Check for updates

Histopathology

Histopathology 2022, **80**, 1121–1127. DOI: 10.1111/his.14659

SHORT REPORT

The future of artificial intelligence in digital pathology – results of a survey across stakeholder groups

Céline N Heinz,[1] Amelie Echle,[1] Sebastian Foersch,[2] Andrey Bychkov[3] & Jakob Nikolas Kather[1,4,5]

[1]Department of Medicine III, University Hospital RWTH Aachen, Aachen, [2]Department of Pathology, University Medical Center Mainz, Mainz, [4]Medical Oncology, National Center for Tumor Diseases (NCT), University Hospital Heidelberg, Heidelberg, Germany, [3]Department of Pathology, Kameda Medical Center, Kamogawa, Chiba, Japan and [5]Pathology and Data Analytics, Leeds Institute of Medical Research at St James', University of Leeds, Leeds, UK

Date of submission 16 January 2022
Accepted for publication 2 April 2022
Published online *Article Accepted* 4 April 2022

Heinz C N, Echle A, Foersch S, Bychkov A & Kather J N
(2022) *Histopathology* **80**, 1121–1127. https://doi.org/10.1111/his.14659

The future of artificial intelligence in digital pathology – results of a survey across stakeholder groups

Aims: Artificial intelligence (AI) provides a powerful tool to extract information from digitised histopathology whole slide images. During the last 5 years, academic and commercial actors have developed new technical solutions for a diverse set of tasks, including tissue segmentation, cell detection, mutation prediction, prognostication and prediction of treatment response. In the light of limited overall resources, it is presently unclear for researchers, practitioners and policymakers which of these topics are stable enough for clinical use in the near future and which topics are still experimental, but worth investing time and effort into.

Methods and results: To identify potentially promising applications of AI in pathology, we performed an anonymous online survey of 75 computational pathology domain experts from academia and industry. Participants enrolled in 2021 were queried about their subjective opinion on promising and appealing subfields of computational pathology with a focus upon solid tumours. The results of this survey indicate that the prediction of treatment response directly from routine pathology slides is regarded as the most promising future application. This item was ranked highest in the overall analysis and in subgroups by age and professional background. Furthermore, prediction of genetic alterations, gene expression and survival directly from routine pathology images scored consistently high throughout subgroups.

Conclusions: Together, these data demonstrate a possible direction for the development of computational pathology systems in clinical, academic and industrial research in the near future.

Keywords: artificial intelligence, digital pathology, survey

Introduction

Digitisation of routine pathology workflows in itself provides efficiency gains[1,2] while maintaining the quality of diagnosis.[3] In addition, a benefit of digitising pathology workflows is that it could enable new biomarkers, potentially expanding the information that can be extracted from tissue slides,[4,5]

Address for correspondence: Jakob Nikolas Kather MD, MSc, Assistant Professor, Department of Medicine III, RWTH University Hospital, 52074 Aachen, Germany. e-mail: jakob-nikolas.kather@alumni.dkfz.de

particularly in cancer.[6–8] Most of these analytical approaches in computational pathology rely upon artificial intelligence (AI),[9] which can be applied for three types of problems (Figure 1A).[4] First, basic applications aim to automate tasks which are usually performed by pathologists, e.g. detection of tumour tissue,[10] subtyping[11] or grading;[12,13] secondly, advanced applications, including the prediction of mutations,[14–16] expression[17] and DNA repair deficiency,[18,19] as well as prediction of survival[20,21] and treatment response[22,23] directly from haematoxylin and eosin (H&E imagess and thirdly, 'enabling tasks' such as quality control,[24] detection of cells (for the purpose of subsequent quantification[25,26]) or quantification of immunohistochemical (IHC) stains.[27] However, it is currently unclear which of these approaches are most likely to bring tangible benefits in the near future.

Methods

ETHICS STATEMENT

We performed an anonymous online survey among professionals in computational pathology who participated voluntarily. No benefit or disadvantage was associated with participation or non-participation. The identity of the participants was unknown to the authors. No patient data were used.

