cancers

MDPI

Review

Artificial Intelligence for Histology-Based Detection of Microsatellite Instability and Prediction of Response to Immunotherapy in Colorectal Cancer

Lindsey A. Hildebrand [1], Colin J. Pierce [1], Michael Dennis [1,2], Munizay Paracha [1] and Asaf Maoz [1,3,*]

[1] Department of Medicine, Boston University School of Medicine and Boston Medical Center, Boston, MA 02118, USA; Lindsey.Hildebrand@bmc.org (L.A.H.); Colin.Pierce@bmc.org (C.J.P.); mjdennis@health.ucsd.edu (M.D.); mparacha@bu.edu (M.P.)

[2] Division of Hematology Oncology, Department of Medicine, University of California San Diego, San Diego, CA 92093, USA

[3] Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA 02215, USA

* Correspondence: asaf_maoz@dfci.harvard.edu

Simple Summary: Defects in a DNA repair pathway called mismatch repair (MMR) can lead to cancer, including colorectal cancer (CRC). The detection of mismatch repair deficiency (dMMR) is based on molecular tests, one of which is microsatellite instability (MSI) testing. Detecting tumors with dMMR/MSI is important for the identification of patients with Lynch Syndrome and determining if patients may benefit from immunotherapy. Recently, artificial intelligence has been evaluated as a method to predict MSI/dMMR directly from tissue slides that are available for most cancer patients. We review the data regarding the utility of machine learning for dMMR/MSI classification, including its accuracy and limitations, focusing on CRC. We also provide an overview of previous efforts to predict MSI from tissue slides and background regarding the use of artificial intelligence for image analyses. We summarize recent efforts to use artificial intelligence for the prediction of MSI and discuss the implications for predicting response to immunotherapy.

Abstract: Microsatellite instability (MSI) is a molecular marker of deficient DNA mismatch repair (dMMR) that is found in approximately 15% of colorectal cancer (CRC) patients. Testing all CRC patients for MSI/dMMR is recommended as screening for Lynch Syndrome and, more recently, to determine eligibility for immune checkpoint inhibitors in advanced disease. However, universal testing for MSI/dMMR has not been uniformly implemented because of cost and resource limitations. Artificial intelligence has been used to predict MSI/dMMR directly from hematoxylin and eosin (H&E) stained tissue slides. We review the emerging data regarding the utility of machine learning for MSI classification, focusing on CRC. We also provide the clinician with an introduction to image analysis with machine learning and convolutional neural networks. Machine learning can predict MSI/dMMR with high accuracy in high quality, curated datasets. Accuracy can be significantly decreased when applied to cohorts with different ethnic and/or clinical characteristics, or different tissue preparation protocols. Research is ongoing to determine the optimal machine learning methods for predicting MSI, which will need to be compared to current clinical practices, including next-generation sequencing. Predicting response to immunotherapy remains an unmet need.

Keywords: colorectal cancer; microsatellite instability; DNA mismatch repair; tumor immunology; immunotherapy; digital pathology; convolutional neural network; deep learning; machine learning; artificial intelligence

check for updates

Citation: 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. https://doi.org/10.3390/cancers13030391

Academic Editor: Marta Lukaszewicz-Zajac

Received: 25 December 2020
Accepted: 18 January 2021
Published: 21 January 2021

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

Copyright: © 2021 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/).

1. Introduction

Colorectal cancer (CRC) is the third most common and second most deadly cancer worldwide, causing an estimated 880,000 deaths in 2018 [1]. Mortality rates for CRC

have been declining in many countries due to improved screening efforts and therapeutic advances [2], but both the incidence and mortality of CRC have been increasing in patients under the age of 50 in high-income countries [3]. CRC is a heterogenous group of diseases (subtypes) with differences in epidemiology, anatomy, histology, genomics, transcriptomics and host immune response [4–7]. This heterogeneity leads to disparate clinical presentation, survival and response to therapy [8–11].

One of the clinically relevant subtypes of CRC is DNA mismatch repair deficient (dMMR) CRC. dMMR occurs due to pathogenic alterations in genes involved in the MMR system (*MLH1*, *MSH2/EPCAM*, *MSH6*, and *PMS2*) [12–14]. Several mechanisms can lead to dMMR, the most common being somatic hypermethylation of the *MLH1* gene promoter [14]. In patients with Lynch Syndrome, who carry a germline pathogenic mutation in one of the MMR genes, an additional somatic alteration can occur ("second hit"), leading to the phenotype of dMMR tumors. Sporadic bi-allelic somatic mutations in the MMR genes can also occur [13]. Deficiencies in MMR cause high rates of mutations throughout the DNA, especially in microsatellites, regions of DNA in which short sequences of nucleotides are repeated in tandem [12]. Thus, dMMR results in microsatellite instability (MSI), which is a highly sensitive marker of dMMR [13].

MSI has diagnostic, prognostic, and therapeutic implications in CRC and other cancers. The detection of dMMR/MSI is recommended as a screening test for Lynch syndrome for every case of CRC [14]. Confirmation requires germline testing and can inform surveillance and treatment decisions for both the patient and their relatives [15]. CRC patients with MSI generally have a better prognosis [13,16,17], which may be explained by a robust host immune response to tumor neoantigens [10,11,18]. Lastly, MSI status can inform treatment decisions, as patients with MSI tumors may be eligible for immune checkpoint inhibitor (ICI) therapy [19], and the benefit of fluorouracil-based chemotherapy regimens for tumors with MSI has been questioned [20–23].

