Mod Pathol 36 (2023) 100335

MODERN PATHOLOGY

USCAP UNITED STATES AND CANADIAN ACADEMY OF PATHOLOGY
Creating a Better Pathologist

Journal homepage: https://modernpathology.org/

Research Article

Pathologist Computer-Aided Diagnostic Scoring of Tumor Cell Fraction: A Swiss National Study

Ana Leni Frei[a,*], Raphaël Oberson[a], Elias Baumann[a], Aurel Perren[a], Rainer Grobholz[b], Alessandro Lugli[a], Heather Dawson[a], Christian Abbet[c], Ibai Lertxundi[a], Stefan Reinhard[a], Aart Mookhoek[a], Johann Feichtinger[d], Rossella Sarro[e], Gallus Gadient[f], Corina Dommann-Scherrer[g], Jessica Barizzi[e], Sabina Berezowska[h], Katharina Glatz[i], Susanne Dertinger[j], Yara Banz[a], Rene Schoenegg[k], Laura Rubbia-Brandt[l], Achim Fleischmann[m], Guenter Saile[n], Pierre Mainil-Varlet[o], Ruggero Biral[p], Luca Giudici[e], Alex Soltermann[p], Audrey Baur Chaubert[q], Sylvia Stadlmann[r], Joachim Diebold[s], Kristof Egervari[l], Charles Bénière[t], Francesca Saro[u], Andrew Janowczyk[v,w,x], Inti Zlobec[a]

[a] *Institute for Tissue Medicine and Pathology, University of Bern, Bern, Switzerland;* [b] *Medical Faculty University of Zurich, Institute of Pathology, Cantonal Hospital Aarau, Aarau, Switzerland;* [c] *Signal Processing Laboratory 5, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland;* [d] *Labor Dr. Ulm GmbH, Wien, Austria;* [e] *Istituto Cantonale di Patologia, Ente ospedaliero cantonale (EOC), Locarno, Switzerland;* [f] *Institute of Pathology, Medica, Zürich, Switzerland;* [g] *Institute of Pathology, Cantonal Hospital Winterthur, Winterthur, Switzerland;* [h] *Institute of Pathology, Lausanne University Hospital, Lausanne, Switzerland;* [i] *Institut of Pathology, University Hospital Basel, Basel, Switzerland;* [j] *Institute of Pathology, Landeskrankenhaus Feldkirch, Feldkirch, Austria;* [k] *Institute of Pathology, Cantonal Hospital St. Gallen, St. Gallen, Switzerland;* [l] *Department of Pathology and Immunology, Geneva University Hospital, Genève, Switzerland;* [m] *Institute of Pathology, Cantonal Hospital Thurgau, Münsterlingen, Switzerland;* [n] *Labor team w AG, St. Gallen, Switzerland;* [o] *Unilabs Bern, Bern, Switzerland;* [p] *Pathologie Länggasse, Ittigen, Switzerland;* [q] *FMH Pathology, Pathology Department of SYNLAB Switzerland SA, Lausanne, Switzerland;* [r] *Institute of Pathology, Cantonal Hospital Baden, Baden, Switzerland;* [s] *Institute of Pathology, Cantonal Hospital Luzern, Luzern, Switzerland;* [t] *Aurigen, Centre de Pathologie, Lausanne Switzerland;* [u] *Institute of Pathology and Molecular Pathology, University Hospital Zürich, Zürich, Switzerland;* [v] *Department of Biomedical Engineering, Emory University, Atlanta, Georgia;* [w] *Department of Oncology, Division of Precision Oncology, University Hospital of Geneva, Geneva, Switzerland;* [x] *Department of Clinical Pathology, Division of Clinical Pathology, University Hospital of Geneva, Geneva, Switzerland*

ARTICLE INFO

Article history:
Received 31 May 2023
Revised 25 August 2023
Accepted 15 September 2023
Available online 22 September 2023

Keywords:
computer-aided diagnostic tool
digital pathology
interobserver variability
tumor cell fraction
artificial intelligence
pathology

ABSTRACT

Tumor cell fraction (TCF) estimation is a common clinical task with well-established large inter-observer variability. It thus provides an ideal test bed to evaluate potential impacts of employing a tumor cell fraction computer-aided diagnostic (TCFCAD) tool to support pathologists' evaluation. During a National Slide Seminar event, pathologists (n = 69) were asked to visually estimate TCF in 10 regions of interest (ROIs) from hematoxylin and eosin colorectal cancer images intentionally curated for diverse tissue compositions, cellularity, and stain intensities. Next, they re-evaluated the same ROIs while being provided a TCFCAD-created overlay highlighting predicted tumor vs non-tumor cells, together with the corresponding TCF percentage. Participants also reported confidence levels in their assessments using a 5-tier scale, indicating no confidence to high confidence, respectively. The TCF ground truth (GT) was defined by manual cell-counting by experts. When assisted, interobserver variability significantly decreased, showing estimates converging to the GT. This improvement remained even when TCFCAD predictions deviated slightly from the GT. The standard deviation (SD) of the estimated TCF to the GT across ROIs was 9.9% vs 5.8% with TCFCAD (*P*

These authors contributed equally: Andrew Janowczyk and Inti Zlobec.

* Corresponding author.
E-mail addresses: ana.frei@unibe.ch (A.L. Frei), inti.zlobec@unibe.ch (I. Zlobec).

https://doi.org/10.1016/j.modpat.2023.100335

< .0001). The intraclass correlation coefficient increased from 0.8 to 0.93 (95% CI, 0.65-0.93 vs 0.86-0.98), and pathologists stated feeling more confident when aided (3.67 ± 0.81 vs 4.17 ± 0.82 with the computer-aided diagnostic [CAD] tool). TCFCAD estimation support demonstrated improved scoring accuracy, interpathologist agreement, and scoring confidence. Interestingly, pathologists also expressed more willingness to use such a CAD tool at the end of the survey, highlighting the importance of training/education to increase adoption of CAD systems.

Introduction

Digitization of glass tissue slides in pathology laboratories is becoming increasingly more common, making histopathology images amenable to computer-aided diagnostic (CAD) support tools. An expert consensus panel consisting of 24 international members from various digital pathology societies on artificial intelligence (AI) in diagnostic pathology agreed that by 2030, the integration of AI will take over certain tasks, including quantitative assessment of diagnostic features, and lead to greater diagnostic accuracy.[1] The use of AI tools will bring more objectivity to certain diagnoses with high interobserver variability and generally bring more standardization, increased quality, and completeness to diagnostic reports. They identified targets for CAD solutions by 2030 and anticipated the adoption of AI tools for increasingly precise automatic measurements (eg, tumor area estimation) to be routinely used by 2030.

The desired increased accuracy provided by CAD tools is likely linked with the rise of precision medicine and personalized therapy approaches. With oncology patients, these approaches often require molecular analysis of their solid tumors to identify mutational profiles used to select the corresponding treatment strategy.[2-4] However, if the minimum number of tumor cells required for a molecular test is not present, the test may yield incorrect results. Hence, before testing, the fraction of tumor cells present in the samples (ie, tumor cell fraction [TCF]) must be precisely assessed by pathologists to ensure accurate downstream analysis and optimal patient care.[5]

Unfortunately, TCF estimation has already been shown to be a task suffering from a high interobserver variability.[6-9] In a study by Smits et al,[6] pathologists were asked to evaluate 47 samples of lung tumors and to record the TCF in 1 of 11 possible categories (0%-5%, 6%-10%, 11%-20%, …, 91%-100%).[6] The corresponding ground truth (GT) was obtained by laboriously manually counting individual cells. In more than 33% of samples, the range of estimates deviated by at least 3 categories and in some instances, by up to 6 categories. In contradiction with the GT, 38% of samples were discarded by pathologists as containing too few tumor cells for downstream molecular testing, potentially impacting patient care. This discordance is likely due to the challenging nature of this task due to: (1) lack of clear guidelines and training sessions on how to assess the TCF; (2) heterogeneity of neoplastic cell presentation, morphology, and size; and (3) the presence of nonneoplastic cells in the tumor regions confounding visual qualitative assessments.[7,8,10]

In this study, we evaluate the potential impact of a CAD solution on the reliability and reproducibility of TCF estimation. On the occasion of a Swiss national "Slide Seminar" for pathologists, a survey was conducted by the Swiss Digital Pathology Consortium whereby respondents were asked to score 10 hematoxylin and eosin (H&E) digital images and evaluate the TCF before and after receiving the support of a CAD tool termed tumor cell fraction computer-aided diagnostic (TCFCAD). Additionally, they reported their level of confidence in CAD solutions in general, as well as their level of confidence in their individual TCF scores. Before and after the survey, questions about the use of AI and CAD systems in medical settings were asked to observe their changing openness to adopting such tools in their clinical practice.

