



Article

Learn to Estimate Genetic Mutation and Microsatellite Instability with Histopathology H&E Slides in Colon Carcinoma

Yimin Guo ^{1,†} , Ting Lyu ^{1,†}, Shuguang Liu ¹, Wei Zhang ¹, Youjian Zhou ¹, Chao Zeng ^{1,*} and Guangming Wu ^{2,*} ¹ The Eighth Affiliated Hospital, Sun Yat-sen University, Shenzhen 518000, China² Center for Spatial Information Science, The University of Tokyo, Kashiwa 277-8568, Japan

* Correspondence: zengch35@mail.sysu.edu.cn (C.Z.); huster-wgm@csis.u-tokyo.ac.jp (G.W.)

† These authors contributed equally to this work.

Simple Summary: Colorectal cancer is one of the most common malignancies and the third leading cause of cancer-related mortality worldwide. Identifying KRAS, NRAS, and BRAF mutations and MSI status are closely related to the individualized therapeutic judgment and oncologic prognosis of CRC patients. In this study, we introduced a cascaded network framework with an average voting ensemble strategy to sequentially identify the tumor regions and predict gene mutations & MSI status from whole-slide H&E images. Experiments on a colorectal cancer dataset indicated that the proposed method can achieve high fidelity in both gene mutation prediction and MSI status estimation. In our testing set, the AUCs for KRAS, NRAS, BRAF, and MSI were ranged from 0.794 to 0.897. The results suggested that the deep convolutional networks have the potential to assist pathologists in prediction of gene mutation & MSI status in colorectal cancer.



Citation: Guo, Y.; Lyu, T.; Liu, S.; Zhang, W.; Zhou, Y.; Zeng, C.; Wu, G. Learn to Estimate Genetic Mutation and Microsatellite Instability with Histopathology H&E Slides in Colon Carcinoma. *Cancers* **2022**, *14*, 4144. <https://doi.org/10.3390/cancers14174144>

Academic Editor: Luca Roncucci

Received: 30 July 2022

Accepted: 22 August 2022

Published: 27 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Colorectal cancer is one of the most common malignancies and the third leading cause of cancer-related mortality worldwide. Identifying KRAS, NRAS, and BRAF mutations and estimating MSI status is closely related to the individualized therapeutic judgment and oncologic prognosis of CRC patients. In this study, we introduce a cascaded network framework with an average voting ensemble strategy to sequentially identify the tumor regions and predict gene mutations & MSI status from whole-slide H&E images. Experiments on a colorectal cancer dataset indicate that the proposed method can achieve higher fidelity in both gene mutation prediction and MSI status estimation. In the testing set, our method achieves 0.792, 0.886, 0.897, and 0.764 AUCs for KRAS, NRAS, BRAF, and MSI, respectively. The results suggest that the deep convolutional networks have the potential to provide diagnostic insight and clinical guidance directly from pathological H&E slides

Keywords: deep convolutional network; H&E slice; gene mutation prediction; microsatellite instability; colon carcinoma

1. Introduction

Colorectal cancer (CRC) is one of the most common lower gastrointestinal malignancies and is currently the third leading cause of cancer-related mortality worldwide [1,2]. Despite the over survival rate of colorectal cancer has increased in recent years due to the improved treatment strategies [3], distant metastasis is still a significant cause of high morbidity and mortality for CRC patients [4]. So far, various predominant environmental risk factors for the development of CRC have been identified, including diet, obesity, lack of physical activity, and inflammatory bowel disease [5]. However, a module formed by the interaction of multiple genetic alterations determines individual differences and tumor progression in CRC patients.

In the past decades, a deep understanding of molecular profiles has been more significant for selecting appropriate therapies for metastatic CRC patients [6]. Numerous frequent

genetic mutations have been identified as critical drivers responsible for comprehensive therapeutic judgment and oncologic prognosis [7]. Mutations of RAS (i.e., exon 2, 3, and 4 of KRAS, exon 2 and 3 of NRAS) are considered negative predictors for targeted therapy with anti-EGFR monoclonal antibodies (e.g., cetuximab and panitumumab) [8,9]. Mutation of BRAF V600E is a worse prognostic biomarker. Patients with BRAF V600E mutation will be less likely to respond to treatment with cetuximab and panitumumab unless combined with a BRAF inhibitor [10,11]. Moreover, the microsatellite instability (MSI) status of CRC patients is also an important marker closely related to the assessment of prognosis, the efficacy of chemotherapeutic and immunity therapy [12,13]. Therefore, all metastatic CRC patients are suggested to detect the KRAS, NRAS, and BRAF mutations and MSI status according to the National Comprehensive Cancer Network (NCCN) clinical practice guidelines in oncology (Colon Cancer, Version 2.2021) [14].

The general diagnosis procedure of molecular pathology includes Sanger sequencing, Next-Generation Sequencing (NGS), ARMS-PCR, and digital PCR, etc. [15]. In recent years, the accuracy and sensitivity of those methods have been significantly improved. However, molecular detection remains limited by various factors such as sample quality, mutated gene abundance, and laboratory conditions. Moreover, in a short period of time, high testing prices are also a heavy burden for most families.

With the development of big data and deep convolutional network, artificial intelligence (AI)-assisted pathological diagnosis has attracted more and more attention. In 2018, Coudray et al. trained a deep convolutional neural network on Whole-Slide Images (WSIs) to predict the cancer subtype and gene mutations in lung cancer [16]. Later, MSI status estimation of CRC from H&E histology was reported [17,18]. Furthermore, Skrede et al. exhibited a promising result in the survival risk interpretation of tumor patients based on artificial intelligence [19]. These methods have significantly extended the application capability of deep convolutional networks. However, genetic mutation prediction from H&E slices in CRC, which has more clinical significance in precision diagnosis, is still very challenging. To fulfill this demand and further explore the potential of H&E slides, we propose a cascaded deep convolutional framework to simultaneously generate gene mutation predicting and MSI status estimation using WSIs in colorectal cancer. The proposed method consists of two tumor region classification models, gene mutation& MSI status estimation models, and an average voting ensemble strategy. The effectiveness of the proposed method is demonstrated by a CRC dataset collected from GDC Data Portal and Eighth Affiliated Hospital, Sun Yat-sen University (see Section 2.1). In qualitative and quantitative evaluation, the proposed method reveals promising accuracy in tumor classification (0.939–0.976 AUC), gene mutation prediction (0.792–0.897 AUC), and MSI status estimation (0.764 AUC).

The main contributions of this study can be summarized as follows:

- We proposed a cascaded deep convolutional framework to simultaneously generate gene mutation prediction and MSI status estimation in colorectal cancer.
- We introduced a simple yet efficient average voting ensemble strategy to produce high fidelity gene mutation prediction and MSI status estimation of the WSI.
- We further analyzed the effectiveness of the number of features selected for model ensembling to understand its effects on the performances of deep CNN models.

The rest of the paper is organized as follows: Firstly, we present the datasets and methods used for this research in Section 2. Then, we illustrate the quantitative and qualitative results in Section 3. Finally, discussion and conclusion are presented in the Sections 4 and 5, respectively.

