A retrospective analysis using deep-learning models for prediction of survival outcome and benefit of adjuvant chemotherapy in stage II/III colorectal cancer

Xingyu Li¹ · Jitendra Jonnagaddala² · Shuhua Yang¹ · Hong Zhang¹ · Xu Steven Xu³

Abstract

Purpose Most of Stage II/III colorectal cancer (CRC) patients can be cured by surgery alone, and only certain CRC patients benefit from adjuvant chemotherapy. Risk stratification based on deep-learning from haematoxylin and eosin (H&E) images has been postulated as a potential predictive biomarker for benefit from adjuvant chemotherapy. However, very limited success has been achieved in using biomarkers, including deep-learning-based markers, to facilitate the decision for adjuvant chemotherapy despite recent advances of artificial intelligence.

Methods We trained and internally validated CRCNet using 780 Stage II/III CRC patients from Molecular and Cellular Oncology. Independent external validation of the model was performed using 337 Stage II/III CRC patients from The Cancer Genome Atlas (TCGA).

Results CRCNet stratified the patients into high, medium, and low-risk subgroups. Multivariate Cox regression analyses confirmed that CRCNet risk groups are statistically significant after adjusting for existing risk factors. The high-risk subgroup significantly benefits from adjuvant chemotherapy. A hazard ratio (chemo-treated vs untreated) of 0.2 (95% Confidence Interval (CI), 0.05–0.65; P = 0.009) and 0.6 (95% CI 0.42–0.98; P = 0.038) are observed in the TCGA and MCO Fluorouracil-treated patients, respectively. Conversely, no significant benefit from chemotherapy is observed in the low- and medium-risk groups (P = 0.2–1).

Conclusion The retrospective analysis provides further evidence that H&E image-based biomarkers may potentially be of great use in delivering treatments following surgery for Stage II/III CRC, improving patient survival, and avoiding unnecessary treatment and associated toxicity, and warrants further validation on other datasets and prospective confirmation in clinical trials.

Keywords Deep learning · Whole-slide Images · MCO dataset · H&E Image · Adjuvant chemotherapy · Overall survival · Colorectal cancer

Introduction

Colorectal cancer (CRC) is the second most common cause of cancer-related death, with 1,880,725 new cases and 915,880 deaths worldwide in 2020 (Sung et al. 2021). For early stage CRC patients, the use of postoperative adjuvant chemotherapy is being debated since surgery is extremely effective, and the majority (approximately 80%) of patients can be cured by surgery alone (Schrag et al. 2002; Weiss et al. 2014; Jemal et al. 2004). In addition, significant side effects and an increased death rate have been associated with adjuvant chemotherapy (Gramont et al. 2012; André et al. 2009; Schmoll et al. 2015; Yothers et al. 2011). In practice, adjuvant chemotherapy is only recommended to patients...
with pathological T4 stage tumors, poorly differentiated tumor, vascular, lymphatic, or perineural invasion, number of lymph nodes sampling less than 12, or clinical presentation with intestinal occlusion or high risk of perforation (Benson et al. 2017; Schmoll et al. 2012).

Recent advances of artificial intelligence based on deep-learning algorithms have greatly improved the current pathological workflows (Bera et al. 2019). Kather et al. proposed CNN-based “deep stroma score” (DSS) prognostic factor for overall survival (OS) in Stage III/IV CRC patients (Kather et al. 2019). DSS could identify subgroups with more aggressive disease in Stage III/IV CRC patients although its prognostic value was limited in Stage I/II CRC. In addition, Danielsen et al. proposed an image-based “tumor-stroma fraction” that could predict prognosis in Stage II CRC (Danielsen et al. 2018). However, a more recent analysis showed that this biomarker was not a significant prognostic factor (Yang et al. 2020). Furthermore, Yao et al. proposed an attention-based multiple instance learning model to predict the survival of CRC patients (Yao et al. 2020). Also, Wulczyn et al. developed a deep learning system (DLS) for predicting disease-specific survival for stage II and III colorectal cancer. Most recently, Skrede et al. developed a prognostic marker to identify high-risk stage II and III CRC patients using deep learning of H&E slides, and postulated that the identified high-risk CRC could benefit from adjuvant chemotherapy (Skrede et al. 2020; Wulczyn et al. 2021). Nevertheless, so far, no analysis has been performed to demonstrate that the model-based risk stratification could predict the treatment effect of adjuvant chemotherapy and identify patients that could benefit from the chemotherapy after surgery. In this study, we use a novel deep-learning model (CRCNet) to integrate predictive imaging phenotypes from different tissue types of whole-slide H&E images (WSIs) from two large international CRC datasets from Molecular and Cellular Oncology (MCO) and The Cancer Genome Atlas (TCGA) studies. CRCNet stratifies patients into different risk categories and predicts prognoses of Stage II/III CRC patients. Additionally, our retrospective analysis provided further evidence that the risk stratification based on H&E images might predict the patients that potentially benefit from postoperative adjuvant chemotherapy and provide a useful tool for individualized guidance for treatment and patient care for Stage II/III CRC patients (Anonymous 2013; Jonnagaddala et al. 2016; Muzny et al. 2012).