SURVEY DESIGN

Participants reported their demographic and professional background in six questions. For each application category, a numerical rating scale was used to indicate the priority participants assigned to a given category. Finally, the survey included two optional free-text questions (Supporting information, Table S1).

IMPLEMENTATION AND DISTRIBUTION OF THE SURVEY

The survey was performed using Google Forms (Google Inc., Mountain View, CA, USA) from 3 June to 20 July 2021. The survey was advertised during professional symposia on digital pathology, including a course about digital pathology run by the French Society for Pathology (SFP), a presentation at the European Congress of Pathology (ECP) and on the social networks Twitter and LinkedIn via professional accounts of the authors. The survey was closed when 75 responses were recorded (predefined stopping criterion).

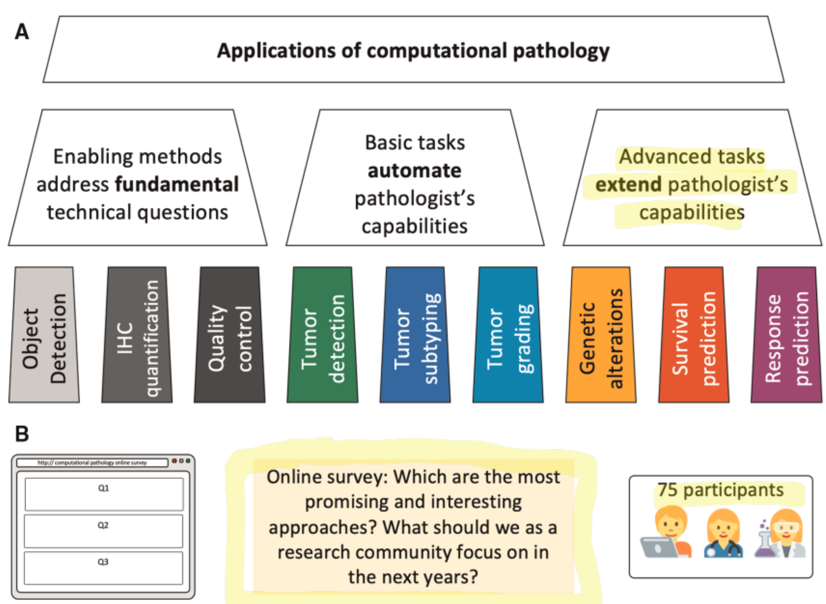

Figure 1. Applications of computational pathology and setup of this survey. **A**, Based on Echle *et al.*,[4] applications of computational pathology can be categorised as 'basic' or 'advanced'. In this study, enabling technologies were introduced as an additional category. **B**, We performed an online survey among 75 participants who are diverse stakeholders in the field. Participants were asked to report their subjective experience or opinion.

DATA ANALYSIS

Responses on a numerical rating scale were analysed by median and interquartile range (IQR). To visualise differences between groups, the mean and standard deviation (SD) was used. To compare responses between categories, the Kruskal–Wallis *U*-test was used. *P*-values below 0.01 were considered significant. Correlation was assessed via Spearman's analysis. Statistical analyses were carried out using Microsoft Excel and Python/SciPy. Analyses were performed for predefined subgroups [older (born before 1980) versus younger participants and medical versus non-medical background].

Results

BASIC DEMOGRAPHIC AND PROFESSIONAL BACKGROUND OF PARTICIPANTS

We report the results of an anonymous survey (Figure 1B, Supporting information, Table S2). Fifty-six (75%) participants were born before and 19 (25%) after 1990 (Supporting information, Figure S1A, Table S3). Thirty-six (48%) participants had a medical while 39 (52%) had a technical background (Supporting information, Figure S1B). Forty-three (57%) participants accessed the survey through social media (Supporting information, Figure S1C). Forty-nine (65%) participants had hands-on-experience in programming (Figure S1D), 42 (56%) participants reported having trained a neural network previously (Supporting information, Figure S1E) and 58 (77%) participants reported having been involved in image analysis projects (Supporting information, Figure S1F).

WHICH APPLICATIONS ARE MOST PROMISING OVERALL?

