



Original Investigation | Pathology and Laboratory Medicine

# Pathologist Validation of a Machine Learning–Derived Feature for Colon Cancer Risk Stratification

Vincenzo L'Imperio, MD; Ellery Wulczyn, MS; Markus Plass, MSc; Heimo Müller, PhD; Nicolò Tamini, MD; Luca Gianotti, MD; Nicola Zucchini, MD; Robert Reihns, MS; Greg S. Corrado, PhD; Dale R. Webster, PhD; Lily H. Peng, MD, PhD; Po-Hsuan Cameron Chen, PhD; Marialuisa Lavitrano, MD, PhD; Yun Liu, PhD; David F. Steiner, MD, PhD; Kurt Zatloukal, MD; Fabio Pagni, MD

## Abstract

**IMPORTANCE** Identifying new prognostic features in colon cancer has the potential to refine histopathologic review and inform patient care. Although prognostic artificial intelligence systems have recently demonstrated significant risk stratification for several cancer types, studies have not yet shown that the machine learning–derived features associated with these prognostic artificial intelligence systems are both interpretable and usable by pathologists.

**OBJECTIVE** To evaluate whether pathologist scoring of a histopathologic feature previously identified by machine learning is associated with survival among patients with colon cancer.

**DESIGN, SETTING, AND PARTICIPANTS** This prognostic study used deidentified, archived colorectal cancer cases from January 2013 to December 2015 from the University of Milano-Bicocca. All available histologic slides from 258 consecutive colon adenocarcinoma cases were reviewed from December 2021 to February 2022 by 2 pathologists, who conducted semiquantitative scoring for tumor adipose feature (TAF), which was previously identified via a prognostic deep learning model developed with an independent colorectal cancer cohort.

**MAIN OUTCOMES AND MEASURES** Prognostic value of TAF for overall survival and disease-specific survival as measured by univariable and multivariable regression analyses. Interpathologist agreement in TAF scoring was also evaluated.

**RESULTS** A total of 258 colon adenocarcinoma histopathologic cases from 258 patients (138 men [53%]; median age, 67 years [IQR, 65-81 years]) with stage II (n = 119) or stage III (n = 139) cancer were included. Tumor adipose feature was identified in 120 cases (widespread in 63 cases, multifocal in 31, and unifocal in 26). For overall survival analysis after adjustment for tumor stage, TAF was independently prognostic in 2 ways: TAF as a binary feature (presence vs absence: hazard ratio [HR] for presence of TAF, 1.55 [95% CI, 1.07-2.25];  $P = .02$ ) and TAF as a semiquantitative categorical feature (HR for widespread TAF, 1.87 [95% CI, 1.23-2.85];  $P = .004$ ). Interpathologist agreement for widespread TAF vs lower categories (absent, unifocal, or multifocal) was 90%, corresponding to a  $\kappa$  metric at this threshold of 0.69 (95% CI, 0.58-0.80).

**CONCLUSIONS AND RELEVANCE** In this prognostic study, pathologists were able to learn and reproducibly score for TAF, providing significant risk stratification on this independent data set. Although additional work is warranted to understand the biological significance of this feature and to establish broadly reproducible TAF scoring, this work represents the first validation to date of human expert learning from machine learning in pathology. Specifically, this validation demonstrates that a computationally identified histologic feature can represent a human-identifiable, prognostic feature with the potential for integration into pathology practice.

JAMA Network Open. 2023;6(3):e2254891. doi:10.1001/jamanetworkopen.2022.54891

**Open Access.** This is an open access article distributed under the terms of the CC-BY-NC-ND License.

JAMA Network Open. 2023;6(3):e2254891. doi:10.1001/jamanetworkopen.2022.54891

## Key Points

**Question** Can a prognostic machine learning–derived histopathologic feature be learned and validated by pathologists?

**Findings** In this prognostic study, 2 pathologists were able to learn a machine learning–derived histopathologic feature and validate its prognostic value for survival among patients with colon cancer.

**Meaning** These findings suggest that computationally identified histopathologic features can provide prognostic value for colon cancer, with the potential for integration into pathology practice.

## + Supplemental content

Author affiliations and article information are listed at the end of this article.

