ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO OPEN SCIENCE FOR OPTIMAL CANCER CARE

ORIGINAL ARTICLE

Artificial intelligence defines spatial patterns of tumor-infiltrating lymphocytes highly associated with outcome — a pan-GI cancer study

C. Chen[1], H. A. Mejbel[2,3], T. Pathak[4], A. Krasinskas[2,3], M. Reid[2], G. Corredor[4], P. Fu[5], J. E. Willis[1,6] & A. Madabhushi[3,4,7]*

[1]Department of Biomedical Engineering, Case Western Reserve University, Cleveland; [2]Department of Pathology and Laboratory Medicine, Emory University, Atlanta; [3]Radiology and Imaging Sciences, Biomedical Informatics (BMI) and Pathology, Georgia Institute of Technology and Emory University, Atlanta; [4]Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta; [5]Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland; [6]Department of Pathology, University Hospitals Cleveland Medical Center and Case Western Reserve University, Cleveland; [7]Atlanta Veterans Administration Medical Center, Decatur, USA

Available online xxx

Background: Gastrointestinal (GI) cancers (including esophagus, stomach, colon, rectum, pancreas and liver) account for more than one-quarter of all cancer diagnoses and 35% of cancer-related fatalities worldwide. Quantification of tumor-infiltrating lymphocytes (TILs) is a known cancer prognostic marker. In this study, we used computer vision and machine learning [artificial intelligence (AI)] approaches to evaluate the prognostic significance of computational pathology features relating to spatial arrangement and diversity in the appearance of TILs and cancer nuclei across five different types of GI cancers: colon, stomach, pancreas, and rectum adenocarcinoma and liver hepatocellular carcinoma.

Patients and methods: The study comprises >1700 patients from four different sites. Pathomic features (2236) were extracted from hematoxylin—eosin stained whole slide images and the top 9 features were selected by Least Absolute Shrinkage and Selection Operator (LASSO) Cox model. The top prognostic features identified were related to the spatial relationships between TILs and the closest cancer nuclei and tumor nuclei shape and texture features captured within local cellular clusters.

Results: Our trained model identified that 'low-risk' patients have significantly better overall survival than those identified as 'high risk' with a hazard ratio (HR) of 2.28 [95% confidence interval (CI) 1.32-3.93, $P = 0.0032$] in liver hepatocellular carcinoma; an HR of 2.79 (95% CI 1.66-4.68, $P = 0.0001$) in pancreatic adenocarcinoma; an HR of 5.85 (95% CI 2.53-15.5, $P = 0.0002$) in rectal adenocarcinoma; an HR of 1.81 (95% CI 1.07-3.07, $P = 0.0268$) in gastric adenocarcinoma. Across three different external validation sets of colorectal cancer (CRC) patients, our model yielded an HR of 2.32 (95% CI 1.67-3.23, $P < 0.0001$) in The Cancer Genome Atlas-Colon Adenocarcinoma (TCGA-COAD), an HR of 2.32 (95% CI 1.67-3.23, $P < 0.0001$) in Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO)-COAD, and an HR of 3.38 (95% CI 1.99-5.71, $P < 0.0001$) in the Emory dataset. Multivariable survival analysis showed that our trained model was prognostic independent of stage, age, race, and sex.

Conclusions: Our findings suggest that the spatial relationships of TILs and cancer nuclei are prognostic of survival across multiple GI cancer types.

Key words: digital pathology, machine learning, gastrointestinal cancer, colon adenocarcinoma, artificial intelligence

INTRODUCTION

Gastrointestinal (GI) tract cancers account for a significant portion of cancer morbidity and mortality in the United States, representing 18% of estimated new USA cases and an estimated 28% of cancer-related deaths in 2024.[1] Worldwide principal malignancies within the GI tract—including stomach (∼1.0 million new cases), liver (840 000 cases), esophagus (570 000 cases), pancreas (460 000 cases), and colorectum (1.8 million cases)—differ markedly in their etiological and epidemiological profiles.[2] Hematoxylin—eosin (H&E)-stained whole slide images (WSIs) from routine pathology tissue sections provide a rich data source for advanced machine learning techniques that define predictive and prognostic signatures across various cancers.[3-11] Multiple approaches, such as foundation models and deep learning (DL) techniques have shown

*Correspondence to: Prof. Anant Madabhushi, Radiology and Imaging Sciences, Biomedical Informatics (BMI) and Pathology, Georgia Institute of Technology and Emory University, 1760 Haygood Drive, Atlanta, Georgia 30322, USA. Tel: +1-267-259-6869

E-mail: anantm@emory.edu (A. Madabhushi).

promise in predicting outcomes across multiple cancer types by analyzing complex biological data.[12] However, these models often lack interpretability due to their 'black box' nature, raising concerns about clinical trust and understanding of the underlying biological mechanisms.[13] A promising alternative involves analyzing immune cells from routine H&E-stained images, focusing not only on cell counts but also on the spatial relationships between tumor-infiltrating lymphocytes (TILs) and tumor cells.[3-5,14,15] While the presence of intra- and peritumoral lymphocytic infiltrates has been long understood to be associated with improved patient prognosis across multiple tumor types,[6-8] inability to accurately assess their spatial arrangements and difficulties with overall quantification have precluded their use in standardized cancer reporting systems.[9] Spatial arrangements offer deeper insights into tumor-immune dynamics and their prognostic significance. Although this approach has shown success in lung and breast cancers,[10,11,16] its application in GI cancers remains underexplored, presenting a valuable opportunity to explore its potential as a prognostic biomarker in these cancers.

We investigate computational features of TIL—tumor spatial relationships and cellular diversity as novel prognostic markers in GI cancers. Our analysis validates these across 1738 patients from four independent datasets, encompassing colon, stomach, pancreatic, rectal, and liver cancers. We focus on microsatellite stable (MSS) GI cancers (~85% of cases) due to their immunotherapy resistance, metastatic risk, and immune infiltrate importance.[17,18] A feature classification model characterizing TIL—tumor nuclei interactions was developed by training a Cox regression model on colorectal WSIs (dataset D1) from University Hospital for disease-free survival (DFS). Results were validated on colon adenocarcinoma from The Cancer Genome Atlas (TCGA), Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO), and Emory University [D2-colon adenocarcinoma (COAD), D3, D4] and broader GI cancers (liver, pancreas, rectum, and stomach) from D2.

METHODS

Dataset description and image preprocessing

An overall flow chart is illustrated in Figure 1 and Supplementary Figure S1 (available at https://doi.org/10.1016/j.esmoop.2025.105757). This study utilized 1702 WSIs across four independent datasets. Our study focused on adenocarcinoma not otherwise specified (NOS) from GI due to its clinical prevalence[17] and sample availability in our cohort. The discovery cohort (D1) comprised 137 stage II/III MSS colorectal cancer (CRC) patients from University Hospitals Cleveland Medical Center, digitized at 40× resolution (0.252 μm/pixel) using a Ventana iScan HT scanner (Roche Diagnostics, Indianapolis, IN). The primary expansion cohort (D2) included 661 MSS GI cancer cases from TCGA, encompassing gastric (STAD, $n = 185$), colon (COAD, $n = 152$), rectal (READ, $n = 109$), pancreatic (PAAD, $n = 78$) adenocarcinomas, and liver hepatocellular carcinomas (LIHC, $n = 137$), with microsatellite status confirmed via cBioPortal (www.cbioportal.org). External validation sets consisted of 726 CRC cases from the PLCO trial (D3)[19] and 178 CRC cases from Emory University (D4), the latter including an MSS subgroup analysis ($n = 51$) due to partial status documentation. MSS cases were prioritized (~85% prevalence) based on their immunotherapy resistance and prognostic relevance,[17,18] with mucinous adenocarcinomas and pre-surgically treated cases excluded throughout. Overall survival (OS) served as the endpoint where disease-free survival (DFS) data were unavailable.