Current testing for dMMR/MSI requires either an immunohistochemical analysis of MMR protein expression or a PCR-based assay of microsatellite markers [14]. While guidelines set forth by multiple professional societies recommend universal testing for dMMR/MSI [24], these methods require additional resources and are not available at all medical facilities, so many CRC patients are not currently tested [25]. Recently, artificial intelligence has been evaluated as a method to predict MSI directly from hematoxylin and eosin (H&E) stained slides (Figure 1). If successful, this approach could have significant benefits, including reducing cost and resource-utilization and increasing the proportion of CRC patients that are tested for MSI.

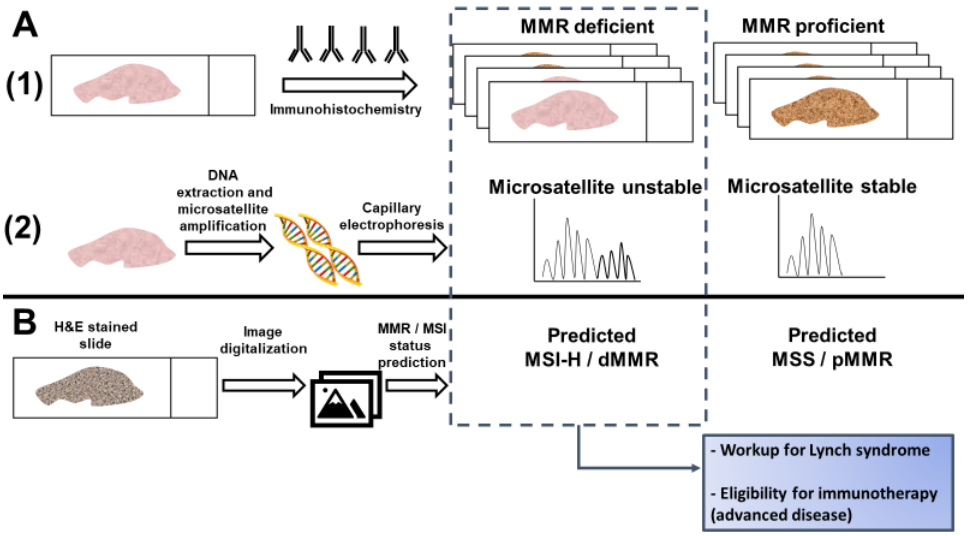

Figure 1. Detection of microsatellite instability (MSI) or mismatch repair (MMR) deficiency is performed by (**A1**) Immunohistochemistry of the mismatch repair proteins or (**A2**) PCR amplification of consensus microsatellite repeats that are analyzed

with capillary electrophoresis. Inference of MSI/MMR status from next generation sequencing (NGS) is not presented. (**B**) MSI/MMR status can be predicted from hematoxylin and eosin (H&E) stained slides, without requiring molecular analyses (see Figure 2). Detection of MSI/dMMR has implications for Lynch Syndrome screening and determining eligibility for immune checkpoint blockade in advanced disease. MSS: microsatellite stable. MSI-H: high microsatellite instability. pMMR: proficient mismatch repair. dMMR: deficient mismatch repair.

We review the emerging data regarding the utility of artificial intelligence for MSI classification, focusing on CRC. We provide (1) an overview of pathologic predictors of MSI, (2) a background regarding the use of artificial intelligence for image analyses, (3) a summary of recent efforts to use artificial intelligence for the prediction of MSI, and (4) a discussion about the implications for predicting response to immunotherapy.

2. Histological and Clinical Predictors of Microsatellite Instability

With the significant cost and non-universal availability of the molecular testing required to determine MMR/MSI status, studies have sought to predict MSI based on routinely available data, such as clinical information and histopathology [26]. CRC tumors with MSI are associated with certain histological features, detectable via standard H&E staining, and clinical data, such as patient age and tumor location [26–28]. Similar observations have been made in other tumors enriched for MSI, such as endometrial cancer [29]. These associations may present a means of identifying the tumors most likely to have the dMMR phenotype, and therefore the patients most likely to benefit from additional testing. They may also help to identify those at low risk who would be less likely to benefit. The targeted deployment of MMR/MSI testing could reduce costs and save resources [26]. Inferring MSI status may be considered in settings where MSI testing is not performed but is unlikely to be adopted in resource-rich settings unless the prediction accuracy is near-perfect.

Several clinicopathologic predictors of MSI have been discovered and several groups have proposed models for MSI prediction (Table 1). Histological features such as signet ring cells, mucinous or medullary morphology, and poor differentiation are significantly associated with MSI status, but show poor sensitivity for MSI prediction on their own [27,30]. Correlations between MSI and immunological features of tumor pathology, such as measurements of tumor infiltrating lymphocytes (TILs) [11,26,28,31] and specific histological structures such as the Crohn's-like lymphoid reaction (CLR), are well established in the literature [18,26,28]. CLR represents CRC-specific tertiary lymphoid aggregates [18]. The host response to MSI tumors is attributed to the high tumor mutational burden (TMB) and the abundance of immunogenic mutations, including insertion-deletion mutations, but other factors may contribute [32–34]. The Revised Bethesda Guidelines for MSI testing in CRC suggested testing tumors with "MSI histology" in patients younger than 60 years of age [35]. MSI histology was defined as the presence of TILs, CLR, mucinous/signet-ring differentiation, or medullary growth pattern. One of the histopathological features most strongly associated with MSI is the density of TILs [26,27,30]. When TIL density was assessed as a potential predictor of MSI, the area under the receiver operating characteristic curve (AUC) was 0.73. With a cutoff value of 40 lymphocytes/0.94 mm^2, MSI status could be predicted with a sensitivity of 75% and a specificity of 67% [30]. However, given that TIL density can vary across tumor area, this study using surgical specimens likely yielded a greater AUC than would be achieved with smaller biopsy specimens, such as those typically available from sites of metastasis.