Methods

Tumor Cell Fraction

Our study followed the consensus recommendations established by a European Delphi survey.[10] TCF content (ie, neoplastic cell percentage) was defined as the count of neoplastic cells divided by the total amount of cells in a predefined area, resulting in the percentage of tumor cells present in the area. Because manual counting of individual cells is too time consuming to be performed routinely in clinical practice, TCF estimation is instead performed by pathologists via visual assessment of the percentage of neoplastic cells vs all nonneoplastic cells in a defined region. Notably, this estimation is differentiated from tumor area percentage as tumor cells may be much larger in size than the surrounding nonneoplastic cells, potentially leading to an overestimation of TCF.

Data Set

The data set consisted of 10 regions of interest (ROIs) of 1000 × 1000 pixels (250 × 250 μm) from H&E-stained colorectal cancer cases scanned at 40× (0.25 μm/pixels) using a P1000 slide scanner (3DHistech, Budapest, Hungary). These ROIs were chosen to exhibit a variety of tumor-to-stroma cellularity levels. ROIs were also selected to contain varying levels of necrosis, immune infiltrates, desmoplastic stroma, and stain variation. Taken together, these confounders are likely to make TCF estimation more difficult and, thus, a potential source of interobserver variability.[8] Necrosis regions were regarded as nontumor.

For each ROI, a cell-by-cell manual GT count was obtained (R.O.) under the supervision of highly experienced gastrointestinal (GI) pathologists (A.L. and H.D.). This process involved the laborious centroid annotation for each individual cell present in the images using QuPath.[11] R.O. was first trained by H.D. for identification of neoplastic cells in the ROIs. All the annotations were then reviewed by A.L..

No patient information was used in this study. ROIs were selected from completely anonymized whole slide images (WSI) that cannot be traced back to patients.

Tumor Cell Fraction Computer-Aided Diagnostic Tool Description

TCFCAD consisted of merging the output of 2 algorithms. To obtain cell centroid coordinates, ROIs were segmented using a pretrained Stardist model.[12] Next, tumor tissue was retrieved using the U-Net model, which was trained for this specific segmentation task. U-Net architecture is composed of a contracting path to capture image features followed by an expanding path to generate the corresponding segmentation map.[13,14] U-Net was first presented for semantic segmentation for biomedical image analysis and has shown great performances for other tasks such as road detection from satellite images and is now widely used for a multitude of segmentation tasks in different fields, including histopathology.[15-17] Cells were classified as tumor cells if their centroid's coordinates were inside the tumor tissue segmentation result. All cells outside the tumor mask were considered nontumor cells (see Fig. 1). Lastly, TCF was computed as the number of predicted tumor cells divided by the total number of cells detected in the ROI, resulting in a TCF score percentage.

Some groups have already proposed digital solutions to estimate TCF in the lung, breast, and colon tissue. Many approaches were based on immunohistochemistry to specifically stain tumor cells,[18] limiting their applicability to typical clinical workflows that focus on H&E. Approaches for H&E images were either semiautomated methods[9,19] or if fully automated, the code was not publicly available,[20-22] necessitating the usage of our previously validated pipeline for fully automated generation of TCF estimates from H&E ROIs. We believe that our approach most closely mirrors a real-world clinically deployed tool, where images would be automatically processed to generate TCF estimations. A fully automated pipeline prevents potential human biases found in semiautomated methods while reducing processing time.

It is also important to note that the goal of the study was not the validation of the TCFCAD model but to observe the impact of presenting pathologists with TCF predictions on their scoring.

Swiss Digital Pathology Consortium Tumor Cell Fraction Estimation Survey Setup

Our survey was conducted on April 30, 2022, during an online Swiss Slide Seminar from the Swiss Society of Pathology hosted by the Swiss Digital Pathology Consortium. The questionnaire was created using SurveyMonkey, and the link to the survey was shared with attending participants. Prior to starting the experiment, the questionnaire was introduced to the participants during a short oral presentation with clear explanations and guidelines (see Supplementary Fig. S1). Before evaluating the ROIs, participants were asked a series of questions related to their current position (resident or attending pathologist), years of experience, and their domain of subspecialization. Moreover, they were asked to denote their familiarity with AI (ranging from "not familiar at all" to "very familiar") and their feelings toward the use of AI as a support for medical practice. The survey questionnaire can be found in Supplementary Figure S1.

The image review component took place in 2 stages: unaided and aided.

In the unaided stage, participants were sequentially shown the ROIs and asked to enter their TCF estimates as a percentage. Additionally, they reported how confident they felt about their score on a scale of 5—(1) very confident, (2) somewhat confident, (3) neither confident nor unconfident, (4) not so confident, and (5) not confident at all.

In the aided stage, participants were presented with the same H&E ROIs again along with the corresponding TCFCAD predicted percentage and an overlay highlighting the cells classified as tumor and nontumor (Supplementary Fig. S1). Participants were again asked to enter their estimation, along with their confidence regarding the score, as well as a rating on how useful the TCFCAD prediction was to their final assigned value. This was on a 5-level scoring—(1) very helpful, (2) somewhat helpful, (3) neither helpful nor not helpful, (4) not so helpful, and (5) not helpful at all.

Throughout these 2 stages, participants were also provided with links to access the high-resolution version of each ROI as these had to be compressed to fulfill SurveyMonkey requirements. This process attempted to mirror clinical practice and the level of information available to pathologists during routine estimation. For each question of the survey, values could not be modified once inputted, and previous pages of the survey could not be navigated back to after values were saved.

At the end of the experiment, participants were asked to provide feedback to determine which TCFCAD output was most

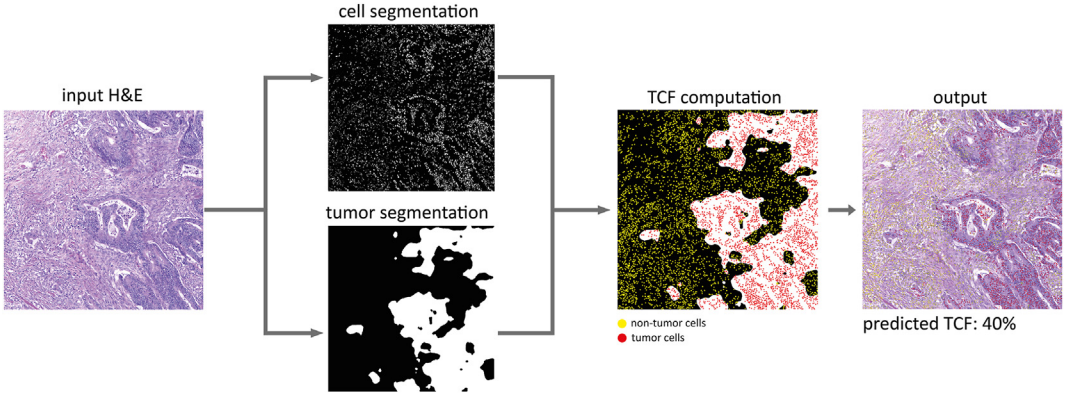

Figure 1.
Illustration of the tumor cell fraction (TCF) computer-aided diagnostic prediction model. A hematoxylin and eosin (H&E) region of interest (ROI) at 0.25 μm/pixel is given as input. Cell segmentation and tumor tissue segmentation are performed in parallel on the input ROI. The cell segmentation results are merged with the tissue tumor segmentation to compute the TCF. Cells inside the tumor tissue are highlighted as red dots and classified as tumor cells, whereas cells outside the tumor segmentation are classified as nontumor cells and highlighted as yellow dots. The classified cells are overlaid on the input H&E ROI for visualization, and the TCF score is computed as the amount of tumor cells divided by the total amount of cells in the ROI.

helpful (score, overlay, and else), whether TCFCAD is useful (yes or no), and if they would be willing to use TCFCAD if it existed (yes or no). Lastly, 2 open-ended questions asked the pathologists how TCFCAD could be improved and for comments and suggestions about the survey.