Methods

The workflow of our convolutional neural networks for processing the whole-slide images and modeling the survival data is illustrated in Fig. 1.

Imaging and clinical data

The analysis included two large-scale datasets. Hematoxylin and eosin (H&E) stained whole-slide images (WSIs, 40x) were collected from both MCO and TCGA CRC studies. The MCO dataset consisted of patients who underwent curative resection for colorectal cancer between 1994 and 2010 in New South Wales, Australia. The Cancer Genome Atlas (TCGA) public dataset included the TCGA-COAD and TCGA-READ datasets. The MCO dataset (v15Jan2021) was used to train the deep learning model while the TCGA dataset obtained in July 2020 was used for external validation.

Deep learning-based risk stratification

The deep learning model (CRCNet) consisted of two sequential components: a tissue-type classifier and a deep multi-instance learning (MIL) survival model. Each whole-slide H&E image was preprocessed to (1) exclude the background area of each image using a U-Net, (2) split into non-overlapping tiles with a size of 224×224 pixels, and color normalized. An Xception model-based tissue-type classifier was fine-tuned and classified each image tile into one of eight tissue classes: adipose tissue (ADI), background (BACK), debris (DEB), lymphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal colon mucosa (NORM), cancer-associated stroma (STR), and colorectal adenocarcinoma epithelium (TUM). For each tissue type on the tissue map, a deep MIL survival model was developed based on a feature matrix (tiles×256) extracted from the last layer of the Xception model. A convolutional one-dimensional layer was used to estimate a score for each tile. The 10 highest and 10 lowest tiles scores of each tissue type were used to predict the patient’s risk score. For the MIL model of each tissue type, the patient was classified into high risk or low risk using the median risk score as a threshold. The top 2 models (tissue types) with the highest C-index were integrated to form an ensemble model and to refine the risk stratification into 3 categories: high risk (both tumor and stroma models = high risk); medium risk (either tumor or stroma model = high risk); and low risk (both tumor and stroma models = low risk).

Statistical analysis

Univariate and multivariate survival analyses for baseline clinical variables, molecular features, and CRCNet score were performed using Cox proportional hazards models (Cox PH model) implemented in the R package survival (Package and for Survival Analysis in R 2020). Log-rank tests were used to evaluate the statistical significance of the difference in survival distributions between subgroups.
C-index was used to assess the model predictive performance and compare different models. A Spearman correlation test was performed to assess the significance of correlation between CRCNet risk classification and existing risk factors. The results were internally validated using the MCO dataset with fivefold cross-validation strategy where folds were stratified based on disease stage. The MCO dataset was split into training (80%) and validation (20%) datasets. External validation was performed using the TCGA database that was kept entirely separate from the training process.

**Results**

**Patient characteristics in MCO and TCGA**

A total of 1,117 UICC TNM Stage II/III CRC patients are included in the analysis (Table 1): 780 patients are from the MCO dataset, and 337 patients are from the TCGA dataset (Compton and Greene 2004). The baseline patient characteristics and demographics are similar between the MCO and TCGA studies. The median age of the MCO dataset is 70 years (range 24–99 years) while the median of the TCGA dataset is 67 years (range 31–90 years). 53% and 51% of the MCO and the TCGA patients, respectively, are male. The MCO population consisted of 53% Stage II patients and 47% Stage III patients, whereas there are 54% and 46% Stage II and III patients, respectively, in the TCGA population. In addition, 16% and 21% of subjects from MCO and TCGA, respectively, have an MSI-H status. After a median follow-up of 59 months, 245 death events are recorded in the MCO dataset, while the number of deaths is 42 in the TCGA dataset following a median follow-up of 25.1 months. The median OS is 100 months (95% Confidence Interval (CI) 83.3 to infinity) for the TCGA patients while the median survival was not reached for the Stage II/III MCO patients at 60 months.

**CRCNet risk classification**

In the MCO dataset, 288 (37%), 185 (24%), and 307 (39%) of 780 patients are stratified into the low-, medium-, and high-risk groups, respectively, according to the risk scores predicted from tumor and stroma tissue compartments (see “Methods”). In the TCGA dataset, 133 (39%), 141(42%), and 63 (19%) of 337 Stage II/III patients are stratified into the low-, medium-, and high-risk groups, respectively (see Fig. 2).