Table 1. Histological predictors of microsatellite instability.

A				
Variable	**Sensitivity % (95% CI)**	**Specificity % (95% CI)**	**Odds Ratio, Univariate (95% CI)**	**Odds Ratio, Multivariate (95% CI)**
Host response features				
Tumor-infiltrating lymphocytes (TILs)	70 [26] 72 (64–78) [27] 21 [30] 60 (50–70) [36]	76 [26] 82 (80–85) [27] 97 [30] 78 (76–79) [36]	7.4 (5.4–10.3) [26] 9.8 (3.5–28.5) [30] 5.2 (3.2–8.5) [36]	3.8 (2.5–5.6) [26] 9.1 (5.9–14.1) [27] 3.7 (2.0–6.8) [36]
Crohn's-like lymphocytic reaction (CLR)	68 [26] 56 (48–63) [27] 49 [30] 69 (59–78) [36]	54 [26] 77 (74–80) [27] 64 [30] 45 (44–46) [36]	2.5 (1.8–3.4) [26] 1.7 (0.9–3.2) [30] 1.8 (1.1–3.0) [36]	2.3 (1.6–3.5) [26] 1.9 (1.2–2.9) [27] 1.1 (0.6–2.0) [36]
Peritumoral lymphocytic reaction	86 (77–92) [36]	42 (40–42) [36]	4.3 (2.3–8.3) [36]	3.7 (1.6–8.6) [36]
Stromal plasma cells	78 (68–86) [36]	48 (46–48) [36]	3.2 (1.9–5.6) [36]	2.1 (1.1–4.1) [36]
Tumor characteristics				
Mucinous morphology *	53 [26] 28 (21–35) [27] ** 22 [30] 51 (41–61) [36]	80 [26] 91 (89–93) [27] ** 93 [30] 78 (77–79) [36]	4.6 (3.4–6.3) [26] 3.7 (1.7–8.0) [30] 3.7 (2.3–6.0) [36]	1.7 (1.1–2.7) [26] 2.8 (1.7–4.8) [27] ** 4.71 (2.1–10.7) [36] 2.13 (1.3–3.4) [37]
Medullary morphology (10–70%)	25 [30]	97 [30]	12.5 (4.6–35.9) [30]	
Grade †	64 [26] † 38 (31–46) [27] 38 [30] 17 (10–26) [36] 32 (23–43) [36] †	81 [26] † 82 (79–84) [27] 87 [30] 90 (89–91) [36] 77 (76–78) [36] †	7.4 (5.4–10.1) [26] † 4.0 (2.2–7.3) [30] 1.8 (0.9–3.5) [36] 1.6 (0.9–2.6) [36] †	3.4 (2.2–5.2) [26] † 1.9 (1.2–3.1) [27]
Signet ring cells	13 [30]	95 [30]	4.3 (2.2–8.7) [26] 2.7 (1.1–6.8) [30]	
Lack of dirty or garland necrosis	59 [26] 26 (18–35) [36]	79 [26] 89 (88–90) [36]	5.4 (3.9–7.4) [26] 2.7 (1.5–4.7) [36]	1.8 (1.1–2.8) [26] 1.4 (0.7–3.0) [36]
Cribriform pattern	13 [30]	72 [30]	0.4 (0.2–0.8) [30]	
Histologic heterogeneity	55 (45–65) [36]	69 (68–70) [36]	4.4 (3.0–6.4) [26] 2.7 (1.7–4.4) [36]	

Table 1. *Cont.*

Clinical/Molecular Features				
Age <50 years	52 (44–60) [27] 21 (13–29) [36]	59 (56–62) [27] 89 (88–90) [36]	2.2 (1.3–3.8) [26] 2.0 (1.1–3.7) [36]	3.1 (1.5–6.5) [26] 1.9 (1.3–2.9) [27] 3.8 (1.8–8.0) [36]
Female	51 (41–62) [36]	63 (62–64) [36]	1.4 (1.0–1.9) [26] 1.8 (1.1–2.8) [36]	1.3 (0.7–2.2) [36] 1.56 (1.0–2.4) [37]
Size > or equal to 60 mm				2.75 (1.8–4.2) [37]
Anatomic site (right sided/proximal)	70 [26] 74 (67–81) [27] 79 (70–87) [36]	63 [26] 70 (67–73) [27] 63 (61–63) [36]	4.1 (2.9–5.7) [26] 6.4 (3.6–11.2) [36]	2.2 (1.5–3.3) [26] 4.7 (3.1–7.3) [27] 5.08 (2.7–9.6) [36] 3.76 (2.4–5.9) [37]
BRAF mutant				13.33 (8.0–22.2) [37]

B

Model	Model Variables	Sensitivity (%)	Specificity (%)	Positive/Negative Predictive Value (%)	AUC or Accuracy (95% CI)
Greenson et al. ** [26]	TIL/HPF, well or poorly differentiated, age < 50, CLR, R-sided, lack of dirty necrosis, any mucinous differentiation	92 **	46 **		AUC 0.850 **
MsPath [27]	Age < 50, proximal location, mucinous/signet ring/undifferentiated, poorly differentiated, CLR, TILs	93	55		AUC 0.890 (0.83– 0.94)
PREDICT [36]	R-sided, mucinous component, age < 50 years, TILs, peritumoral reaction, increased stromal plasma cells	96.9	76.6	35.2/99.5	AUC 0.924
Fujiyoshi et al. [37]	Female, mucinous component, tumor size > or equal to 60 mm, proximal location, *BRAF* mutation	76	77		AUC 0.856 (0.806 –0.905)

