MODERN PATHOLOGY

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

Research Article

Fully Automated Tumor Bud Assessment in Hematoxylin and Eosin-Stained Whole Slide Images of Colorectal Cancer

John-Melle Bokhorst[a,*], Francesco Ciompi[a], Sonay Kus Öztürk[a], Ayse Selcen Oguz Erdogan[a], Michael Vieth[b], Heather Dawson[c], Richard Kirsch[d], Femke Simmer[a], Kieran Sheahan[e], Alessandro Lugli[b], Inti Zlobec[b], Jeroen van der Laak[a,f], Iris D. Nagtegaal[a]

[a] Department of Pathology, Radboud University Medical Center, Nijmegen, The Netherlands; [b] Klinikum of Pathology, Bayreuth University, Bayreuth, Germany; [c] Institute of Tissue Medicine and Pathology, University of Bern, Bern, Switzerland; [d] University of Toronto, Mount Sinai Hospital, Toronto, Canada; [e] Department of Pathology, St Vincent's Hospital, Dublin, Ireland; [f] Center for Medical Image Science and Visualization, Linköping University, Linköping, Sweden

ARTICLE INFO

ABSTRACT

Tumor budding (TB), the presence of single cells or small clusters of up to 4 tumor cells at the invasive front of colorectal cancer (CRC), is a proven risk factor for adverse outcomes. International definitions are necessary to reduce interobserver variability. According to the current international guidelines, hotspots at the invasive front should be counted in hematoxylin and eosin (H&E)-stained slides. This is time-consuming and prone to interobserver variability; therefore, there is a need for computer-aided diagnosis solutions. In this study, we report an artificial intelligence-based method for detecting TB in H&E-stained whole slide images. We propose a fully automated pipeline to identify the tumor border, detect tumor buds, characterize them based on the number of tumor cells, and produce a TB density map to identify the TB hotspot. The method outputs the TB count in the hotspot as a computational biomarker. We show that the proposed automated TB detection workflow performs on par with a panel of 5 pathologists at detecting tumor buds and that the hotspot-based TB count is an independent prognosticator in both the univariate and the multivariate analysis, validated on a cohort of n = 981 patients with CRC. Computer-aided detection of tumor buds based on deep learning can perform on par with expert pathologists for the detection and quantification of tumor buds in H&E-stained CRC histopathology slides, strongly facilitating the introduction of budding as an independent prognosticator in clinical routine and clinical trials.

Introduction

Colorectal cancer (CRC) is among the most common cancers, responsible for almost 2,000,000 new cases yearly and over 900,000 deaths globally.[1] Staging of CRC is based on invasion depth (T), number of involved lymph nodes (N), and the presence of distant metastases (M).[2] However, to accurately predict the disease progression of individual patients, this staging system is insufficient and additional biomarkers are required to decide on therapeutic strategies and prevent possible under- and overtreatment. Among the most promising histologic biomarkers in CRC are tumor budding (TB)[3] and poorly differentiated clusters (PDCs).[4] TB is defined as isolated single cells or small clusters of up to 4 tumor cells located at the invasive tumor front. In contrast,

* Corresponding author.
  E-mail address: john-melle.bokhorst@radboudumc.nl (J.-M. Bokhorst).

clusters of 5 or more tumor cells without gland formation are defined as PDCs. Both biomarkers are associated with lymph node and distant metastasis and increased patient mortality in CRC.[5]

International harmonization of TB scoring was achieved by the International Tumor Budding Consensus Conference (ITBCC) recommendations in 2016[3]: identification of the budding hotspot (measuring 0.785 mm^2) at the invasive front of the tumor and counting of the number of buds results in a score that can be classified as Bd1 (0-4 buds, low budding), Bd2 (5-9 buds, intermediate budding) or Bd3 (10 or more buds, high budding). Although pan-cytokeratin immunohistochemical (IHC) staining might be helpful in recognizing TB, the definition is based on hematoxylin and eosin (H&E) slide scoring. Despite these clear definitions, the interobserver agreement is moderate to substantial.[6]

In recent years, the application of machine learning approaches based on deep learning has increased the accuracy, reproducibility, and efficiency of histopathologic slide analysis.[7] When presented with sufficient high-quality annotated training data, convolutional neural networks, a special type of deep learning based on artificial neural networks, can learn complex histologic patterns from structured data such as medical images. Current applications of artificial intelligence in computational pathology include recognition, segmentation, and classification of morphologic structures in digital pathology whole slide images (WSIs), such as tumor (peripheral) regions or the detection and classification of several types of cells, including tumor cells, to be used as a base for the development of computational biomarkers.[8]

Despite advances in the field of current applications of artificial intelligence in computational pathology, computer-aided detection of tumor buds in H&E remains a challenging task, mostly due to the following: (1) the small size of the objects to detect (ie, the tumor buds), (2) a possible variation in phenotype due to an epithelial-to-mesenchymal transition,[9] and (3) the very heterogeneous composition of the microenvironment in which they are located. In addition, observer variability in identifying individual buds makes the collection of a large training set of TB and PDC challenging. To the best of our knowledge, to date, most of the (semi-)automatic TB detection methods work on immunohistochemically stained slides (reviewed in Studer et al[10]), and only a few H&E-based algorithms have been proposed,[11,12] either based on single-center slides,[11] limiting their generalizability, or not distinguishing TB from PDC.[12]

In this study, we developed a novel, fully automated pipeline for computer-aided detection and quantification of tumor buds in WSIs from H&E-stained CRC specimens. Following the recommendations of the ITBCC, our method automatically identifies the tumor-invasive front, detects tumor buds, and quantifies them in a hotspot-driven fashion. We compared the detection performance of our computer-aided detection model with a panel of 5 pathologists and analyzed the prognostic value of our computer-based hotspot-derived TB count on a large external independent cohort or CRC cases.