Participants were asked to rank different applications of AI, where 1 represents the least important and 10 the most interesting and/or promising application in their subjective opinion. Accordingly, the most promising application for AI in digital pathology was predicting treatment response directly from H&E images of solid tumours, reaching an overall arithmetic mean of 9.09 (median = 10, Figure 2A). In a pairwise comparison of categories, prediction of treatment response was scored significantly ($P < 0.01$) higher than all items in 'enabling' or 'basic' categories (Supporting information, Figure S2A). Also, predicting genetic mutations as well as gene expression directly from H&E images were considered second and third most interesting with an almost equal score of mean 8.64 (median = 10) and mean 8.63 (median = 10), respectively (Figure 2A). Taken together, these data show that 'advanced' tasks according to Echle *et al.*[4] (Figure 1A) reached the highest scores. In contrast, 'enabling' technologies scored lower: with an arithmetic mean of 6.83 (median = 8), AI for detection of large structures was considered to be the overall least promising application of AI in digital pathology and the scores were significantly ($P < 0.01$) lower than all scores of items in the 'advanced' category (Figure 2). In addition, we analysed the correlation between responses and found that individual items within the categories 'enabling technologies', 'automation' and 'advanced' (Figure 1A) generally clustered together, i.e. were often rated similarly by the participants (Supporting information, Figure S2B).

SUBGROUP ANALYSIS BY PROFESSIONAL AND DEMOGRAPHIC BACKGROUND

Among participants with a professional background in medicine and from a non-medical field (Supporting information, Figure S1B), as well as for younger and older participants (Supporting information, Table S3), prediction of treatment response was consistently ranked highest. Detection of large structures in histopathology images was consistently ranked lowest or second-to-lowest. In all four subgroups, prediction of survival was among the top 50% ranked targets and prediction of molecular features was also consistently ranked high. Interestingly, prediction of survival was ranked higher by participants with a medical (Figure 2B) than with a non-medical (Figure 2C, Supporting information, Table S4) background. Another interesting finding is related to the quantification of immunostains, which was ranked second-highest by the younger subgroup (Figure 2D, Supporting information, Table S5). Apart from these observations, no striking deviation from the overall ranking results (Figure 2A) was made for any of the subgroups.

ANALYSIS OF FREE TEXT RESPONSES

Participants were asked for free-text suggestions on application of AI in computational pathology, and 32 answers were given to this question. Most participants ($n = 8$) suggested using AI for time-consuming and laborious routine tasks in the pathological field (e.g. virtual staining and automated measurement) to integrate AI-based procedures into clinical practice