Table 1. *Cont.*

RERTest6 ** [38]	Proximal location, expansive growth pattern, CLR, solid pattern %, mucinous pattern %, cribriform pattern	78.01 **	93.39 **	51.8/97.9 **	Accuracy 0.921 **

Table 1 summarizes individual features and composite models used for the prediction of microsatellite instability (MSI) from histology. (A) Individual variables associated with MSI are listed. * Jenkins et al. [27] included signet cell and medullary morphology with mucinous morphology. † These studies included either poorly or well differentiated tumors as associated with MSI. (B) Multivariable models and their performance in differentiating between tumors with and without MSI are presented. ** Values are presented for validation cohorts when available. Greenson et al. and RERTest6 did not include a separate validation cohort. CI: confidence interval. TIL: tumor infiltrating lymphocytes. CLR: Crohn's-like lymphoid reaction. HPF: high power field. R-sided: right sided. AUC: area under the curve.

Multiple histological and clinical variables have been incorporated into algorithms designed to predict MSI status. The MsPath score was developed to predict MSI in patients under the age of 60 [27]. Using a scoring system incorporating age, anatomical site of the primary tumor, histologic type, tumor grade, and the presence or absence of TILs and CLR, an AUC of 0.89 was achieved when the model was tested against a separate validation cohort (Table 1). Validation of the MsPath score in a population based-cohort showed that its accuracy was insufficient for the selection of patients for Lynch Syndrome germline testing, misclassifying 18% (2/11) of patients with a pathogenic mutation in *MLH1/MSH2* [39]. Another scoring scheme by Greenson et al. incorporated similar variables but included lack of dirty necrosis in the model and was derived from a population that included patients of all ages [26]. The features associated with MSI all had a negative predictive value >90%. This model yielded an AUC of 0.85 based on the study cohort alone (no validation cohort was tested) (Table 1). Over half of the tumors analyzed had less than 5% chance of harboring MSI, presenting the potential for significant cost savings [26]. In another cohort, the model by Greenson et al. detected 93% of tumors with MSI and outperformed MsPath [40].

The PREDICT score was developed to improve on MsPath and other models [36]. It included variables that were significantly associated with MSI in a multivariable regression model, including age <50, right sided location, TILs, a peritumoral lymphocytic reaction, any mucinous component and increased stromal plasma cells [36]. PREDICT reported a sensitivity of 97% for the detection of MSI with an AUC of 0.924 in the validation cohort (Table 1). The RERtest6 model was developed to maximize the negative predictive value and included tumor location, growth pattern, solid and mucinous pattern, TIL and CLR [38]. The model had an accuracy of 92% in the global cohort and a negative predictive value of 97.9% (Table 1). The prevalence of MSI was 8.5% in this study. If this model were applied as screening for MSI in this study population, only 10% of patients would need confirmatory testing [38].

Another large study of MSI prediction from commonly available clinico-pathologic data included over three thousand patients over 50 years of age in Japan [37]. Female sex, proximal location, tumor size larger than 60 mm, mucinous component and *BRAF* mutation were associated with MSI and were included in a composite score used for prediction. CLR and TILs were not evaluated. In the validation cohort, the AUC was 0.856. Patients with *MLH1* promoter hypermethylation had higher scores than patients with Lynch Syndrome, as a result of the known association between *BRAF* mutations and *MLH1* hypermethylation and the high score given to *BRAF* mutations in the model. Overall, the performance of the model was disappointing, with approximately 25% of MSI tumors misclassified at the proposed threshold [37].

The encouraging performance of certain histology-based prediction models has not been sufficient to supersede universal testing for MSI/dMMR. Measurement of the variables for MSI prediction requires significant effort and expertise by pathologists, and inter-rater differences may affect the perceived reliability of histology-based scoring systems [41,42]. However, this work is fundamental to the premise that MSI can be predicted from histology, which has now been proposed as a task for deep learning from digital pathology [43] (Figure 1).

3. What Is Deep Learning and How Does It Apply to Digital Pathology?

Artificial intelligence is a broad term that characterizes the ability of machines to mimic intelligent human actions. *Machine learning* is a subset of artificial intelligence that allows computer systems to improve their performance ("learn") without being explicitly programmed [44]. *Deep learning* is a branch of machine learning that incorporates several layers of computational operation for the execution of complex tasks [45]. In the context of computer vision, deep learning often utilizes *convolutional neural networks* (CNNs) [44,46,47]. CNNs are designed to process raw data in the form of multiple arrays, such as color images; their structure is inspired by architecture found in the human brain's visual cortex [45]. To

achieve their goals of classifying images (e.g., is this a tumor or not?) or identifying objects, CNNs are trained on datasets that have been labeled with the desired output [41,44].

The layers of CNNs are arranged such that deeper layers represent increasingly synthesized features of an image. For example, the first layer typically detects edges; the second represents motifs related to the arrangement of edges; subsequent layers combine motifs into a representation of objects, and so on [45]. A simplified version of a CNN is presented in Figure 2. Images are represented as red, green and blue (RGB) color arrays such that each pixel in the image is represented by three numbers. RGB arrays are then subjected to filters. Filters are matrices that are used to learn specific features that are not prespecified but will help the CNN perform its task. For example, first layer filters often detect object edges in different orientations. This happens in the following fashion. The RGB arrays are convolved (a mathematical operation) with filters to create a multi-dimensional convolutional layer (Figure 2). To mimic the physiological "firing" of a neuron, a non-linear activation function is applied to the results of the convolution operation. Next, a pooling, or subsampling, procedure can be used to summarize the features from the convolutional layer and reduce the number of parameters, such that a pooling layer is created (Figure 2).