Materials

This section describes the data sets used in this work, namely (1) a single-center model development data set containing H&E slides with manual annotations used to train a deep learning algorithm to segment epithelial regions; (2) a multicentric technical validation data set used to validate the tumor-bud detection performance of the developed method; and (3) a multicentric clinical validation data set used to assess the prognostic value of the automated TB count produced by the presented method.
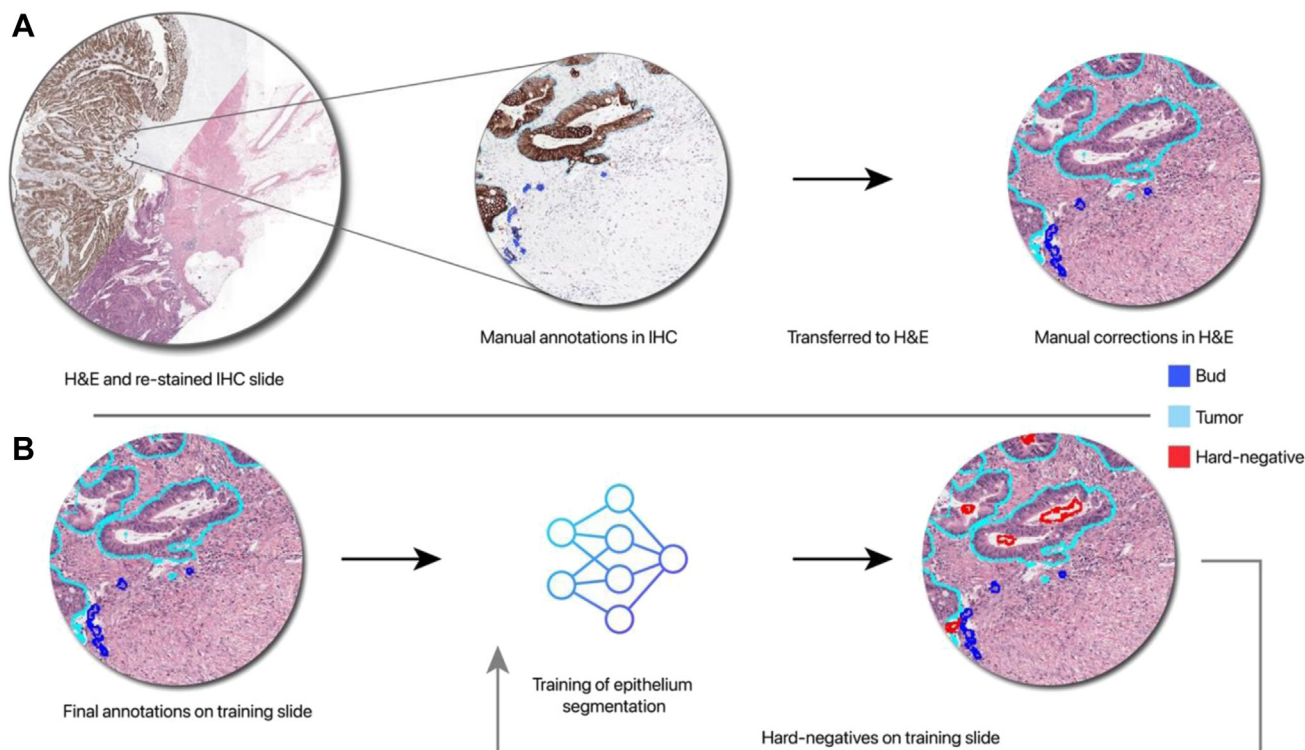
Model Development Data set

Between 1 and 5 slides from patients with CRC (n = 37) with the presence of TB reported during the initial sign-out were collected from the Radboud University Medical Center (Radboudumc), Nijmegen (The Netherlands). In total, 60 slides were included. Only slides where the invasive front was clearly visible were selected, stained with H&E, digitized, and subsequently restained with a cytokeratin 8-18 (CK8-18) IHC marker after scanning, according to our established protocol,[13] resulting in 2 versions of the same slide that were digitized using the Pannoramic P250 Flash II scanner (3D-Histech), at 40 X magnification (spatial resolution of 0.24 µm/px). On each slide, regions of interest (ROI) near the invasive front were manually selected and transferred to CK8-18 slides, where all positive cells were delineated. These annotations were transferred back to H&E slides and removed when no bud was visible in the H&E slide. Subsequently, the remaining nonepithelium pixels within the ROI were labeled as either necrosis or background. Annotations were made using the in-house developed open-source software ASAP (https://github.com/computationalpathologygroup/ASAP). The data set was randomly split into a training (n = 42) and validation set (n = 18). We will refer to this data set as Dev. An overview of the annotating procedure can be found in Figure 1A.

Technical Validation Data set

For technical validation, ie, the comparison between the detection performance of the proposed computer aided diagnosis (CAD) system and a panel of pathologists, we included a set of n = 15 WSIs selected from 4 different centers, including Dublin University Hospital (Dublin, Ireland), Bayreuth University Hospital (Bayreuth, Germany), Bern University, Institute of Pathology (Bern, Switzerland), and Mount Sinai Hospital (Toronto, Canada). Cases were selected based on the presence of high budding reported in the original diagnostic report. Glass slides were stained with H&E at the pathology laboratory of each center, therefore generating variation in the H&E staining, scanning, restaining with CK8-18 IHC marker, and rescanning at Radboudumc using a Pannoramic P250 Flash II scanner (3D-Histech), at 40 X magnification. Manual annotations were made as described above. We will refer to this data set as Val-t.