## Introduction

Colorectal adenocarcinoma represents the third most common cancer and the second leading cause of cancer mortality. The management of these cases relies primarily on classic histopathology-based prognostic markers,<sup>1</sup> including tumor budding,<sup>2</sup> lymphovascular invasion,<sup>3</sup> tumor differentiation, and TNM staging.<sup>4,5</sup> Prognostic markers are of significant clinical interest in colorectal cancer, as some patients with stage II disease may benefit from adjuvant chemotherapy, and, for patients with stage III disease, improved prognostic information can inform treatment regimen and duration.<sup>6–8</sup> Better risk stratification and prognostic markers within stage II and stage III colorectal cancer, therefore, offer opportunities to improve therapy decisions and patient care. In this setting, the use of digital pathology tools (eg, machine learning) has recently demonstrated the capability to provide prognostic information about colon cancer with the use of routine histopathologic slides.<sup>9–11</sup> This led to the identification of the tumor adipose feature (TAF), moderately to poorly differentiated tumor cells in close proximity to adipocytes, as a machine learning–derived feature that demonstrated promising, independent prognostic value in stage II and III colorectal cancer cases.<sup>10</sup> In the present study, we evaluated and validated this feature via traditional histopathologic review, assessing whether the machine learning–derived TAF retains its prognostic value when assessed by human pathologists in an external cohort of colorectal adenocarcinoma cases.

## Methods

### Data Source

This prognostic study used consecutive, archived colorectal cancer cases retrieved from the ASST Monza, San Gerardo Hospital, University of Milano-Bicocca (UNIMIB), Monza, Italy, from January 2013 to December 2015. Inclusion criteria included untreated cases of primary stage II or III cancer, availability of tumor-containing slides, and availability of information on clinical outcomes. Institutional review of pathology reports and clinical notes was performed to obtain the clinicopathologic cohort characteristics described. Cohort characteristics in the **Table** were ascertained from pathology reports. These data represent a validation cohort independent from the initial feature discovery and test cohort described previously.<sup>10</sup> Institutional review board approval for this retrospective study using deidentified slides was obtained from the ethics committee of UNIMIB and the ethics committee of the Medical University of Graz with a waiver of informed consent because deidentified data were used. This study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline for diagnostic and prognostic studies.

### TAF Scoring

Tumor adipose feature was scored independently from December 2021 to February 2022 by 2 pathologists (V.L. and F.P.) with 8 and 12 years, respectively, of experience in gastrointestinal pathology. Pathologists were blinded to the patient outcomes per case (see the Prognostic Evaluation subsection of the Methods section). All glass slides from the cases were evaluated independently by the 2 observers to assess the presence or absence of TAF and the extent of the feature, when present, resulting in a summary score for each case. Additional histopathologic features were not reviewed at the time of retrospective TAF scoring. Prior to case review and TAF scoring, the pathologists completed training via review and discussion of the TAF patch examples identified in the initial feature discovery cohort.<sup>10</sup> The images in eFigures 1, 2, 3, and 4 in [Supplement 1](#) were also reviewed to enable a better understanding of TAF appearance and histologic context at different magnifications. Pathologists also reviewed an initial set of archived cases from UNIMIB to align and calibrate on TAF classification. Based on both initial pathologist discussion and the quantitative significance of TAF observed previously,<sup>10</sup> TAF was classified semiquantitatively at the case level as absent, unifocal, multifocal, or widespread (**Figure 1**). For the prespecified primary

analysis, the threshold for presence of TAF was defined as multifocal or widespread. All tumor-containing slides from each case were used for case-level scoring using archived glass slides and traditional microscopy. After independent review, cases in the final study set with discrepant scoring were jointly reviewed to resolve disagreement.

### Prognostic Evaluation

Planned analysis was conducted to evaluate the association of the presence of TAF with overall survival (OS) and colorectal cancer disease-specific survival (DSS) after adjustment for tumor stage (stage II or III in this study). Overall survival and DSS were ascertained from electronic health records as of December 2021. Disease-free survival time was not consistently available for analysis.