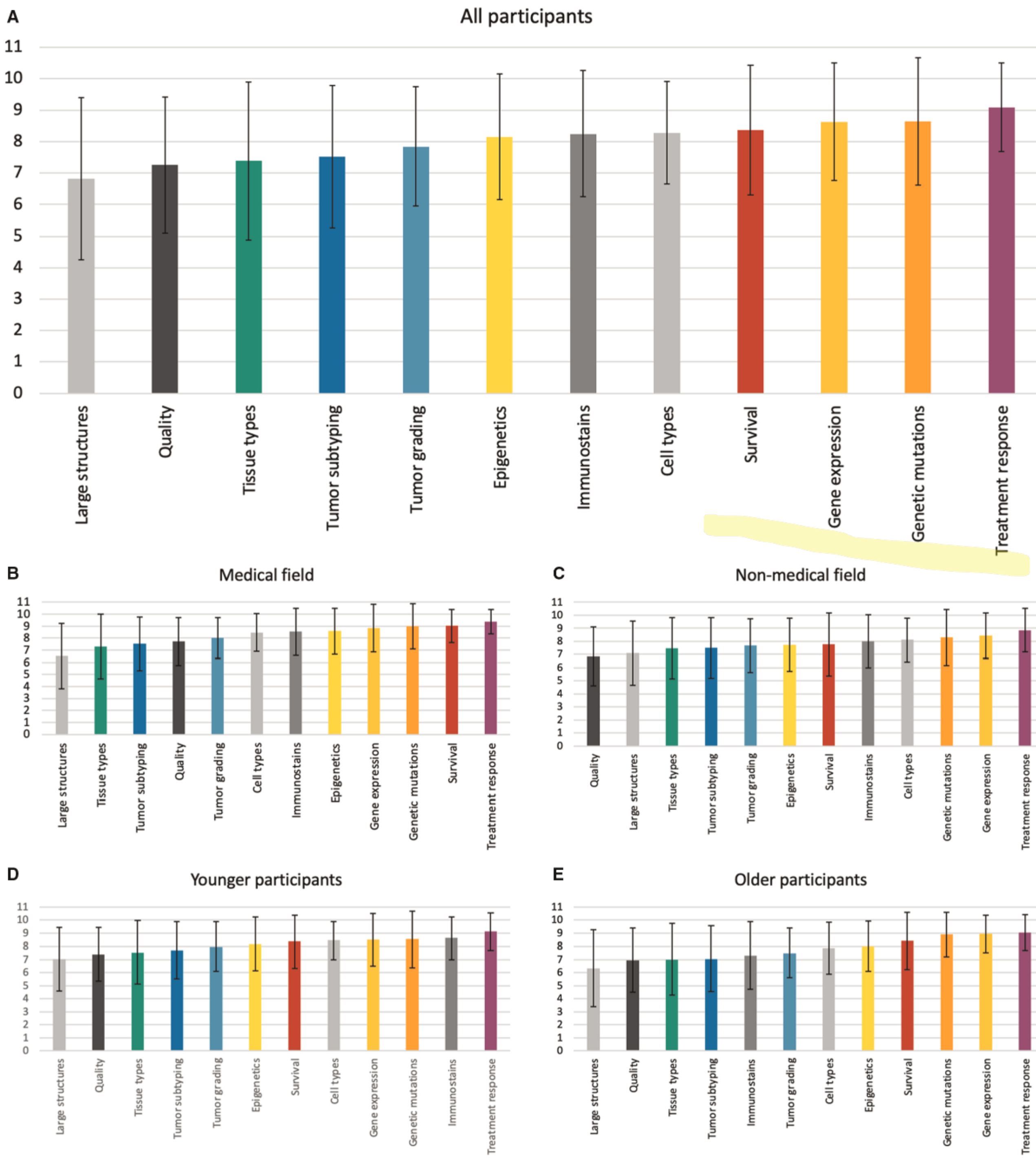

Figure 2. Ranking of responses in all participants and subgroups. **A**, Arithmetic mean and standard deviation for all participants. **B**, For participants with a medical professional background. **C**, For participants with a non-medical professional background. **D**, Younger participants born between 1980 and 1999. **E**, Older participants born between 1960 and 1979.

and improve the efficiency of pathologists and medical researchers in their daily workflow (Supporting information, Figure S3A, Table S6). Another common response ($n = 6$) was to use AI for quality control purposes, especially quality of the tissue, quality of staining and quality of scans. Furthermore, respondents showed their interest in using AI for detection of subtle morphological features such as a more detailed cell and tissue characterisation (e.g. in shape and structure) ($n = 3$). In addition, participants were asked for suggestions of how AI could improve medical research in general (Supporting information, Figure S3B, Table S7) and 36 answers were given to this question. Common suggestions were to reinforce integration of AI into clinical real-life settings ($n = 13$) to improve availability of large data sets, clinical cohorts and open-source AI codes ($n = 11$). The importance of comparing AI mined results with the real existing clinical data, and thereby increasing reproducibility, was another common suggestion ($n = 9$). Six respondents suggested interpretability of AI models for medical professionals as an important area, arguing that this could create stronger interconnections between technical and non-technical fields. Equally represented was the implementation of prospective and multicentre studies.

Discussion

LIMITATIONS

A limitation of our method is a possible non-response bias. Participants in the survey were recruited by disseminating invitations at professional conferences and on social networks targeting a particular audience. The decision of individuals in the target population to participate in the survey could bias the results.[28,29] Pathology experts not familiar with or even critical to digital pathology and AI applications might have a different opinion concerning what the research should focus upon in this regard. Another limitation is that this survey provides just a snapshot in time and topics of subjective relevance could change over time. In this sense, it may serve as a starting-point for the follow-up studies over a certain time-period. In particular, it would be interesting to perform a similar survey in another population: pathologists with no or very little experience from image analysis.