Clinical Validation Data Set

For clinical validation, ie, the analysis of the prognostic value of an automated TB score, n = 557 patients with CRC from Radboudumc and n = 568 patients with CRC from Mount Sinai Hospital, Toronto (Canada) were included. Based on the pathologists' assessment of multiple sections, the clinical validation set was established by visualizing a single slide per case in line with the ITBCC guidelines. Radboudumc slides were scanned using the same 3DHistech scanner used in Val-t; the Canadian cases were scanned with the Aperio AT2 scanner (Leica Biosystems). All slides were scanned at 40 X magnification, yielding a spatial resolution of 0.24 µm/px. In the Canadian cohort, a tumor bud count was established by an expert, according to ITBCC recommendations; n = 144 patients received neoadjuvant treatment and were

Figure 1.
Schematic overview of model development. (A) Within corresponding hematoxylin and eosin (H&E)-stained and cytokeratin (CK)-stained images, manual regions of interest (ROI) were selected. Within the ROIs, all CK-positive cells were annotated as either Bud or Tumor in the CK-stained image. These were subsequently checked in the H&E-stained tissue and removed if not visible or modified to follow the border in H&E. (B) Manual annotations were used to train the deep learning network, after which the network was applied to the training and validation set to identify hard-negatives and false-positives. In the next training session, we sampled more of these difficult objects. IHC, immunohistochemical.

therefore excluded from this study. As a result, n = 981 patients with CRC were included in the clinical validation for a total of n = 1125 WSIs.

We will refer to these sets in the clinical validation data set as cohorts C and D for the Canadian and Radboudumc cohorts, respectively.

Methods

Our deep learning pipeline for automated quantification of tumor buds in H&E WSIs consisted of 4 steps (Fig. 2). First, we determined the region of the invasive front of the tumor; second, we identified all the epithelial regions in the proximity of the invasive front via segmentation of epithelial structures at high resolution; third, we detected and characterized TB and PDC by detecting and counting neoplastic cell nuclei within the segmented epithelium compartments. Fourth, we computed the density of TBs in the entire invasive front and reported the TB count as the number of TBs in the hotspot. All steps of the proposed pipeline are described in the next sections.
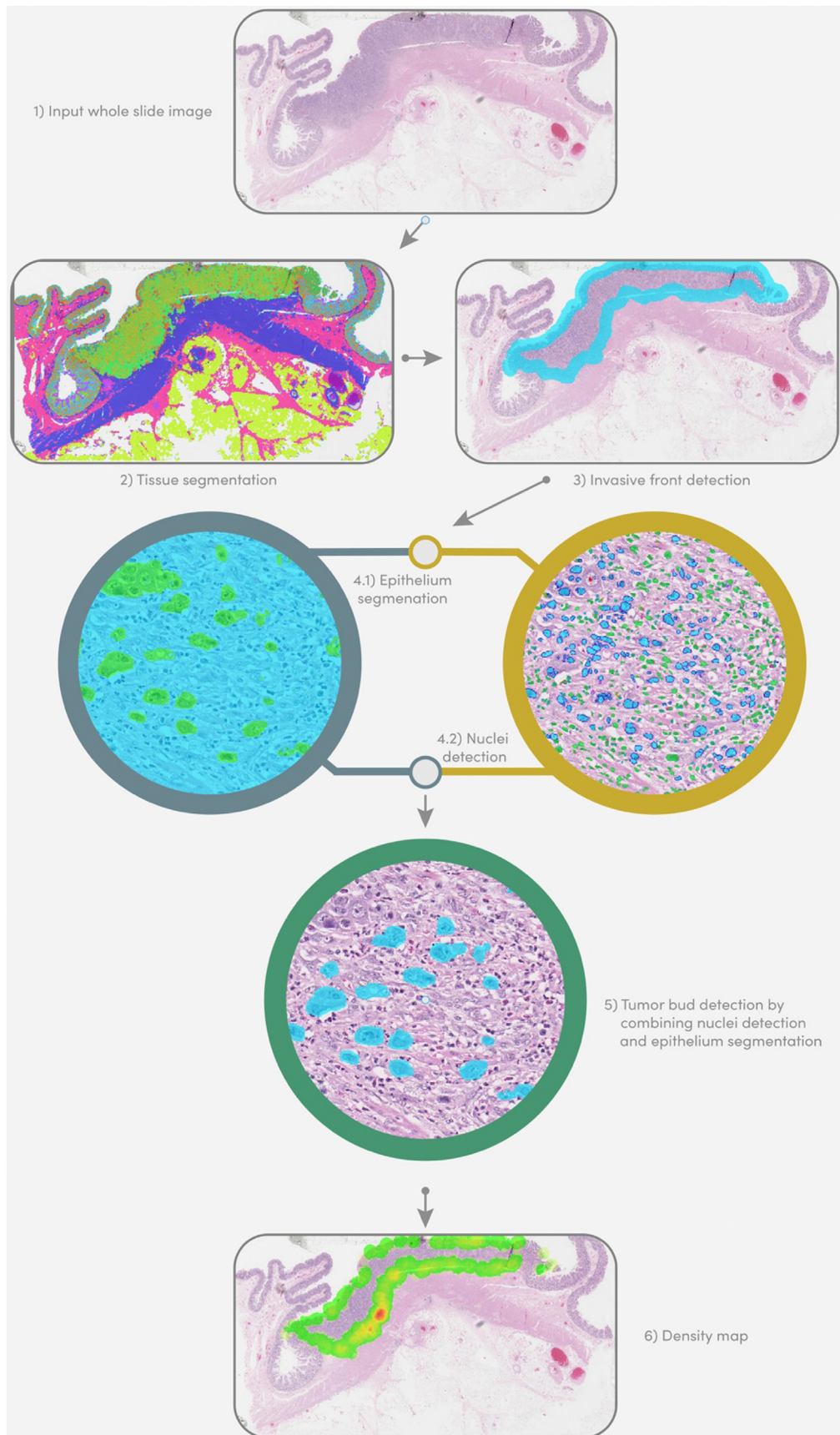
Step 1: Find the Tumor Border

In the ITBCC guidelines, tumor buds are quantified within the region of the invasive front of the tumor. For this reason, the first step of our approach consisted of the fully automated delineation of the tumor border, which we achieved in 3 steps.