Image preprocessing was initiated with HistoQC[20] to generate artifact-free tissue masks by excluding regions with pen markings, blurring, or bubbles. Tumor regions were subsequently segmented using Self-Rule to Adapt (SRA),[21] a self-supervised DL segmentation model for domain-adaptive histopathology analysis. Pathologist evaluation of 15 randomly selected WSIs confirmed ≥70% tumor segmentation accuracy across all datasets. The intersection of HistoQC tissue masks and SRA tumor masks guided extraction of 1024 × 1024-pixel tiles for cellular analysis. Cerberus, a deep learning-based cell segmentation approach, was deployed to segment nuclei as well as epithelium from the individual image patches.[18] TIL identification was carried out via a pre-trained classifier leveraging texture, shape, and color features.[15] Pathologist assessment of 70 randomly sampled patches demonstrated robust TIL classification performance, with 92.8% achieving ≥70% concordance (good) and 7.2% showing 50%-70% accuracy (fair).

Feature extraction

Building on the concept of cell cluster graphs (CCG) where nodes represent nuclei groups rather than individual cells,[19,20] our group's prior work accounted for cellular heterogeneity by generating feature-driven clusters—aggregating nuclei based on morphometric attributes (e.g. size, texture, detailed in Supplementary Figure S9-S13, Supplementary Table S3 and feature descriptions section of Supplementary Material, available at https://doi.org/10.1016/j.esmoop.2025.105757).[4,21,22] In this study, our approach specifically focuses on tumor-immune spatial interactions in GI cancers, inspired by SpaTIL.[15] We modified the CCG to include only TILs and tumor nuclei (Supplementary Figures S14 and S15, available at https://doi.org/10.1016/j.esmoop.2025.105757), retaining local clusters containing both cell types for analysis. This modified CCG generated 2260 features capturing spatial relationships (e.g. TIL—tumor distances) and local morphology/texture within the graph detailed in Supplementary Material; Section 4, feature descriptions available at https://doi.org/10.1016/j.esmoop.2025.105757. Following feature extraction per tile, a kernel density map was created using two dimensional principal component analysis (2D-PCA) embeddings (Supplementary Figure S2, available at https://doi.org/10.1016/j.esmoop.2025.105757).[23] The top 20th percentile density tiles per WSI were selected as representative for subsequent analysis.

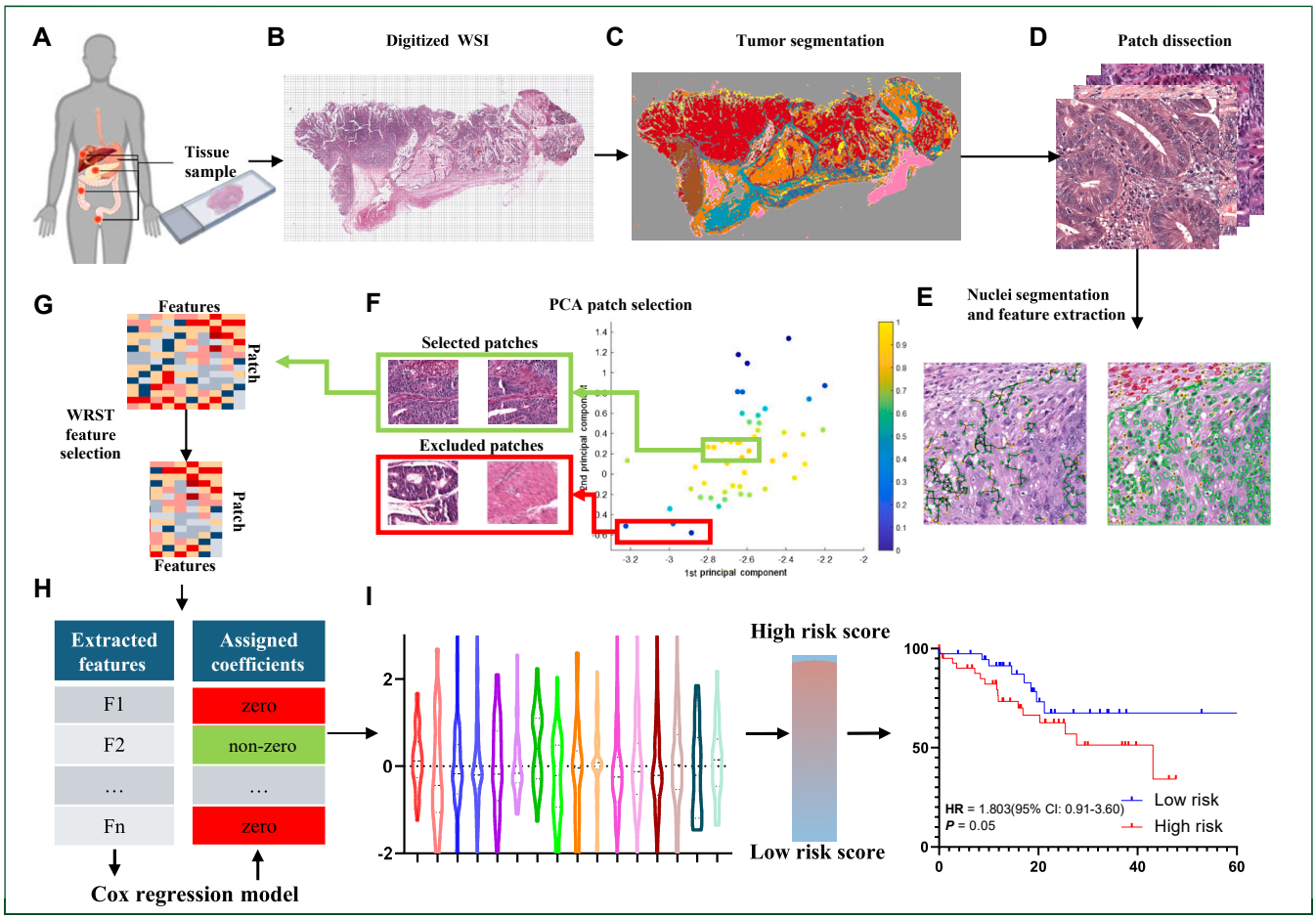

Figure 1. Flowchart. The flowchart illustrates the experimental design. (A, B) Digitized colorectal cancer whole slide images (WSIs) were collected from University Hospitals Cleveland Medical Center and used as the training set [dataset 1 (D1)]. WSIs were collected from The Cancer Genome Atlas with five common GI cancers [gastric, rectal, pancreatic, and colon adenocarcinomas (COAD), and liver hepatocellular carcinoma] was the independent validation set (D2). External validation WSIs were also acquired from PLCO-COAD (D3) and Emory University (D4). (C) Deep learning (DL) model combined with post image processing techniques were utilized to segment the tumor region on WSI (red: tumor). (D) WSIs were then dissected into tiles based on the final tumor mask. (E) DL and machine learning model were combined to segment out epithelium tumor nuclei and tumor-infiltrating lymphocytes (TILs) from each dissected tile (red: stroma nuclei; yellow: TIL; green: tumor epithelium nuclei). Histomorphometric features as well as spatial arrangement features of TIL and tumor nuclei were extracted. We here showcase a modified cell cluster graph which contains both TIL and tumor epithelium nuclei. Principal component analysis (PCA) was then applied to select the most representative tiles within each WSI (Supplementary Figure S2, available at https://doi.org/10.1016/j.esmoop.2025.105757). (G) Highly correlated features were removed. A Wilcoxon ranked sum test (WRST) was then applied to further reduce the feature number into 100. (H). A Cox regression model was trained on D1 to identify the top nine features using LASSO. (I) Top features were plotted out and a Kaplan–Meier survival analysis was carried out based on the risk score given by the trained model.