COMPARISON TO PREVIOUS STUDIES

Unlike previous surveys about AI in digital pathology,[30–32] we aimed to identify research areas of interest which could help stakeholders to sensibly allocate resources in academic and industry settings. Based on previous work,[4] we divided the fields of application into enabling technologies, automation of pathology workflows and advanced tasks (Figure 1). We found that advanced applications of AI were consistently rated as most interesting across all subgroups of participants (Figure 2). Responses in free-text questions pointed out the need of using AI to automate time-consuming diagnostic tasks and real-world validation, which is in line with initiatives to increase the level of evidence of computational pathology systems.[33] Another prominent point was improvement of data sharing, which is in line with approaches to standardise reporting of AI biomarker studies.[34]

OUTLOOK

The results of our survey could be used to guide future research in academia and industry. Importantly, there are strong interdependencies of the different areas; e.g. AI-driven quality control is useful for an AI-enhanced diagnosis and decision support. Overall, our study shows the need for computational pathology solutions which go beyond workflow automation and provide true new biomarkers for outcome and response prediction. We suggest that more pathologists could be trained in AI to enable them to devise and test their own ideas for new biomarkers. Also, this could help overcome fears of AI, as it was shown in radiology that the more a person knows about AI, the less they fear it.[35] Finally, beyond its powers in image analysis, AI also excels at natural language processing (NLP). Combining vision and language AI models would open the possibility to interact with AI through natural language, which could further improve the integration into clinical workflows.

Acknowledgements

J.N.K. is supported by the German Federal Ministry of Health (DEEP LIVER, ZMVI1-2520DAT111) and the Max-Eder-Programme of the German Cancer Aid (grant no. 70113864). No other specific funding for this work is declared by any of the authors.

Conflicts of interest

J.N.K. declares consulting services for Owkin, France and Panakeia, UK. No other potential conflicts of interest are reported by any of the authors. None of the sponsors or funding agencies had any role in study design, collection, analysis or interpretation of data.

Data availability statement

The data that supports the findings of this study are available in the supplementary material of this article.