2. Materials and Methods

2.1. Data

To explore the possibility of estimating somatic mutations and microsatellite instability (MSI) using Hematoxylin-Eosin(H&E) stained whole-slide image (WSI), we downloaded diagnostic slides and corresponding clinical data of the TCGA-COAD cohort from GDC Data Portal (<https://portal.gdc.cancer.gov/projects/TCGA-COAD>, accessed at 20 Febru-

ary 2022). The pre-compiled somatic mutation data and MSI status data were acquired from UCSC Xena (<https://xenabrowser.net/datapages/>, accessed at 10 March 2022) and MSIsensor-pro [20], respectively. The original WSIs were formatted in a magnification ratio of either 20× or 40×. Prior to performing our experiments, we manually resize the 40× images to 20× using libvips (<https://github.com/libvips/libvips>) (see Figure 1A–D). There were 292 WSIs with corresponding somatic mutations and MSI statuses in the TCGA-COAD dataset. To achieve better generalization, we also collected the SYSU8H dataset with the cooperation of The Eighth Affiliated Hospital, Sun Yat-sen University. The selected pathological specimens were fixed in formalin, embedded in paraffin wax block, and cut by several consecutive slices in 3–5 μm by a Leica HistoCore Autocut. Later, the slices were used for Hematoxylin-Eosin (H&E) staining, IHC staining, or gene sequencing, separately. Compared with the scanned H&E slices, the tumor areas for the sequencing slices are in micron-level drifts that tumor genomic heterogeneity among these slices is negligible. There were total 104 WSIs captured with 20× magnification ratio by PANNORAMIC 1000, 3DHISTECH Ltd. (see Figure 1E–H). Unlike next-generation sequencing (NGS) of TCGA-COAD, in the SYSU8H dataset, the genetic information was obtained by sanger sequencing. The binary masks of tumor areas of the WSIs were carefully annotated by experienced pathologists using QGIS (v3.22.7 LTR, <https://qgis.org/>).

As shown in Table 1, the 396 WSIs samples were randomly divided into training, validation, and testing groups with the ratios of 70%(278), 15%(59), and 15%(59), respectively. At 5× magnification WSIs, there were 283,126, 49,988, and 55,787 tiles within the corresponding training, validating, and testing set. At 10× magnification WSIs, 1,152,481, 203,183, and 2,275,595 tiles were within the corresponding training, validating, and testing set. In our experiment, the size of each tile was set to 512 × 512 pixels.

Table 1. Distribution of patients and whole-side images samples.

		WSIs			Overall (n = 396)
		Train (n = 278)	Val (n = 59)	Test (n = 59)	
Age(year)	Min.	22	29	36	22
	Max.	90	90	90	90
	Median	65.5	67	68	66
Gender	Male	138	30	33	201
	Female	140	29	26	195
KRAS	W.T.	162	24	35	221
	M.T.	116	35	24	175
NRAS	W.T.	267	57	57	381
	M.T.	11	2	2	15
BRAF	W.T.	251	51	51	353
	M.T.	27	8	8	43
MSI	MSI-H	236	49	52	337
	MSS/MSI-L	42	10	7	59
Tiles	5× Mag.	283,126	49,988	55,787	388,901
	10× Mag.	1,152,481	203,183	227,595	1,583,259

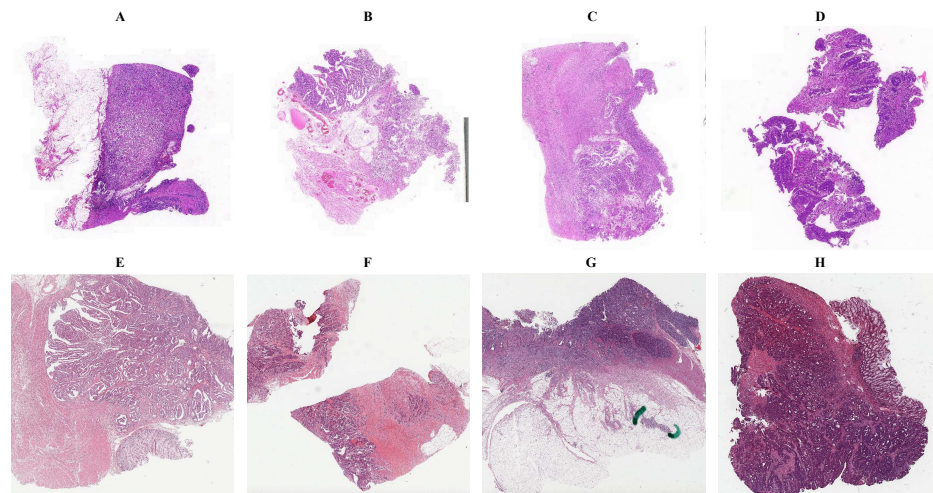


Figure 1. Representative H&E stained whole-side images (WSIs) from SYSU8H and TCGA-COAD dataset. The (A–D) and (E–H) samples are randomly selected from SYSU8H and TCGA-COAD datasets, respectively.

2.2. Methodology

In this study, we proposed a cascaded network framework to directly estimate somatic gene mutation and microsatellite instability status from the H&E stained whole-side image.

As shown in Figure 2, at the training stage, WSIs and corresponding binary masks of the training and validation set were partitioned into $5\times$ or $10\times$ tiles for training and validating the tumor classifier. The annotated tumor tiles and their somatic gene mutations or microsatellite instability (MSI) were used for training a binary classifier to discriminate wild type (i.e., W.T.) vs. mutant type (i.e., M.T.) of the gene or MSI-H vs. MSS/MSI-L, respectively. The top N highest probabilities of all tiles within a WSI were used to generate the final prediction for the patient.

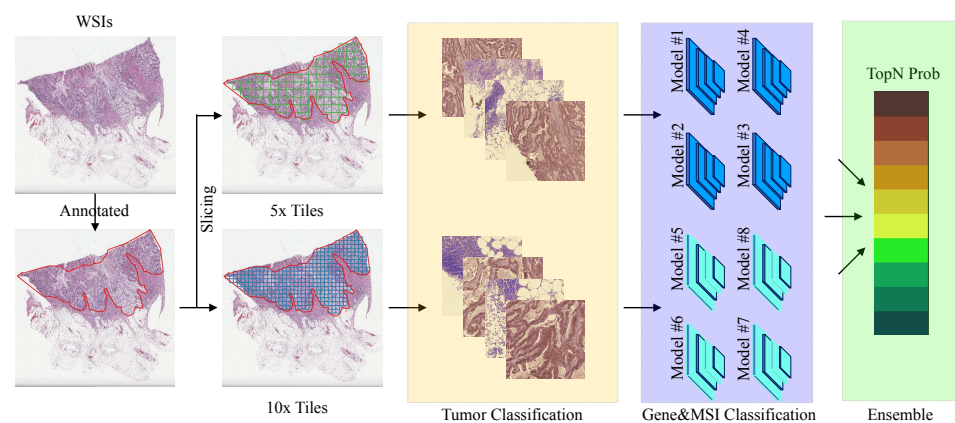


Figure 2. Experimental workflow for estimating somatic gene mutation and microsatellite instability with H&E stained whole-side images. The $5\times$ or $10\times$ tiles from WSIs will be accessed by a tumor classifier, a gene&MSI classifier, and a TopN ensemble classifier.