First, we segmented the entire WSI into multiple morphologic categories by applying a multiclass tissue segmentation algorithm previously developed by our group.[14] In brief, a U-Net deep learning model was trained to segment n = 14 different morphologic regions in the entire WSI, namely (1) normal glands, (2) low-grade dysplasia, (3) high-grade dysplasia/tumor, (4) submucosal stroma, (5) desmoplastic stroma, (6) stroma lamina propria, (7) mucus, (8) necrosis and debris, (9) lymphocytes, (10) erythrocytes, (11) adipose tissue, (12) muscle, (13) nerve, and (14) background. The model was trained to operate at 10× magnification to guarantee a fast yet accurate interpretation of the slide morphology. Additional details about this model can be found in Bokhorst et al.[14]

Second, we identified the tumor bulk region by running a convex-hull algorithm on the tumor segmentation mask obtained by considering regions segmented as tumors by the multiclass algorithm. As a result, a polygon identifying the tumor bulk border was obtained. Third, following the approach proposed by Galon et al,[15] we defined the border as the region within 500 μm from the outer side of the tumor borderline.

Step 2: Segment Epithelial Regions

Once the tumor border region was defined, we restricted our focus to epithelial regions within the invasive front area. For this, we developed a binary segmentation model based on the U-Net architecture[16] with an EfficientNetB4[17] to segment epithelial regions in H&E at the maximum resolution available, ie, 0.24 μm/

1) Input whole slide image

2) Tissue segmentation

3) Invasive front detection

4.1) Epithelium segmenation

4.2) Nuclei detection

5) Tumor bud detection by combining nuclei detection and epithelium segmentation

6) Density map

px (40 X magnification) to differentiate between very small epithelial regions, potentially containing tumor buds, and non-epithelium particles, such as activated fibroblasts to construct a binary map of all epithelial regions, which was then intersected with the invasive tumor map to generate potential candidates of TB and PDC. An overview of the training procedure can be found in Supplementary 1.

Step 3: Find Tumor Buds Via Nuclei Detection

Within epithelial regions segmented inside the tumor-invasive front, we identified TBs and PDCs by detecting neoplastic cell nuclei. For this purpose, we used the nuclei segmentation method based on the HoVerNet model presented by Graham et al.[18] This network can segment and classify cell nuclei from 6 different cell types, including neoplastic cells. Here, we processed HoVerNet only in epithelial regions in the invasive front and only considered neoplastic cells within epithelial regions. In this way, we defined TBs as connected components of epithelial regions containing up to 4 neoplastic cells and the rest as PDCs. This approach allowed us to (1) use HoVerNet efficiently, solely running it in selected regions, guaranteeing a fast inference time; and (2) enable the analysis of the role of different number of tumor cells in TBs and their potential impact on prognosis. Step 4: Count the buds and find the hotspot.

Step 4: Find the Tumor Budding Hotspot

The final step of our method consisted of finding the hotspot of tumor buds in the entire slide and producing both the location and TB count in the hotspot. For this, we first move a circular region of 0.785 mm^2 over every location (within the tumor border). For each location, we considered the TBs inside the circle and associated the TB count with the central location of the circle within the invasive front. As a result, a TB density map within the invasive front was obtained (Fig. 3C).

Reader Study

In order to assess the performance of the proposed framework at the level of TB detection, we involved 5 pathologists in a reader study and asked them to manually mark TBs within predefined hotspots. The goal was to use the results of this study to assess human performance and interobserver variability at the specific task of TB detection and compare our CAD system with a panel of experienced pathologists (A.L., M.V., R.K., S.O., and A.E.), with an average of 10 years of experience in the field of TB in both clinical and research setting.

We asked each pathologist to manually mark each tumor bud in the manually predefined hotspots of each WSI in the Val-t data set (n = 15). The hotspots were selected in line with the ITBCC guidelines and contained a high number of tumor buds. Additionally, more difficult regions were also selected (ie, regions with high inflammation). Slides were uploaded to the Reader Study section of the grand-challenge.org platform, and pathologists

were shown both the H&E slide and the corresponding CK8-18 slide side by side.

Statistical Analysis

Statistical analyses for clinical validation were performed after entry in an anonymized database using R 4.1.2 (R Foundation for Statistical Computing[19]). Both patient cohorts were analyzed separately. The relationship of TB with the outcome (overall survival [OS] and disease-free survival [DFS]) was performed using Kaplan-Meier curves and Cox regression analysis. A complete overview of the patient information can be found in the Table.

Cut-off values for TB categories were adjusted because of the increased numbers compared to manual counting. To create 3 risk categories, we used the 30th and 60th percentiles from cohort C, which was subsequently validated in cohort D. A P value of .05 was considered significant.

The comparison of manual and automatic scoring was evaluated by Cox's regression analysis in combination with the Akaike Informational Criterion (AIC, based on the log-likelihood of the model). The AIC score is a number that describes how the models compare in terms of performance given a certain data set. When comparing 2 methods, this value helps to identify how identical the models are.

Results

The performance of the epithelium segmentation algorithm was assessed on data set Val-t using the Dice coefficient. We applied the algorithm to the 15 WSIs and calculated the Dice scores on the manually annotated regions. A Dice coefficient of 0.86 and 0.97 was found for the epithelium and nonepithelium classes, respectively. Some examples of the segmented epithelium are shown in Figure 3A.

Validation of the Tumor Bud Detection Pipeline

To evaluate the performance of the full detection pipeline, we compared the results to the manual detections made by the 5 observers. Because pathologists' annotations may vary, we calculated the recall, precision, and F1 scores of all observers (the algorithm considered observer 5) relative to each other in a one-vs-one fashion. The algorithm had an F1 value of 0.58 at maximum and recall of 0.95, but precision was suboptimal (Fig. 4), with, on average, an amount of detected TB 7 times as the other observers. An example of manual vs computer outputs can be found in Figure 3D.

In addition, we determined per hotspot how many TBs were missed by the algorithm (false negative number) compared with cumulative TB annotations of all 5, all 5 minus 1, all 5 minus 2, etc, observers. In this way, we found a recall percentage of 1.0, 0.90, and 0.84 for objects that have been TB annotated by at least 5, 4, and 3 observers, respectively.