**Table. Cohort Characteristics**

Characteristic	All cases, No. (%) (N = 258)
Age, median (IQR), y	67 (65-81)
Events	
Overall survival	123 (47)
Disease-specific survival	36 (14)
Sex	
Female	120 (47)
Male	138 (53)
Stage	
IIA	102 (40)
IIB	14 (5)
IIC	3 (1)
IIIA	11 (4)
IIIB	99 (38)
IIIC	29 (11)
Grade	
Low	189 (73)
High	69 (27)
Tumor location	
Right colon	160 (62)
Left colon	98 (38)
Histologic subtype	
Adenocarcinoma NOS	235 (91)
Mucinous	23 (9)
Lymphovascular invasion	
Absent	87 (34)
Present	171 (66)
Margin status	
Negative	253 (98)
Positive	5 (2)
MMR or MSI	
Stable	71 (28)
Loss of MMR or MSI	14 (5)
NA	173 (67)
TAF	
None observed	138 (54)
Unifocal	26 (10)
Multifocal	31 (12)
Widespread	63 (24)

Abbreviations: MMR, mismatch repair; MSI, microsatellite instability; NA, not available; NOS, not otherwise specified; TAF, tumor adipose feature.

Disease-specific death was defined on the basis of documented tumor metastasis or progression being present at or near the time of death (either via clinical or pathologic reports) and thus may underrepresent actual disease-specific events if such documentation was not present in the available data. These end points occurred between 2013 and 2021, corresponding to a mean (SD) follow-up period for events of 61 (34) months. On the basis of pilot data for a cohort of 200 cases, planned analysis was also conducted to evaluate the hazard ratio (HR) associated with increasing amounts of TAF as scored by pathologist review (see the TAF Scoring subsection). Additional analysis was performed to investigate the association of TAF with prognosis after adjustment for additional available baseline variables, including grade, histologic subtype, mismatch, lymphovascular invasion status, and tumor location.

**Statistical Analysis**

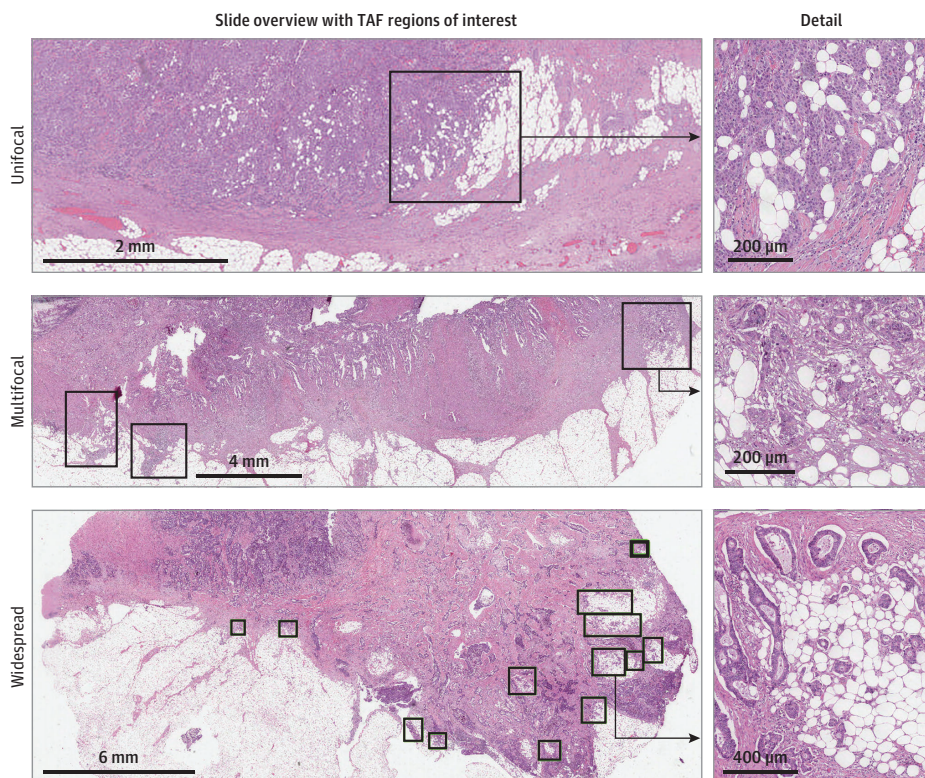
All survival analyses (eg, Cox proportional hazards regression, Kaplan-Meier estimation) were performed using the lifelines, version 0.26.0, software package.<sup>12</sup> All *P* values were from 2-sided tests and results were deemed statistically significant at *P* < .05. The sample size was determined via power analysis assuming an HR of 2.5 (estimated on prior data). At least 250 cases were required to ensure power exceeding 0.8.