Through several cycles of training and validation, the hyperparameters, including batch size, the number of iterations, and learning rate, were optimized with the Adam stochastic optimizer [21]. Subsequently, the predictions generated by the optimized models were evaluated using the WSIs of the test set (see details in Table 1). For performance evaluations, we carefully measured the area under the receiver operator characteristic (ROC) curve [22] and its confidence interval (CI) [23].

2.2.1. Data Preprocessing

At first, the 396 pairs of whole-side images (WSIs) and their corresponding clinical records were shuffled and partitioned into three groups: training (70%), validating (15%), and testing (15%). Within each pair, a binary tumor mask of WSI was generated through polygon rasterization of its manually created tumor annotation. Later, a square window of 512×512 pixels was applied to the whole-side image and the corresponding tumor mask to extract paired tiles of WSI and mask. Then, each tile of WSI was labeled according to the positive ratio of pixels of the tumor mask. To focus on the tumor regions, tiles with positive ratios less than 80% were marked as 0. Otherwise, tiles were marked as 1. There were 388,901 and 1,583,259 tiles extracted from $5\times$ and $10\times$ magnification. As shown in Table 1, at $5\times$ magnification, there were 283,126, 49,988, and 55,787 tiles within the training, validating, and testing set. While at $10\times$ magnification, the number of tiles used for training, validation, and testing was 1,152,481, 203,183, and 2,275,595, respectively.

2.2.2. Network Architectures

For simplicity and efficiency, we adopted an advanced convolutional neural network (CNN) architecture, i.e., EfficientNet [24], as a backbone for tumor classification and gene&MSI classification.

In 1998, Lecun et al. introduced the classic CNN architecture, LetNet-5 [25], which consists of two sets of convolutional & pooling layers, a flattening convolutional layer, and two fully-connected layers. The CNN reveals two important concepts, sparse connectivity and shared weights, significantly reducing memory occupation and promoting computational efficiency. With the growing complexity of the dataset and rapid development of computational capacity, computer scientists have proposed more advanced CNN architectures for better generalization capacity and computational efficiency [26]. These architectures significantly promote CNN performance by introducing well-designed novel strategies, such as network in network (i.e., NIN) [27], residual learning (i.e., ResNet) [28], inception architecture [29], and dense connection (i.e., DenseNet) [30]. Differ from the above-mentioned models, which mainly focus on model accuracy, the EfficientNet architecture is designed to get a present accuracy level with limited computational operations. The EfficientNet introduces a uniformed scaling method that scales all dimensions of depth, width, and resolution with a set of fixed scaling coefficients [24].

In our experiments, we chose an ImageNet-1K [31] pretrained EfficientNet B0 (https://pytorch.org/vision/master/models/generated/torchvision.models.efficientnet_b0.html, accessed at 4 March 2022) as the backbone for both tumor classification and Gene&MSI classification. As shown in Table 2, we introduced a dropout layer ($p = 0.5$) [32] to prevent overfitting. Then, we replaced the dimensions of fully-connected (FC) layer from 1280×1000 to 1280×1 .

Subsequently, the activation function was changed from softmax to sigmoid.

$$z_i = b + \sum_{j=1}^c w_j \times x_{i,j}$$

$$p_i = \frac{1}{1 + e^{-z_i}}$$
(1)

The $w \in \mathbb{R}^c$ and $b \in \mathbb{R}^1$ denote the weights and bias, respectively. The range of prediction p_i is limited to $[0, 1]$.

Instead of binary cross entropy [33], we adopted focal loss [34] as our object function to focus learning on hard misclassified examples and address class imbalance. The equation can be formulated as:

$$p_t = \begin{cases} p_i, & \text{if } y_i = 1 \\ 1 - p_i, & \text{if } y_i = 0 \end{cases}$$

$$Loss_{focal} = -(1 - p_t)^\gamma \log(p_t)$$
(2)

where p_i and y_i is the i th prediction and corresponding ground truth. The value of p_t is p_i if the observation is in class 1; otherwise, the value is $1 - p_i$. The γ (≥ 0) is a tunable focusing parameter which reduces the relative loss for well-classified examples (i.e., $p_t > 0.5$) and puts more focus on hard, misclassified examples.

Table 2. The backbone network for both tumor classification and Gene&MSI classification. Each row describes the stage, operation, input resolution, output channel, and the number of layers.

Stage	Operator	Resolution	Channels	Layers
0		512 × 512	3	0
1	Conv3 × 3	512 × 512	32	1
2	MBCConv1, k3 × 3	256 × 256	16	1
3	MBCConv6, k3 × 3	256 × 256	24	2
4	MBCConv6, k5 × 5	128 × 128	40	2
5	MBCConv6, k3 × 3	64 × 64	80	3
6	MBCConv6, k5 × 5	32 × 32	112	3
7	MBCConv6, k5 × 5	32 × 32	192	4
8	MBCConv6, k3 × 3	16 × 16	320	1
9	Conv1 × 1&Pooling	16 × 16	1280	1
10	Dropout&FC	1280 × 1	1	1

With all of the above layers being trained by mini-batch stochastic gradient descent (SGD) [35] to minimize the focal loss, the model learns how to map from the input 512 × 512 RGB image to a binary prediction.

2.2.3. Model Ensemble

To make a decisive conclusion on the whole-slide-image (WSIs) using the separated predictions of 5 × and 10 × tiles, we introduced a simple yet efficient average voting strategy using the top N number of features to ensemble models. To ensure the high fidelity of selected features, a high threshold (i.e., 0.8) was used to filter out tiles with a low probability of being a tumor region. Later, tiles with a high probability of being tumor regions were passed to corresponding gene&MSI classification models to generate predictions of 5 × tiles (P_{x5}) and 10 × tiles (P_{x10}). Then, the top N highest probabilities of predictions from both 5 × and 10 × tiles were selected for the final estimation of the WSI (P_{wsi}). Finally, the P_{wsi} and corresponding ground truth (Y_{wsi}) were used to calculate the area under the curve (AUC) for performance estimation.

$$P_{topN} = \max([P_{x5}, P_{x10}], N)$$

$$P_{wsi} = \frac{1}{N} \sum_{i=1}^N P_{topN} \quad (3)$$

3. Results

A total of 396 colorectal cancer (CRC) patients with various gene mutations and MSI status from the SYSU8H and TCGA-COAD datasets were recruited in this study. The collected WSIs were randomly split into three sets: training, validation, and testing with the ratio of 70%, 15%, and 15%, respectively. The tiles extracted from the training and validation set were used for training and optimizing hyperparameters of the proposed classification models. In order to estimate the performance of the proposed classification models, we have conducted heavy quantitative and qualitative comparisons on the testing set. All experiments were performed on the same dataset and processing platform.