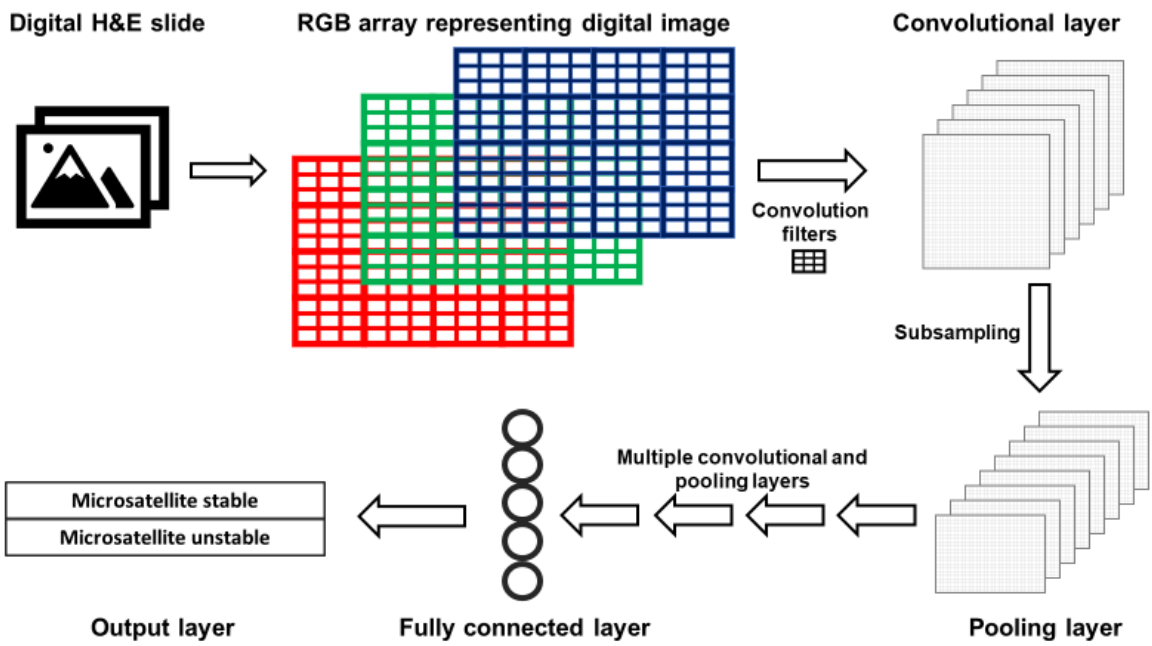

Figure 2. Simplified architecture of a convolutional neural network (CNN). Images acquired with digital pathology are composed of pixels. The color of each pixel can be represented with values of the red, green and blue (RGB) scheme. RGB arrays are subjected to filters to create a convolutional layer. Filters detect specific features from an image (e.g., lines, edges). A pooling layer with a reduced number of parameters is created by summarizing (subsampling) the input of the convolutional layer. After a defined number of convolutional and pooling layers, fully connected layers are created. Fully connected layers are uni-dimensional layers from which the output is predicted.

After a series of convolutional and pooling layers, fully connected layers are created. A fully connected layer is typically a unidimensional layer that is used to create the output prediction function (Figure 2). When the CNN output is generated, it is compared to the "true" label assigned to the data. When the CNN output is wrong, the CNN modifies its filters to improve the prediction accuracy of the CNN, thus learning the features associated with the desired output.

Training well-performing CNNs requires large datasets for training, testing and validation [47,48]. For the purpose of supervised learning such as image classification, these data need to be labeled according to the desired output. Many of the CNNs used for deep learning from digital pathology were originally developed for object detection and

image classification as part of the ImageNet challenge [47,49–51]. ImageNet is an annotated database of over a million non-medical images for which increasingly efficient CNNs were designed. These CNNs are robust to diverse classification and image recognition tasks and have been applied to digital pathology tasks [52]. Another advantage of applying CNNs to new tasks is the ability to use *transfer learning*, building on previously trained CNNs to perform a new task. Computationally, this means that not all layers of the CNN need to be trained again, and adequate performance can be achieved using a smaller dataset.

The availability of digital pathology datasets, annotated with clinical and molecular data, has led to a growing number of studies to evaluate the performance of CNNs on digitalized histology slides [41,46,52]. Such datasets include The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) project. The tasks that have been assigned to CNNs are diverse, including predicting clinical outcome and response to therapy, identifying molecular features, and others [52]. While CNNs have similar features, as described above, they differ from one another in their architecture, including the size, sequence and number of layers and filters, their number of parameters, and the connections between the layers of the CNN. As a result, CNNs vary in their computational efficiency and performance [52]. Common CNNs for the purpose of image classification include VGG [49], ResNet [50] and Inception [51].

4. Application of Deep Learning to Digital Pathology in Oncology

Deep learning is being explored for a myriad of research and clinical uses. Emerging studies in oncology have suggested that deep learning can be used to predict the diagnosis, prognosis, and response to treatment using histopathology digital slides as input [41]. For example, a deep learning algorithm has been developed for the prediction of the histology-based Gleason score in prostate cancer [53]. The deep learning algorithm outperformed general pathologists, but accuracy in assigning Gleason scores was only 0.70 when compared with reference scores provided by genitourinary pathologists [53]. In CRC, a deep learning method predicted five-year CRC-specific survival from spot images of H&E tumor slides, independent of tumor stage and grade [54]. MSI and the immune response to the tumor were not reported, and pathologists were provided only with spot images for risk stratification. Although conceptually intriguing, the comparison with pathologist performance in this study does not reflect standard pathologic evaluation. A "deep stroma score" has also been generated by using CNN transfer learning for the prediction of survival in CRC [55].