**Results**

**Cohort Characteristics**

A total of 258 colon adenocarcinoma histopathologic cases from 258 patients (138 men [53%] and 120 women [47%]; median age, 67 years [IQR, 65-81 years]) with stage II (119 [46%]) and stage III (139 [54%]) cancer were included. Additional cohort characteristics are summarized in the Table. From an initial consecutive cohort of 371 cases, 11 cases from patients who underwent neoadjuvant

Figure 1. Examples of Tumor Adipose Feature (TAF) Identification and Semiquantitative Scoring



Representative regions from each possible scoring category (unifocal, multifocal, and widespread) are shown. Highlighted regions (black boxes) represent areas identified with TAF involvement and arrows represent higher magnification of select regions as indicated. Regions not highlighted represent areas without TAF identified.

therapy, 5 cases representing recurrent disease, 91 cases with either stage I (n = 75) or IV (n = 16) disease, 3 cases without clinical data available, and 3 cases with rectal cancer were excluded.

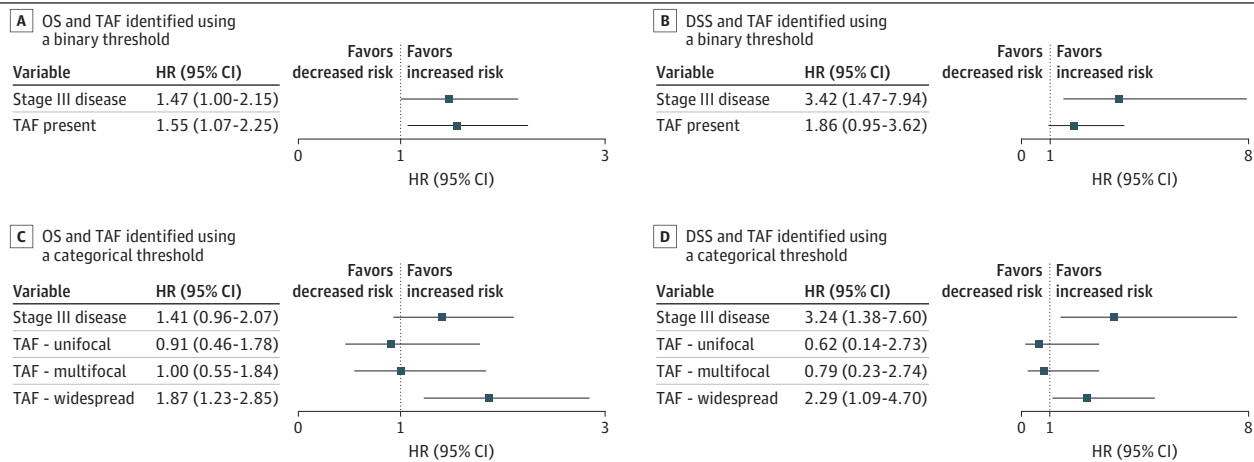
**TAF Evaluation**

The final TAF classification for the cases (none observed, unifocal, multifocal, or widespread) in this study are summarized in the TAF rows of the Table, with example images in Figure 1. Tumor adipose feature was identified in 120 cases (47%), with multifocal involvement in 31 cases (12%) and widespread involvement in 63 cases (24%). Overall pathologist agreement on initial review was 72% across all TAF scores and 90% for widespread TAF vs other classifications, corresponding to a κ metric of 0.69 (95% CI, 0.58-0.80) at this threshold. Concordance between pathologists for TAF classification and resolution of initial disagreements are summarized in eTable 1 in Supplement 1.

**TAF Prognostic Value**

Planned analysis was conducted to evaluate the association of TAF with OS and DSS (Figure 2 and Figure 3). Significant prognostic value of pathologist-identified TAF using a binary threshold was observed for OS (HR, 1.55 [95% CI, 1.07-2.25]; P = .02) (Figure 2A), but not for DSS (HR, 1.86 [95% CI, 0.95-3.62]; P = .07) (Figure 2B). In addition, a quantity-dependent association of TAF was observed for both OS and DSS, with widespread TAF demonstrating an HR of 1.87 (95% CI, 1.23-2.85; P = .004) for OS (Figure 2C and Figure 3A) and an HR of 2.29 (95% CI, 1.09-4.70; P = .03) for DSS in stage-adjusted analysis (Figure 2D and Figure 3B). Finally, we analyzed the association of TAF with prognosis after adjustment for additional available baseline variables, including age, sex, grade, histologic subtype, lymphovascular invasion status, and tumor location. These results are summarized in eTable 2 and eFigure 5 in Supplement 1. For OS, only age (HR, 1.07 [95% CI, 1.05-1.09]; P < .001), stage (HR 1.60 [95% CI, 1.03-2.51]; P = .04), and widespread TAF (HR 1.79 [95% CI, 1.14-2.81]; P = .01) remained independently prognostic in multivariable analysis. For DSS, only stage (HR, 3.57 [95% CI, 1.39-9.18]; P = .008) and widespread TAF (HR, 2.19 [95% CI, 1.01-4.75]; P = .047) remained independently prognostic. Univariable hazard ratios are provided in eTable 3 in Supplement 1.