3.1. Tumor Classification

The tumor regions annotated by the pathologist and probability maps generated by the tumor classification models using 5 × and 10 × tiles of WSIs are presented in Figure 3. Both 5 × and 10 × models display high fidelity in tumor recognition compared to manual

annotations. Compared with the $5\times$ model, the model trained with $10\times$ tiles shows fewer false positives (e.g., orange and red patches outside the blue dashed curve of A, B, and C), fewer false negatives (e.g., blue and green patches inside the blue dashed curve of E and F), and better boundaries (e.g., around the blue dashed curve of A, B, and D). We selected 5 tiles from each of the four randomly selected whole slide images in the testing set, which present the highest probabilities to be the tumor regions according to our trained $5\times$ or $10\times$ tumor classification models (Figures 4 and 5). The selected tiles show high consensus with the annotations by the pathologist. The receiver operator characteristic (ROC) curve and area under the curve (AUC), are used to evaluate the performance of tumor classification models using $5\times$ and $10\times$ tiles of the WSIs (Figure 6). The AUCs of $5\times$ classification model have achieved 0.939 (95% CI of 0.937–0.940), 0.910 (95% CI of 0.905–0.914), and 0.959 (95% CI of 0.957–0.961) for training, validating, and testing set, respectively. Slightly better than the $5\times$ model, the AUCs of $10\times$ classification model are up to 0.971 (95% CI of 0.971–0.972), 0.973 (95% CI of 0.972–0.973), and 0.976 (95% CI of 0.975–0.977) for training, validating, and testing set, respectively. These values are consistent with our observation in Figures 3–5.

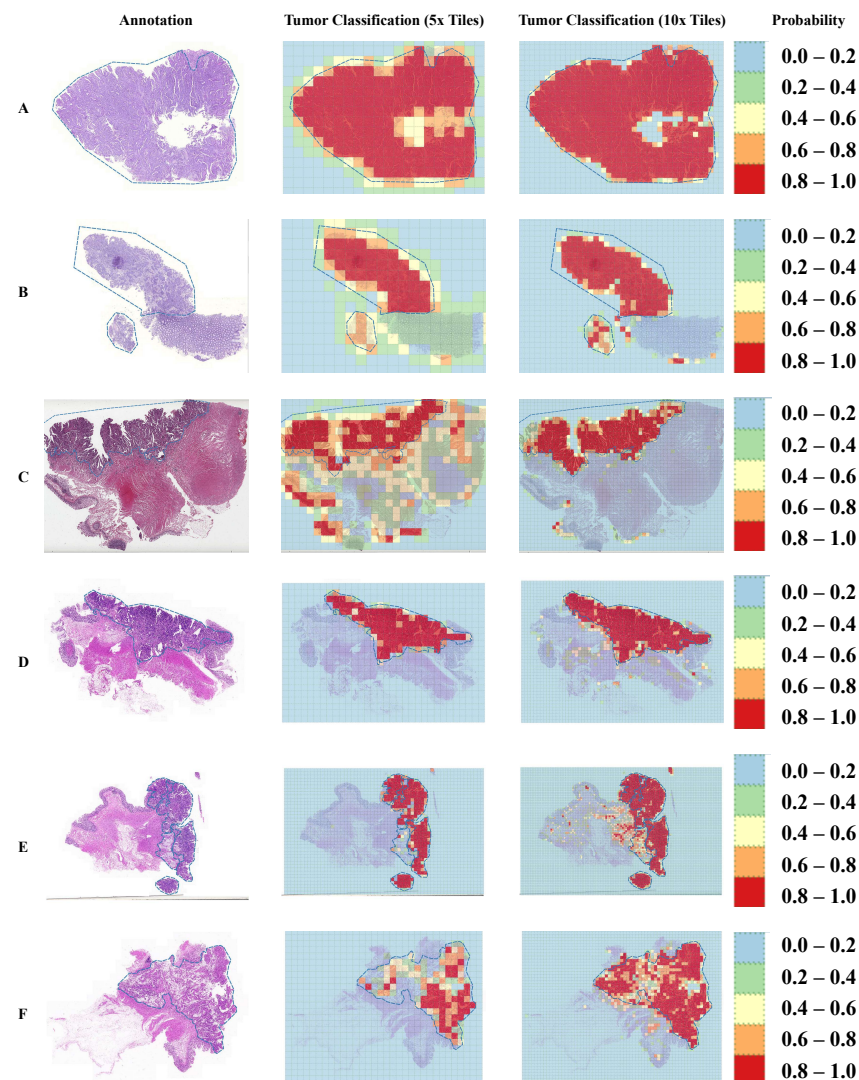


Figure 3. Probability maps of tumor classification using $5\times$ and $10\times$ tiles of the whole slide images (WSIs). The annotations created by the pathologist are marked with the blue dashed curve. The probability values are categorized into five groups with different color representations. The (A–C) and (D–F) samples are randomly selected from the testing set of TCGA-COAD and SYSU8H, respectively.

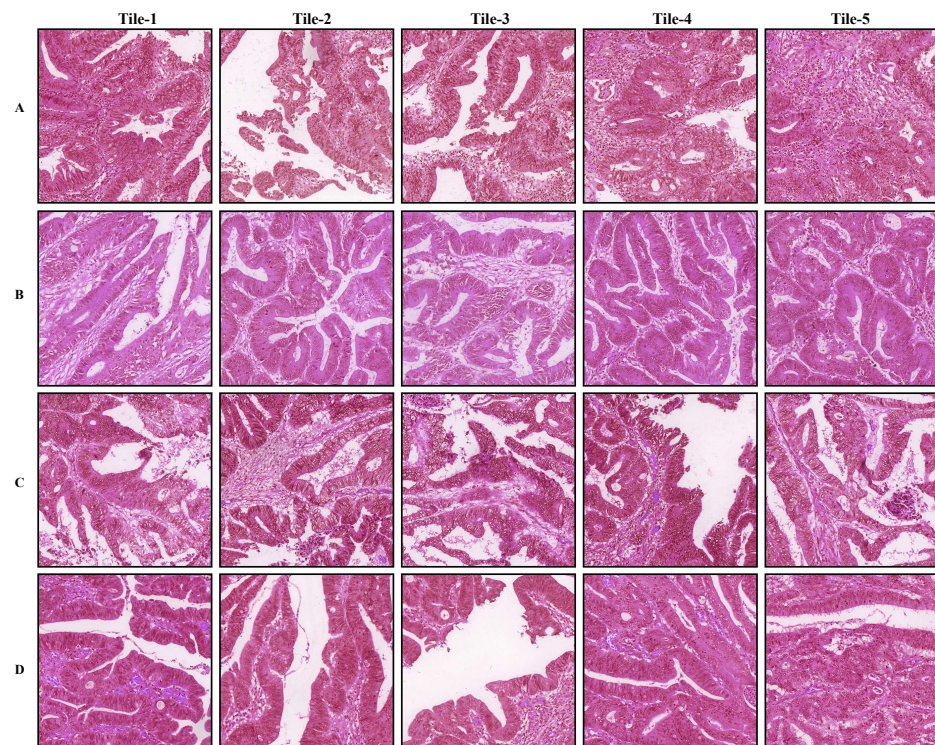


Figure 4. Representative tiles of tumor classification using $5\times$ tiles of the whole slide images (WSIs). The (A–D) samples are randomly selected from the testing set. In each row, tiles 1–5 are patches from the same WSI.