Recently, ten CNNs were used to generate an independent prognostic biomarker for CRC-specific survival, termed "DoMore-v1-CRC" [56]. This biomarker was associated with several clinical and molecular features and was trained on different resolutions of H&E images. The immune response to tumors was not reported, and the biomarker was not compared with the Immunoscore [57], a validated prognostic marker measuring immune cells in the CRC microenvironment using digital pathology. A prospective trial of tailoring therapy to prognostic subgroups is planned based on these results [56].

Deep learning has also been used to predict tumor molecular features, such as genomics, transcriptomics and proteomics. Using an Inception CNN, researchers classified non-small cell lung cancer into histological subtypes and predicted the mutational status of several genes in lung adenocarcinoma [58]. Histological subtype classification achieved high AUC (0.97) when trained on TCGA data, but performance was worse for independent datasets, requiring manual identification of tumor areas by pathologists. Some of the misclassifications included the labeling of blood vessels, clots, inflammation, and necrosis as lung adenocarcinoma and the labeling of cartilage as lung squamous cell carcinoma. For the purpose of genomic prediction, mutations in *STK11, EGFR, KRAS*, *TP53* and other genes were predicted with AUCs of 0.733 to 0.856.

Using the ResNet CNN, the prediction of PD-L1 expression was performed from H&E slides in non-small cell lung cancer [59]. Prediction was good for adenocarcinomas (AUC = 0.83) but did not perform well for squamous cell histology (AUC = 0.64). This difference

highlights potential challenges in the generalizability of such CNNs to different histological subtypes and different tumor sites, and how the composition of the training dataset may influence CNN performance. In breast cancer, the VGG CNN was used to predict tumor grade, estrogen receptor status, histological subtype, and RNA-based molecular subtype and recurrence risk score [60]. Accuracy was highest for the task of histological subtype classification, raising the possibility that output based on visual patterns may be more amenable to prediction than transcriptomic data. Consensus molecular subtypes of CRC, derived from transcriptomic data, have also been predicted from histology using an Inception CNN [61]. This study used three datasets and suggested that the CNN learned features that are specific to the dataset, thus potentially biasing the learning process and limiting generalizability. To overcome this, the authors implemented adversarial training to minimize the weight of dataset-specific features. Histological predictors of the consensus molecular subtype 1 were underrepresented in a dataset comprised of rectal cancer biopsies, requiring adjustment of the classification probabilities.

These examples demonstrate the broad applications of machine learning methods to digital pathology in oncology as well as some of the caveats to their performance. Careful evaluation of individual studies is required, as methods differ considerably and at times include modifications that may hamper the feasibility or generalizability of the methods proposed. Most studies are retrospective and have not been evaluated in prospective clinical trials with standard of care comparator methods. In addition, CNNs learn features that are biased and do not reflect biological differences. This can be partially mitigated if features that may cause bias are known, but since the features determining CNN output are often unknown, validation in additional cohorts is crucial.

5. Predicting MSI Status with Deep Learning

Recently, several studies have investigated the potential for CNNs to predict MSI from H&E stained histological samples. Kather et al. trained and tested CNNs on gastric, endometrial, and colorectal samples that were snap-frozen or formalin-fixed paraffin-embedded (FFPE) [43]. FFPE slides are routinely used for histological diagnosis and immunohistochemistry. Fixation with formalin and embedding with paraffin are performed to maintain tissue architecture and morphology, and to allow long-term preservation at room temperature. The process of generating an FFPE slide requires many hours and the fixation process results in the cross-linking of DNA and proteins that can impair the performance of molecular analyses. Snap-frozen tissue is not routinely obtained but can be used for intraoperative diagnoses because it can be rapidly reviewed by a pathologist. Snap-frozen tissue can also be used for extensive molecular analyses [62,63]. The morphological quality of snap-frozen tissue is not considered sufficient to render a definitive diagnosis, and confirmation using FFPE slides is typically required [64–66]. All CNNs in the study by Kather et al. had been pretrained on the ImageNet database, and only the last ten layers of the CNNs were trainable. After assessing the performance of five CNNs in differentiating tumor tissue from healthy tissue, the CNN ResNet-18 (a ResNet with 18 layers) was selected for further evaluation based on its strong performance and smaller number of parameters. The advantage of a CNN with a smaller number of parameters is a decreased risk of overfitting the data and increased likelihood of maintaining performance when applied to a validation cohort. ResNet-18 was trained with two sets of CRC (fresh frozen and FFPE slides) and one gastric cancer dataset (FFPE) from TCGA (Table 2). Tumor tissue was divided into smaller tiles, each of which was separately analyzed and assigned a predicted MSI score. Predicted MSI status for each slide was determined by the predicted MSI status of the majority of its constituent tiles.

Table 2. Deep learning for prediction of MSI from digital pathology.