**Figure 2. Association of Tumor Adipose Feature (TAF) With Overall Survival (OS) and Disease-Specific Survival (DSS)**



Forest plots for hazard ratios (HRs) based on Cox proportional hazards regression after adjustment for stage. A, Overall survival and TAF analyzed as a binary variable (none observed or unifocal as the reference vs multifocal or widespread as TAF present). B, Disease-specific survival and TAF analyzed as a binary variable. C, Overall survival and

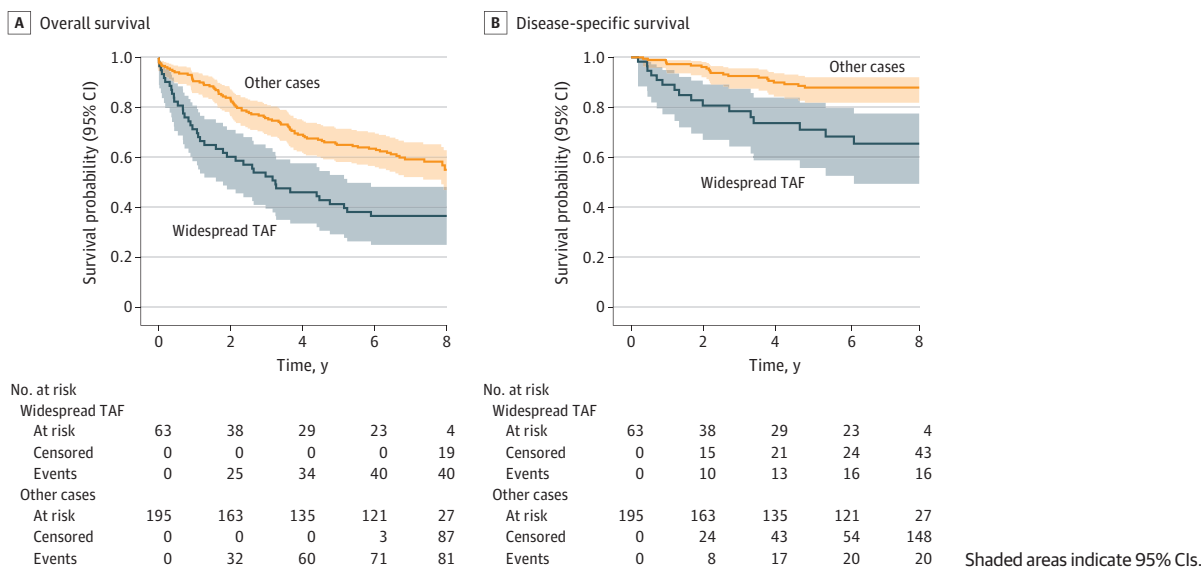
TAF analyzed as a categorical variable using semiquantitative TAF scoring and "none observed" as the reference category. D, Disease-specific survival and TAF analyzed as a categorical variable.

## Discussion

Our study validates that pathologists can learn a novel morphologic feature discovered by a machine learning model, and that this human-based scoring can be reproducible and prognostic. There has been substantial anticipation that artificial intelligence (AI) can serve as a discovery tool for novel histologic features associated with disease biology or prognosis. However, with many top-performing prognostic models relying on hard-to-interpret approaches and with interpretability efforts that often focus on identification of already established concepts, learning novel features from machine learning remains particularly challenging. By providing proof of concept that a specific, machine learning–derived morphologic feature can provide prognostic value when learned and scored by pathologists, this work represents a milestone for AI in pathology. Specifically, this study further validates the prognostic significance of TAF across an external cohort from a different country and institution by using pathologist TAF scoring of complete cases on glass histology slides. In particular, the cases with widespread TAF were detected with substantial interobserver agreement; these cases demonstrated significantly poorer prognosis compared with cases without TAF or those with lower amounts of TAF.