Figure 2.

Overview of the workflow of the algorithm. First, the input whole slide image is segmented into 14 tissue types. Based on the detected tumor, a 500 μm invasive front border is drawn on both sides of the edge of the tumor. Within the invasive front, we identify each nucleus and segment all epithelium. After combining the 2 outputs, we can discard all objects with more than 4 nuclei. Based on the detected buds, we can create a density map to show the regions with the most tumor budding.
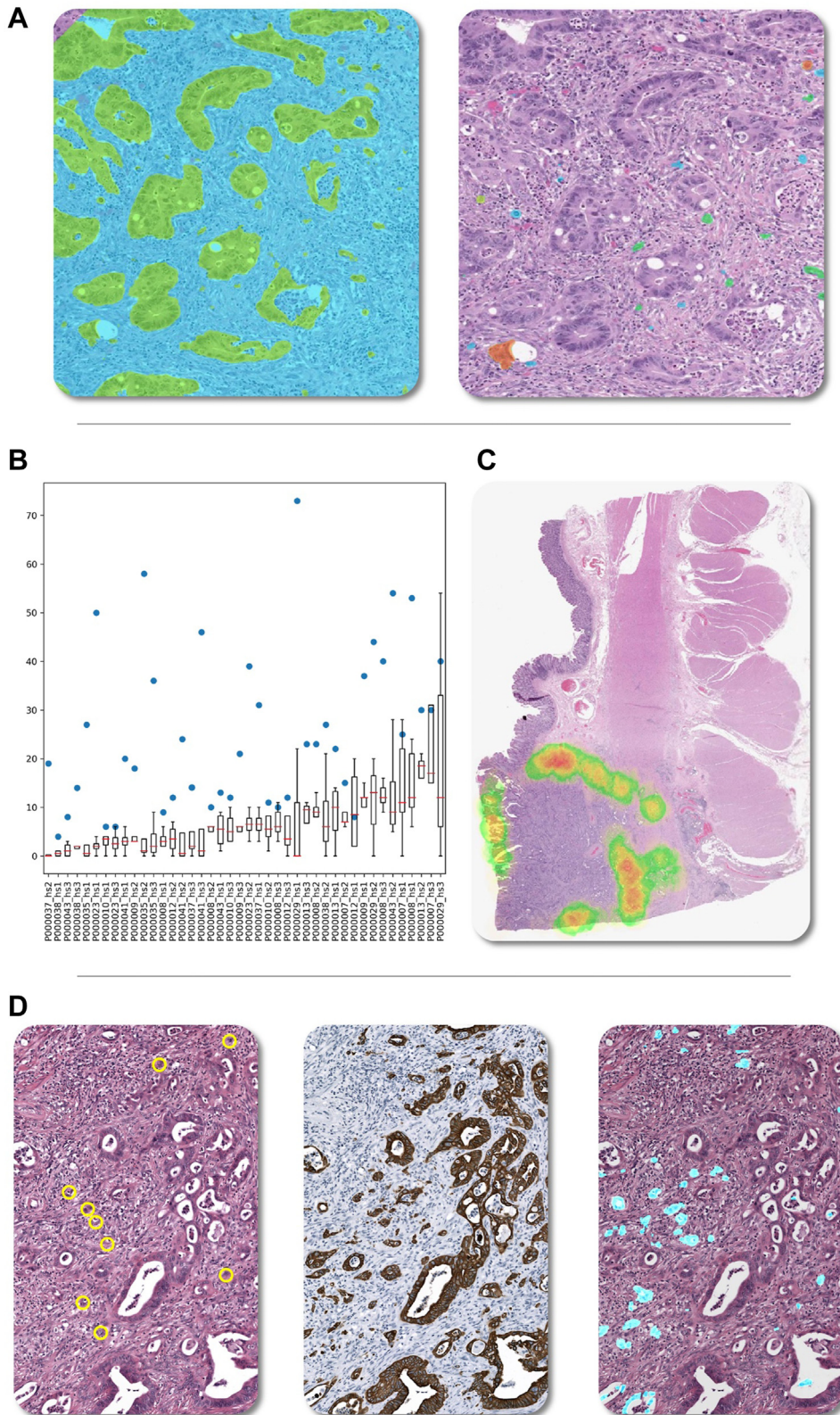
Figure 3.
(A) An example of the output of the epithelium segmentation network (left) and the combined results (right). (B) Manually (5 observers) and per algorithm obtained tumor bud (TB) counts per hotspot. The manually obtained TB range is shown from minimum to maximum with median values (marked in red), boxed by 25th and 75th percentile values. For the algorithm, detected numbers are shown as blue dots. The hotspots are sorted according to the degree of spread, from low (left) to high (right). (C) An example of a density heatmap generated based on automatic TB detections. (D) An example of technical validation results with hematoxylin and eosin-stained image with manual annotations (left), corresponding immunohistochemistry image (middle), and the segmented epithelium in blue (right).

Table
Clinicopathologic characteristics of cohorts C and D

Feature	Cohort C		Cohort D	
	n	%	n	%
Age, y				
<65	226	57	425	66
>65	169	43	219	34
Sex				
M	215	54	339	53
F	180	46	305	47
Invasion depth				
T1	27	7	18	3
T2	74	19	90	14
T3	218	55	402	62
T4	76	19	134	21
Nodal status				
N0	224	57	371	58
N+	171	43	273	42
Synchronous metastases				
M0	351	89	618	96
M1	44	11	26	4
Perineural invasion				
No	83	21	n/a	n/a
Yes	312	79	n/a	n/a
Grade				
Low grade	350	89	462	71
High grade	45	11	182	29
Recurrence				
No	318	81	502	78
Yes	77	19	142	22
Death				
No	292	74	414	64
Yes	103	26	230	36

The Size of Tumor Buds

For each patient group, we determined the size of TBs present (Fig. 5) and compared this with the manual TB annotations on the Val-t data set. In the algorithm groups, there was an over-representation of single-cell TB in all groups; however, in the manual scores, fewer single-cell TBs were scored. Two-cell TBs (45%) and 3-cell TBs (28%) were more common in the manual scoring. There was no effect of TB size on the outcome.