**Prediction of overall survival**

The CRCNet consistently predicts the OS in Stage II/III CRC in both MCO and TCGA datasets (Fig. 2). Compared to patients in the low-risk group, patients with both medium and high CRCNet score have significant shorter OS in the MCO

---

**Table 1 Baseline patient characteristics of the training (MCO) and validation (TCGA) dataset populations**

| Characteristic          | MCO cohort (n = 780) | TCGA cohort (n = 337) |
|-------------------------|---------------------|-----------------------|
| Age, years              | 70 (24–99)          | 68 (31–90)            |
| Sex                     |                     |                       |
| Female                  | 361 (46%)           | 166 (49%)             |
| Male                    | 413 (53%)           | 171 (51%)             |
| Missing                 | 6 (1%)              | –                     |
| UICC Stage              |                     |                       |
| II                      | 416 (53%)           | 182 (54%)             |
| III                     | 364 (47%)           | 155 (46%)             |
| pN stage                |                     |                       |
| N0–N1                   | 647 (83%)           | 283 (84%)             |
| N2                      | 127 (16%)           | 54 (16%)              |
| Missing                 | 6 (1%)              | –                     |
| pT stage                |                     |                       |
| pT1–pT3                 | 595 (76%)           | 306 (91%)             |
| pT4                     | 179 (23%)           | 31 (9%)               |
| Missing                 | 6 (1%)              | –                     |
| Location                |                     |                       |
| Colon                   | 501 (64%)           | 254 (75%)             |
| Rectum                  | 279 (36%)           | 83 (25%)              |
| Adjuvant treatment      |                     |                       |
| No treatment            | 466 (60%)           | 135 (49%)             |
| Treatment               | 470 (60%)           | 111 (41%)             |
| 5-FU                    | 305 (39%)           | –                     |
| Other treatment         | 165 (12%)           | –                     |
| Missing                 | 243 (31%)           | 27 (10%)              |
| MSI                     |                     |                       |
| MSI–H                   | 124 (16%)           | 59 (18%)              |
| MSS                     | 642 (82%)           | 246 (73%)             |
| Missing                 | 14 (2%)             | 32 (9%)               |
| KRAS                    |                     |                       |
| Wild type               | 511 (66%)           | 168 (50%)             |
| Mutated                 | 251 (32%)           | 106 (31%)             |
| Missing                 | 18 (2%)             | 63 (19%)              |
| BRAF                    |                     |                       |
| Wild type               | 659 (84%)           | 237 (70%)             |
| Mutated                 | 100 (13%)           | 37 (11%)              |
| Missing                 | 20 (3%)             | 63 (19%)              |
| Venous invasion         |                     |                       |
| Absent                  | 585 (75%)           | 226 (67%)             |
| Present                 | 190 (25%)           | 74 (22%)              |
| Missing                 | –                   | 37 (11%)              |
| Lymphatic invasion      |                     |                       |
| Absent                  | 474 (61%)           | 174 (52%)             |
| Present                 | 306 (39%)           | 145 (43%)             |
| Missing                 | –                   | 18 (5%)               |

Data are median(range) or n (%)
Fig. 1 Flow-chart and global methodology of the study. The deep learning model (CRCNet) consisted of two sequential components: a tissue-type classifier and a deep multi-instance learning (MIL) survival model. Each whole-slide H&E image was preprocessed to (1) exclude the background area of each image using a Unet, (2) split into non-overlapping tiles with a size of 224 × 224 pixels, and color normalized. An Xception model-based tissue-type classifier was fine-tuned and classified each image tile into one of eight tissue classes: adipose tissue (ADI), background (BACK), debris (DEB), lymphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal colon mucosa (NORM), cancer-associated stroma (STR), and colorectal adenocarcinoma epithelium (TUM). For each tissue type on the tissue map, a deep MIL survival model was developed based on a feature matrix (tiles × 256) extracted from the last layer of the Xception model. A convolutional one-dimensional layer was used to estimate a score for each tile. The 10 highest and 10 lowest tiles scores of each tissue type were used to predict the patient’s risk score. For the MIL model of each tissue type, the patient was classified into high risk or low risk using the median risk score as a threshold. The top 2 models (tissue types) with the highest C-index (tumor and stroma: C-index = 0.61), were integrated to form an ensemble model and to refine the risk stratification into 3 categories: high risk (both tumor and stroma models = high risk); medium risk (either tumor or stroma model = high risk); and low risk (both tumor and stroma models = low risk).