Statistical analysis

The Pearson correlation coefficient was computed for features within D1, leading to the exclusion of highly correlated features (correlation coefficient >0.90). The identified correlated features were also removed from D2, D3, and D4. Feature values were standardized within D1 and D2 to facilitate consistent interpretation and analysis. For evaluating the prognostic ability of the features, an OS defined as the time from diagnosis to death or the date of last follow-up for survivors was chosen as the event of interest, as DFS data was unavailable for D3. DFS is calculated from the date of diagnosis until the date of disease recurrence or death, whichever comes first. D1 was used as the training set, whereas D2 was utilized for external validation. A Wilcoxon rank-sum test was then applied to reduce the feature set to 100. The Least Absolute Shrinkage and Selection Operator (LASSO) technique was used to select the top nine features (details in Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105757) to train a Cox regression model,[24] M1, on D1. Patients in D2 were stratified into low- or high-risk categories based on the 55th percentile of the D1-derived risk score, and Kaplan–Meier (KM) curves were plotted for each GI tumor type. Sub-stage analyses were carried out for organ sites with >50 WSIs, and violin plots were generated to visualize the distribution of critical features. The independent prognostic value of our trained model was assessed using multivariable Cox regression, adjusting for clinicopathological variables to evaluate OS. OS was used for validation across D2, D3, and D4. The trained model M1, was validated on D2 using the same feature coefficients, and KM survival analysis was carried out to compare OS between high-risk (risk score >55th percentile) and low-risk groups. Hazard ratios (HR) and KM curves were produced, along with violin plots to display the distribution of key features. Subgroup analysis was

conducted on MSS patients from D4 (D4_MSS). UMAP was utilized to project extracted features from higher-dimensional space into a 2D space, enabling the investigation of potential batch effects.

RESULTS

Clinicopathologic features and multivariable analysis

As shown in Supplementary Table S2, available at https://doi.org/10.1016/j.esmoop.2025.105757, the median follow-up time for D1, D3, and D4 was significantly longer (78, 180, and 78 months, respectively) compared with D2, which had a follow-up time of 11 months. Additionally, 48% of patients in D2 were younger than 66 years old, whereas a lower percentage of patients in D1, D3, and D4 fell into this age group. D1 had a larger population of female patients, whereas male patients dominated the other three datasets. Furthermore, D2, D3, and D4 had a low percentage of stage IV patients ($\leq$10%). Patients with stage II to III GI cancer were the majority across all datasets.

Figure 2 presents multivariable Cox regression results assessing clinicopathological variables and our trained model for OS prediction. After adjusting for covariates (e.g. age, stage), the model emerged as a robust independent prognostic indicator across GI cancers: STAD [HR = 1.62, 95% confidence interval (CI) 1.02-2.25, $P < 0.05$], COAD (HR = 1.77, 95% CI 1.30-2.33, $P < 0.05$), READ (HR = 1.49, 95% CI 1.18-2.16, $P < 0.05$)., PAAD (HR = 1.97, 95% CI 1.26-2.87, $P < 0.05$), and LIHC (HR = 1.83, 95% CI 1.21-2.65, $P < 0.01$), and by contrast, variables like age and sex showed non-significant associations, aligning with prior evidence that spatial immune-cancer pattern may better reflect tumor-immune dynamics than most clinicopathologic parameters.

Experiment 2: evaluating the prognostic significance of spatial and diversity patterns of immune and cancer cells in gastrointestinal cancers

A total of nine features were identified through LASSO Cox regression modeling with DFS on D1. The top selected feature was: Skewness-CCG:Number of Edges. This feature measures how unevenly immune cells (TILs) and cancer cells interact spatially. A 'skewed' value indicates that some regions have many more connected edges than others, suggesting areas of dense immune infiltration or immune exclusion. Feature 3, Mean-Arch:Density of Polygons, quantifies the compactness of clustering between cancer and immune cells. A higher polygon count (geometric shapes defined by adjacent cells) indicates complex cellular groupings, whereas a lower count suggests sparse or disordered arrangements. Increased density potentially signifies immune cells surrounding tumors to attack them, reflecting potential antitumor immunity. Apart from spatial architecture between TILs and tumor nuclei, features characterize TIL-adjacent tumor nuclei's shape and texture were also selected. For instance, feature 2, Mean-CCM-Area:std (info-measure1), quantifies how much cancer nuclei vary in size within a tissue region. A higher value means nuclei sizes are highly inconsistent (e.g. some very large, some very small). The top nine features are described in Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105757. The violin plots of the top feature in low-risk and high-risk cancers in each data set is shown in Supplementary Figure S3, available at https://doi.

Clinic Variables	Organ Site	HR (95% CI)	P value
Age	COAD	0.99 (0.96−1.02)	0.831
	LIHC	1.01 (0.97−1.03)	0.901
	PAAD	1.00 (0.98−1.03)	0.816
	READ	0.99 (0.98−1.02)	0.926
	STAD	1.00 (0.98−1.03)	0.88
Sex (male versus female)	COAD	0.96 (0.93−1.02)	0.674
	LIHC	1.02 (0.71−1.32)	0.261
	PAAD	0.82 (0.51−1.32)	0.416
	READ	0.90 (0.63−1.29)	0.559
	STAD	1.07 (0.64−1.79)	0.806
Race (Caucasian vs African American)	COAD	0.94 (0.81−1.10)	0.528
	LIHC	0.87 (0.72−1.27)	0.343
	PAAD	0.83 (0.56−1.22)	0.333
	READ	0.93 (0.76−1.14)	0.482
	STAD	0.93 (0.49−1.80)	0.839
Stage	COAD	0.98 (0.92−1.06)	0.656
	LIHC	1.04 (0.85−1.24)	0.579
	PAAD	0.94 (0.55−1.59)	0.7355
	READ	1.03 (0.91−1.16)	0.609
	STAD	1.05 (0.81−1.33)	0.707
Spatial TIL and tumor nuclei (low versus high risk)	COAD	1.77 (1.30−2.33)	0.019
	LIHC	1.83 (1.21−2.65)	0.01
	PAAD	1.97 (1.26−2.87)	0.007
	READ	1.49 (1.18−2.16)	0.022
	STAD	1.62 (1.02−2.25)	0.014

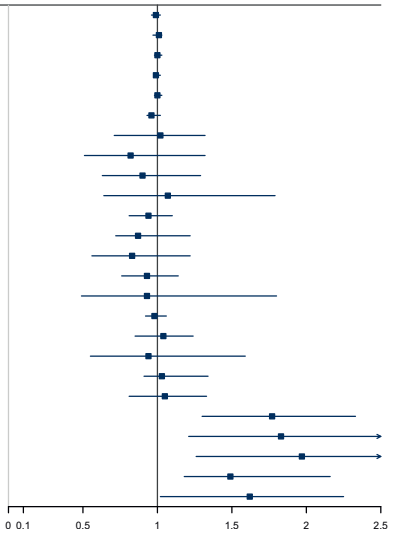

Figure 2. Multivariable analysis. Our model emerged as an independent prognostic indicator for overall survival and a robust independent prognostic indicator across gastrointestinal cancers from dataset 2: STAD (HR = 1.62, 95% CI 1.02-2.25, $P < 0.05$), COAD (HR = 1.77, 95% CI 1.30-2.33, $P < 0.05$), READ (HR = 1.49, 95% CI 1.18-2.16, $P < 0.05$), PAAD (HR = 1.97, 95% CI 1.26-2.87, $P < 0.05$), and LIHC (HR = 1.83, 95% CI 1.21-2.65, $P < 0.01$), and by contrast, variables like age and sex showed non-significant associations, aligning with prior evidence that spatial immune-cancer pattern may better reflect tumor-immune dynamics than most clinicopathologic parameters.
CI, confidence interval; COAD, colorectal adenocarcinoma; HR, hazard ratio; LIHC, liver hepatocellular carcinoma; PAAD, pancreatic adenocarcinoma; READ, rectal adenocarcinoma; STAD, gastric adenocarcinoma.

