI platform. A test subset of the data was withheld to assess model accuracy, sensitivity, and specificity.

Results: The AI model displayed a high accuracy rate, identifying tubular adenoma and hyperplastic polyps with 100% success and normal colon with 97% success. Sensitivity and specificity error rates were very low.

Conclusions: This study demonstrates how accessible AutoML algorithms can readily be used in digital pathology to develop diagnostic AI models using whole-slide images. Such models could be used by pathologists to improve diagnostic efficiency.

Introduction

Many important pathological diagnoses are made by expert pathologists' careful examination of formalin-fixed paraffin-embedded tissue slides. Advances in digital microscopy have enabled large-scale digitization of glass slides at high resolution, and the adoption of whole-slide images (WSIs) for primary sign-out of pathology is increasing [1-3]. One of the main benefits of digital pathology is improved diagnostic efficiency, which is increasingly important as the field of

pathology deals with increased volumes while also struggling with the recruitment of new pathologists [4,5]. The combination of a decreased workforce coupled with an increase in the volume of cases secondary to an aging population means that pathologists need to become more efficient to meet future demand. Artificial intelligence (AI) tools applied to WSIs can be used to improve the efficiency of pathologists [6-8]. Advancements in slide scanners [9], which facilitate large-scale digitization of slides, have brought about a paradigm shift in pathology. Digitization not only enhances the efficiency of pathological examinations but also bridges the gap between conventional techniques and the ever-evolving field of AI. Ultimately, the creation of WSIs is the first step in incorporating AI into the field of pathology.

One advantage of the digitization of WSIs will be the creation of libraries of high-quality labeled training data for use with machine learning (ML) algorithms [10]. Recent developments in the fields of ML and AI, such as deep learning, for computer vision and object detection–related tasks [11-13] have led to a rapid uptake of the use of these tools in computational pathology research, where their utility has been widely recognized [14-17]. ML has traditionally required massive computational power and advanced knowledge of computer science and programming languages such as Python and R [18-20]. However, with the large-scale deployment of Machine Learning as a Service (MLaaS) platforms, these barriers to entry are minimized, allowing domain-specific experts (ie, pathologists and medical imaging professionals) to make use of advanced AI/ML tools [21]. "AutoML" algorithms and cloud-based ML platforms such as Amazon's Sagemaker and Google's VertexAI provide affordable, easy-to-access options that reduce overall costs by allocating centralized computer resources on demand to end users [22].

Pathologists are well-positioned with the expertise and tools required to build high-quality training datasets, which are the bedrock of effective AI models, and develop real-world uses for the production of ML models. Our project examined whether a small dataset for common colon polyp entities could be used to develop a robust and accurate ML model for diagnostic purposes using an AutoML model from Google's Vertex AI. Colon polyps are a precursor to invasive carcinoma, and a high-volume sample is encountered in the pathology lab. As many jurisdictions use screening programs to detect and remove polyps for cancer surveillance, accurate and efficient pathology diagnosis is a key part of colon cancer screening programs [23,24]. As there are relatively few diagnostic entities for colon polyps, this area is well-suited to AI screening algorithms to assist pathologists in making a rapid diagnosis. This led to the formation of our project, aiming at evaluating an MLaaS model, trained on our own institutional data, to evaluate both the ease of model development as well as model performance when compared to pathologist interpretation.

Methods

Overview

Alberta Precision Laboratories has a robust digital pathology slide set, used primarily for teaching. This slide set contains images of previous cases (hematoxylin and eosin–stained histology slides) that have been scanned using an Aperio GT450. The bulk of our case images came from this dataset. In addition, to increase variability within our dataset, both for hematoxylin and eosin stain quality, as well as scanner type, we also used publicly accessible WSIs from Leeds University, University of Michigan, and the Cancer Imaging Archive [25-27]. Cases (n=494) were randomly divided into 3 allocations: training (75%), validation (10%), and test (15%, Table 1).