When stratified by the UICC stage, clear separation of survival between different CRCNet risk groups are still observed in Stage II or III colorectal cancer patients from both MCO dataset (medium risk: HR (Hazard Ratio) = 1.67; 95% CI 1.16–2.40; \( P=0.0056 \); high risk: HR = 2.41; 95% CI 1.77–3.28; \( P<0.0001 \)) (Supplementary Table 1 and Fig. 2a). Similarly, the statistical significance is demonstrated in the high-risk group of the TCGA dataset (HR = 2.84; 95% CI 1.26–6.40; \( P=0.01 \)) (Supplementary Table 1 and Fig. 2b). The univariate analysis indicates the consistency and superiority in prediction performance of CRCNet compared to other reported risk factors such as age, sex, lymphovascular invasion, pT stage, pN stage, UICC stage, KRAS mutation, BRAF mutation as well as MSI status (Supplementary Table 1).
and TCGA studies (Supplementary Fig. 2). The CRCNet classifier also identifies significantly different survival within the pT stage (pT1–pT3 or pT4 disease; Supplementary Fig. 3). Similarly, CRCNet provides a consistent prediction in the subgroups of pN0—N1 and pN2 (Supplementary Fig. 4), where a substantial difference in OS is identified among different risk groups. In addition, our CRCNet can further separate the risk for MSS patients, and the high-risk MSS patients demonstrate a significantly poorer prognosis (Supplementary Fig. 5c and 5d). Within the MSI-H status, the survival is similar irrespective of risk groups (Supplementary Fig. 5a and 5b). The predictive ability of OS by CRCNet in subgroups of established risk parameters indicates that features beyond the current clinical and biomarker risk factors can be captured using our deep learning algorithm implemented in CRCNet.

In the multivariate Cox PH models (Table 2) combining CRCNet with other established clinical and molecular risk factors (i.e., pT stage, pN stage, MSI status, lymphovascular invasions, BRAF mutation, KRAS mutation, age, and sex), the CRCNet risk classifier remains statistically significant in both MCO (high risk vs low risk: HR = 2.14; 95% CI 1.56–2.93; P < 0.0001) and TCGA datasets (HR = 2.76; 95% CI 1.17–6.50; P = 0.02). The statistical significance of CRCNet in the multivariate setting confirms that CRCNet is a robust predictor of OS and provides extra predictive information in addition to the other existing risk parameters in the Stage II/III CRC patients.

More detailed information regarding prediction of overall survival is in Supplementary materials.

**Prediction of survival benefit from adjuvant chemotherapy**

The classification of high-risk Stage II/III CRC patients using deep-learning models may facilitate identification of patients who can benefit from adjuvant chemotherapy and help the treatment decision after surgery. We retrospectively examine the effect of postoperative adjuvant chemotherapy between patients who did or did not receive adjuvant chemotherapy in each CRCNet risk group (Fig. 3).

Overall, in the MCO dataset, among the Stage II/III CRC patients, 334 patients received adjuvant chemotherapy while 280 patients did not have chemotherapy. In the TCGA dataset, 111 Stage II/III patients received adjuvant chemotherapy and 135 patients did not have postoperative chemotheraphy. However, analysis shows that patients who received postoperative adjuvant chemotherapy do not have a significant survival benefit from the treatment compared to those who did not receive adjuvant chemotherapy in both MCO (HR = 0.9; 95% CI 0.67–1.17; P = 0.4; Fig. 3) and TCGA (HR = 0.5; 95% CI 0.26–1.12; P = 0.1; Fig. 3) studies. The 4-year survival is 74.8% with chemotherapy vs 70.8% without chemotherapy in the MCO dataset, while the 4-year survival is 85.8% with chemotherapy vs 77.7% without chemotherapy in the TCGA dataset. This suggests that there is still room for improvement in terms of current clinical decision for adjuvant chemotherapy.

CRCNet-based risk classification clearly predicts the benefit of chemotherapy in the TCGA dataset (Fig. 3a). The CRCNet low-risk patients who received postoperative chemotheraphy do not show any survival benefit compared to the low-risk patients not receiving chemotherapy (HR = 1.1; 95% CI 0.28–4.52; P = 0.87). The CRCNet medium-risk patients who had chemotherapy appear to have a numerical reduction in risk (HR = 0.9; 95% CI 0.32–2.6; P = 0.86), whereas in the CRCNet high-risk subgroup, adjuvant chemotherapy significantly reduces the risk of death (HR = 0.2; 95% CI 0.05–0.65; P = 0.009). No detailed information regarding drug regimens/classes for the adjuvant chemotherapy is provided in the TCGA dataset.
Similar trends are observed in the MCO dataset. The Forest plot analysis (Fig. 3b) reveals the predictive value of CRCNet for treatment effect of chemotherapy in terms of OS, where the CRCNet risk groups (low-to-high-risk) demonstrate a distinct, incremental response to chemotherapy. In the low-risk groups, adjuvant chemotherapy does not provide significant benefit of survival (HR = 1.0; 95% CI 0.57–1.77; \( P = 0.99 \)). Numerical improvement of survival is observed in the chemotherapy-treated medium-risk patients (HR = 0.8; 95% CI 0.47–1.47; \( P = 0.52 \)). In contrast, in the CRCNet high-risk group, there is a statistically borderline association of chemotherapy with survival (HR = 0.7; 95% CI 0.46–1.01; \( P = 0.057 \)). Further analysis of the MCO dataset demonstrate that patients in the CRCNet high-risk subgroup benefited from Fluorouracil (5FU) the most and 5FU is significantly associated with survival benefit (HR = 0.6; 95% CI 0.42–0.98; \( P = 0.038 \)). In patients who have medium risk according to CRCNet, there is a numerical improvement in survival when receiving 5FU (HR = 0.7; 95% CI 0.33–1.32; \( P = 0.24 \)).