Even in this initial study, the pathologist agreement for feature scoring was on par with that of well-established prognostic features with scoring guidelines that have been refined over years or decades.<sup>12,13</sup> For example, at the most relevant prognostic threshold for TAF, widespread vs other (including absent), the  $\kappa$  value was 0.69 (corresponding to 90% agreement). For widespread vs unifocal or multifocal TAF, the  $\kappa$  value was 0.58. These results appear in line with or even exceed the interobserver agreement for established prognostic factors routinely evaluated in colorectal cancer, such as tumor budding and lymphovascular invasion. For example, the estimated interobserver agreement for lymphovascular invasion has been reported to be a  $\kappa$  value ranging from 0.23 to 0.28 on hematoxylin-eosin staining,<sup>3,14</sup> with only mild improvement when using immunohistochemistry for endothelial markers ( $\kappa$ , 0.26-0.42)<sup>3</sup> or elastin ( $\kappa$ , 0.41).<sup>14</sup> For tumor budding, another well-established prognostic feature in colorectal cancer, estimated interobserver variability reports a  $\kappa$  value ranging from 0.41 to 0.93, again with immunohistochemistry associated with only marginal improvement ( $\kappa$ , 0.53-0.87).<sup>13</sup> In summary, these results suggest that TAF scoring can achieve substantial interobserver reproducibility consistent with the potential to be clinically usable. However, just as was necessary for tumor budding and other prognostic features used in clinical

**Figure 3. Kaplan-Meier Curves Stratified on the Presence of Widespread Tumor Adipose Feature (TAF) vs All Other Cases**



practice, additional efforts to define appropriate scoring systems and thresholds will be required to refine and improve the prognostic value reported here.

This study demonstrates an important proof of concept for AI-based feature discovery and validation, addressing the important connection between explainability and utility for AI models in medicine.<sup>15</sup> These findings raise the biological significance of TAF as an important topic for further investigation, a notion that is further supported by the recent AI-based identification of inflammatory adipose tissue associated with lymph node metastasis in colon cancer<sup>16</sup> and a proposed mechanism linking adipose tissue to colorectal cancer metastasis and epithelial to mesenchymal transition.<sup>17</sup> In addition, while TAF is not specifically associated with depth of tumor invasion (based on presence and prognostic value across stage and T categories) and appears to be distinct from the presence of adipose more generally as a component of the tumor microenvironment,<sup>18</sup> efforts to understand its possible association with obesity and colon cancer features such as tumor border configuration<sup>19</sup> and tumor budding are also warranted.

### Limitations

This work has some limitations, including the single-institution nature of this validation study as well as the limited number of pathologists involved to demonstrate a reproducible scoring strategy. In addition, the availability of molecular information for the cases from this period was limited. Validation in data sets with more complete information regarding molecular covariates, such as mismatch repair status or *BRAF* variants, could also be useful.

### Conclusions

This prognostic study represents a milestone for AI in pathology and medicine, demonstrating both the feasibility and prognostic potential for pathologist-based integration of a feature identified via machine learning. Although much work is still necessary to establish and validate reproducible scoring systems for TAF and to further understand generalizability of the results reported here, the ability to identify, learn, and validate a machine learning–derived feature offers opportunities for feature discovery and hypothesis generation using AI. After the demonstration of generalizable prognostic value and consistent scoring strategies across pathologists, AI-derived prognostic features can potentially be used along with well-established features in prospective cases to enable further validation and clinical integration.

### ARTICLE INFORMATION

**Accepted for Publication:** December 13, 2022.

**Published:** March 14, 2023. doi:[10.1001/jamanetworkopen.2022.54891](https://doi.org/10.1001/jamanetworkopen.2022.54891)

**Open Access:** This is an open access article distributed under the terms of the [CC-BY-NC-ND License](https://creativecommons.org/licenses/by-nc-nd/4.0/). © 2023 L'Imperio V et al. *JAMA Network Open*.

**Corresponding Authors:** David F. Steiner MD, PhD, Google Health, 3400 Hillview Ave, Palo Alto, CA 94304 ([davesteiner@google.com](mailto:davesteiner@google.com)); Fabio Pagni, MD, Department of Medicine and Surgery, University of Milano-Bicocca, 20900 Monza, Italy ([fabio.pagni@unimib.it](mailto:fabio.pagni@unimib.it)).