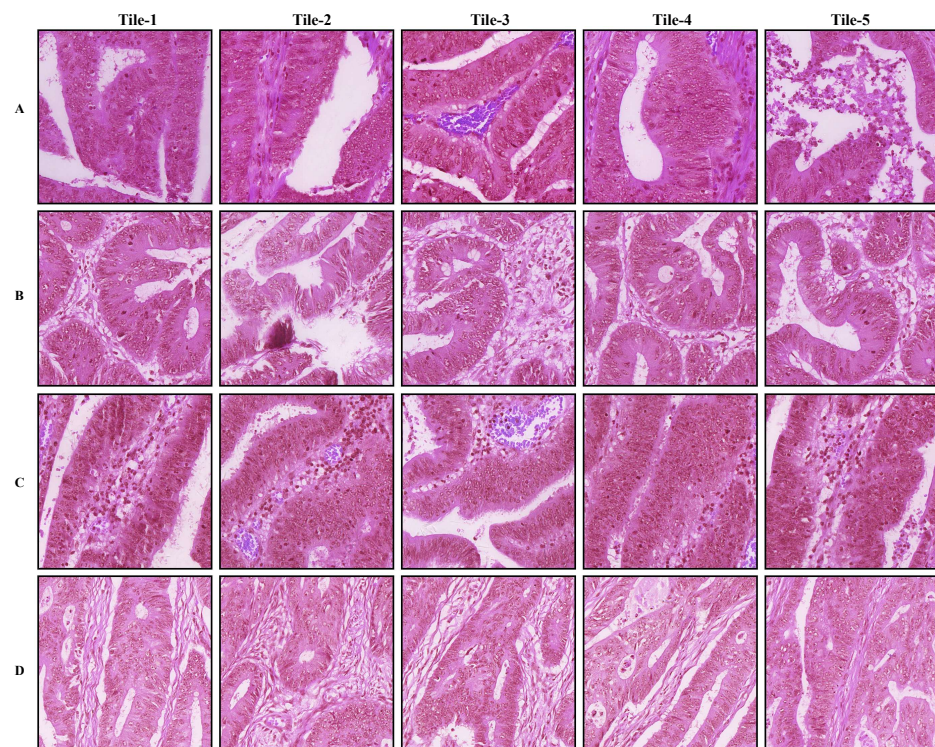


Figure 5. Representative tiles of tumor classification using $10\times$ tiles of the whole slide images (WSIs). The (A–D) samples are randomly selected from the testing set. In each row, tiles 1–5 are patches from the same WSI.

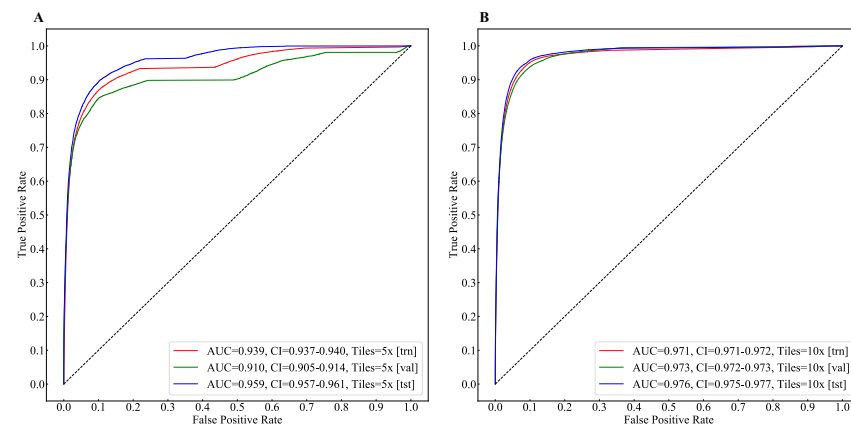


Figure 6. The receiver operator characteristic (ROC) curve and area under the curve(AUC) of tumor classification using $5\times$ and $10\times$ tiles of the whole slide images (WSIs). **(A)** The curves of training, validating, and testing set using $5\times$ tiles. **(B)** The curves of training, validating, and testing set using $10\times$ tiles.

3.2. Gene&MSI Classification

After model ensembling, the proposed method generates probabilities of gene mutations (i.e., KRAS, NRAS, and BRAF) and MSI status of every WSI.

As shown in Figure 7a–c, in the testing set, the proposed method reaches 0.792 (95% CI of 0.669–0.914), 0.886 (95% CI of 0.688–1.00), and 0.897 (95% CI of 0.800–0.994) AUCs for gene mutation predictions of KRAS, NRAS, and BRAF, respectively. In Figure 7d, our method shows high accuracy (i.e., 0.764 AUC, 95% CI 0.563–0.965) on the MSI status estimating in colorectal cancer.

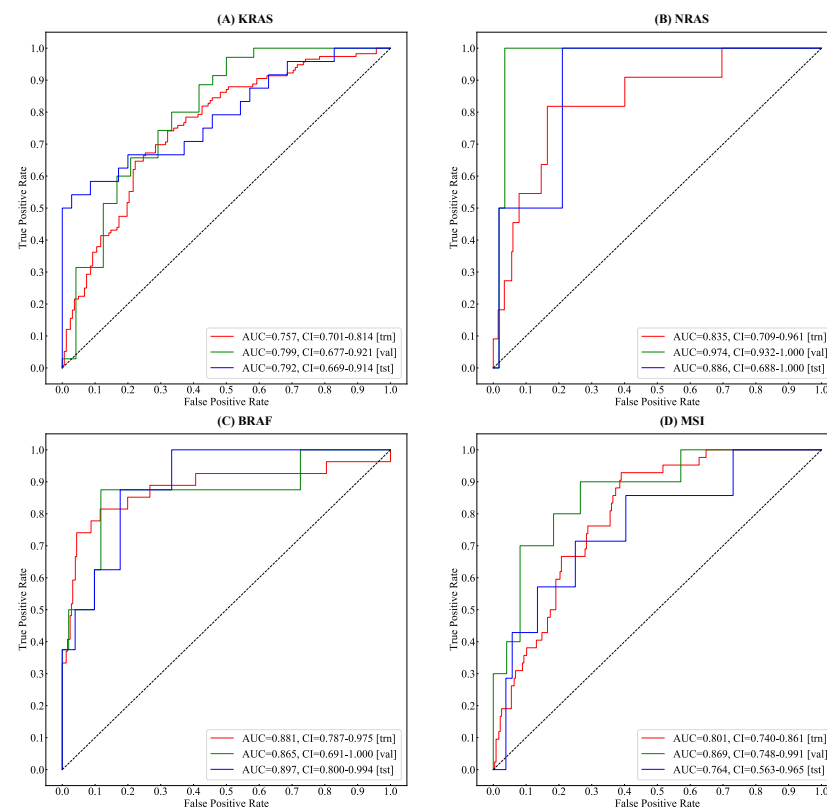


Figure 7. The receiver operator characteristic (ROC) curve and area under the curve(AUC) of Gene&MSI classification using $5\times$ and $10\times$ tiles of the whole slide images (WSIs). **(A)** The curves of KRAS gene mutation classification. **(B)** The curves of NRAS gene mutation classification. **(C)** The curves of BRAF gene mutation classification. **(D)** The curves of MSI status classification.

To investigate the effect of the selected number of features (i.e., topN) used for model ensembling, we conducted a comparison experiment on the testing set using sequential values (i.e., [1, 3, 5, 7, 9]) of topN. Figure 8 shows the trend of the AUC values under sequential values of topN in the testing set. Among all values, the proposed method achieves the highest KRAS, NRAS, and BRAF gene mutation prediction accuracy while topN equals 7. In gene mutation predictions, as the value of topN increases, the AUC value will firstly increase and then decrease. In MSI status estimation, the AUC increases gradually as the value of topN increases. As the value of topN passes 7, the increment of AUC narrows down.