The chemotherapy-treated patients in both MCO and TCGA data are significantly younger and more fit (73% and 78% < 70 year in MCO and TCGA, respectively; Supplementary Table 4). In addition, among patients who received adjuvant chemotherapy in both studies, only 8–9% patients are with an MSI-H status, and approximately 90% are MSS patients. Significantly, more pN2 patients are treated with adjuvant chemotherapy in both MCO (27% in treated group vs 7% in untreated) and TCGA (25% in treated vs 9% in untreated). However, even though more pT4 patients are allocated to chemotherapy (28% vs 17% in the untreated group), a similar proportion of pT4 patients are treated (11%) and untreated (7%) groups.

Due to the potential confounding between treatment allocations and clinical risk factors, we perform additional examination of the survival benefit of postoperative chemotherapy for different CRCNet risk groups after stratifying the patients by pTstage, pNstage, MSI status, and lymphovascular invasion. Given the small sample size in subpopulations, we pool the data from MCO and TCGA together for this analysis. The further Forest plot analysis within the subgroups of these clinical risk factors reveals that the trend between CRCNet risk groups and adjuvant chemotherapy remained (Supplementary Fig. 7). In general, more profound treatment benefit is observed in the CRCNet high-risk group in all subgroups (i.e., pT1–T3/pT4, pN0–pN1/pN2, MSI-H/MSS, age (≤70 years/>70 years), and lymphovascular invasion (yes/no)) compared to those in the CRCNet low- and medium-risk groups. However, it is worth mentioning that MSI-H patients have minimal benefit from chemotherapy regardless of the risk status for these patients based on the CRCNet model. This supports the NCCN recommendation that adjuvant chemotherapy is not needed for MSI-H patients.

**Table 2** Multivariate Cox Regression Model in MCO and TCGA Studies

| Variable          | MCO HR (95% CI) | P value | TCGA HR (95% CI) | P value |
|-------------------|-----------------|---------|------------------|---------|
| CRCNet score      |                 |         |                  |         |
| Low               | 1 (ref)         | –       | 1 (ref)          | –       |
| Medium            | 1.77 (1.22,2.55)| 0.002** | 1.89 (0.85,4.20) | 0.117   |
| High              | 2.14 (1.56,2.93)| < 0.0001*** | 2.76 (1.17,6.50) | 0.020*  |
| KRAS Mutated      | 0.84 (0.63,1.10)| 0.667   | 1.66 (0.81,3.41) | 0.166   |
| Wild              | 1 (ref)         | –       | 1 (ref)          | –       |
| MSI Status        |                 |         |                  |         |
| MSS               | 1 (ref)         | –       | 1 (ref)          | –       |
| MSI-H             | 1.20 (0.76,1.90)| 0.442   | 1.41 (0.58,3.43) | 0.454   |
| Sex               |                 |         |                  |         |
| Male              | 1 (ref)         | –       | 1 (ref)          | –       |
| Female            | 0.69 (0.53,0.90)| 0.005** | 1.10 (0.58,2.10) | 0.767   |
| Age               | 1.04 (1.03,1.05)| 0.0001*** | 1.02 (1.00,1.05) | 0.074   |
| Lymphovascular invasion |     |         |                  |         |
| No                | 1 (ref)         | –       | 1 (ref)          | –       |
| Yes               | 1.75 (1.33,2.30)| < 0.0001*** | 1.36 (0.68,2.73) | 0.390   |
| pT stage          |                 |         |                  |         |
| T1–T3             | 1 (ref)         | –       | 1 (ref)          | –       |
| T4                | 1.58 (1.20,2.08)| 0.001** | 3.10 (1.37,7.03) | 0.007** |
| pN stage          |                 |         |                  |         |
| N0–N1             | 1 (ref)         | –       | 1 (ref)          | –       |
| N2                | 1.95 (1.46,2.62)| < 0.0001*** | 2.28 (1.09,4.79) | 0.029*  |