CNN and Additional Methods	Other CNNs Evaluated	Training Cohort	Test Cohort(s) with AUC (95% CI) or Accuracy	External Validation Cohort(s) with AUC (95% CI)
ResNet-18 Whole-slide image classified per majority of image tiles [43]	AlexNet, VGG-19, InceptionV3, SqueezeNet	TCGA CRC FFPE	TGCA CRC FFPEAUC 0.77 (0.62–0.87)	DACHS CRC FFPE AUC 0.84 (0.72–0.92)
		TCGA CRC frozen	TCGA CRC frozen AUC 0.84 (0.73–0.91)	DACHS CRC FFPE 0.61 (0.50–0.73)
		TCGA gastric FFPE	TCGA gastric FFPE AUC 0.81 (0.69–0.90)	DACHS CRC FFPE AUC 0.60 (0.48–0.69) KCCH gastric FFPE AUC 0.69 (0.52–0.82)
		TCGA uterine FFPE	TCGA uterine FFPE AUC 0.75 (0.63–0.83)	
ShuffleNet [67]	AlexNet, InceptionV3, ResNet-18, DenseNet201	TCGA CRC	TCGA CRC AUC 0.805	DACHS CRC AUC 0.89 (0.88–0.92)
ResNet-18 Whole slide image classified using two multiple instance learning pipelines integrated into an ensemble classifier [68]	none	TCGA CRC frozen	TCGA CRC frozen AUC 0.885	Asian CRC AUC FFPE 0.650
		TCGA CRC frozen with 10% Asian CRC FFPE	Asian CRC FFPEAUC 0.850	
		TCGA CRC frozen with 70% Asian CRC FFPE	Asian CRC FFPEAUC 0.926	
Custom multilayer perceptron (HE2RNA) applied after feature extraction by ResNet-50; with and without transcriptomic representation of histology [69]	none	TCGA CRC FFPE, with transcriptomic representation and 20% of training cohort	TCGA CRC FFPEAUC ~0.80 *	
		TCGA CRC FFPE, using >80% of training cohort	TCGA CRC FFPEAUC ~0.80 *	

Table 2. *Cont.*

CNN and Additional Methods	Other CNNs Evaluated	Training Cohort	Test Cohort(s) with AUC (95% CI) or Accuracy	External Validation Cohort(s) with AUC (95% CI)
Inception-V3 with and without adversarial learning [70]	VGG-19, ResNet-50	TCGA CRC	TCGA CRC Accuracy 98.3%	TCGA endometrial Accuracy 53.7%
		TCGA CRC and endometrial	TCGA CRC Accuracy 72.3% TCGA endometrial Accuracy 84.2%	TCGA gastric Accuracy 34.9%
		TCGA CRC and endometrial with adversarial learning	TCGA CRC Accuracy 85.0%TCGA endometrial Accuracy 94.6%	TCGA gastric Accuracy 57.4%
InceptionResNetV1 [71]	InceptionV1-3, InceptionResnetV1-2, Panoptes1-4 (multibranch custom InceptionResnet)	TCGA and CPTAC endometrial carcinoma	TCGA and CPTAC endometrial carcinoma AUC 0.827 (0.705–0.948)	
ShuffleNet [72]	none	MSIDETECT CRC (color normalized)	MSIDETECT CRC AUC 0.92 (0.90–0.93)	YCR-BCIP CRC surgical samples AUC 0.96 (0.93–0.98)
				YCR-BCIP CRC biopsy samples AUC 0.78 (0.75–0.81)
		YCR-BCIP CRC biopsy samples	YCR-BCIP CRC biopsy samples	

Table 2 summarizes the methods and performance of available data regarding deep learning for MSI prediction from pathology. * Exact AUC not reported, estimate based on an included graphical representation. CNN: convolutional neural network. TCGA: The Cancer Genome Atlas, CRC: colorectal cancer, DACHS: CRC prevention through screening study (abbreviation in German), FFPE: Formalin-Fixed Paraffin-Embedded, AUC: Area under the receiver operating characteristic curve, KCCH: Kanagawa Cancer Center Hospital (Japan), CPTAC: Clinical Proteomic Tumor Analysis Consortium, MSIDETECT: A consortium comprised of whole slide images from TCGA, DACHS, the United Kingdom-based Quick and Simple and Reliable trial (QUASAR), and the Netherlands Cohort Study (NLCS), YCR-BCIP: Yorkshire Cancer Research Bowel Cancer Improvement Programme.

Using this process, the CNN was able to detect MSI in snap-frozen and FFPE TCGA samples with similar AUCs to those achieved with previous pathology-based scoring systems such as MsPath and the model by Greenson et al. (0.84 for snap-frozen CRC samples, 0.77 for FFPE CRC samples, and 0.81 in FFPE gastric adenocarcinoma samples). This level of performance was maintained when the CNN trained on FFPE CRC samples was tested on an external validation cohort from the DACHS (Darmkrebs: Chancen der Verhütung durch Screening) study (Table 2), which consisted of FFPE CRC samples from Germany (AUC 0.84). The authors also tested the classification performance of the ResNet when applied to slides with limited tissue, finding that performance plateaued with a quantity of tissue that is available from standard needle biopsies [43].

To attempt to identify what pathological features the ResNet used to make its classifications, tumor regions that were assigned high or low MSI scores were visually inspected. Areas predicted by the CNN to represent MSI often showed characteristics consistent with known pathological correlates of MSI, such as poor differentiation and lymphocytic infiltration. PD-L1 expression and an interferon gamma transcriptomic signature were correlated with the proportion of a sample's tiles predicted to have MSI. This finding is consistent with previous data showing high expression of PD-L1 and interferon gamma in CRC with MSI [73,74].

Despite encouraging performance for MSI classification in similar cohorts, testing against different cohorts revealed some limitations. CNNs trained on snap-frozen CRC samples or gastric adenocarcinoma samples did not perform as well as the CNN both trained and tested on FFPE CRC samples. When the CNN was trained to detect MSI in endometrial cancers, its performance was significantly reduced to an AUC of 0.75, raising the possibility that the CNN is learning tissue-specific features associated with MSI. Additionally, the CNN trained on TCGA gastric adenocarcinomas did not perform as well when tested on a Japanese gastric adenocarcinoma cohort (AUC 0.69), possibly due to distinctive histological patterns seen in gastric adenocarcinomas in this cohort [43].