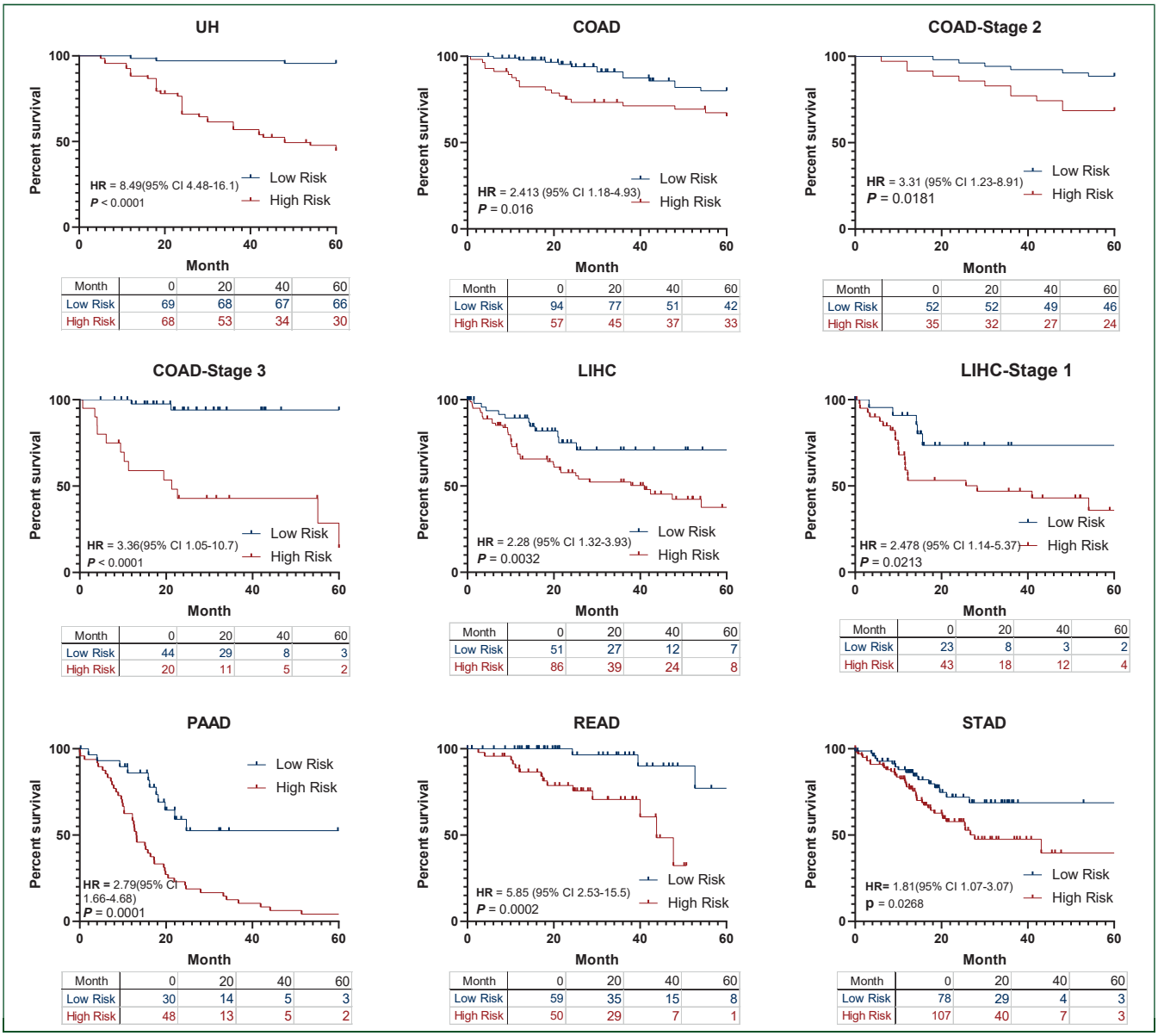

Figure 3. Kaplan–Meyer (KM) curves of experiment 1. (A) The KM survival curves of University Hospital (UH) for disease-free survival (training) and overall survival (validation) between high-risk and low-risk groups are plotted for datasets 1 (D1) and D2. A significantly favorable outcome in the low-risk group compared with the high-risk group was observed with a hazard ratio (HR) = 8.49 [95% confidence interval (CI) 4.48-16.1, $P < 0.0001$] on D1 and HR = 2.41 (95% CI 1.18-4.93, $P = 0.016$) on D2-COAD. Our model achieved an HR of 2.28 (95% CI 1.32-3.93, $P = 0.0032$) in liver hepatocellular carcinoma (LIHC); an HR of 2.79 (95% CI 1.66-4.68, $P = 0.0001$) in pancreatic adenocarcinoma (PAAD); an HR of 5.85 (95% CI 2.53-15.5, $P = 0.0002$) in rectal adenocarcinoma (READ); and an HR of 1.81 (95% CI 1.07-3.07, $P = 0.0268$) in gastric adenocarcinoma (STAD). For stagewise analysis, we only analyzed stages with a number ≥50 patients. The model yielded an HR of 3.31(95% CI 1.67-3.23, $P = 0.0181$) in stage 2 of D2-COAD; an HR of 3.36 (95% CI 1.05-10.7, $P < 0.0001$) in stage 3 of D2-COAD. The trained model achieved statistically significant separation of patient groups in stage 1 of LIHC (HR = 2.48, 95% CI 1.14-5.73, $P = 0.021$).
COAD, colon adenocarcinoma.

org/10.1016/j.esmoop.2025.105757. The violin plots of features 2 to 9 are shown in Supplementary Figure S4, available at https://doi.org/10.1016/j.esmoop.2025.105757. We also applied UMAP to examine the presence of potential batch effects among the selected features. Supplementary Figure S5, available at https://doi.org/10.1016/j.esmoop.2025.105757 demonstrates that no distinct clusters are evident within the whole dataset, indicating no obvious batch effect was observed among the selected features. Figures 3 and 4 present the KM plots for the second experiment, which illustrate the differences between the high-risk and low-risk groups for datasets D1, D2, D3, and D4 as defined by the cutoff risk score (55th percentile of the training risk score). The curves demonstrate a marked distinction between the two groups, with the low-risk group exhibiting a significantly more favorable outcome compared with the high-risk group. In the training set D1, the HR was 8.49 (95% CI 4.48-16.1, $P < 0.0001$), indicating a much higher risk of disease recurrence or progression in the high-risk group. Similarly, in D2-COAD, the HR values were 3.31 (95% CI 1.67-6.63, $P = 0.0181$) for stage II and 3.36 (95% CI 1.05-10.7, $P < 0.0001$) for stage III, showing a substantial increase in the risk for the high-risk group as the disease stage advances. In LIHC, the