Collected data were cleaned and quality-controlled to remove incomplete or erroneous answers (eg, participants entering the same TCF score for each image). This resulted in a set of 69 complete responses that formed the data set for analysis in this study.

Statistical Analysis

Descriptive statistics (ie, minimum, maximum, mean, SDs, and variance) were computed for the distribution of all TCF estimation scores for each ROI. Discordance between GT values and participants' estimates, as well as between individual unaided and aided scores, was analyzed. Under the assumption of paired non-normally distributed samples, Wilcoxon signed-rank test was applied to test differences in means and SDs. When testing confidence levels (categorical variables), Spearman rank test was applied. For unpaired samples, the Mann-Whitney U test was used. These tests were conducted using Python's SciPy library.[23] The interclass correlation coefficients (ICCs) were used to determine the intraobserver variability using the intraclass correlation function from Python's pingouin library.[24]

Results

Interobserver Variability and Statistics

The 10 H&E ROIs included in the survey, along with the TCFCAD predictions, annotation overlays, and GT annotations can be found in Figure 2 and Supplementary Figure S2.

A complete set of scores was available for 69 participants, with the descriptive statistics summarized in the Table. Marked variability in the scores attributed to an individual ROI was observed. For example, ROI 1 was assigned TCF scores as low as 20% and as high as 90% by some respondents, with an average TCF of 55%.

Overall, without CAD support, the average difference in the range of values across all 10 ROIs was 52% ± 12.1%, which decreased significantly to 28% ± 8.4% after use of TCFCAD, indicating a convergence of scores toward the mean. Moreover, the scores also significantly converged toward the GT, reflected by a decrease of the SD of estimated TCF to the GT from 9.9% ± 1.4% without CAD to 5.8% ± 1.7% with CAD (for all TCF scores over the 10 ROIs, $P = 3.8 \times 10^{-20}$), and the average mean moved from 5.5% to 4.1% toward the GT. Across all ROIs, on average, 69% (477/690) of the scores were changed (99.4% (474/477) were changed by at least 5%) when using TCFCAD, and of these, 71% (341/477), on average, were closer to the GT with assistance. The ICC increased from 0.8 to 0.93 (95% CI, 0.65-0.93 vs 0.86-0.98), reflecting an improved interobserver agreement. These results are further highlighted in Figure 3, which depicts the distribution of the TCF estimates for each ROI with and without TCFCAD together with the corresponding TCFCAD predictions and GT values.

Despite the strong interobserver variability of individual scores, the mean values of TCF with and without TCFCAD ($P = .8748$) and the values with and without TCFCAD vs the GT values ($P = .6219$ and $P = .7415$, respectively) were not statistically different. However, the mean TCF values do not reflect the score variability since the positive and negative deviations balance each other out. Over the 10 ROIs, the positive and negative deviations to the GT without TCFCAD were 7.5% ± 5% and 9% ± 5.5% and with TCFCAD were 3.7% ± 1.9% and 5.9% ± 3.7%, respectively. Strong linear correlations between average GT TCF scores with ($r^2 = 0.978$) and without TCFCAD ($r^2 = 0.912$) were observed.

Impact of Imprecise Tumor Cell Fraction Computer-Aided Diagnostic Predictions

We can observe that TCFCAD predictions were less than or equal to the corresponding GT in all cases. This is likely due to the observation that the Stardist model employed for cell segmentation occasionally oversegments elongated cells, such as fibroblast, into more than 1 cell. This yields a slight overestimation of stroma cellularity and, thus, a lower TCF prediction (see Fig. 4). However, between the 10 ROIs, the average variance between TCFCAD predictions and GT was only 4.1% ± 3.1%, suggesting that the TCFCAD predictions were overall very accurate and did not suffer from this potential oversegmentation.

In some ROIs, for example, ROIs 2 and 9, the TCFCAD predictions were >6% different than the GT. Here, the tumor segmentation algorithm appears to have missed parts of the tissue, with an observable overdetection of cells in the stroma (see Supplementary Fig. S3). On these ROIs, the mean TCF score from pathologists was slightly closer to the GT before using TCFCAD (1.1% decrease in mean to GT for both ROIs, see the Table), suggesting an imprecise TCFCAD value negatively impacts accuracy. That said, the average SD of the scores still significantly converged to the GT after using TCFCAD ($P = 1.42 \times 10^{-6}$ and $P = .0083$ for ROIs 2 and 9, respectively), suggesting that even with this muted performance, the overall value of using TCFCAD remained strong. In short, pathologists aided by TCFCAD were more concordant than TCFCAD alone or pathologists alone. This highlights the benefit of "collaboration" between pathologists and CAD tools even when CAD predictions may be imprecise.

Confidence in Results Before and After Computer-Aided Diagnostic Support

Figure 5 shows the self-evaluated confidence in the TCF scoring for all 10 ROIs without and with TCFCAD. In the unaided stage of the study, the participants felt "somewhat confident" in 60% of the cases (413/690). After TCFCAD support was given, the pathologists claimed feeling more confident in 80% of the assessments (551/690), where they stated feeling "very more confident" in 51% of the cases (281/551) and "somewhat more confident" in 49% (270/551). Moreover, when pathologists stated feeling less confident during the first stage, their scoring confidence increased with the TCFCAD support. Supplementary Table S1 depicts the distribution of confidence level after TCFCAD for cases where pathologists felt unconfident during the unaided stage. Of these 74 assessments, 15 were rated as very more confident, 53 were rated as somewhat more confident, 6 were rated as neither more confident nor less confident, and none were rated as less confident when being provided with TCFCAD support. An increase in accuracy accompanied this increase in confidence. With TCFCAD support, pathologists were significantly closer in their estimations to the GT ($P = .0014$, Supplementary Figure S4). This indicates that TCFCAD