Figure 9 shows the top weighted tiles of whole slide images (WSIs) in gene mutation prediction and MSI status estimation by the proposed models.

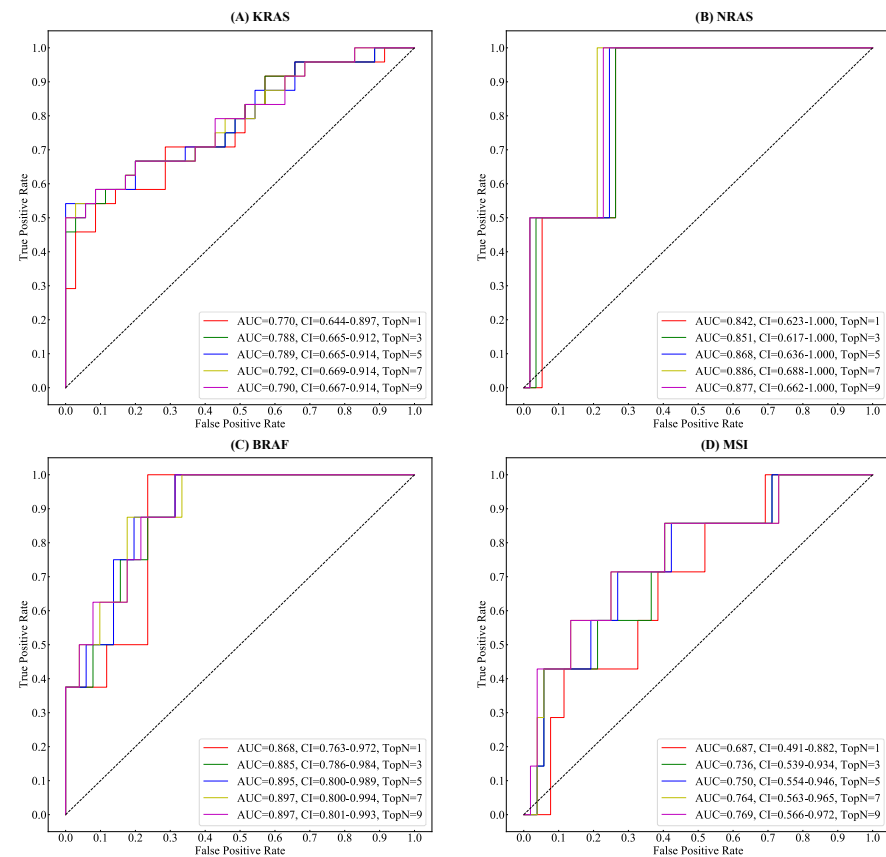


Figure 8. The receiver operator characteristic (ROC) curve and area under the curve(AUC) of Gene&MSI classification using sequential values of topN. (A) The trend of KRAS gene mutation classification. (B) The trend of NRAS gene mutation classification. (C) The trend of BRAF gene mutation classification. (D) The trend of MSI status classification.

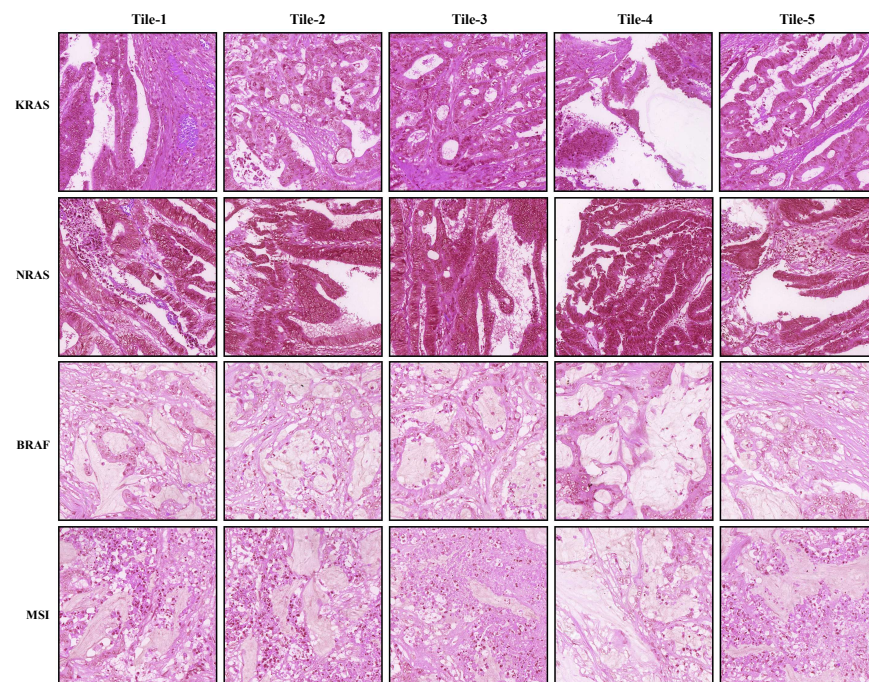


Figure 9. Top weighted tiles of whole slide images (WSIs) in gene mutation and MSI status estimation. In each row, tiles 1–5 are either 5× or 10× tiles extracted from the same WSI.

4. Discussion

4.1. Regarding the Cascaded Framework

In recent years, deep convolutional networks have demonstrated their potential in computer-aided cancer identification using clinical images such as CT scan [36], ultrasonic [37], and MRI images [38]. Other than tumor recognition, a growing number of researches are trying to look deeper into microsatellite instability estimation [39,40], gene mutation prediction [41] or survival risk evaluation [19], which are vital for precision pathological diagnosis and treatment.

To the best of our knowledge, the proposed cascaded framework is the first end-to-end method that simultaneously generates gene mutation prediction and MSI status estimation using the whole slide image (WSI) in colorectal cancer. Our method can produce high-fidelity gene mutation prediction and MSI status estimation for each WSI through a simple yet efficient average voting strategy to ensemble models. Predicting the gene mutations (KRAS, NRAS, and BRAF) and MSI status from deep convolutional networks provides pathologists with a more convenient way to evaluate prognosis and guide medication. For example, advanced metastatic CRC patients with KRAS and NRAS mutations are not recommended to choose anti-EGFR monoclonal drugs (cetuximab and panimab) for treatment. The evaluation of BRAF mutation can stratify the prognosis and guide clinical treatment. Patients with BRAF genetic mutation are unlikely to respond to the treatment of cetuximab or panimab. MSI is a predictor of the efficacy of immune checkpoint inhibitors, CRC patients with MSI-H are more likely to benefit from the treatment of immune checkpoint inhibitors (e.g., pabrolizumab). Qualitative and quantitative results of the experiment data demonstrated the effectiveness of our proposed framework. These results suggest that the deep learning models have the potential to provide diagnostic insight and clinical guidance directly from pathological H&E slides. Additionally, as the gene mutation prediction and MSI status estimation are directly computed from histopathology H&E slides, in principle, the proposed method should apply not only to colorectal cancer but also to other malignant cancers (e.g., lung, breast, and liver cancer).