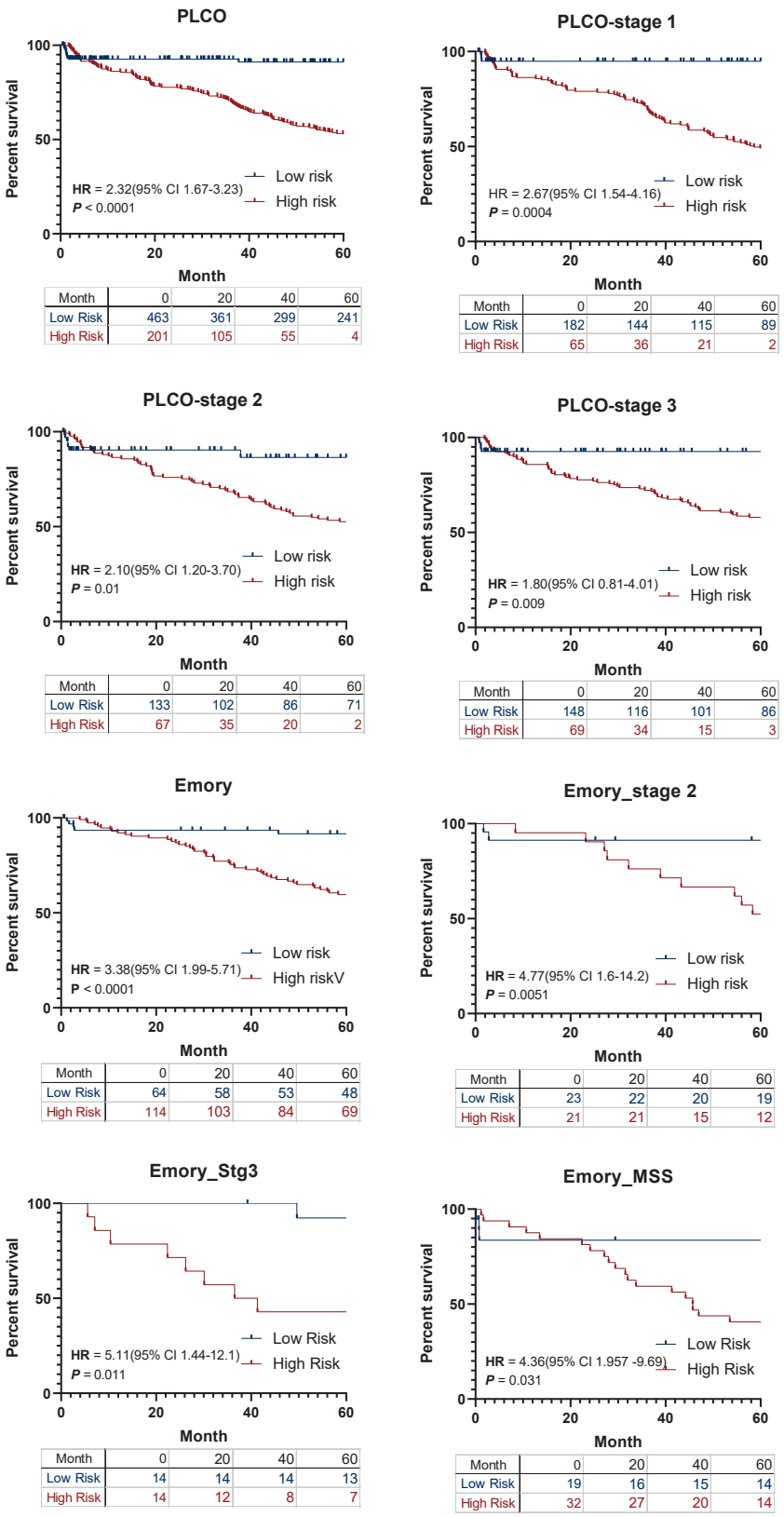

Figure 4. Kaplin–Meyer curves of datasets 3 (D3) and D4. (A) The Kaplin–Meyer survival curves of overall survival between high-risk and low-risk groups are plotted for all patients from D3 (PLCO) and D4 (Emory). Our model achieved a hazard ratio (HR) of 2.32 [95% confidence interval (CI) 1.67-3.23, $P < 0.0001$] in D3 using a log-rank test. A statistically significantly favorable outcome in the low-risk group compared with the high-risk group was observed with an HR = 3.38 (95% CI 4.48-16.1, $P < 0.0001$) on all patients from D4. For stagewise results, the model yielded an HR of 2.67 (95% CI 1.54-4.16, $P = 0.0004$) in stage 1 of D3; an HR of 2.10 (95% CI 1.20-3.70, $P = 0.01$) in stage 2 of D3; an HR of 1.80 (95% CI 0.81-4.01, $P = 0.009$) in stage 3 of D3; an HR of 4.77 (95% CI 1.60-14.2, $P = 0.0051$) in stage 2 of D4; and an HR of 3.36 (95% CI 1.44-12.1, $P = 0.011$) in stage 3 of D4. We further analyzed the microsatellite stable (MSS) only patients from D4 and yielded an HR of 4.36 (95% CI 1.96-9.69, $P = 0.031$).

model yielded an HR of 2.28 with a 95% CI of 1.32 to 3.93, and $P = 0.0032$, suggesting a more than twofold increase in the risk of mortality for the high-risk group compared with the low-risk group. In PAAD, the HR was 2.79, with a 95% CI of 1.66 to 4.68, and $P = 0.0001$, indicating a higher risk of mortality in the high-risk group. The model showed particularly striking results in READ, with an HR of 5.85, a 95% CI of 2.53 to 15.5, and $P = 0.0002$, reflecting a nearly six-fold increased risk for the high-risk group. In STAD, the HR was 1.81, with a 95% CI of 1.07 to 3.07, and $P = 0.0268$, suggesting a significant but lower relative increase in mortality risk for the high-risk group, though the absolute risk of death due to STAD is high. For a more detailed analysis, the model was applied to stages with a patient count of 50 or more. In stage I LIHC, the model achieved significant separation of patient groups, with an HR of 2.48, a 95% CI of 1.14 to 5.73, and $P = 0.021$. This indicates a more than twofold increase in the risk of mortality for the high-risk group in the early stages of the disease.

In D3, the HR for the overall comparison was 2.32 (95% CI 1.67-3.23, $P < 0.0001$), and when broken down by stages, the HR was 2.67 (95% CI 1.54-4.16, $P = 0.0004$) for stage I, 2.10 (95% CI 1.20-3.70, $P = 0.01$) for stage II, and 1.80 (95% CI 0.81-4.01, $P = 0.009$) for stage III. The KM curves of OS between high-risk and low-risk groups were plotted for all patients from D4, D4-stage II, D4-stage III, and D4_MSS. A significantly favorable outcome was observed in the low-risk group compared with the high-risk group with an HR of 3.38 (95% CI 4.48-16.1, $P < 0.0001$) for all patients from D4. For stage-wise results: the model yielded an HR of 4.77 (95% CI 1.60-14.2, $P = 0.0051$) in stage II of D4; an HR of 3.36 (95% CI 1.44-12.1, $P = 0.011$) in stage III of D4. We further analyzed the MSS only CRCs from D4 and obtained an HR of 4.36 (95% CI 1.96-9.69, $P = 0.031$). These results suggest that the risk associated with the high-risk group is elevated throughout the progression of the disease, with a particularly significant increase at later stages. The log-rank test was used to statistically compare the survival distributions of the two groups, confirming that the observed differences are statistically significant.

Heatmaps (Supplementary Figures S6 and S7, available at https://doi.org/10.1016/j.esmoop.2025.105757) of the top selected feature (Skewness-CCG:Number of Edges), standardized patch-wise, was generated. Gaussian blur was subsequently applied to reduce spatial granularity and improve visualization. Figure 5 and Supplementary Figure S8, available at https://doi.org/10.1016/j.esmoop.2025.105757, display the feature maps overlaid with our modified CCGs, comparing low-risk and high-risk patient cohorts. Low-risk patients exhibited a higher density of epithelial TILs embedded within cancer cell regions, fostering greater connectivity to tumor nuclei in localized cell network diagrams. These connections are represented by edges in the graphs, with TILs labeled with red dots and tumor nuclei labeled with green dots. This figure also demonstrates that low-risk patches have more TILs in proximity to the tumor epithelium.