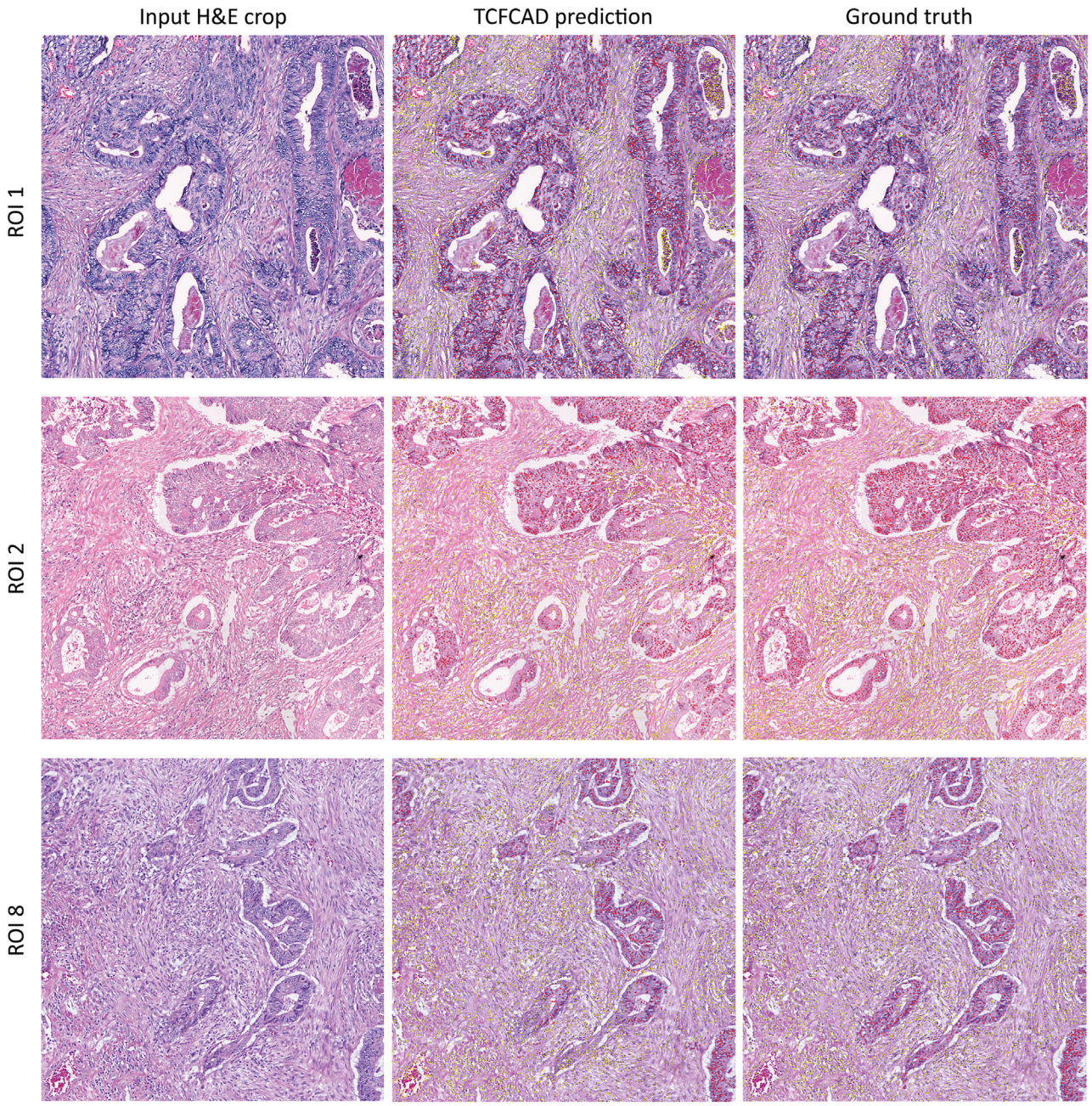

Figure 2.
Example hematoxylin and eosin crops with tumor cell fraction computer-aided diagnostic (TCFCAD) predictions and ground truth annotations. The first column shows 3 of the 10 hematoxylin and eosin digital colorectal cancer regions of interest (ROIs) that were scored during the survey, the second column shows the corresponding overlayed TCFCAD predictions, and the third column shows the ground truth. Red dots corresponds to tumor cells and yellow dots nontumor cells. All regions of interest, corresponding predictions, and ground truth images can be found in the Supplementary Figure S2.

helped to increase pathologists' confidence in their assessments while shifting the scores toward the GT.

Perceptions in Clinical Adoption of Artificial Intelligence Before and After the Survey

To explore changes in the perception of CAD systems before and after the experiment, the study survey began by asking participants about their familiarity with AI and their confidence in AI algorithms for clinical adoption. These results were then contrasted after the survey, by asking participants whether they found TCFCAD useful and whether they would use such a tool if available (Fig. 6). The majority of respondents reported being not at all or not very familiar (38/69; 50.88%) with AI, while 22% (15/69) reported that they were somewhat familiar, and only 2.94% (2/69) were very familiar with AI. In terms of confidence in AI for clinical adoption, 17.39% (12/69) felt very confident and 42% (29/69) felt somewhat confident, while 13.04% (9/69) and 2.94% (2/69) felt not very and not at all confident, respectively. The association between familiarity with AI and confidence in its clinical adoption was statistically significant ($P = .045$).

Table
Description of the tumor cell fraction scoring with and without tumor cell fraction computer-aided diagnostic support

Scoring regimen		ROIs										Average
		1	2	3	4	5	6	7	8	9	10	
Pathologists without TCFCAD	Mean (%)	55.0	41.3	49.0	43.5	36.2	67.5	21.3	14.7	75.8	45.2	
	SD to mean	13.1	10.3	11.5	11.2	15.1	10.9	7.5	6.5	10.2	12.4	10.8 ± 2.4
	SD to GT	12.4	10.6	11.5	9.2	11.9	8.4	8.4	8.4	8.6	9.8	9.9 ± 1.4
	Min (%)	20	20	20	15	10	35	5	5	50	20	
	Max (%)	90	75	80	70	70	90	40	35	90	80	
	Variance (max to min) (%)	70	55	60	55	60	55	35	30	40	60	52 ± 12.1
Pathologists with TCFCAD	Mean (%)	58.9	40.2	49.1	37.7	29.5	70.6	23.9	16.9	69.1	40.4	
	SD to mean	6.0	5.1	5.4	6.2	6.7	4.9	4.8	3.5	6.1	5.9	5.5 ± 1.1
	SD to GT	5.4	9.5	8.3	4.8	4.9	3.0	5.1	5.1	5.9	6.4	5.8 ± 1.7
	Min (%)	40	20	30	20	15	60	10	10	50	20	
	Max (%)	80	50	60	50	50	80	30	20	85	50	
	Variance (max to min) (%)	40	30	30	30	35	20	20	10	35	30	28 ± 8.4
GT	%	63 (4439/7060)	49 (3663/7458)	57 (4164/7319)	39 (3242/8286)	31 (2822/9044)	71 (4793/6727)	28 (1486/5222)	22 (1332/5941)	73 (5397/7497)	44 (2437/5499)	
TCFCAD prediction	%	62	39	54	34	25	71	27	18	65	42	
	SD to GT	1	10	4	5	6	0	1	4	8	2	4.1 ± 3.1

For each region of interest, the mean (%), minimum (%), and maximum (%) of the scores were reported. The variance (%) together with the SD of the scores regarding the mean and the ground truth were computed. The tumor cell fraction computer-aided diagnostic predictions and the ground truth values are also reported.
GT, ground truth; Max, maximum; Min, minimum; ROI, region of interest; TCFCAD, tumor cell fraction computer-aided diagnostic.