Table 1. Data allocation.

Image allocation	Number of images used, n	Percent of total
Training	1110 (370 cases)	75%
Validation	150 (50 cases)	10%
Test	222 (74 cases)	15%

In Table 1, cases and their associated images were split into 75% training, 10% validation, and 15% test. Slightly increased test cases were used (that the traditional 80/10/10 split) to better evaluate model performance, and to compensate for a relatively small training set.

We focused on 3 common entities seen in colon polyps: hyperplastic polyps, tubular adenomas, and normal colon. All cases were opened in Aperio Imagescope, and manual image/patch extraction was completed at 4× and 10× objective power (40× and 100× magnification, respectively), focusing on the most representative tumor/diagnostic areas. A total of one 4× extraction/tile and two 10× extraction/tiles were used per case, for a total of 1482 images (Table 1). Tiles were chosen by a pathologist to ensure that the most representative areas of the slide were used. The extraction was carried out using Aperio-Imagescope's built-in image extraction tool, using the

embedded International Color Consortium (ICC) profile and exported in .jpeg format. The ICC profile is a method of color normalization used by Aperio (and other digital pathology vendors) to ensure the image generated from the slide matches as closely as possible to the real-world color profile of the slide. Both the scanner and image profile used in this study are fully validated and accredited for clinical use.

Images in the dataset were rescreened by a blinded pathologist to ensure that the appropriate diagnostic label had been applied and to ensure that no images from test cases were present in either the training or validation data. A total of 3 users (2 pathologists and 1 pathology resident) were involved in the tile and review process to prevent bias in the tile selection. Instructions for tile selection were to select the most appropriate tile for the diagnosis. Only tiles with perfect consensus (3 of 3 agreed) labels were chosen to be used.

A single-label image classification model with 3 labels ("hyperplastic polyp," "tubular adenoma," and "normal") was then developed using Vertex AI and an autoML algorithm. Other model parameters were set to default to demonstrate a general yet easy to use model for nontechnical experts (model details are as follows: training method: "AutoML," Objective: "Image classification (single-label)," Data-split: "Manual," Budget: 48 node hours, Actual: 43.04 node hours, Training time: 5 h 33 min).

The AI model was tested on the "test" allocation of cases that were not part of the training or validation datasets. XRAI overlay [28] was used to view explainability and ensure that the algorithm was identifying the pertinent areas of the image (Figure 1). Model output was evaluated on a per-image basis, as opposed to a per-case basis. Results from the test data were used to evaluate precision and recall. VertexAI incorporates model evaluation and automatically generates both the precision-recall curve as well as precision-recall by threshold. An overall confidence threshold of 0.50 was chosen for the evaluation of our model.