Hotspot Overlap

We visually evaluated cases to ensure that the automatically chosen hotspots were in regions that could be used for clinical assessment according to the ITBCC guidelines. We discovered that in roughly 4 out of 5 cases, the top 5 automatically identified hotspots were on the invasive front. In the other cases, the top 5 hotspots were incorrectly located. For example, they were not located in the invasive front but on the luminal side of the WSI.

*Clinical Validation: Prognostic Value of Tumor Budding*

The presence of low TB was associated with better DFS, both in the manual scoring (cohort C: HR, 2.1; 95% CI, 1.6-2.8) and the automatic scoring (cohort C: HR, 1.8; 95% CI, 1.2-2.6; cohort D: HR, 1.3; 95% CI, 1.0-1.6). In the multivariable analysis, after we corrected for age, sex, T-stage, N-stage, M-stage, histologic grade, and perineural invasion status for cohort C, the prognostic value of TB was still retained (cohort C manual: HR, 1.8; 95% CI, 1.3-2.5; cohort C automatic: HR, 1.6; 95% CI, 1.03-2.4; cohort D: HR, 1.2; 95% CI, 0.9-1.6) (Fig. 6).

Similarly, improved OS was observed in patients with low TB, both in the manual scoring (cohort C: HR, 2.0; 95% CI, 1.6-2.6) and the automatic scoring (cohort C: HR, 2.6; 95% CI, 1.7-3.8; cohort D: HR, 1.3; 95% CI, 1.1-1.5). In the multivariable analysis, after we corrected for age, sex, T-stage, N-stage, M-stage, and histologic grade, the prognostic value of TB was still retained (cohort C manual: HR, 1.3; 95% CI, 1.0-1.7; cohort C automatic: HR, 1.5; 95% CI, 1.0-2.3; cohort D: HR, 1.2; 95% CI, 1.0-2.3).

To compare the manual and automatic scoring systems, we used the AIC of cohort C. For both DFS and OS, the AIC was slightly lower in the manual group (790 vs 806 and 1145 vs 1148).