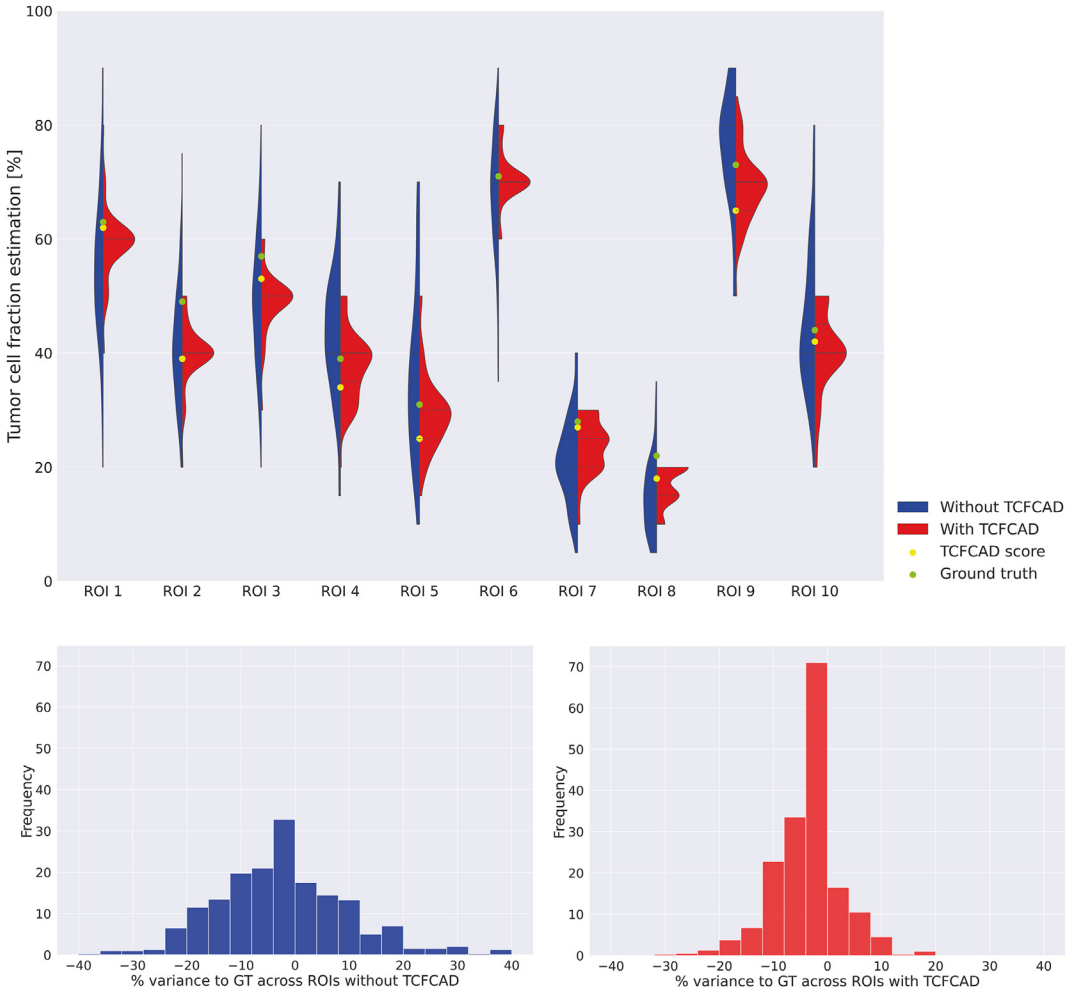

Figure 3.
Tumor cell fraction estimation distribution with and without tumor cell fraction computer-aided diagnostic (TCFCAD) support. Top violin plot shows the scoring distribution by pathologists for the 10 regions of interest (ROIs) with and without TCFCAD assistance (blue and red, respectively). Yellow dots represent the TCFCAD score that was given to the pathologists and green dots corresponds to the tumor cell fraction ground truth (GT). Bottom bar plots represent the mean variance across the 10 ROIs to the GT value. Upon TCFCAD assistance, the mean variance to GT decreased from 9.9% to 5.8%.

The overwhelming majority of respondents (65/68; 95.6%) answered "yes" when asked whether TCFCAD was useful, and 63/69 (91.3%) said that they would use this tool if available. Interestingly, of the 11 respondents who answered that they were not at all or not so confident in the clinical adoption of AI before the survey, 8 would use the TCFCAD if made available ($P = .0047$). This indicates the importance of training and educational sessions with pathologists to familiarize them with CAD systems in order to increase motivation regarding adoption of clinical CAD tools.

Tumor Cell Fraction Computer-Aided Diagnostic Scored as Not Helpful

Interestingly, in most cases where pathologists rated TCFCAD as not helpful at all or not so helpful, they entered a different score from their first evaluation (23/32; 72%). Indeed, a significant convergence of the scores toward the GT ($P = 0.028$) was observed. The SD from the GT was 9.4% without TCFCAD vs 6.6% with TCFCAD. This suggests that TCFCAD also helped in cases even when participants thought the support was not helpful. The details of score variations for each image can be seen in the Supplementary Figure S5.

Role of Specialty and Years of Experience

Respondents' field of diagnostic specialty and years of experience are reported in Supplementary Figure S6. These 2 characteristics represent possible biases in our results and are thus investigated below.

No significant difference was observed when comparing scoring distribution to GT from GI pathologists and non-GI pathologists without and with TCFCAD assistance ($P = .26$ and $P = .39$, respectively). Also, GI pathologists did not report feeling significantly more confident in their assessments than non-GI pathologists ($P = .19$), with the score variation without and with TCFCAD support between GI and non-GI pathologists showing no statistical difference ($P = .13$). This suggests that pathologists benefited from the CAD tool independent of their field of diagnostic specialty.

When looking at the impact of years of experience on TCF scoring estimates, no significant difference between the

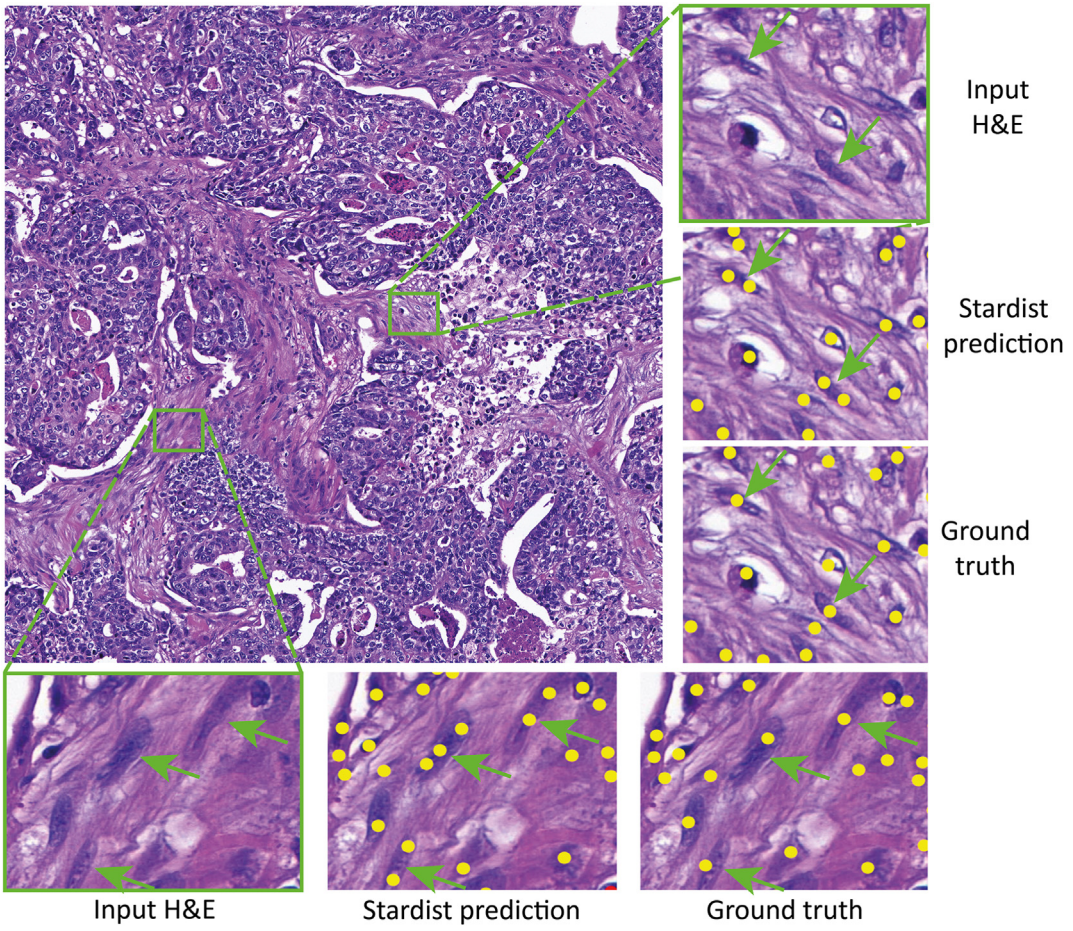

Figure 4.
Stardist oversegmentation of elongated cell. Two regions from region of interest 9, highlighted by the green boxes, were extracted for comparison between Stardist segmentation and ground truth annotations. Cells segmented by Stardist are highlighted by a yellow dot on their centroid in the Stardist prediction panels. The ground truth panels also show the manual annotations for the same regions as yellow dots. Green arrows indicate elongated single cells that were segmented as more than on cell by Stardist.

different experience categories (all $P > .05$) was observed (see Supplementary Figure S6). Furthermore, the score changes when provided TCFCAD support also showed no significant difference between experience categories (all $P > .05$). This might result from the lack of training and feedback of pathologists when performing TCF estimations in clinical routine, as showed by Mikubo et al.[7] However, pathologists with more than 20 years of experience felt significantly more confident in their scoring ($P = .00022$). This suggests that increased years of experience appear to increase reported confidence. Regardless, all pathologists appear to have benefited from TCFCAD support, regardless of their experience level, which is in line with the results from the study by Kazdal et al.[9]

Feedbacks of the Participants

One major concern outlined by participants was the ability to compare the initial TCF scores (without TCFCAD) to the TCFCAD predictions. Indeed, during the second stage of the survey, pathologists were not shown their previous scores. Some respondents stated they would feel more confident in employing TCFCAD if their scores were similar to their initial estimate. Moreover, by comparing the scores, they would be able to observe their own subjectivity and receive immediate personalized feedback.