Figure 1. (A) Normal colon sample image. (B) Normal colon sample image with XRAI (4) overlay. (C) Hyperplastic polyp sample image. (D) hyperplastic polyp sample image with XRAI overlay. Green intensity correlates with areas of increased positive probability, that is, image segments that contribute most strongly to a given class prediction.
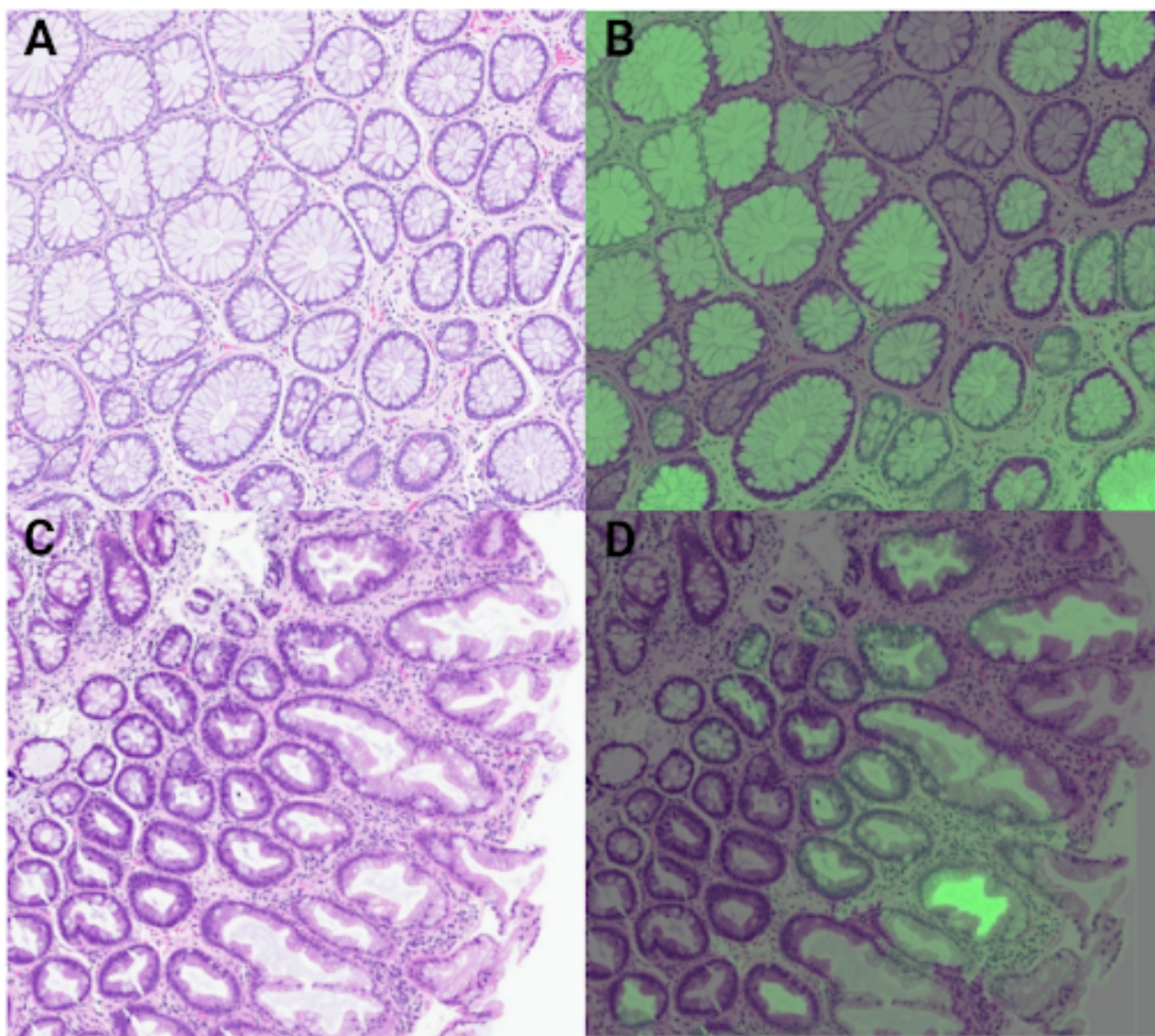
Ethical Considerations

Ethics approval and a waiver of consent were obtained for using deidentified case images from our institutional database. This study was approved by the Health Research Ethics Board of Alberta (HREBA.CC-23-0347).

Results

Using a CI of 0.5, the overall accuracy of the model on the test dataset was 98.4% (Table 2). Tubular adenomas and hyperplastic polyps were identified 100% of the time (66/66 and 48/48, respectively). Normal colon was identified with 97% accuracy (102/105, with 3/105 being misclassified as "hyperplastic"). Visual inspection of XRAI overlays demonstrates that the model

is identifying the pertinent areas (Figure 1). Results from the classification of the test data were used to calculate both recall as well as precision, in addition to a precision-recall by threshold curve (Figure 2), and an area under the curve value of 0.99.

Table 2. Tubular adenoma and hyperplastic polyps were identified 100% of the time (66/66 and 48/48, respectively). Normal colon was identified with 97% accuracy (102/105).