Discussion

In this study, we used a convolutional neural network for the segmentation of epithelium in WSIs in combination with a nuclei detection network to develop a new deep learning-based automated TB assessment tool for H&E-stained WSIs, which we compared to the ITBCC recommended method of TB assessment, performed by 5 gastrointestinal pathologists from different countries in an initial validation setting. We showed that our method is technically sound and clinically relevant. As such, our method is not yet optimized for clinical use and could be further improved for the time-efficient processing of slides. One of the main improvements that can be made is increasing the efficiency of the epithelium segmentation and nuclei detection part of the pipeline. For example, both networks are currently applied to the invasive front at the maximum image resolution, ie, 40 X magnification. Future work will be dedicated to investigating lowering the resolution (eg, to 20 X) for some parts of the pipeline
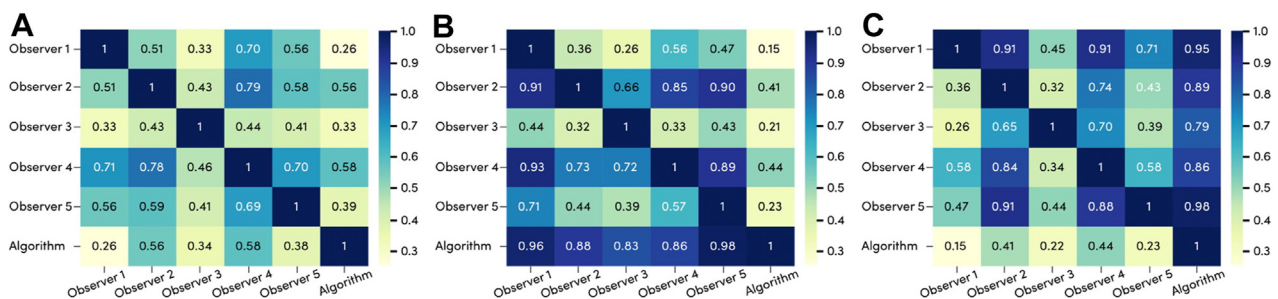
Figure 4.
F1 scores (A), precision (B), and recall (C) in a one-vs-one setup. For every score, one observer was set as a reference and compared with the other observers + algorithm.
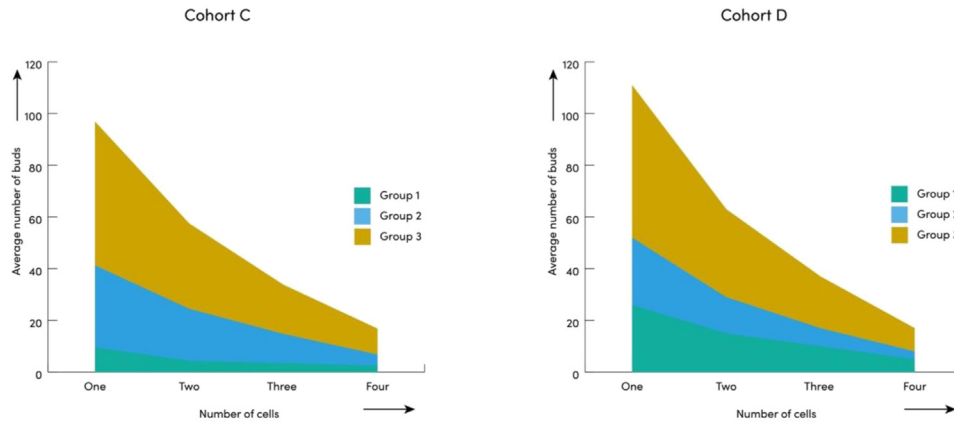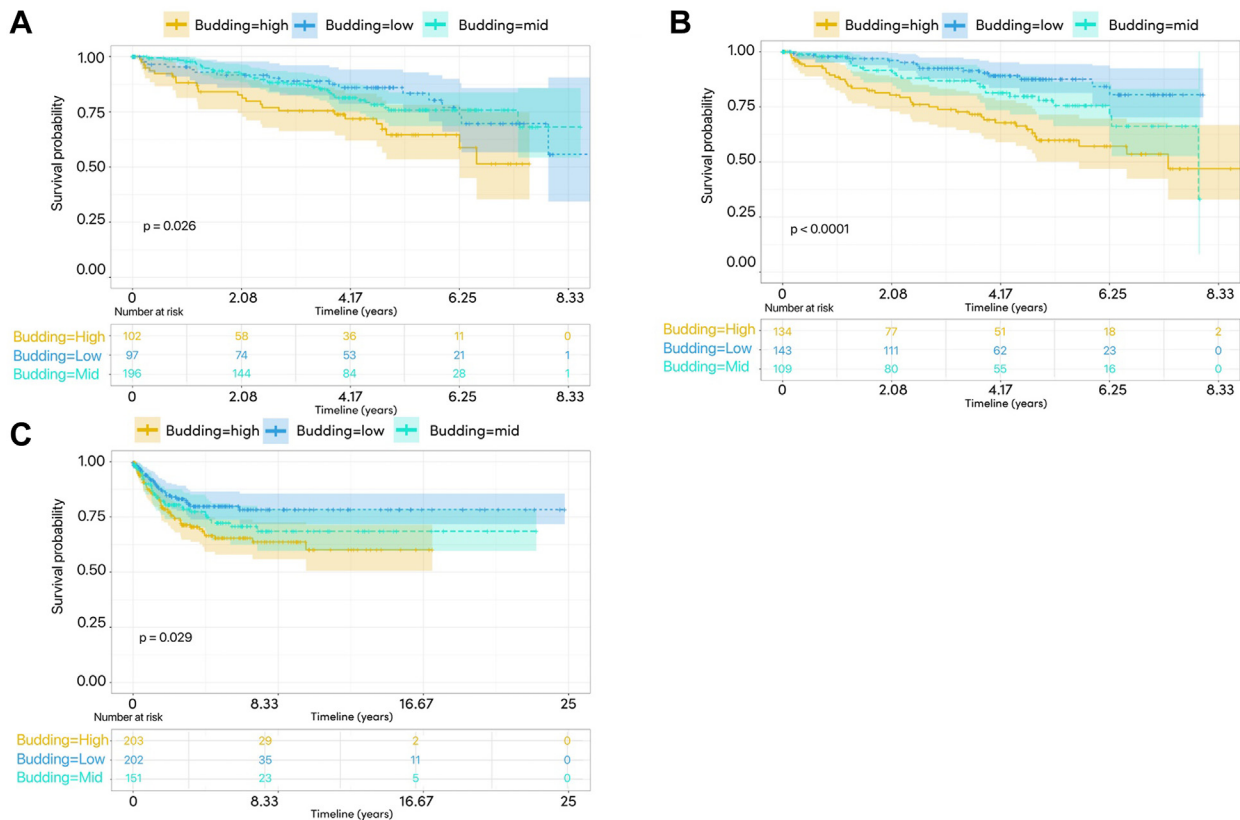
Figure 5.
The absolute number of 1, 2, 3, and 4 cell tumor buds found by the algorithm per patient risk group in cohorts C (left) and D (right).

(eg, epithelium segmentation) without reducing overall performance, which would reduce processing time.

The evident benefit of automatic assessment is standardization and minimization of interobserver variability.[10] Interobserver variability for TB is considerable,[5] even when single objects should be graded.[20] In the current study, variation was observed between the observers for both sensitivity/recall and specificity/precision. The algorithm was in line with the 2 observers who had the highest degree of interobserver agreement.

The determination of the invasive front remains a point of discussion between pathologists and, consequently, the application of the algorithm. In this study, we opted for a fully automated selection of the tumor border to reduce the possible observer variability in hotspot selection. Although in most of the cases the hotspot was in a region pathologists would use for tumor bud scoring, the automatically selected hotspot is, in some cases, not in the correct location. A practical approach to circumvent this issue is a visual assessment of TB heatmaps (as shown in Fig. 4C) to

Figure 6.
Kaplan-Meier curves on disease-free survival for (A) automatic tumor bud count in cohort C with 30th and 60th percentiles as cut-off values. (B) Manual tumor bud counts in cohort C with cut-off values according to the International Tumor Budding Consensus Conference. (C) Automatic tumor bud counts in cohort D, with cut-off values determined by the 30th and 60th percentiles of cohort C.

determine the adequacy of the selected hotspot and manual correction.

The ITBCC guidelines define 3 different budding grades based on H&E slides, which consist of 0-4 (BD1), 5-9 (BD2), and 10 or more (BD3) buds in a hotspot. Three categories are essential since, for different clinical questions, different cut-off levels are required. For example, high risk in pT1 CRC is characterized by BD2 and BD3. In contrast, only BD3 is an indication for adjuvant therapy in stage II CRC. We also assessed 3-tier classifications for TB (Supplemental data) with similar results to the 2-tiered classification (Fig. 6). The formal TB score, according to the guidelines, is manual for H&E slides. For IHC, other cut-off levels likely need to be defined, such as for automatic scoring. A recent addition to the 3-tiered score is the BD0 category,[21] characterized by the complete absence of TB and very good outcomes. Because we trained our algorithm on TB-containing slides and might overcall TB compared with manual scoring, we cannot guarantee adequate automatic BD0 detection. Additional testing on cases in this part of the spectrum is necessary.