Another comment concerned the overlay, suggesting the transparency of the dots, their colors, and size could be improved for visualization. Highlighting the tumor boundary with lines, tumor infiltrates, empty spaces, and blood-filled vessels were also mentioned as useful possible improvements. The combination of scores and overlays was appreciated to reduce the decision-energy cost of assigning a score while increasing TCFCAD trust by seeing the actual cell-level prediction. Importantly, these feedbacks are all presently feasible for implementation.

Discussion

TCF is an important parameter for both diagnostics and research and, therefore, was selected as a use case for our survey. In molecular pathology diagnostics, the proportion of tumor cells within a sample has an influence on the identification of mutations and the clinical interpretation of next-generation sequencing results.[25] Unfortunately, these TCF estimates are only available by pathologist visual assessment of glass slides, subjecting them to intraobserver and interobserver variability. As a result, if the TCF is overestimated, an inconclusive and yet expensive, tissue-destructive, molecular test may be performed

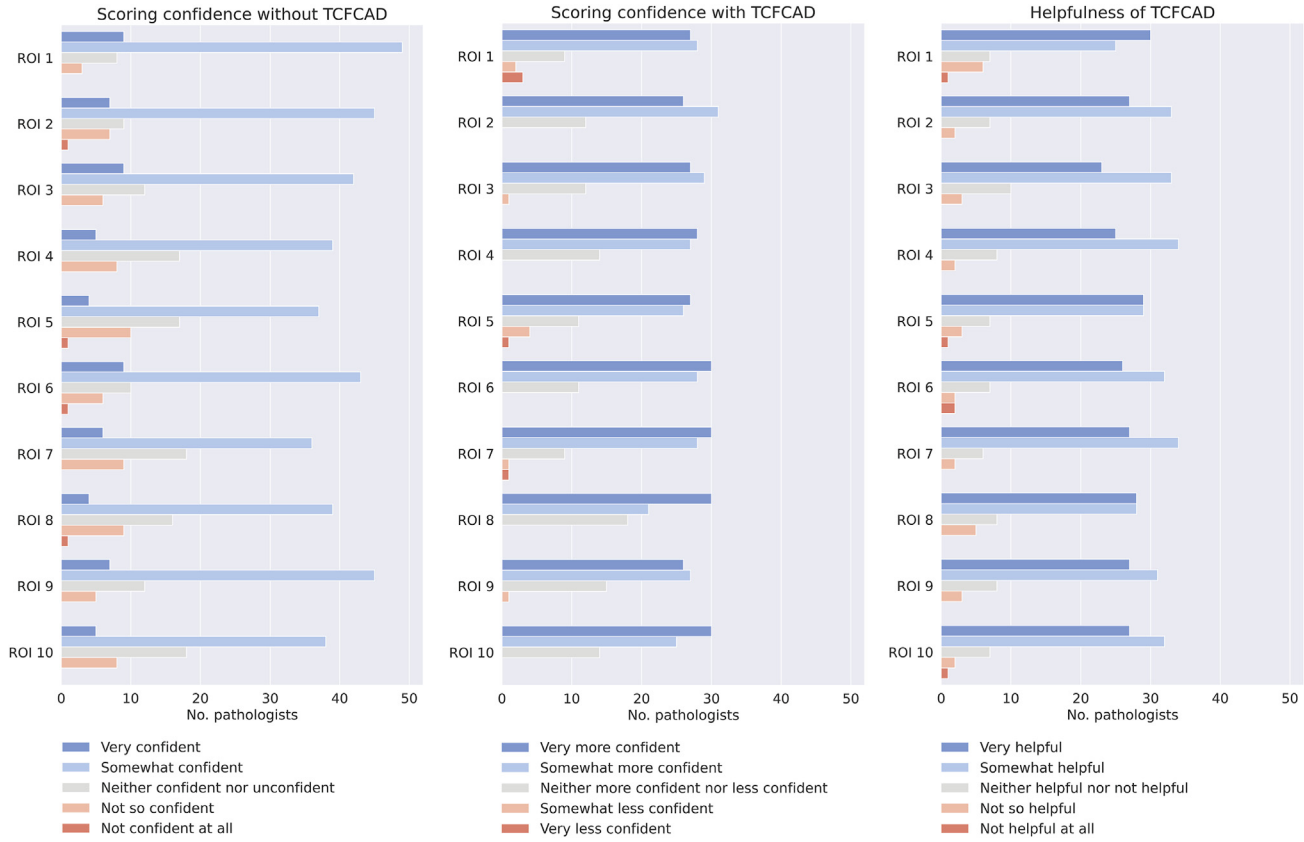

Figure 5.
Confidence level in scoring and tumor cell fraction computer-aided diagnostic (TCFCAD) support. Left bar plot: confidence scores assigned by pathologists during the first stage of the study, when assessing tumor cell fraction without the computer-aided diagnostic support. Middle bar plot: increase or decrease in scoring confidence upon TCFCAD assistance during the second stage of the study. Right bar plot: scoring of the level of helpfulness of the TCFCAD support. ROI, region of interest.

leading to potential false negative results and mutations might be missed.[6] Correct TCF scores could further help to improve understanding of intratumoral heterogeneity and subclonal populations to get a better evaluation of driver vs passenger mutations and optimal patient care.[10]

In this national survey on the use of CAD for TCF evaluation, our findings indicate that CAD solutions can have a marked impact on reducing interobserver variability between pathologists and encourage convergence of the TCF estimates toward a laboriously manually established GT. No significant difference in TCF means was observed with and without the TCFCAD support tool. This was expected considering that the scores close to the average were not anticipated to exhibit notable variations. Conversely, severely deviating scores were expected to converge toward the average, leading to increased scoring agreement, as shown by the increase in ICC. That said, we cannot exclude that by increasing the number of participants, the change in means might become significant when looking at means differences smaller or equal to 1%. However, we have to keep in mind that a minor change in average TCF score is not the focus point of this work; the primary clinical benefit of using a TCFCAD support tool would be to rectify scores exhibiting large variance, thereby guiding them to converge toward the mean and increase interobserver agreement.

The use of CAD tools has been shown to benefit pathologists for other tasks as well, such as the analysis of HER2 by immunohistochemistry and fluorescence in situ hybridization,[26] Ki-67 scoring in breast[27] and pancreatic neuroendocrine tumors,[28] as well as prostate cancer detection and Gleason grading.[29] Recently, Sakamoto et al[30] studied the benefits of a collaborative workflow between pathologists and CAD for the TCF estimation in lung adenocarcinoma. Consistent with our observations, the adjusted scores with the CAD support were more accurate. Such TCFCAD tools could be further expanded to other clinical settings, such as estimating the number of vital tumor cells for fluorescence in situ hybridization analysis or tumor fraction scoring in the context of neoadjuvant treatments.

Our results are in line with these findings. We also observed that although TCFCAD was not always 100% concordant with GT counts, respondents still asserted that TCFCAD was helpful in their scoring decisions and their scores accordingly still converged toward the GT. Importantly, the estimations by pathologists aided by the TCFCAD tool were more accurate than those of pathologists alone or TCFCAD alone. These collaborations between pathologists and TCFCAD improved performance, and a large majority of the participants felt more confident in their scoring when aided. This gain in confidence resulted from the TCFCAD, presenting a TCF score together with cell-level prediction overlays that were reported by pathologists to reduce the decision-energy cost when performing quantitative assessments, as stated by participants' feedbacks. At the end of the survey, almost all participants would be willing to use such a tool for clinical practice. Interestingly, the confidence in the use of CAD before and after the survey was significantly different, in line with the idea that education or training with CAD increases the confidence in its adoption.