True label ↓/Predicted label →	Tubular adenoma, n (%)	Hyperplastic polyp, n (%)	Normal colon, n (%)
Tubular adenoma	66 (100)	0 (0)	0 (0)
Hyperplastic polyp	0 (0)	48 (100)	0 (0)
Normal colon	0 (0)	3 (3)	105 (97)

Figure 2. Precision-recall curve by threshold. Using a confidence value of 0.25 results in a precision of 98.7%, and recall of 100%. Using a confidence value of 0.5 results in both a precision and recall of 98.6%. Using a confidence value of 0.75 results in a precision of 100% and a recall of 97.3%. Precision sharply falls below a confidence threshold of 0.13. Recall sharply falls above a confidence threshold of 0.80. Our model output data are based on a confidence value of 0.50.
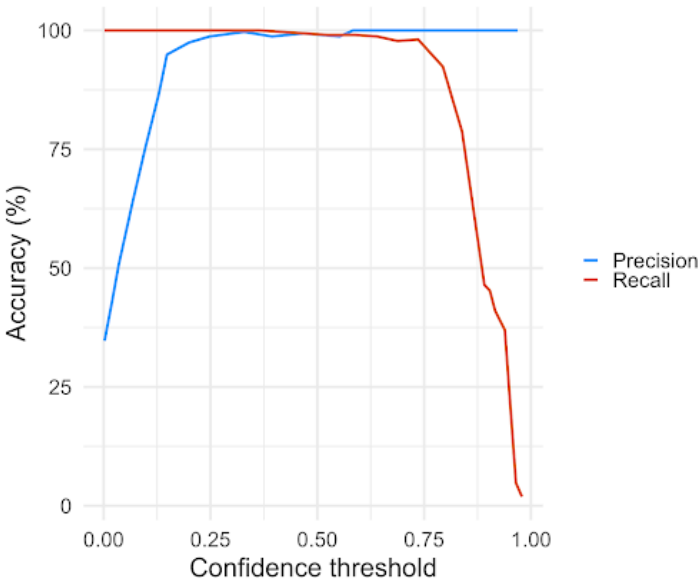
Discussion

Principal Findings

The integration of AI in the field of digital pathology is not merely a technological advancement; it is becoming a necessity driven by the contemporary challenges that the medical community faces. Here we have demonstrated that it is relatively easy to train your own AI algorithms on your own data, as a pathologist. As ML continues to revolutionize diagnostic pathology, our study demonstrates how simple off-the-shelf AI tools can readily be used to develop effective models for improving diagnostic efficiency. The recall and precision of the developed AI model in detecting colon polyps were remarkable, approaching 100%. This exceptional performance underscores the potential of ML to transform diagnostic pathology, especially given the rapid advancements in technology, availability, and cost-effectiveness of MLaaS platforms such as Google's VertexAI.

Previous AI applications in digital pathology, specifically in colorectal cancer screening, have shown great potential in increasing the efficiency and accuracy of diagnosis. For instance, Korbar et al [29] demonstrated that deep learning models could classify colorectal histology slides with high accuracy, bridging the gap between manual microscopic evaluations and automated assessments. Our study builds upon these foundations, offering further evidence for the efficiency and effectiveness of AI in this domain. One could say we have crafted a model that emulates the performance of an early-stage pathology resident, as the 3 entities used in this model are relatively simple; the difference is that this model took days to train, versus years for the average pathology resident. Nevertheless, there are certain limitations to consider.

First, our study had a limited sample size and did not include a complete range of diagnostic entities. With only tubular adenoma, hyperplastic polyp, and normal colon as labels, this model does not account for other critical entities like serrated adenomas, high-grade dysplasia, or carcinoma, an important histologic feature of high-risk polyps [30]; this would be a good area for future projects. AI models, particularly deep learning models, generally require large training datasets in order to generalize well in real-world scenarios [18,31]. A more extensive dataset might allow for more refined models that capture nuances and rare presentations of colon polyps, such as high-grade dysplasia. This limitation is significant, as evidenced by the emphasis on broad scope in successfully used deep neural networks for detecting colorectal cancer on WSIs [32].