4.2. Accuracies, Uncertainties, and Limitations

The proposed framework revealed high values of area under the curve (AUC) in both tumor classification and gene&MSI classification tasks. In tumor classification, the 5× and 10× classification models achieved 0.959 (95% CI of 0.957–0.961) and 0.976 (95% CI of 0.975–0.977) AUCs in the testing set, respectively. The values show a very close judgment between the pathologist and the proposed method, which suggest that the AI-algorithm can potentially serve as a pre-screening tool. The performance will be further evaluated using a larger dataset with multiple tissue samples collected from varied pathology departments.

In gene mutation prediction, the proposed method achieved 0.792 (95% CI of 0.669–0.914), 0.886 (95% CI of 0.688–1.00), and 0.897 (95% CI of 0.800–0.994) AUCs for gene mutation predictions of KRAS, NRAS, and BRAF, respectively. Because of the extremely biased ratio of mutant type / wild type distribution (i.e., 15 vs. 381 of NRAS, 43 vs. 353 of BRAF), the value of AUCs fluctuates in a large range within 95% confidence interval (see details in Figure 7). In terms of MSI status estimation, recent researches [39,40] had reported higher performance than ours (i.e., 0.764 AUC, 95% CI 0.563–0.965). Compared with these methods, our method is able to simultaneously generate gene mutation prediction (KRAS, NRAS, and BRAF) and MSI status estimation, which are all mandatory for metastatic CRC patients. As for future clinical application, improving the accuracy level of our algorithm remains one of the main future goals.

With the current cascaded classification-based scheme, the models are trained to generate tile-to-label predictions using features extracted from sequential convolutional layers. The lack of internal connectivity with adjacent tiles within the same WSI might lead to partial misclassification (e.g., red patches outside the blue dashed curve and green patches within the blue dashed curve in Figure 1B,D). Since the models are trained and optimized separately, the proposed framework requires extra computational time and storage for training and saving checkpoints of multiple models. Considering the computational efficiency, a unified model with shared parameters and object functions should be explored in further work.

Considering the type of H&E used for staining, varied types of hematoxylin have certain differences in stability, durability, and dyeing time, which may lead to distinct visual patterns. In the SYSU8H dataset, the H&E slices were stained using an identical form of hematoxylin (i.e., Harris hematoxylin) to make sure both the nucleus and cytoplasm can be clearly visible and discriminated. Due to the fact that the TCGA-COAD dataset was collected from multiple centers, the forms of hematoxylin used for staining were very likely to be different. However, as shown in Figure 3, in tumor classification, prediction accuracies among slices were not so significant. The result indicates that our method can be adapted to different forms of H&E staining approaches.

Another issue that should not be ignored is the tumor heterogeneity of the primary and metastatic lesions. Clinically, whether it is pathological diagnosis or target gene detection, the tumor specimen of the primary lesion is the first choice. However, there may be discrepancies in the gene mutation between the primary and metastatic tumor. For advanced metastatic tumors, when the target gene mutation of the primary tumor is negative, the target gene detection of the metastatic tumor can be carried out if conditions permitted, which can increase the opportunity for patients to receive one more targeted drug treatment. In this study, limited by the publicly available clinical samples attached with the gene mutation information of the primary tumor and the corresponding metastases, our method focused exclusively on primary tumors. Further evaluation is still necessary to clarify the reliability and generalization of our model performance.

5. Conclusions

For colon carcinoma, we design a cascaded deep convolutional framework to simultaneously generate gene mutation predicting and MSI status estimation based on the whole-slide images. The proposed method introduces a simple yet efficient average voting ensemble strategy to produce a high-fidelity prediction of the WSI. In gene mutation&MSI

status classification task, the proposed method achieves 0.792 (95% CI of 0.669–0.914), 0.886 (95% CI of 0.688–1.00), 0.897 (95% CI of 0.800–0.994), and 0.764 (95% CI 0.563–0.965) AUCs for KRAS, NRAS, BRAF, and MSI, respectively. These results suggest that the deep learning models have the potential to provide diagnostic insight and clinical guidance directly from pathological H&E slides. We plan to improve the architecture of the framework and apply it to other data sources to achieve better generalization capacity and diagnostic reliability.

Author Contributions: Conceptualization, C.Z. and G.W.; Formal analysis, S.L., W.Z. and Y.Z.; Funding acquisition, C.Z.; Investigation, Y.G. and T.L.; Methodology, Y.G. and T.L.; Writing—original draft, Y.G. and T.L.; Writing—review & editing, G.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of The Eighth Affiliated Hospital, Sun Yat-sen University (protocol code 2022d013, 20220222).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank all investigators and contributing pathologists from the TCGA (<http://portal.gdc.cancer.gov>), UCSC Xena (<https://xenabrowser.net>), and SYSU8H (<http://www.sysu8h.com.cn/>) for providing us samples and tools in this study.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Samples of the SYSU8H whole-side images and corresponding clinical information are available from the authors.

Abbreviations

The following abbreviations are used in this manuscript:

CRC	Colorectal Cancer
MSI	Microsatellite Instability
WSI	Whole-Side Image
CNN	Convolutional Neural Network

References

1. Wang, Z.; Zhou, C.; Feng, X.; Mo, M.; Shen, J.; Zheng, Y. Comparison of cancer incidence and mortality between China and the United States. *Precis. Cancer Med.* **2021**, *4*. [[CrossRef](#)]
2. Xia, C.; Dong, X.; Li, H.; Cao, M.; Sun, D.; He, S.; Yang, F.; Yan, X.; Zhang, S.; Li, N.; et al. Cancer statistics in China and United States, 2022: Profiles, trends, and determinants. *Chin. Med. J.* **2022**, *135*, 584–590. [[CrossRef](#)] [[PubMed](#)]
3. Lee, R.M.; Cardona, K.; Russell, M.C. Historical perspective: Two decades of progress in treating metastatic colorectal cancer. *J. Surg. Oncol.* **2019**, *119*, 549–563. [[CrossRef](#)] [[PubMed](#)]
4. Zhou, Z.; Mo, S.; Dai, W.; Xiang, W.; Han, L.; Li, Q.; Wang, R.; Liu, L.; Zhang, L.; Cai, S.; et al. Prognostic nomograms for predicting cause-specific survival and overall survival of stage I–III colon cancer patients: A large population-based study. *Cancer Cell Int.* **2019**, *19*, 1–15. [[CrossRef](#)]
5. Carr, P.R.; Weigl, K.; Edelmann, D.; Jansen, L.; Chang-Claude, J.; Brenner, H.; Hoffmeister, M. Estimation of absolute risk of colorectal cancer based on healthy lifestyle, genetic risk, and colonoscopy status in a population-based study. *Gastroenterology* **2020**, *159*, 129–138. [[CrossRef](#)]
6. Xie, Y.H.; Chen, Y.X.; Fang, J.Y. Comprehensive review of targeted therapy for colorectal cancer. *Signal Transduct. Target. Ther.* **2020**, *5*, 1–30. [[CrossRef](#)]
7. Mizukami, T.; Izawa, N.; Nakajima, T.E.; Sunakawa, Y. Targeting EGFR and RAS/RAF signaling in the treatment of metastatic colorectal cancer: From current treatment strategies to future perspectives. *Drugs* **2019**, *79*, 633–645. [[CrossRef](#)]
8. Van Cutsem, E.; Köhne, C.H.; Hitre, E.; Zaluski, J.; Chang Chien, C.R.; Makhson, A.; D’Haens, G.; Pintér, T.; Lim, R.; Bodoky, G.; et al. Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N Engl. J. Med.* **2009**, *360*, 1408–1417. [[CrossRef](#)]