DISCUSSION

GI tract cancers, including stomach, liver, esophagus, pancreas, and colorectum, are major health concerns worldwide, accounting for 26% of global cancer incidence and 35% of cancer-related deaths.[1] Multiple investigations over many decades have identified large numbers of prognostic markers in GI cancers.[25-34] A disadvantage to most of these biomarkers is the cost and technical expertise needed for widespread validation and adoption. H&E-stained WSIs are a widely available resource from routine pathology. With the advent of digital pathology and machine learning, there is potential to predict molecular traits, predictive signatures and outcomes in cancers from clinically acquired and processed pathology specimens,[3-11] including the potential prognostic significance of the spatial relationship of lymphocytes and tumor cells.[4-8] To date there has been only a small number of studies directly addressing this concept in GI cancers.[35] To address this issue in a pan-GI cancer concept, we drew inspiration from the SpaTIL[15] as well as cell cluster graph[4] frameworks and developed a novel approach that combines these two feature families to investigate the spatial relationship between TILs and adjacent tumor nuclei. By focusing on handcrafted features that capture the structural and spatial organization of these cellular components, we aimed to create an interpretable and computationally efficient model for prognosticating GI cancer patients. To our knowledge, this is the first handcrafted feature-based model to comprehensively address pan-GI cancers, offering an innovative alternative to black box DL models.

Our study aimed to establish the prognostic value of TIL—tumor nuclei orientation across various GI cancers (stomach, colon, rectum, pancreas, and liver). Our model demonstrated a strong prognostic association between the computational pathology derived signatures and GI cancers (stomach, colon, rectum, pancreas, and liver) from D2, with HRs indicating significant prognostic value across different cancer types and stages. These findings are particularly notable for stomach, colon, rectum, pancreas, and liver, showcasing the model's robustness and broad applicability. Moreover, our subgroup analysis on MSS only patients from D4 indicates that our model is capable of risk-stratifying MSS only. These results highlight the model's ability to stratify patients into risk groups with significant differences in OS across various GI cancers. The findings underscore the potential utility of the model in clinical settings for prognostic assessment and treatment planning. The independent prognostic significance of our model, validated through multivariable analysis, indicates its utility as a tool for identifying high-risk patients. This approach addresses a critical gap in current GI cancer management, especially considering the interpretability challenges associated with DL-based models and the lack of focus on pan-GI cancer in existing research. Furthermore, our spatial risk model consistently identifies high-risk MSS stage II CRC patients across independent cohorts (Figures 3 and 4). This stratification directly addresses a major clinical dilemma.

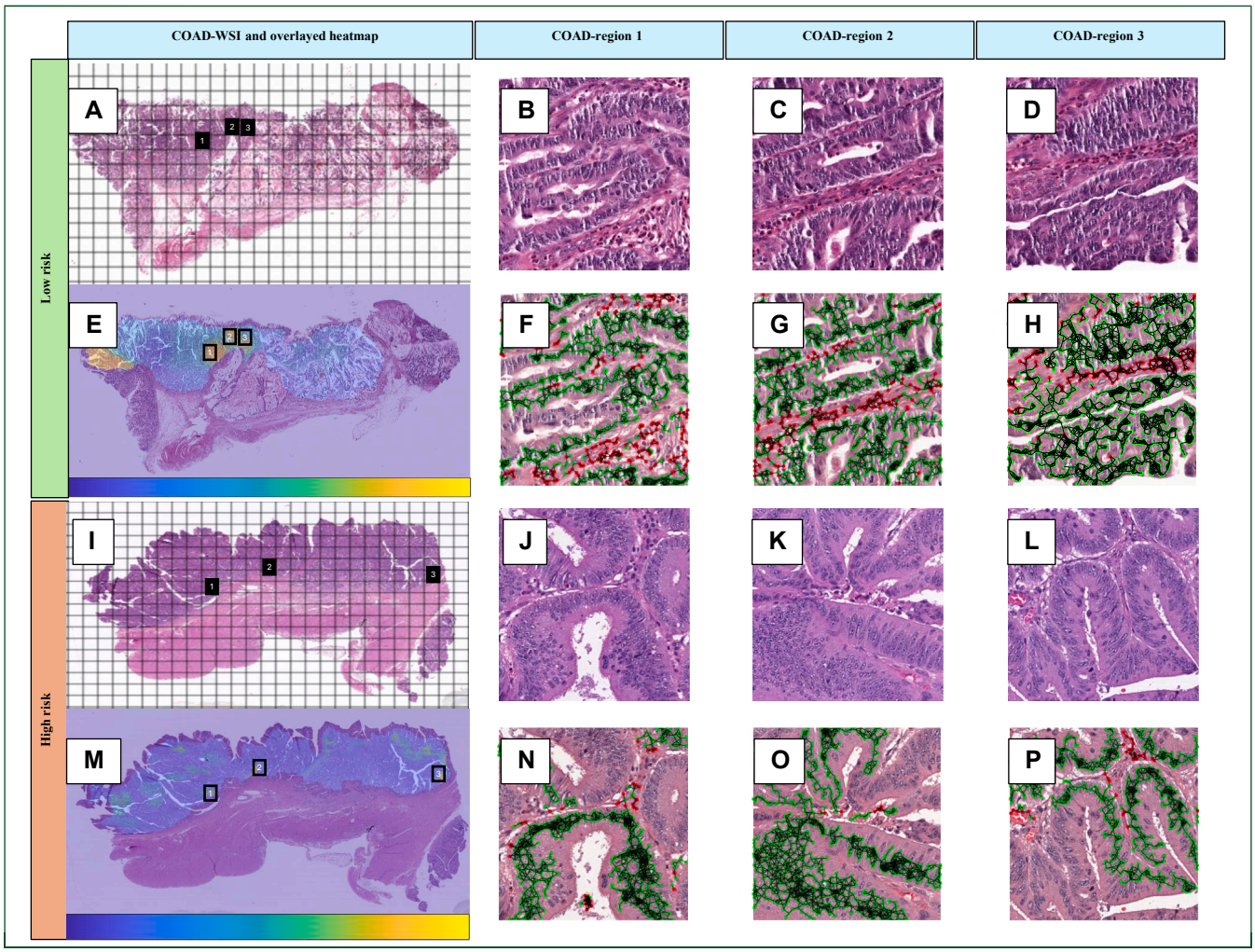

Figure 5. Illustration of feature maps for colon adenocarcinoma (COAD). Whole slide images (WSI) with a low-risk (top) and high-risk (bottom) patient. Features including number of edges of graphs connecting tumor-infiltrating lymphocytes (TIL) (labeled with red dot) and tumor nuclei (labeled with green dot) appeared significantly different between short-term and long-term survivors. (A and I) tiled WSI at 40× magnification, (B-D) representative patches from low-risk patients (zoomed area 1-3 from panel A), (J-L) representative patches from high-risk patients (zoomed area 1-3 from panel I), (E and M) color-coded WSI using features described in row 1 of Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105757, respectively, with shades of blue, green and yellow, (F-H), and (N-P) modified local cell graphs overlayed on representative patches shown in (B-D) and (J-L), In high-risk patients, the epithelial TILs are lower than in low-risk patients. In local cell graphs from identified low-risk cases, the presence of epithelial TILs dispersed among cancer cells created more connections to tumor nuclei. There are more connected TIL clusters in local cell graphs from defined low-risk than high-risk patients.

High-risk stage II MSS patients, identified by their deficient TIL—tumor spatial architecture, represent prime candidates for future prospective studies evaluating adjuvant chemotherapy benefit. Conversely, low-risk patients may be candidates for de-escalated surveillance.