Second, while our model shows promise, integrating it into real-world workflows would necessitate the inclusion of a more diverse range of clinical criteria beyond just the image data. We did not include certain clinical characteristics that would be important, like location (ie, ascending colon), which can affect diagnosis, especially for serrated lesions [23].

Third, the need to manually annotate our dataset means that we are effectively working with a "best-case scenario" data input. Real-world scenarios might present slides with artifacts, suboptimal staining, or overlapping tissues that could challenge the model's predictions [33]. The success of AI models in clinical settings largely depends on the quality and diversity of the input data. As we only used tiles with consensus between our 3 subject matter experts (2 pathologists and 1 pathology resident), this meant a small number of imperfect tiles were excluded. These "imperfect" samples, of course, exist in the real world, and these must be interpreted as well.

Fourth, our model misclassified a number of "normal" as "hyperplastic polyp," which appeared to be on normal colon where glands were not perfectly round. Likely inclusion of more "normal" images with different-shaped lumens would help tune the model for proper identification.

However, despite these limitations, our study exemplifies the promise of using ML as a service in histopathology. The near-perfect performance of our model in a controlled environment suggests that with adequate training data and a broader range of diagnostic entities, such platforms could soon play a significant role in augmenting the capabilities of pathologists.

Our project has shown that widely available "AutoML" algorithms, such as Google's Vertex AI, can be applicable to the medical field, specifically in digital pathology. Our training model was robust with a sensitivity approaching 100% and a specificity of 98%. This aligns with other authors' findings of success with AutoML algorithms in the medical field [31]. Such AI tools will help to improve the efficiency of front-line pathologists, allowing them to keep up with increasing demand and so helping to mitigate workforce constraints. For example, our model could be used to auto-generate a report, saving time for the pathologist in dictating/typing.

Conclusion

Overall, we show that cloud-based ML platforms can produce accurate models that are specific to the field of pathology and WSIs. Furthermore, they are easy to use, with only a relatively basic level of ML knowledge required. As the medical community strives for implementation of precision medicine that should lead to improved patient outcomes [34], integration of ML tools into the realm of histopathology may soon become indispensable. Future studies with larger datasets, diverse diagnostic entities, and more real-world scenarios are warranted to further elucidate the potential of AI in this domain.

Authors' Contributions

DB was involved in conceptualization, data curation, and writing of the initial draft. LM was involved in supervision, validation, and writing (review and editing). ED was involved in the validation and writing (review and editing).

Conflicts of Interest

None declared.