C-index of the MCO multivariate model = 0.715; C-index of the TCGA multivariate model = 0.709

\( * P < 0.05, ** P < 0.01, *** P < 0.0001 \)

**Discussion**

The most challenging question in the adjuvant setting for CRC is which patients should receive chemotherapy, as most of the patients can be cured by surgery alone. Currently, the
treatment decision is mainly based on clinical and pathological staging. Although novel biomarkers based on genetic mutation status, gene expression profiling, and immunohistochemistry have been developed to facilitate the decision for adjuvant chemotherapy, limited success has been achieved due to moderate prognostic accuracy and/or lack of prediction ability for treatment benefit (Guinney et al. 2015; Galon et al. 2014). Skrede et al. indicated that H&E image-based high-risk CRC might benefit better from adjuvant chemotherapy (Skrede et al. 2020). In this manuscript, we developed a novel deep learning model based on whole-slide H&E images to predict clinical outcome for Stage II/III CRC and provided further evidence that image-based risk stratification may predict the treatment effect of adjuvant chemotherapy.

In both MCO and TCGA studies, the CRCNet identified high-risk subgroups benefited from adjuvant chemotherapy the most, and the OS in these high-risk patients who received chemotherapy (particularly 5FU in the MCO dataset) after surgery was significantly better than that in those who were not given postoperative chemotherapy. In contrast, the CRCNet identified low- and medium-risk patients did not respond to adjuvant chemotherapy or only had numerical improvement with chemotherapy vs. without chemotherapy. In the CRCNet identified high-risk subgroup, 61% patients from the TCGA dataset and 51% from the MCO dataset

*Fig. 3* Forest plot showing the effect of adjuvant chemotherapy on overall survival among patients with different risk categories according to CRCNet: **a** TCGA and **b** MCO
were not offered any chemotherapy after surgery. Those subjects could have longer survival if they were given adjuvant chemotherapy. On the other hand, those patients who were in the CRCNet low- and medium-risk subgroups and received adjuvant chemotherapy might be able to avoid unnecessary treatment and chemo-associated toxicity. Therefore, our model strongly suggests that H&E image-based biomarkers like CRCNet may be a predictive biomarker that may facilitate selection of high-risk Stage II/III CRC patients who can benefit from adjuvant chemotherapy. However, due to the retrospective nature of this analysis, further research and validation of H&E image-based biomarkers on other datasets and prospective, randomized clinical trials is warranted.

Several assays based on gene expression profiling (OncoType DX, ColoPrint, and ColDx), immunohistochemistry (e.g., validated Immunoscore according to CD3 + and CD8 + immune cell densities), and post-surgical circulating tumor DNA (ctDNA) have been developed to provide prognostic and predictive information to facilitate the decisions regarding adjuvant therapy in Stage II/III CRC patients (Gray et al. 2011; O’Connell et al. 2010; Yothers et al. 2013; Mlecnik et al. 2020; Tie et al. 2019; Reinert et al. 2019). However, all these assays may be limited by the complex sample preparation processes, high cost, and issues like large batch effects, etc. In clinical practice, histological slides of tumor tissue, particularly H&E slides, are available for almost every subject. Therefore, image-based approaches can be a great alternative to genomic, transcriptomic, or immunohistochemistry assays due to its accessibility and low cost. Furthermore, automation of image-based workflows largely eliminates human involvement, shortens the image processing time, and allows quick turnaround for clinical decisions.

Taken together, we developed a CNN-based deep-learning model on whole-slide images from CRC patients. This model integrates a supervised tissue type classification and a deep-learning approach to extract prognostic features from small H&E image tiles, which allows for prediction of survival of Stage II/III CRC patients and identification of high-risk Stage II/III CRC patients who may potentially benefit from adjuvant chemotherapy. The retrospective analysis provides further evidence that coupling deep-learning models and H&E images can potentially reveal novel predictive biomarkers that offer better decision-making for therapeutic allocations for Stage II/III CRC patients.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00432-022-03976-5.

Acknowledgements The research of Xingyu Li, Shuhua Yang, and Hong Zhang was partially supported by National Natural Science Foundation of China (No. 11771096, 72091212), Anhui Center for Applied Mathematics, and Special Project of Strategic Leading Science and Technology of CAS (No. XDC08010100). Jitendra Jonnagaddala is funded by the Australian National Health and Medical Research Council (No. GNT1192469), Jitendra Jonnagaddala is also supported by the Google Cloud Research Credits program (No. GCP19980904). We thank Michelle Xu (Princeton Day School) for writing assistance and language editing.