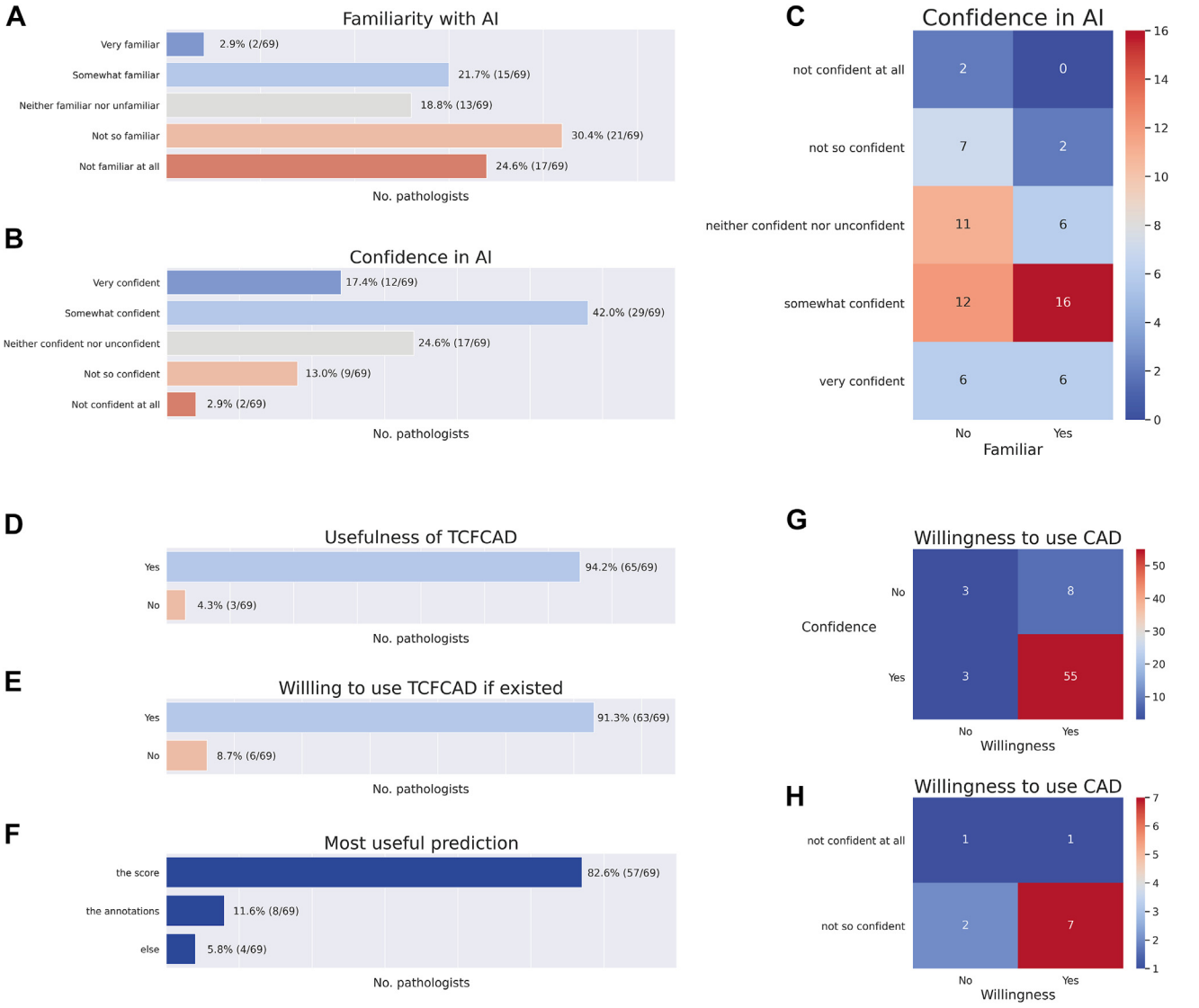

Figure 6.
Familiarity and confidence in using tumor cell fraction computer-aided diagnostic (TCFCAD) systems in medical practice. Pathologists' responses when answering to (A) "How familiar are you with artificial intelligence (AI)?" and (B) "In general, how do you feel about using AI help in medical practice?." (C) The correlation between familiarity ("Yes" represents "very familiar," "somewhat familiar," and "neither familiar nor unfamiliar," and "No" represents "not so familiar" and "not familiar at all") and confidence level of using AI in medical practice. Pathologists' answers for the questions (D) "Do you think such an AI model for tumor cell fraction prediction could be useful?," (E) "Would you be willing to use such an AI help if it existed?," and (F) "What AI output did you find more helpful?." (G) Heat map of the correlation between confidence level in AI for medical use (see panel B) and willingness to use TCFCAD (panel E). (H) Heat map of the correlation between pathologists who did not feel confident in the adoption of medical computer-aided diagnostic (CAD) tools and their willingness to use TCFCAD.

Our study benefits from a large number of respondents, which we believe accurately reflects the current experience of pathologists with CAD solutions in Switzerland.[31] There are, however, some limitations to our survey. One limitation is the absence of a wash-out period between the 2 stages. Although pathologists first scored the 10 ROIs unaided, the recall of their previous scores may have influenced their reported TCFCAD-aided values. Despite this, the intraobserver differences between the first and second stages are large, as seen by the increased ICCs, and thus, it is unlikely that this had a large impact on our study. Another limitation was that due to time restrictions, the TCF assessments took place at an ROI level as opposed to a WSI level. However, the ROI-based observations made in this study likely represent an idealized lower-bound discordance as compared to assessments at the WSI level, where nonuniform region selection will further increase disparity between readers. Finally, ROIs scored only originated from colorectal cancer cases, imparting potential experiential bias, but no statistical confounding for specialty or years of experience was witnessed. To further analyze experiential bias, one could record how often participants assess TCF in their daily clinical routine and how the TCFCAD impacts their assessments, which should be explored in future work. Although this study was not aimed at validating the TCFCAD model performance but was aimed at evaluating the impact of a CAD support tool on pathologists, the TCFCAD predictions were accurate and reflected a life-like scenario by presenting pathologists' predictions occasionally slightly deviating from the GT. Future assessment studies should measure potential efficiency improvements in unaided vs aided scoring, which were not feasible here given the nature of the slide seminar organization.

To conclude, CAD solutions seem to have a positive impact not only on reducing interobserver and intraobserver variability but also on the level of confidence that pathologists may have in their scores and to reduce the decision-energy cost of quantitative assessments. Employing CAD sees scores converging toward the GT TCF, a benefit in terms of reliability and reproducibility, eventually likely leading to improved patient care. When the CAD slightly diverged from the GT, the rescoring by assisted pathologists was better than pathologists alone or CAD alone, highlighting the benefits of collaboration between pathologists and AI. The participants further demonstrated, however, that if this synergy is desired, additional targeted training with CAD systems is needed in order to increase pathologists' trust in CAD solutions.

Acknowledgments

A.J. reports support by the National Institutes of Health under awards U01CA239055, 1R01LM013864, 1U01DK133090, U01CA248226, and 2R01DK118431. The author also thank Irene Centeno Ramos for her input.

Author Contributions

A.J., I.Z., A.P., A.L.F., E.B., S.R., and A.M. contributed to the study design and conception. A.L.F., E.B., and C.A. contributed to the data selection. A.L.F. and I.L. contributed to the model development and predictions, and R.O., H.D., and A.L. contributed to the ground truth generation. J.F., S.R., G.G., C.D.-S., J.B., S.B., K.G., S.D., Y.B., R.S., L.R.-B., A.F., G.S., P.M.-V., R.B., L.G., A.S., A.B.C., S.S., J.D., K.E., C.B., and F.S. participated among others in the study. A.L.F., E.B., I.Z., and A.J. contributed to literature search. A.L.F., E.B., and I.Z. contributed to the data analysis. A.L.F., I.Z., A.J., and R.G. contributed to the interpretation of the findings. A.L.F. contributed to the figures. A.L.F., I.Z., and A.J. wrote the first draft of the manuscript. I.Z., A.J., and R.G. reviewed and edited the manuscript. I.Z. contributed to funding acquisition. All authors read and approved the final manuscript.

Data Availability

All data can be made available upon request.

Funding

This work was funded by the Swiss National Science Foundation (CRSII5 193832).