References

1. Louis DN, Feldman M, Carter AB, Dighe AS, Pfeifer JD, Bry L, et al. Computational pathology: a path ahead. Arch Pathol Lab Med. 2016;140(1):41-50. [FREE Full text] [doi: 10.5858/arpa.2015-0093-SA] [Medline: 26098131]
2. Zarella MD, Bowman D, Aeffner F, Farahani N, Xthona A, Absar SF, et al. A practical guide to whole slide imaging: a white paper from the digital pathology association. Arch Pathol Lab Med. 2019;143(2):222-234. [FREE Full text] [doi: 10.5858/arpa.2018-0343-RA] [Medline: 30307746]
3. Rizzo PC, Caputo A, Maddalena E, Caldonazzi N, Girolami I, Dei Tos AP, et al. Digital pathology world tour. Digit Health. 2023;9:20552076231194551. [FREE Full text] [doi: 10.1177/20552076231194551] [Medline: 37654717]
4. Ho J, Ahlers SM, Stratman C, Aridor O, Pantanowitz L, Fine JL, et al. Can digital pathology result in cost savings? A financial projection for digital pathology implementation at a large integrated health care organization. J Pathol Inform. 2014;5(1):33. [FREE Full text] [doi: 10.4103/2153-3539.139714] [Medline: 25250191]
5. Lujan G, Quigley JC, Hartman D, Parwani A, Roehmholdt B, Meter BV, et al. Dissecting the business case for adoption and implementation of digital pathology: a white paper from the digital pathology association. J Pathol Inform. 2021;12:17. [FREE Full text] [doi: 10.4103/jpi.jpi_67_20] [Medline: 34221633]
6. Hanna MG, Reuter VE, Ardon O, Kim D, Sirintrapun SJ, Schüffler PJ, et al. Validation of a digital pathology system including remote review during the COVID-19 pandemic. Mod Pathol. 2020;33(11):2115-2127. [FREE Full text] [doi: 10.1038/s41379-020-0601-5] [Medline: 32572154]
7. Schüffler PJ, Geneslaw L, Yarlagadda D, Hanna M, Samboy J, Stamelos E, et al. Integrated digital pathology at scale: a solution for clinical diagnostics and cancer research at a large academic medical center. J Am Med Inform Assoc. 2021;28(9):1874-1884. [FREE Full text] [doi: 10.1093/jamia/ocab085] [Medline: 34260720]
8. Marletta S, Eccher A, Martelli FM, Santonicco N, Girolami I, Scarpa A, et al. Artificial intelligence-based algorithms for the diagnosis of prostate cancer: a systematic review. Am J Clin Pathol. 2024;161(6):526-534. [doi: 10.1093/ajcp/aqad182] [Medline: 38381582]

9. Rizzo PC, Girolami I, Marletta S, Pantanowitz L, Antonini P, Brunelli M, et al. Technical and diagnostic issues in whole slide imaging published validation studies. Front Oncol. 2022;12:918580. [FREE Full text] [doi: 10.3389/fonc.2022.918580] [Medline: 35785212]

10. Hanna MG, Singh R, Parwani AV. Whole slide imaging: applications in education. In: Whole Slide Imaging. Cham. Springer; 2022:95-103.

11. Bankhead P. Developing image analysis methods for digital pathology. J Pathol. 2022;257(4):391-402. [FREE Full text] [doi: 10.1002/path.5921] [Medline: 35481680]

12. Ben Hamida A, Devanne M, Weber J, Truntzer C, Derangère V, Ghiringhelli F, et al. Deep learning for colon cancer histopathological images analysis. Comput Biol Med. 2021;136:104730. [doi: 10.1016/j.compbiomed.2021.104730] [Medline: 34375901]

13. Rakha EA, Toss M, Shiino S, Gamble P, Jaroensri R, Mermel CH, et al. Current and future applications of artificial intelligence in pathology: a clinical perspective. J Clin Pathol. 2021;74(7):409-414. [doi: 10.1136/jclinpath-2020-206908] [Medline: 32763920]

14. Acs B, Rantalainen M, Hartman J. Artificial intelligence as the next step towards precision pathology. J Intern Med. 2020;288(1):62-81. [FREE Full text] [doi: 10.1111/joim.13030] [Medline: 32128929]

15. Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. J Pathol Inform. 2016;7:29. [FREE Full text] [doi: 10.4103/2153-3539.186902] [Medline: 27563488]

16. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. Radiographics. 2017;37(2):505-515. [FREE Full text] [doi: 10.1148/rg.2017160130] [Medline: 28212054]

17. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med. 2018;24(10):1559-1567. [FREE Full text] [doi: 10.1038/s41591-018-0177-5] [Medline: 30224757]

18. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]

19. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med. 2019;25(8):1301-1309. [FREE Full text] [doi: 10.1038/s41591-019-0508-1] [Medline: 31308507]

20. Ho C, Zhao Z, Chen XF, Sauer J, Saraf SA, Jialdasani R, et al. A promising deep learning-assistive algorithm for histopathological screening of colorectal cancer. Sci Rep. 2022;12(1):2222. [FREE Full text] [doi: 10.1038/s41598-022-06264-x] [Medline: 35140318]