References

1. Hanna Matthew G, Reuter Victor E, Hameed Meera R *et al.* Whole slide imaging equivalency and efficiency study: experience at a large academic center. *Mod. Pathol.* 2019; **32**; 916–928. https://doi.org/10.1038/s41379-019-0205-0.
2. Vodovnik A. Diagnostic time in digital pathology: a comparative study on 400 cases. *J. Pathol. Inform.* 2016; **7**; 4. https://doi.org/10.4103/2153-3539.175377.
3. Pantanowitz L, Sinard John H, Henricks Walter H *et al.* validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch. Pathol. Lab. Med.* 2013; **137**; 1710–1722. https://doi.org/10.5858/arpa.2013-0093-cp.
4. Echle A, Timon RN, Josef BT, Tom L, Thomas PA, Nikolas KJ. Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br. J. Cancer* 2021; **124**; 686–696. https://doi.org/10.1038/s41416-020-01122-x.
5. Nam D, Chapiro J, Paradis V, Seraphin TP, Kather JN. Artificial intelligence in liver diseases: improving diagnostics, prognostics and response prediction. *JHEP Rep.* 2022; **4**; 100443.
6. Sultan AS, Elgharib MA, Tavares T, Jessri M, Basile JR. The use of artificial intelligence, machine learning and deep learning in oncologic histopathology. *J. Oral Pathol. Med.* 2020; **49**; 849–856.
7. Noorbakhsh J, Saman F, Foroughi pour A *et al.* Deep learning-based cross-classifications reveal conserved spatial behaviors within tumor histological images. *Nat. Commun.* 2020; **11**(1). https://doi.org/10.1038/s41467-020-20030-5.
8. Litjens G, Thijs K, Ehteshami BB *et al.* A survey on deep learning in medical image analysis. *Med. Image Anal.* 2017; **42**; 60–88. https://doi.org/10.1016/j.media.2017.07.005.
9. Kather JN, Calderaro J. Development of AI-based pathology biomarkers in gastrointestinal and liver cancer. *Nat. Rev. Gastroenterol. Hepatol.* 2020; **17**; 591–592.
10. Campanella G, Hanna Matthew G, Luke G *et al.* Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* 2019; **25**; 1301–1309. https://doi.org/10.1038/s41591-019-0508-1.
11. Lu MY, Williamson Drew FK, Chen Tiffany Y, Chen Richard J, Matteo B, Faisal M. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* 2021; **5**; 555–570. https://doi.org/10.1038/s41551-020-00682-w.
12. Peter S, Kimmo K, Henrik O *et al.* Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol.* 2020; **21**; 222–232. https://doi.org/10.1016/s1470-2045(19)30738-7.
13. Wouter B, Hans P, van Hester B *et al.* Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol.* 2020; **21**; 233–241. https://doi.org/10.1016/s1470-2045(19)30739-9.
14. Nikolas KJ, Heij Lara R, Grabsch Heike I *et al.* Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat. Can.* 2020; **1**; 789–799. https://doi.org/10.1038/s43018-020-0087-6.
15. Yu F, Jung Alexander W, Viñas TR *et al.* Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat. Can.* 2020; **1**; 800–810. https://doi.org/10.1038/s43018-020-0085-8.
16. Lavinia LCM, Nadina OB, Max J *et al.* Artificial intelligence–based detection of FGFR3 mutational status directly from routine histology in bladder cancer: a possible preselection for molecular testing? *Eur. Urol. Focus* 2021. https://doi.org/10.1016/j.euf.2021.04.007.
17. Benoît S, Alberto R, Elodie P *et al.* A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nat. Commun.* 2020; **11**(1). https://doi.org/10.1038/s41467-020-17678-4.
18. Sophie MH, Rosaline HL, Gisela K *et al.* Development and validation of deep learning classifiers to detect Epstein-Barr virus and microsatellite instability status in gastric cancer: a retrospective multicentre cohort study. *The Lancet Digital Health* 2021; **3**; e654–e664. https://doi.org/10.1016/s2589-7500(21)00133-3.
19. Amelie E, Irmgard GH, Philip Q *et al.* Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. *Gastroenterology* 2020; **159**; 1406–1416.e11. https://doi.org/10.1053/j.gastro.2020.06.021.
20. Ole-Johan S, Sepp DR, Andreas K *et al.* Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet* 2020; **395**; 350–360. https://doi.org/10.1016/s0140-6736(19)32998-8.
21. Nikolas KJ, Johannes K, Pornpimol C *et al.* Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. *PLoS Med.* 2019; **16**; e1002730. https://doi.org/10.1371/journal.pmed.1002730.
22. Saman F, Fernandez Aileen I, Shabbir AF *et al.* Deep learning trained on hematoxylin and eosin tumor region of Interest predicts HER2 status and trastuzumab treatment response in HER2+ breast cancer. *Mod. Pathol.* 2022; **35**; 44–51. https://doi.org/10.1038/s41379-021-00911-w.
23. Bychkov D, Nina L, Aleksei T *et al.* Deep learning identifies morphological features in breast cancer predictive of cancer ERBB2 status and trastuzumab treatment efficacy. *Sci. Rep.* 2021; **11**(1). https://doi.org/10.1038/s41598-021-83102-6.
24. Andrew J, Ren Z, Hannah G, Michael F, Anant M. HistoQC: an open-source quality control tool for digital pathology slides. *JCO Clin. Cancer Inform.* 2019; **3**; 1–7. https://doi.org/10.1200/cci.18.00157.
25. Yajie H, Feng S, Kun D *et al.* Deep learning system for lymph node quantification and metastatic cancer identification from whole-slide pathology images. *Gastric Cancer* 2021; **24**; 868–877. https://doi.org/10.1007/s10120-021-01158-9.
26. Basavanhally AN, Ganesan S, Agner S *et al.* Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology. *IEEE Trans. Biomed. Eng.* 2010; **57**; 642–653. https://doi.org/10.1109/tbme.2009.2035305.
27. Kamal SMM, Yasmine M, Craig Stephanie G *et al.* A means of assessing deep learning-based detection of ICOS protein expression in colon cancer. *Cancer* 2021; **13**; 3825. https://doi.org/10.3390/cancers13153825.
28. Eaden J, Mayberry MK, Mayberry JF. Research methods: questionnaires: the use and abuse of social survey methods in medical research. *Postgrad. Med. J.* 1999; **75**; 397–400. https://doi.org/10.1136/pgmj.75.885.397.
29. Kelley K. Good practice in the conduct and reporting of survey research. *Int. J. Qual. Health Care* 2003; **15**; 261–266. https://doi.org/10.1093/intqhc/mzg031.
30. Rosaria GM, Sara C, Livia C, Daniele G. Artificial intelligence in digital pathology: what is the future? part 2: an investigation on the insiders. *Healthcare* 2021; **9**; 1347. https://doi.org/10.3390/healthcare9101347.
31. Shihab S, Anglin D, Kevin F *et al.* Physician perspectives on integration of artificial intelligence into diagnostic pathology.