9. De Roock, W.; Claes, B.; Bernasconi, D.; De Schutter, J.; Biesmans, B.; Fountzilias, G.; Kalogeras, K.T.; Kotoula, V.; Papamichael, D.; Laurent-Puig, P.; et al. Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: A retrospective consortium analysis. *Lancet Oncol.* **2010**, *11*, 753–762. [[CrossRef](#)]
10. Tol, J.; Nagtegaal, I.D.; Punt, C.J. BRAF mutation in metastatic colorectal cancer. *N. Engl. J. Med.* **2009**, *361*, 98–99. [[CrossRef](#)]
11. Taïeb, J.; Lapeyre-Prost, A.; Laurent Puig, P.; Zaanani, A. Exploring the best treatment options for BRAF-mutant metastatic colon cancer. *Br. J. Cancer* **2019**, *121*, 434–442. [[CrossRef](#)]
12. Copija, A.; Waniczek, D.; Witkoś, A.; Walkiewicz, K.; Nowakowska-Zajdel, E. Clinical significance and prognostic relevance of microsatellite instability in sporadic colorectal cancer patients. *Int. J. Mol. Sci.* **2017**, *18*, 107. [[CrossRef](#)]
13. Battaglin, F.; Naseem, M.; Lenz, H.J.; Salem, M.E. Microsatellite instability in colorectal cancer: Overview of its clinical significance and novel perspectives. *Clin. Adv. Hematol. Oncol. H&O* **2018**, *16*, 735.
14. Benson, A.B.; Venook, A.P.; Al-Hawary, M.M.; Arain, M.A.; Chen, Y.J.; Ciombor, K.K.; Cohen, S.; Cooper, H.S.; Deming, D.; Farkas, L.; et al. Colon cancer, version 2.2021, NCCN clinical practice guidelines in oncology. *J. Natl. Compr. Cancer Netw.* **2021**, *19*, 329–359. [[CrossRef](#)]
15. McDonough, S.J.; Bhagwate, A.; Sun, Z.; Wang, C.; Zschunke, M.; Gorman, J.A.; Kopp, K.J.; Cunningham, J.M. Use of FFPE-derived DNA in next generation sequencing: DNA extraction methods. *PLoS ONE* **2019**, *14*, e0211400. [[CrossRef](#)]
16. Coudray, N.; Ocampo, P.S.; Sakellaropoulos, T.; Narula, N.; Snuderl, M.; Fenyö, D.; Moreira, A.L.; Razavian, N.; Tsirigos, A. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **2018**, *24*, 1559–1567. [[CrossRef](#)]
17. Kather, J.N.; Pearson, A.T.; Halama, N.; Jäger, D.; Krause, J.; Loosen, S.H.; Marx, A.; Boor, P.; Tacke, F.; Neumann, U.P.; et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* **2019**, *25*, 1054–1056. [[CrossRef](#)]
18. Park, J.H.; Kim, E.Y.; Luchini, C.; Eccher, A.; Tizaoui, K.; Shin, J.I.; Lim, B.J. Artificial Intelligence for Predicting Microsatellite Instability Based on Tumor Histomorphology: A Systematic Review. *Int. J. Mol. Sci.* **2022**, *23*, 2462. [[CrossRef](#)]
19. Skrede, O.J.; De Raedt, S.; Kleppe, A.; Hveem, T.S.; Liestøl, K.; Maddison, J.; Askautrud, H.A.; Pradhan, M.; Nesheim, J.A.; Albrechtsen, F.; et al. Deep learning for prediction of colorectal cancer outcome: A discovery and validation study. *Lancet* **2020**, *395*, 350–360. [[CrossRef](#)]
20. Jia, P.; Yang, X.; Guo, L.; Liu, B.; Lin, J.; Liang, H.; Sun, J.; Zhang, C.; Ye, K. MSIsensor-pro: Fast, accurate, and Matched-normal-sample-free detection of microsatellite instability. *Genom. Proteom. Bioinform.* **2020**, *18*, 65–71. [[CrossRef](#)]
21. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
22. Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **1997**, *30*, 1145–1159. [[CrossRef](#)]
23. Sun, X.; Xu, W. Fast implementation of DeLong’s algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Process. Lett.* **2014**, *21*, 1389–1393. [[CrossRef](#)]
24. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
25. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
26. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
27. Lin, M.; Chen, Q.; Yan, S. Network in network. *arXiv* **2013**, arXiv:1312.4400.
28. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
29. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
30. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
31. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*. [[CrossRef](#)]
32. Baldi, P.; Sadowski, P.J. Understanding dropout. *Adv. Neural Inf. Process. Syst.* **2013**, *26*.
33. Shore, J.; Johnson, R. Properties of cross-entropy minimization. *IEEE Trans. Inf. Theory* **1981**, *27*, 472–482. [[CrossRef](#)]
34. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
35. Hinton, G.; Srivastava, N.; Swersky, K. Overview of mini-batch gradient descent. *Neural Netw. Mach. Learn.* **2012**, *575*.
36. Jin, Q.; Meng, Z.; Sun, C.; Cui, H.; Su, R. RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans. *Front. Bioeng. Biotechnol.* **2020**, *1471*. [[CrossRef](#)]
37. Liu, Z.; Yang, C.; Huang, J.; Liu, S.; Zhuo, Y.; Lu, X. Deep learning framework based on integration of S-Mask R-CNN and Inception-v3 for ultrasound image-aided diagnosis of prostate cancer. *Future Gener. Comput. Syst.* **2021**, *114*, 358–367. [[CrossRef](#)]
38. Ismael, S.A.A.; Mohammed, A.; Hefny, H. An enhanced deep learning approach for brain cancer MRI images classification using residual networks. *Artif. Intell. Med.* **2020**, *102*, 101779. [[CrossRef](#)]

39. Hildebrand, L.A.; Pierce, C.J.; Dennis, M.; Paracha, M.; Maoz, A. Artificial intelligence for histology-based detection of microsatellite instability and prediction of response to immunotherapy in colorectal cancer. *Cancers* **2021**, *13*, 391. [[CrossRef](#)]
40. Echle, A.; Laleh, N.G.; Quirke, P.; Grabsch, H.; Muti, H.; Saldanha, O.; Brockmoeller, S.; van den Brandt, P.; Hutchins, G.; Richman, S.; et al. Artificial intelligence for detection of microsatellite instability in colorectal cancer—A multicentric analysis of a pre-screening tool for clinical application. *ESMO Open* **2022**, *7*, 100400. [[CrossRef](#)]
41. Chen, M.; Zhang, B.; Topatana, W.; Cao, J.; Zhu, H.; Juengpanich, S.; Mao, Q.; Yu, H.; Cai, X. Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning. *NPJ Precis. Oncol.* **2020**, *4*, 1–7.