TB is defined based on size, varying between 1 and 4 tumor cells. Automatic detection shows an overrepresentation of single-cell TB (45% of TB), partly explaining the difference in the total number of TB on manual scoring. Previous studies have indicated that TB detection in IHC generates approximately 3 to 6 times more TB than TB detection in H&E,[22] most likely because of higher numbers of single-cell TB. In line, semi-automatic and automatic detection methods, mainly based on IHC, report higher TB.[10] With the high sensitivity of our detection tool, we detected, on average, about 7 times as much TB as the gastrointestinal pathologists per hotspot. This can be expected because we annotated our H&E slides based on cytokeratin staining patterns. However, the increased TB can also be attributed to so-called pseudobudding[6] that occurs in areas with inflammation and necrosis. These areas are avoided by pathologists. Proper segmentation is necessary to overcome this problem, illustrating that we need a combined approach in which the algorithm considers the TB microenvironment.

Indeed, the overcalling of TB might be responsible for the slight loss of information when comparing manual versus automatic TB assessment, as is evident from the AIC. However, our clinical validation confirms the prognostic value of automated TB count using the proposed algorithm in 2 different cohorts with different (local) staining protocols. TB has independent prognostic value for both DFS and OS. Larger cohorts are required to determine the prognostic value when correcting for more extensive clinicopathologic parameters (eg, mismatch repair status and extramural venous invasion).

The proposed pipeline for automated tumor bud detection opens the door to validate the role of TB in different stages or clinicopathologic parameters on large multicentric cohorts. Furthermore, it might help to understand the budding phenomenon better, as we can now start to investigate the dynamics with the surrounding macrosystem and the spatial distribution throughout a single or multiple section(s). This is now possible because we can identify TB in larger fields compared with the current detection area of 0.785 mm^2.

In conclusion, TB is an independent and relevant prognostic factor in CRC that can be assessed by automatic methods. High TB scores show a low miss rate by the algorithm, but precision might be improved by considering the TB microenvironment and ignoring parts with pseudobudding. However, the application of novel scoring methodology requires a reevaluation of current definitions and cut-off values.

Supplementary Material

The online version contains supplementary material available at https://doi.org/10.1016/j.modpat.2023.100233

References

1. Sung H, Ferlay J, Siegel L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209−249.
2. Brierley JD, Gospodarowicz MK, Wittekind C. *TNM Classification of Malignant Tumours*. John Wiley & Sons; 2017.
3. Lugli A, Kirsch R, Ajioka Y, et al. Recommendations for reporting tumor budding in colorectal cancer based on the International Tumor Budding Consensus Conference (ITBCC) 2016. *Mod Pathol*. 2017;30(9):1299−1311.

4. Shivji S, Conner JR, Barresi V, Kirsch R. Poorly differentiated clusters in colorectal cancer: a current review and implications for future practice. *Histopathology*. 2020;77(3):351−368.

5. Lugli A, Zlobec I, Berger MD, Kirsch R, Nagtegaal ID. Tumour budding in solid cancers. *Nat Rev Clin Oncol*. 2021;18(2):101−115.

6. Haddad TS, Lugli A, Aherne S, et al. Improving tumor budding reporting in colorectal cancer: a Delphi consensus study. *Virchows Arch*. 2021;479(3):459−469.

7. Litjens G, Ciompi F, Laak J. A decade of GigaScience: the challenges of gigapixel pathology images. *GigaScience*. 2022;11.

8. Gertych A, Swiderska-Chadaj Z, Ma Z, et al. Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides. *Sci Rep*. 2019;9(1):1483.

9. Zlobec I, Lugli A. Epithelial mesenchymal transition and tumor budding in aggressive colorectal cancer: tumor budding as oncotarget. *Oncotarget*. 2010;1(7):651−661.

10. Studer L, Blank A, Bokhorst JM, et al. Taking tumour budding to the next frontier—a post International Tumour Budding Consensus Conference (ITBCC) 2016 review. *Histopathology*. 2021;78(4):476−484.

11. Liu S, Zhang Y, Ju Y, et al. Establishment and clinical application of an artificial intelligence diagnostic platform for identifying rectal cancer tumor budding. *Front Oncol*. 2021;11: 626626.

12. Pai RK, Hartman D, Schaeffer DF, et al. Development and initial validation of a deep learning algorithm to quantify histological features in colorectal carcinoma including tumour budding/poorly differentiated clusters. *Histopathology*. 2021;79(3):391−405.

13. Van den Brand M, Hoevenaars BM, Sigmans JH, et al. Sequential immunohistochemistry: a promising new tool for the pathology laboratory. *Histopathology*. 2014;65(5):651−657.

14. Bokhorst JM, Nagtegaal I, Fraggetta F, et al. (2021). Automated risk classification of colon biopsies based on semantic segmentation of histopathology images. *arXiv*. 2021:2109.07892.

15. Galon J, Marincola F, Angell H, et al. Cancer classification using the Immunoscore: a worldwide task force. *J Transl Med*. 2012;10:205.

16. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention−MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing; 2015:234−241.

17. Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. International conference on machine learning. Long Beach, CA: PMLR; 2019:6105−6114.

18. Graham S, Vu QD, Raza SEA, et al. Hover-net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med Image Anal*. 2019;58: 101563.

19. Team, R. R core team: a language and environment for statistical computing R foundation for statistical computing. 2013.

20. Bokhorst J, Blank A, Lugli A, et al. Assessment of individual tumor buds using keratin immunohistochemistry: moderate interobserver agreement suggests a role for machine learning. *Mod Pathol*. 2020;33(5):825−833.

21. Zlobec I, Bächli M, Galuppini F, et al. Refining the ITBCC tumor budding scoring system with a "zero-budding" category in colorectal cancer. *Virchows Arch*. 2021;479(6):1085−1090.

22. Fisher NC, Loughrey MB, Coleman HG, Gelbard MD, Bankhead PD, Dunne P. Development of a semi-automated method for tumor budding assessment in colorectal cancer and comparison with manual methods. *Histopathology*. 2022;80(3):485−500.