**Author Affiliations:** Department of Medicine and Surgery, Pathology, University of Milano-Bicocca, IRCCS (Scientific Institute for Research, Hospitalization and Healthcare) Fondazione San Gerardo dei Tintori, Monza, Italy (L'Imperio, Zucchini, Lavitrano, Pagni); Google Health, Google LLC, Palo Alto, California (Wulczyn, Corrado, Webster, Peng, Chen, Liu, Steiner); Medical University of Graz, Diagnostic and Research Institute of Pathology, Graz, Austria (Plass, Müller, Reihls, Zatloukal); Department of Surgery, San Gerardo Hospital, Monza, Italy (Tamini, Gianotti).

**Author Contributions:** Drs L'Imperio and Pagni had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Dr L'Imperio and Messrs Wulczyn and Plass contributed equally as co-first authors. Drs Steiner, Zatloukal, and Pagni contributed equally as co-senior authors.

**Concept and design:** Plass, Reihs, Corrado, Webster, Peng, Chen, Lavitrano, Liu, Steiner, Zatloukal, Pagni.

**Acquisition, analysis, or interpretation of data:** L'Imperio, Wulczyn, Plass, Müller, Tamini, Gianotti, Zucchini, Reihs, Lavitrano, Liu, Zatloukal, Pagni.

**Drafting of the manuscript:** L'Imperio, Wulczyn, Zucchini, Liu, Steiner, Zatloukal, Pagni.

**Critical revision of the manuscript for important intellectual content:** L'Imperio, Plass, Müller, Tamini, Gianotti, Reihs, Corrado, Webster, Peng, Chen, Lavitrano, Liu, Steiner, Zatloukal, Pagni.

**Statistical analysis:** Wulczyn, Plass, Müller, Reihs, Liu, Steiner.

**Obtained funding:** Corrado, Webster, Peng, Chen, Zatloukal.

**Administrative, technical, or material support:** L'Imperio, Plass, Reihs, Corrado, Peng, Chen, Liu, Steiner.

**Supervision:** Plass, Gianotti, Corrado, Peng, Chen, Lavitrano, Liu, Steiner, Zatloukal, Pagni.

**Conflict of Interest Disclosures:** Dr Müller reported receiving grants from Google LLC during the conduct of the study. Dr Corrado reported a patent pending for machine learning applications in digital pathology; and receiving grants from and holding stock in Google LLC. Dr Chen reported holding stock in Google LLC. Dr Liu reported holding stock in Google LLC during the conduct of the study. Dr Steiner reported holding stock in Google LLC during the conduct of the study and having a patent pending for machine learning applications in digital pathology (not directly involving this work). Dr Zatloukal reported receiving grant No. 676550 from European Union's Horizon 2020 research and innovation programme (H2020-INFRADEV ADOPT BBMRI-ERIC), Google LLC, Kapsch BusinessCom AG, and the Austrian Federal Ministry of Education, Science and Research Hochschulraumstrukturmittel outside the submitted work and being founder and CEO of Zatloukal Innovations GmbH. No other disclosures were reported.

**Funding/Support:** Parts of this work were supported by Google LLC.

**Role of the Funder/Sponsor:** Google LLC was involved in the design and conduct of the study; analysis and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. Google LLC and the authors affiliated with Google LLC (Mr Wulczyn and Drs Corrado, Peng, Chen, Liu, and Steiner) did not have access to complete data and did not play a direct role in the collection or management of data.

**Data Sharing Statement:** See Supplement 2.

**Additional Contributions:** Rory Sayres, PhD, and Michael Howell, MD (both with Google Health), provided feedback on the manuscript. They were not compensated for their contributions.

**Additional Information:** Drs L'Imperio, Zucchini, and Pagni and Prof Lavitrano contributed to this work during their Italian MUR Dipartimento di Eccellenza Impact Medicine 2023-2027 project.