Previous work[36] in DL-based GI cancer research has demonstrated strong predictive performance but often suffers from limited interpretability, which poses challenges for clinical adoption. Unlike the opaque 'black box' nature of DL models, our handcrafted approach provides interpretability by identifying graph-based and histomorphologic features that are prognostic across multiple GI cancers. By contrast, our model directly characterizes the tumor microenvironment through features that pathologists can understand and validate. In terms of prognostic performance, Kather et al. trained a VGG16 model and achieved HRs of 1.63 (1.14-2.33) and 2.29 (1.5-3.48) in the prediction of OS and disease-specific survival in CRC patients.[37] Geessink et al. also trained a ConvNets to prognosticate patients with rectal cancer and achieved an HR of 2.05 (1.11-3.78) in DFS.[38] Our handcrafted feature-based model demonstrated comparable accuracy to DL models for outcome prediction. However, it offers a critical advantage in interpretability, allowing for direct insights into the biological underpinnings of risk stratification. The spatial arrangement of tumor nuclei—TIL graphs revealed distinct patterns between low-risk and high-risk patients. Low-risk patients exhibited a higher density of epithelial TILs, forming more connections with tumor nuclei, which in turn suggests active immune targeting of cancer cells (Figure 5 and Supplementary Figure S8, available at https://doi.org/10.1016/j.esmoop.2025.105757). In contrast, high-risk patients displayed sparse epithelial TILs, fragmented TIL clusters, and fewer connections to tumor nuclei, suggesting a weaker immune defense and correlating with a greater risk of recurrence or progression. Supplementary

Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105757 highlights additional features characterizing the nuclei—TIL graph: feature 3 quantifies the number of polygons within local cellular clusters, revealing higher complexity and disorganization in high-risk patients, whereas feature 4 evaluates the ratio of cell density heterogeneity, further underscoring chaotic and fragmented nuclei distribution in high-risk cases. Beyond graph-based features, nuclei shape and texture (features 2, 5-9) also differentiate risk groups, with high-risk patients comprising tumor nuclei of larger average area, wider minor axis ranges, chaotic orientations, and higher intensity. Collectively, these findings suggest that high-risk patients exhibit a more disorganized TIL spatial arrangement, whereas low-risk patients demonstrate a more structured and interconnected TIL—tumor cluster, suggestive of a stronger antitumor-immune response.

Our study builds on concepts from SpaTIL,[15] which emphasized the spatial distribution of TILs, and extended them by integrating tumor nuclei morphology and graph-based features into a pan-GI cancer framework. Compared to SpaTIL, which focused primarily on spatial arrangements of TILs, our model captures the spatial relationship between TILs and tumor nuclei and evaluates additional histomorphologic attributes, making it uniquely equipped to address the heterogeneity in GI cancers. Original CCG[19] methodologies often overlooked cellular diversity. To overcome this, our group pioneered the use of morphometric feature-based clustering (grouping nuclei by quantitative traits like size and texture),[4,25,26] demonstrating that the proximity and arrangement of these phenotypically similar nuclear clusters held valuable prognostic information in cancer. While this represented a significant advance in characterizing tumor architecture, it treated immune and tumor cells largely as separate spatial entities. Crucially, it did not investigate the localized, bidirectional interactions occurring at the immune—tumor frontier—specifically, how the presence and spatial distribution of TILs might actively shape or be shaped by the cancer nuclei immediately surrounding them. Our present study bridges this gap by developing novel metrics to quantify these specific TIL-neighboring tumor nuclei relationships, hypothesizing they are central to immune evasion or recognition. Furthermore, our approach leverages a multi-institutional dataset, enhancing its generalizability and robustness across diverse populations. Overall, our approach combines interpretability, pan-GI cancer applicability, and multi-institutional validation to provide a practical and effective alternative to DL models. By elucidating key risk factors through interpretable features, our study contributes not only to accurate prognostication but also to guiding personalized treatment strategies, potentially improving patient outcomes.

We acknowledge that our study did have its limitations. Firstly, the validation of our findings was conducted retrospectively rather than prospectively. This approach may limit the generalizability of our results to future patient cohorts and clinical scenarios. Secondly, our study focused exclusively on prognostic factors, emphasizing outcome prediction without addressing predictive factors such as treatment response or therapeutic benefits. This limitation restricts the scope of our findings and prevents a comprehensive assessment of the model's potential clinical utility, particularly in guiding therapeutic decision-making. Thirdly, we did not investigate potential population-specific variations in the prognostic signature across different racial or ethnic groups, such as black and white patients. This limitation represents a critical gap, as understanding these differences is essential to ensure the equitable application of our model across diverse populations. Moreover, the exclusion of mucinous adenocarcinomas ensures architectural homogeneity for feature development but limits applicability to this subtype. Future work should explicitly evaluate and potentially adapt these spatial features for mucinous CRC, which may require specialized computational approaches due to their distinct histomorphology. Furthermore, our study focused on treatment-naïve tumors to define baseline spatial prognostic features. The exclusion of neoadjuvant treated cases limits immediate applicability in modern therapeutic contexts where such treatments are used. Validating and potentially refining these spatial features in cohorts receiving neoadjuvant immunotherapy or chemotherapy represents a critical future step, particularly given promising trials like NICHE[39,40] and FOxTROT.[41] While our study primarily focused on adenocarcinoma NOS due to its prevalence,[17] we recognize the biological diversity of GI cancer histotypes (e.g. mucinous, micropapillary) and their potential impact on our model's performance. Future directions include validating our models on histotype-specific datasets to address morphological heterogeneity. Future studies could also incorporate lymphocyte subtyping via multiplexed assays to dissect how specific immune subsets (e.g. immunosuppressive regulatory T cells or cytotoxic natural killer cells) influence the spatial dynamics quantified here. Finally, while this study provides a foundational framework, further validation in the context of completed clinical trials is necessary. Such validation efforts are essential to confirm the predictive utility of our methods in real-world clinical settings, thereby enhancing their practical relevance and impact on patient care.

In conclusion, our findings support the integration of digital pathology and machine learning into clinical practice, offering a cost-effective, accessible, and interpretable method for prognostication in GI cancers. We validated the prognostic value of TIL—tumor nuclei metrics across the most common GI cancers, identifying nine features with a primary focus on the spatial relationship between TILs and cancer nuclei. Of particular interest was the ability of this methodology to stratify stage II CRC patients into risk categories, demonstrating its potential to address the clinical need for better post-operative risk assessment. Our model leverages H&E-

stained WSIs to potentially streamline and economize prognostication. Its independent prognostic significance positions our approach as an innovative tool for identifying high-risk GI cancer patients, circumventing the drawbacks of current DL-based models and enhancing clinical decision-making.

DISCLOSURE

AM holds equity in Picture Health, Elucid Bioimaging, and Inspirata Inc.; serves on advisory boards for Picture Health, Aiforia Inc., and SimBioSys; and has research agreements with Astra Zeneca, Boehringer-Ingelheim, Eli-Lilly, and Bristol Myers-Squibb. All other authors have declared no conflicts of interest.

FUNDING

This work was supported by the National Cancer Institute [grant numbers R01CA268287A1, U01CA269181, R01CA26-820701A1, R01CA249992-01A1, R01CA202752-01A1, R01C-A208236-01A1, R01CA216579-01A1, R01CA220581-01A1, R01CA257612-01A1, 1U01CA239055-01, 2P50CA150964-06A1, 1U01CA248226-01, 1U54CA254566-01]; National Heart, Lung and Blood Institute [grant numbers 1R01HL15127701A1, R01HL15807101A1]; National Institute of Biomedical Imaging and Bioengineering [grant number 1R43EB028736-01]; VA Merit Review Award [grant number IBX004121A] from the United States Department of Veterans Affairs Biomedical Laboratory Research and Development Service the Office of the Assistant Secretary of Defense for Health Affairs, through the Breast Cancer Research Program [grant number W81XWH-19-1-0668], the Prostate Cancer Research Program [grant number W81XWH-20-1-0851], the Lung Cancer Research Program [grant numbers W81XWH-18-1-0440, W81XWH-20-1-0595], the Peer Reviewed Cancer Research Program [grant numbers W81XWH-18-1-0404, W81XWH-21-1-0345, W81XWH-21-1-0160], the Kidney Precision Medicine Project Glue Grant and sponsored research agreements from Bristol Myers-Squibb, and Astra Zeneca.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the United States Department of Veterans Affairs, the Department of Defense, or the United States Government.