Declaration of Competing Interest

A. Janowczyk reports receiving consulting fees for Merck, Roche, and Lunaphore unrelated to this work. C. Bénière reports a disclosure to Franklin.ai unrelated to this work.

Supplementary Material

The online version contains supplementary material available at https://doi.org/10.1016/j.modpat.2023.100335

References

1. Berbís MA, McClintock DS, Bychkov A, et al. Computational pathology in 2030: a Delphi study forecasting the role of AI in pathology within the next decade. *EBioMedicine*. 2023;88:104427. https://doi.org/10.1016/j.ebiom.2022.104427
2. Au TH, Wang K, Stenehjem D, Garrido-Laguna I. Personalized and precision medicine: integrating genomics into treatment decisions in gastrointestinal malignancies. *J Gastrointest Oncol*. 2017;8(3):387–404. https://doi.org/10.21037/jgo.2017.01.04
3. VanderLaan PA, Rangachari D, Majid A, et al. Tumor biomarker testing in non-small-cell lung cancer: a decade of change. *Lung Cancer*. 2018;116:90–95. https://doi.org/10.1016/j.lungcan.2018.01.002
4. Loree JM, Pereira AAL, Lam M, et al. Classifying colorectal cancer by tumor location rather than sidedness highlights a continuum in mutation profiles and consensus molecular subtypes. *Clin Cancer Res*. 2018;24(5):1062–1072. https://doi.org/10.1158/1078-0432.CCR-17-2484
5. Dufraing K, De Hertogh G, Tack V, Keppens C, Dequeker EMC, van Krieken JH. External quality assessment identifies training needs to determine the neoplastic cell content for biomarker testing. *J Mol Diagn*. 2018;20(4):455–464. https://doi.org/10.1016/j.jmoldx.2018.03.003
6. Smits AJ, Kummer JA, De Bruin PC, et al. The estimation of tumor cell percentage for molecular testing by pathologists is not accurate. *Mod Pathol*. 2014;27(2):168–174. https://doi.org/10.1038/modpathol.2013.134
7. Mikubo M, Seto K, Kitamura A, et al. Calculating the tumor nuclei content for comprehensive cancer panel testing. *J Thorac Oncol*. 2020;15(1):130–137. https://doi.org/10.1016/j.jtho.2019.09.081
8. Lhermitte B, Egele C, Weingertner N, et al. Adequately defining tumor cell proportion in tissue samples for molecular testing improves interobserver reproducibility of its assessment. *Virchows Arch*. 2017;470(1):21–27. https://doi.org/10.1007/s00428-016-2042-6
9. Kazdal D, Rempel E, Oliveira C, et al. Conventional and semi-automatic histopathological analysis of tumor cell content for multigene sequencing of lung adenocarcinoma. *Transl Lung Cancer Res*. 2021;10(4):1666–1678. https://doi.org/10.21037/tlcr-20-1168
10. Dufraing K, van Krieken JH, De Hertogh G, et al. Neoplastic cell percentage estimation in tissue samples for molecular oncology: recommendations from a modified Delphi study. *Histopathology*. 2019;75(3):312–319. https://doi.org/10.1111/HIS.13891
11. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: open source software for digital pathology image analysis. *Sci Rep*. 2017;7(1):16878. https://doi.org/10.1038/s41598-017-17204-5
12. Schmidt U, Weigert M, Broaddus C, Myers G. Cell detection with star-convex polygons. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*. 2018:265–273. https://doi.org/10.1007/978-3-030-00934-2_30
13. Ronneberger O, Fischer P, Brox T. Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference Munich, Germany, October 5-9, 2015 proceedings, part III. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2015;9351(Cvd):12–20. https://doi.org/10.1007/978-3-319-24574-4
14. Jvanvugt. pytorch-unet. Published online 2019. Accessed October 1, 2021 https://github.com/jvanvugt/pytorch-unet
15. Bulten W, Bándi P, Hoven J, et al. Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard. *Sci Rep*. 2019;9(1):864. https://doi.org/10.1038/s41598-018-37257-4
16. Hermsen M, de Bel T, den Boer M, et al. Deep learning-based histopathologic assessment of kidney tissue. *J Am Soc Nephrol*. 2019;30(10):1968–1979. https://doi.org/10.1681/ASN.2019020144
17. Khan A, Brouwer N, Blank A, et al. Computer-assisted diagnosis of lymph node metastases in colorectal cancers using transfer learning with an ensemble model. *Mod Pathol*. 2023;36(5):100118. https://doi.org/10.1016/J.MODPAT.2023.100118
18. Hershkovitz T, Shenhav A, Sabo E, Ben-Izhak O, Hershkovitz D. Development of a computerized morphometry application for assessment of the tumor fraction in colon carcinoma tissue samples. *Appl Immunohistochem Mol Morphol*. 2013;21(1):54–58. https://doi.org/10.1097/PAI.0b013e318256d9bd
19. Greene C, O'Doherty E, Abdullahi Sidi F, et al. The potential of digital image analysis to determine tumor cell content in biobanked formalin-fixed, paraffin-embedded tissue samples. *Biopreserv Biobank*. 2021;19(4):324–331. https://doi.org/10.1089/bio.2020.0105
20. Azimi V, Chang YH, Thibault G, et al. Breast cancer histopathology image analysis pipeline for tumor purity estimation. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE; 2017:1137–1140. https://doi.org/10.1109/ISBI.2017.7950717
21. Hamilton PW, Wang Y, Boyd C, et al. Automated tumor analysis for molecular profiling in lung cancer. *Oncotarget*. 2015;6(29):27938–27952. https://doi.org/10.18632/oncotarget.4391
22. Rakhlin A, Tiulpin A, Shvets AA, Kalinin AA, Iglovikov VI, Nikolenko S, Breast tumor cellularity assessment using deep neural networks. *Proceedings of the*

2019 International Conference on Computer Vision Workshops. ICCVW; 2019: 371–380. https://doi.org/10.1109/ICCVW.2019.00048
23. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17(3):261–272. https://doi.org/10.1038/s41592-019-0686-2
24. Vallat R. Pingouin: statistics in Python. *J Open Source Softw*. 2018;3(31):1026. https://doi.org/10.21105/joss.01026
25. Siegmund SE, Manning DK, Davineni PK, Dong F. Deriving tumor purity from cancer next generation sequencing data: applications for quantitative ERBB2 (HER2) copy number analysis and germline inference of BRCA1 and BRCA2 mutations. *Mod Pathol*. 2022;35(10):1458–1467. https://doi.org/10.1038/s41379-022-01083-x
26. Palm C, Connolly CE, Masser R, et al. Determining HER2 status by Artificial Intelligence: an investigation of primary, metastatic, and HER2 low breast tumors. *Diagnostics (Basel, Switzerland)*. 2023;13(1):168. https://doi.org/10.3390/diagnostics13010168
27. Varga Z, Cassoly E, Li Q, et al. Standardization for Ki-67 assessment in moderately differentiated breast cancer. A retrospective analysis of the SAKK 28/12 study. *PLoS One*. 2015;10(4):e0123435. https://doi.org/10.1371/journal.pone.0123435
28. Luchini C, Pantanowitz L, Adsay V, et al. Ki-67 assessment of pancreatic neuroendocrine neoplasms: systematic review and meta-analysis of manual vs. digital pathology scoring. *Mod Pathol*. 2022;35(6):712–720. https://doi.org/10.1038/s41379-022-01055-1
29. Pantanowitz L, Quiroga-Garza GM, Bien L, et al. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study. *Lancet Digit Health*. 2020;2(8):e407–e416. https://doi.org/10.1016/S2589-7500(20)30159-X
30. Sakamoto T, Furukawa T, Pham HHN, et al. A collaborative workflow between pathologists and deep learning for the evaluation of tumour cellularity in lung adenocarcinoma. *Histopathology*. 2022;81(6):758–769. https://doi.org/10.1111/his.14779
31. Unternaehrer J, Grobholz R, Janowczyk A, Zlobec I. Current opinion, status and future development of digital pathology in Switzerland; Swiss Digital Pathology Consortium (SDiPath). *J Clin Pathol*. 2020;73(6):341–346. https://doi.org/10.1136/jclinpath-2019-206155