## REFERENCES

1. Schneider NI, Langner C. Prognostic stratification of colorectal cancer patients: current perspectives. *Cancer Manag Res*. 2014;6:291-300.
2. Martin B, Schäfer E, Jakubowicz E, et al. Interobserver variability in the H&E-based assessment of tumor budding in pT3/4 colon cancer: does it affect the prognostic relevance? *Virchows Arch*. 2018;473(2):189-197. doi:10.1007/s00428-018-2341-1
3. Harris EI, Lewin DN, Wang HL, et al. Lymphovascular invasion in colorectal cancer: an interobserver variability study. *Am J Surg Pathol*. 2008;32(12):1816-1821. doi:10.1097/PAS.0b013e3181816083
4. Amin MB, Greene FL, Edge SB, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin*. 2017;67(2):93-99. doi:10.3322/caac.21388
5. Puppa G, Sonzogni A, Colombari R, Pelosi G. TNM staging system of colorectal carcinoma: a critical appraisal of challenging issues. *Arch Pathol Lab Med*. 2010;134(6):837-852. doi:10.5858/134.6.837
6. Gray R, Barnwell J, McConkey C, Hills RK, Williams NS, Kerr DJ; Quasar Collaborative Group. Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. *Lancet*. 2007;370(9604):2020-2029. doi:10.1016/S0140-6736(07)61866-2
7. Kannarkatt J, Joseph J, Kurniali PC, Al-Janadi A, Hrinchenko B. Adjuvant chemotherapy for stage II colon cancer: a clinical dilemma. *J Oncol Pract*. 2017;13(4):233-241. doi:10.1200/JOP.2016.017210
8. Yothers G, O'Connell MJ, Allegra CJ, et al. Oxaliplatin as adjuvant therapy for colon cancer: updated results of NSABP C-07 trial, including survival and subset analyses. *J Clin Oncol*. 2011;29(28):3768-3774. doi:10.1200/JCO.2011.36.4539
9. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol*. 2019;16(11):703-715. doi:10.1038/s41571-019-0252-y



10. Wulczyn E, Steiner DF, Moran M, et al. Interpretable survival prediction for colorectal cancer using deep learning. *NPJ Digit Med*. 2021;4(1):71. doi:10.1038/s41746-021-00427-2
11. Skrede O-J, De Raedt S, Kleppe A, et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet*. 2020;395(10221):350-360. doi:10.1016/S0140-6736(19)32998-8
12. Dawson H, Galuppini F, Träger P, et al. Validation of the International Tumor Budding Consensus Conference 2016 recommendations on tumor budding in stage I-IV colorectal cancer. *Hum Pathol*. 2019;85:145-151. doi:10.1016/j.humpath.2018.10.023
13. Mitrovic B, Schaeffer DF, Riddell RH, Kirsch R. Tumor budding in colorectal carcinoma: time to take notice. *Mod Pathol*. 2012;25(10):1315-1325. doi:10.1038/modpathol.2012.94
14. Dawson H, Kirsch R, Driman DK, Messenger DE, Assarzadegan N, Riddell RH. Optimizing the detection of venous invasion in colorectal cancer: the Ontario, Canada, experience and beyond. *Front Oncol*. 2015;4:354. doi:10.3389/fonc.2014.00354
15. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov*. 2019;9(4):e1312. doi:10.1002/widm.1312
16. Brockmoeller S, Echle A, Ghaffari Laleh N, et al. Deep learning identifies inflamed fat as a risk factor for lymph node metastasis in early colorectal cancer. *J Pathol*. 2022;256(3):269-281. doi:10.1002/path.5831
17. Di Franco S, Bianca P, Sardina DS, et al. Adipose stem cell niche reprograms the colorectal cancer stem cell metastatic machinery. *Nat Commun*. 2021;12(1):5006. doi:10.1038/s41467-021-25333-9
18. Tabuso M, Homer-Vanniasinkam S, Adya R, Arasaradnam RP. Role of tissue microenvironment resident adipocytes in colon cancer. *World J Gastroenterol*. 2017;23(32):5829-5835. doi:10.3748/wjg.v23.i32.5829
19. Koelzer VH, Lugli A. The tumor border configuration of colorectal cancer as a histomorphological prognostic indicator. *Front Oncol*. 2014;4:29. doi:10.3389/fonc.2014.00029

#### SUPPLEMENT 1.

**eTable 1.** TAF Scoring per Pathologist

**eTable 2.** Expanded Multivariable Analysis for Association With Clinical Outcome

**eTable 3.** Expanded Univariable Analysis for Association With Clinical Outcome

**eFigure 1.** Example 1 Used for Illustrating the Tumor Adipose Feature

**eFigure 2.** Example 2 Used for Illustrating the Tumor Adipose Feature

**eFigure 3.** Example 3 Used for Illustrating the Tumor Adipose Feature

**eFigure 4.** Example 4 Used for Illustrating the Tumor Adipose Feature

**eFigure 5.** Forest Plots for Expanded Multivariable Analysis Corresponding to eTable 2

#### SUPPLEMENT 2.

**Data Sharing Statement**