Author contributions XSX, XL, and HZ contributed to design of the research; JJ, XL, and XSX contributed to data acquisition; XL, XSX, and SY contributed to data analysis. XL, XSX, JJ, and HZ contributed to data interpretation. XL, XSX, JJ, and HZ wrote the manuscript; and all authors critically reviewed the manuscript and approved the final version.

Funding National Natural Science Foundation of China (No. 11771096, 72091212, 12171451), Anhui Center for Applied Mathematics, and Special Project of Strategic Leading Science and Technology of CAS (No. XDC08010100). Australian National Health and Medical Research Council (No. GNT1192469). Google Cloud Research Credits program with the award (No. GCP19980904).

Data availability The TCGA dataset is publicly available at the TCGA portal (https://portal.gdc.cancer.gov). The public TCGA clinical data is available at the website (https://xenabrowser.net/datapages/). Xception model weights are available at (https://github.com/collet/dip_learning-models/releases/download/v0.4/xception_weights_tf_dim_ordering_kernels_notop.h5). The MCO dataset is available through the SREDH Consortium (www.sredhconsortium.org), which was used with permission in our current study.

Code availability Source code is available at https://github.com/1996lixingyu1996/CRCNet.

Declarations

Conflict of interest The authors declare no potential conflicts of interest.

Software All statistical analysis were conducted in R (version 4.1.0) unless otherwise specified. The following libraries were used: survminer (version 0.4.9), survival (version 3.2–13), and ggplot2 (version 3.3.5). The U-Net model, tissue classifier, risk group predictor were trained with Python (version 3.7.9), TFI-nightly-gpu (version 2.5.0.dev20210209), scipy (version 1.6.1), scikit-learn (version 0.24.1), openslide-python (version 1.1.2), openencv-python (version 4.5.1.48), numpy (version 1.19.5), numba (version 0.52.0), matplotlib (version 3.3.4), pandas (version 1.2.2), and torchvision (version 0.8.2). All training parameters were provided in the source code available at https://github.com/1996lixingyu1996/CRCNet. Source code is available at https://github.com/1996lixingyu1996/CRCNet.