21. Ribeiro M, Grolinger K, Capretz MAM. MLaaS: machine learning as a service. 2015. Presented at: IEEE 14th International Conference on Machine Learning and Applications (ICMLA); 2015 December 09-11:896-902; Miami, FL, USA. [doi: 10.1080/15265161.2015.1028658]

22. Berg G. Image classification with machine learning as a service: - a comparison between Azure, SageMaker, and Vertex AI. Linnaeus University; 2022. URL: https://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-113829 [accessed 2025-05-16]

23. Bujanda L, Cosme A, Gil I, Arenas-Mirave JI. Malignant colorectal polyps. World J Gastroenterol. 2010;16(25):3103-3111. [FREE Full text] [doi: 10.3748/wjg.v16.i25.3103] [Medline: 20593495]

24. Rex DK, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, Kaltenbach T, et al. Colorectal cancer screening: recommendations for physicians and patients from the U.S. multi-society task force on colorectal cancer. Am J Gastroenterol. 2017;112(7):1016-1030. [doi: 10.1038/ajg.2017.174] [Medline: 28555630]

25. Slide Library, Virtual Pathology at the University of Leeds. 2022. URL: http://www.virtualpathology.leeds.ac.uk/slides/library/ [accessed 2022-01-12]

26. Slide Library, University of Michigan Virtual Slide Box. 2022. URL: https://www.pathology.med.umich.edu/apps/slides/ [accessed 2022-01-12]

27. Slide Library, Histopathology imaging on TCIA (The Cancer Imaging Archive). 2022. URL: https://www.cancerimagingarchive.net/histopathology-imaging-on-tcia/ [accessed 2022-01-12]

28. Kapishnikov A, Bolukbasi T, Viégas FB, Terry M. XRAI: better attributions through regions. In: Ann Nutr Metab. IEEE; 2019. Presented at: IEEE/CVF International Conference on Computer Vision (ICCV); 2019 October 27-02 November; Seoul, Korea (South). [doi: 10.1159/000508056]

29. Korbar B, Olofson AM, Miraflor AP, Nicka CM, Suriawinata MA, Torresani L, et al. Looking under the hood: deep neural network visualization to interpret whole-slide image analysis outcomes for colorectal polyps. 2017. Presented at: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); July 21-26, 2017:821-827; Honolulu, HI, USA.

30. Gschwantler M, Kriwanek S, Langner E, Göritzer B, Schrutka-Kölbl C, Brownstone E, et al. High-grade dysplasia and invasive carcinoma in colorectal adenomas: a multivariate analysis of the impact of adenoma and patient characteristics. Eur J Gastroenterol Hepatol. 2002;14(2):183-188. [doi: 10.1097/00042737-200202000-00013] [Medline: 11981343]

31. Basu S, Mitra S, Saha N. Deep learning for screening COVID-19 using chest x-ray images. medRxiv. 2020. [doi: 10.1101/2020.05.04.20090423]

32.   Schwen LO, Schacherer D, Geißler C, Homeyer A. Evaluating generic autoML tools for computational pathology. Inform Med Unlocked. 2022;29:100853. [doi: 10.1016/j.imu.2022.100853]

33.   Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. Comput Struct Biotechnol J. 2018;16:34-42. [FREE Full text] [doi: 10.1016/j.csbj.2018.01.001] [Medline: 30275936]

34.   Aeffner F, Zarella MD, Buchbinder N, Bui MM, Goodman MR, Hartman DJ, et al. Introduction to digital image analysis in whole-slide imaging: a white paper from the digital pathology association. J Pathol Inform. 2019;10:9. [FREE Full text] [doi: 10.4103/jpi.jpi_82_18] [Medline: 30984469]

Abbreviations

AI: artificial intelligence
ICC: International Color Consortium
ML: machine learning
MLaaS: Machine Learning as a Service
WSI: whole-slide image