NPJ Digital Medicine 2019; **2**. https://doi.org/10.1038/s41746-019-0106-0.

32. Jill WB, Jessica L, Oien Karin A, Darren T. Digital pathology access and usage in the UK: results from a national survey on behalf of the National Cancer Research Institute's CM-Path initiative. *J. Clin. Pathol.* 2018; **71**; 463–466. https://doi.org/10.1136/jclinpath-2017-204808.
33. Andreas K, Ole-Johan S, Sepp DR, Knut L, Kerr David J, Danielsen Håvard E. Designing deep learning studies in cancer diagnostics. *Nat. Rev. Cancer* 2021; **21**; 199–211. https://doi.org/10.1038/s41568-020-00327-9.
34. Viknesh S, Hutan A, Ravi A *et al*. Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: the STARD-AI Steering Group. *Nat. Med.* 2020; **26**; 807–808. https://doi.org/10.1038/s41591-020-0941-1.
35. Merel H, Erik R, William P *et al*. An international survey on AI in radiology in 1,041 radiologists and radiology residents part 1: fear of replacement, knowledge, and attitude. *Eur. Radiol.* 2021; **31**; 7058–7066. https://doi.org/10.1007/s00330-021-07781-5.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Questionnaire description and reference.

Table S2. Median (med) and interquartile range (IQR) for all subgroups for all questionnaire items. Questionnaire items were judged on a scale from 1 to 10, 10 indicating highest agreement/importance.

Table S3. Precise representation of the individual age subgroups defined by a ten-year range of birth years respectively. Birth years are listed in ascending order.

Table S4. Detailed overview of the professional background subgroup analysis including the two secondary established subgroups .medical. and .non-medical. and showing arithmetic mean, median and interquartile range (IQR) for the top three ranked applications respectively.

Table S5. Detailed overview of age subgroup analysis including the two secondary established age subgroups .younger. and .older. and showing arithmetic mean, median and interquartile range (IQR) for the top three ranked applications respectively.

Table S6. Overview of respondents' most frequently given answers for open question number 1 ('Which other applications of AI in computational pathology would you find promising or interesting?'). Different answer categories were established during free text question analysis and are listed in decreasing order.

Table S7. Overview of respondents' most frequently given answers for open question number 2 ('In your opinion, how can the use of AI in medical research be further improved?'). Different answer categories were established during free text question analysis and are listed in decreasing order.

Figure S1. Basic demographics and previous experience of survey participants. (**A**) Self-reported age among participants, (**B**) Professional background of participants. (**C**) Distribution of the survey (**D**) Computer programming experience among participants, (**E**) Hands-on machine learning experience among participants, (**F**) project experience of participants.

Figure S2. (A) Statistical significance of pairwise differences between scores in categories (pairwise Kruskal *P* value), (B) Correlation between scores in categories (Spearman correlation coefficient) in N=75 participants.

Figure S3. Qualitative analysis of responses to free text questions. **(A)** 'Which other applications of AI in computational pathology would you find promising or interesting?', **(B)** 'In your opinion, how can the use of AI in medical research be further improved?'