References

A Package for Survival Analysis in R, 2020
André T, Boni C, Navarro M et al (2009) Improved overall survival with oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment in stage II or III colon cancer in the MOSAIC trial. J Clin Oncol 27(19):3109–3116. https://doi.org/10.1200/jco.2008.20.6771
Anonymous (2013) Data from: MCO study tumour collection. https://doi.org/10.4225/53/5559205bea135
Benson AB III, Venook AP, Cederquist L et al (2017) Colon Cancer, Version 1.2017, NCCN Clinical Practice Guidelines in Oncology.
J Nat Comprehensive Cancer Netw JNCCN. 15(3):370–398. https://doi.org/10.6004/jnccn.2017.0036
Bera K, Schalper KA, Rimm DL, Velchetti V, Madabhushi A (2019) Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. Nat Rev Clin Oncol 16(11):703–715. https://doi.org/10.1038/s41571-019-0252-y
Compton CC, Greene FL (2004) The staging of colorectal cancer: 2004 and beyond. CA Cancer J Clin 54(6):295–308. https://doi.org/10.3322/canclin.j.54.6.295
Danielsen HE, Hveem TS, Domingo E et al (2018) Prognostic markers for colorectal cancer: estimating ploidy and stroma. Ann Oncol 29(3):616–623. https://doi.org/10.1093/annonc/mdx794
de Gramont A, Van Cutsem E, Schmoll HJ et al (2012) Bevacizumab plus oxaliplatin-based chemotherapy as adjuvant treatment for colon cancer (AVANT): a phase 3 randomised controlled trial. Lancet Oncol 13(12):1225–1233. https://doi.org/10.1016/s1470-2045(12)70509-0
Galon J, Mlecnik B, Bindea G et al (2014) Towards the introduction of the “Immunoscore” in the classification of malignant tumours. J Pathol 232(2):199–209. https://doi.org/10.1002/path.4287
Gray RG, Quirke P, Handley K et al (2011) Validation study of a quantitative multigene reverse transcriptase-polymerase chain reaction assay for assessment of recurrence risk in patients with stage II colon cancer. J Clin Oncol 29(35):4611–4619. https://doi.org/10.1200/jco.2010.32.8732
Guinney J, Dienstmann R, Wang X et al (2015) The consensus molecular subtypes of colorectal cancer. Nat Med 21(11):1350–1356. https://doi.org/10.1038/nm.3967
Jemal A, Tiwari RC, Murray T et al (2004) Cancer statistics, 2004. CA Cancer J Clin 54(1):8–29. https://doi.org/10.3322/canclin.j.54.1.8
Jonnagaddala J, Croucher JL, Jue TR et al (2016) Integration and diction of colorectal cancer outcome: a discovery and validation study. Lancet Oncol 17(11):e478–e486. https://doi.org/10.1016/s1470-2045(16)30101-x
Kather JN, Krisam J, Charoentong P et al (2019) Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. PLoS Med 16(1):e1002730. https://doi.org/10.1371/journal.pmed.1002730
Mlecnik B, Bifulco C, Bindea G et al (2020) Multicenter international society for chemotherapy of cancer study of the consensus immunoscore for the prediction of survival and response to chemotherapy in stage III colon cancer. J Clin Oncol 38(31):3638–3651. https://doi.org/10.1200/jco.19.03205
Murray DM, Bainbridge MN, Chang K et al (2012) Comprehensive molecular characterization of human colon and rectal cancer. Nature 487(7407):330–337. https://doi.org/10.1038/nature11252
O’Connell MJ, Lavery I, Yothers G et al (2010) Relationship between tumor gene expression and recurrence in four independent studies of patients with stage II/III colon cancer treated with surgery alone or surgery plus adjuvant fluorouracil plus leucovorin. J Clin Oncol 28(25):3937–3944. https://doi.org/10.1200/jco.2010.28.9538
Reinert T, Henriksen TV, Christensen E et al (2019) Analysis of plasma cell-free DNA by ultradepth sequencing in patients with stages I to III colorectal cancer. JAMA Oncol 5(8):1124–1131. https://doi.org/10.1001/jamaoncol.2019.0528
Schmoll HJ, Van Cutsem E, Stein A et al (2012) ESMO Consensus Guidelines for management of patients with colon and rectal cancer. A personalized approach to clinical decision making. Ann Oncol 23(10):2479–2516. https://doi.org/10.1093/annonc/mds236
Schmoll HJ, Tabernero J, Maroun J et al (2015) Capecitabine plus oxaliplatin compared with fluorouracil/folinic acid as adjuvant therapy for stage III colon cancer: final results of the NO16968 randomized controlled phase III trial. J Clin Oncol 33(32):3733–3740. https://doi.org/10.1200/jco.2015.60.9107
Schrag D, Rifa-Shiman S, Salz L, Bach PB, Begg CB (2002) Adjuvant chemotherapy use for Medicare beneficiaries with stage II colon cancer. J Clin Oncol 20(19):3999–4005. https://doi.org/10.1200/jco.2002.11.084
Skrede OJ, De Raedt S, Kleppe A et al (2020) Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. Lancet (london, England) 395(10221):350–360. https://doi.org/10.1016/j.lancet.2019.32998-8
Sung H, Ferlay J, Siegel RL et al (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 71(3):209–249. https://doi.org/10.3322/caac.21660
Tie J, Cohen JD, Wang Y et al (2019) Circulating tumor DNA analyses as markers of recurrence risk and benefit of adjuvant therapy for stage III colon cancer. JAMA Oncol 5(12):1710–1717. https://doi.org/10.1001/jamaoncol.2019.3616
Weiss JM, Schumacher J, Allen GO et al (2014) Adjuvant chemotherapy for stage II right-sided and left-sided colon cancer; analysis of SEER-medicare data. Ann Surg Oncol 21(6):1781–1791. https://doi.org/10.1245/s10434-014-3631-8
Wulczyn E, Steiner DF, Moran M et al (2021) Interpretable survival prediction for colorectal cancer using deep learning. Npj Digit Med 4(1):71. https://doi.org/10.1038/s41746-021-00427-2
Yang L, Chen P, Zhang L et al (2020) Prognostic value of nucleotyping, DNA ploidy and stroma in high-risk stage II colon cancer. Br J Cancer 123(6):973–981. https://doi.org/10.1038/s41416-020-0974-8
Yao J, Zhu X, Jonnagaddala J, Hawkins N, Huang J (2020) Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. Med Image Anal 65:101789. https://doi.org/10.1016/j.media.2020.101789
Yothers G, O’Connell MJ, Allegra CJ et al (2011) Oxaliplatin as adjuvant therapy for colon cancer: updated results of NSABP C-07 trial, including survival and subset analyses. J Clin Oncol 29(28):3768–3774. https://doi.org/10.1200/jco.2011.36.4539
Yothers G, O’Connell MJ, Lee M et al (2013) Validation of the 12-gene colon cancer recurrence score in NSABP C-07 as a predictor of recurrence in patients with stage II and III colon cancer treated with fluorouracil and leucovorin (FU/LV) and FU/LV plus oxaliplatin. J Clin Oncol 31(36):4512–4519. https://doi.org/10.1200/jco.2012.47.3116

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.