REFERENCES

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin*. 2024;74(1):12-49.
2. Arnold M, Abnet CC, Neale RE, et al. Global burden of 5 major types of gastrointestinal cancer. *Gastroenterology*. 2020;159(1):335-349.e15.
3. Lu C, Romo-Bucheli D, Wang X, et al. Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers. *Lab Invest*. 2018;98(11):1438-1448.
4. Lu C, Wang X, Prasanna P, et al. Feature driven local cell graph (FeDeG): predicting overall survival in early stage lung cancer. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. *Lecture Notes in Computer Science*. Springer: International Publishing; 2018. p. 407-416.
5. Lu C, Lewis JS, Dupont WD, Plummer WD, Janowczyk A, Madabhushi A. An oral cavity squamous cell carcinoma quantitative histomorphometric-based image classifier of nuclear morphology can risk stratify patients for disease-specific survival. *Mod Pathol*. 2017;30(12):1655-1665.
6. Galon J, Costes A, Sanchez-Cabo F, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science*. 2006;313(5795):1960-1964.
7. Yao W, He JC, Yang Y, et al. The prognostic value of tumor-infiltrating lymphocytes in hepatocellular carcinoma: a systematic review and meta-analysis. *Sci Rep*. 2017;7(1):7525.
8. Lee HE, Chae SW, Lee YJ, et al. Prognostic implications of type and density of tumour-infiltrating lymphocytes in gastric cancer. *Br J Cancer*. 2008;99(10):1704-1711.
9. Willis J, Anders RA, Torigoe T, et al. Multi-institutional evaluation of pathologists' assessment compared to immunoscore. *Cancers*. 2023;15(16):4045.
10. Kather JN, Calderaro J. Development of AI-based pathology biomarkers in gastrointestinal and liver cancer. *Nat Rev Gastroenterol Hepatol*. 2020;17(10):591-592.
11. Wulczyn E, Steiner DF, Moran M, et al. Interpretable survival prediction for colorectal cancer using deep learning. *NPJ Digit Med*. 2021;4:71.
12. Shreve JT, Khanani SA, Haddad TC. Artificial intelligence in oncology: current capabilities, future opportunities, and ethical considerations. *Am Soc Clin Oncol Educ Book*. 2022;42:1-10.
13. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med*. 2021;13:152.
14. Tsai PC, Lee TH, Kuo KC, et al. Histopathology images predict multi-omics aberrations and prognoses in colorectal cancer patients. *Nat Commun*. 2023;14(1):2102.
15. Corredor G, Wang X, Zhou Y, et al. Spatial architecture and arrangement of tumor-infiltrating lymphocytes for predicting likelihood of recurrence in early-stage non-small cell lung cancer. *Clin Cancer Res*. 2019;25(5):1526-1534.
16. Jiang D, Liao J, Duan H, et al. A machine learning-based prognostic predictor for stage III colon cancer. *Sci Rep*. 2020;10:10333.
17. Ono Y, Yilmaz O. Emerging and under-recognised patterns of colorectal carcinoma morphologies: a comprehensive review. *J Clin Pathol*. 2024;77(7):439-451.
18. Graham S, Vu QD, Jahanifar M, et al. One model is all you need: multi-task learning enables simultaneous histology image segmentation and classification. *Med Image Anal*. 2023;83:102685.
19. Lewis JS, Ali S, Luo J, Thorstad WL, Madabhushi A. A quantitative histomorphometric classifier (QuHbIC) identifies aggressive versus indolent p16-positive oropharyngeal squamous cell carcinoma. *Am J Surg Pathol*. 2014;38(1):128-137.
20. Shin D, Protano MA, Polydorides AD, et al. Quantitative analysis of high-resolution microendoscopic images for diagnosis of esophageal squamous cell carcinoma. *Clin Gastroenterol Hepatol*. 2015;13(2):272-279.e2.
21. Lu C, Bera K, Wang X, et al. A prognostic model for overall survival of patients with early-stage non-small cell lung cancer: a multicentre, retrospective study. *Lancet Digit Health*. 2020;2(11):e594-e606.
22. Lu C, Koyuncu C, Corredor G, et al. Feature-driven local cell graph (FLocK): new computational pathology-based descriptors for prognosis of lung cancer and HPV status of oropharyngeal cancers. *Med Image Anal*. 2021;68:101903.
23. Ding R, Prasanna P, Corredor G, et al. Compactness measures of tumor infiltrating lymphocytes in lung adenocarcinoma are associated with overall patient survival and immune scores. In: Tomaszewski JE, Ward AD, editors. *Medical Imaging 2020: Digital Pathology*. Houston, Texas, United States: SPIE Medical Imaging; 2020 11320.
24. Cox DR. Regression models and life-tables. *J R Stat Soc Series B*. 1972;34(2):187-202.
25. Alnakli AAA, Mohamedali A, Heng B, et al. Protein prognostic biomarkers in stage II colorectal cancer: implications for post-operative management. *BJC Rep*. 2024;2(1):13.
26. Lişcu HD, Verga N, Atasiei DI, et al. Biomarkers in colorectal cancer: actual and future perspectives. *Int J Mol Sci*. 2024;25(21):11535.

27. Nunes L, Li F, Wu M, et al. Prognostic genome and transcriptome signatures in colorectal cancers. *Nature*. 2024;633(8028):137-146.
28. Cheng E, Shi Q, Shields AF, et al. Association of inflammatory biomarkers with survival among patients with stage III colon cancer. *JAMA Oncol*. 2023;9(3):404-413.
29. Bass AJ, Thorsson V, Shmulevich I, et al. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014;513(7517):202-209.
30. Zhang Y, Xue Y, Gao Y, Zhang Y. Prognostic and predictive value of pathohistological features in gastric cancer and identification of SLITRK4 as a potential biomarker for gastric cancer. *Sci Rep*. 2024;14(1):29241.
31. Burz C, Pop V, Silaghi C, Lupan I, Samasca G. Prognosis and treatment of gastric cancer: a 2024 update. *Cancers*. 2024;16(9):1708.
32. Li Y, Pan L, Mugaanyi J, et al. Pathomic and bioinformatics analysis of clinical-pathological and genomic factors for pancreatic cancer prognosis. *Sci Rep*. 2024;14(1):27769.
33. Yu J, Park R, Kim R. Promising novel biomarkers for hepatocellular carcinoma: diagnostic and prognostic insights. *J Hepatocell Carcinoma*. 2023;10:1105-1127.
34. Liu Y, Sethi NS, Hinoue T, et al. Comparative molecular analysis of gastrointestinal adenocarcinomas. *Cancer Cell*. 2018;33(4):721-735.e8.
35. Lim Y, Choi S, Oh HJ, et al. Artificial intelligence-powered spatial analysis of tumor-infiltrating lymphocytes for prediction of prognosis in resected colon cancer. *NPJ Precis Oncol*. 2023;7(1):124.
36. Kuntz S, Krieghoff-Henning E, Kather JN, et al. Gastrointestinal cancer classification and prognostication from histology using deep learning: systematic review. *Eur J Cancer*. 2021;155:200-215.
37. Kather JN, Krisam J, Charoentong P, et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. *PLoS Med*. 2019;16(1):e1002730.
38. Geessink OGF, Baidoshvili A, Klaase JM, et al. Computer aided quantification of intratumoral stroma yields an independent prognosticator in rectal cancer. *Cell Oncol*. 2019;42(3):331-341.
39. Chalabi M, Fanchi LF, Dijkstra KK, et al. Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. *Nat Med*. 2020;26(4):566-576.
40. Chalabi M, Verschoor YL, Tan PB, et al. Neoadjuvant immunotherapy in locally advanced mismatch repair-deficient colon cancer. *N Engl J Med*. 2024;390(21):1949-1958.
41. Morton D, Seymour M, Magill L, et al. Preoperative chemotherapy for operable colon cancer: mature results of an International Randomized Controlled Trial. *J Clin Oncol*. 2023;41(8):1541-1552.