Other studies have attempted to improve upon these results using other CNNs and machine learning techniques (Table 2). In a follow up study by Kather et al., the prediction of MSI was performed as a benchmark task by various CNNs, which were pretrained on the ImageNet database [67]. The ResNet and Inception CNNs were outperformed by the DenseNet [75] and ShuffleNet [76] architectures. ShuffleNet, a CNN optimized for mobile devices, was able to achieve an AUC of 0.89 when trained on a CRC cohort from TCGA and validated on the DACHS CRC cohort (Table 2). The ResNet used for the previous study by Kather et al. achieved an AUC of 0.84 [43,67].

Another group reports improvement upon the results by Kather et al. in terms of overall predictive accuracy and generalizability to different cohorts [68]. This study also used ResNet-18 to assign each tile within the tumor area an MSI likelihood. However, multiple instance learning was used to train the CNN to classify the whole slide image. Multiple instance learning assumes that not all tumor regions contribute the same amount of information to the task of classification of the tumor as a whole [77]. Certain regions or patterns found in limited areas of a sample may be more important to determining the likelihood of the tumor being MSI. For example, any mucinous differentiation increases the likelihood of a tumor harboring dMMR/MSI [26,28]; this may be focal and not seen in the majority of tumor areas. Two different multiple instance learning methods were used in this study, and their input was integrated into a final ensemble predictor (Table 2). This ensemble classifier achieved an AUC of 0.885 [68], which was better than the performance reported by Kather et al. [43].

This group also found a significant reduction in AUC (0.650) when the TCGA-trained ensemble classifier was tested on a cohort of Asian patients with samples acquired with a different slide preparation protocol [68]. They were able to overcome this reduction in performance by transfer learning. By adding increasing proportions of data from the Asian cohort to the training set, they were able to achieve an AUC of 0.850 with 10% samples from the Asian cohort, with continued improvement up to an AUC of 0.926 with 70% samples

from the Asian cohort (Table 2) [68]. Pathologic signatures were derived from the model and were associated with known features of MSI, including TMB and insertion-deletion mutational burden, as well as transcription signatures of immune activation.

A conference paper by Wang et al. also assessed an alternative technique, Patch Likelihood Histogram (PALHI), for integrating tile-level MSI predictions into patient-level predictions using whole slide images from a TCGA endometrial cancer cohort [78]. First, a ResNet-18 pre-trained on ImageNet was trained to predict MSI for individual tiles on a subset of the TCGA cohort. PALHI then generated a histogram of the patch-level estimated MSI likelihoods, which were used to train a machine learning classifier called XGBoost to make patient-level predictions. The performance of a pipeline using PALHI to make patient-level predictions was compared to pipelines using another machine learning method, Bag of Words (BoW) and the "majority voting" method, using another subset of the TCGA cohort as a testing set. The three methods were each trained on both patches assigned binary "hard labels" and patches assigned "soft labels," or MSI probabilities. The PALHI method trained using "soft labels" yielded the best performance on the test set, with an AUC of 0.75. By comparison, the AUCs for BoW and the majority method using "soft labels" were 0.71 and 0.56, respectively [78].

Transcriptomic prediction from H&E slides has also been used to improve MSI prediction when limited training data are available [69]. First, features were extracted from each tissue tile using the ResNet-50, pretrained on the ImageNet database. These features served as the input for a custom multilayer perceptron, which was trained to predict gene expression from RNA-Seq data. *Multilayer perceptrons* are neural networks composed of fully connected layers, typically without convolutional layers. This neural network was trained on pan-cancer and tissue-specific TCGA datasets and was able to predict several expression signatures, including adaptive immune response signatures [69]. For MSI prediction, the authors simulated a situation where a limited number of training slides are available at two sites. They showed that, using the transcriptomic representation trained at one site, they could improve MSI prediction at the second site. However, when increasing proportions of data at the second site were used for MSI prediction without integrating transcriptomic representation, this advantage was largely lost. Neither method achieved an AUC > 0.85 and no external validation set was used (Table 2) [69]. It is unclear if this approach would be applicable in real-life settings.

In a conference paper submitted to the 1st Conference on Medical Imaging with Deep Learning (MIDL 2018) [70] and a related patent [79], adversarial learning was used to improve the generalizability of CNNs for MSI prediction across different cancers. The Inception-V3, ResNet-50 and VGG-19 CNNs were compared; Inception-V3 was chosen for downstream analysis. TCGA samples were used for both testing and training; this study did not use an external validation dataset. MSI status was categorized as stable, low instability or high instability. Inception-V3 was trained on CRC samples and achieved a slide-level accuracy of 98.3% with an internal validation set of 10% of TCGA slides. It is unclear if this level of accuracy represents overfitting. Accuracy was poor when applied to endometrial carcinoma samples at 54%, whereas training the CNN on both CRC and endometrial carcinoma decreased the accuracy of MSI prediction for CRC to 72% (Table 2). This CNN also performed poorly at classifying MSI in gastric adenocarcinoma with a slide-level accuracy of 35%. Next, a tumor type classifier was added to the CNN with an adversarial objective—to decrease the ability of the model to predict tumor type. The rationale for creating this adversarial objective is to remove tissue-specific features that are learned by the CNN, such that the model will recognize the features associated with MSI better. Adversarial training improved MSI classification across the three cancer types, but accuracy remained poor for gastric adenocarcinoma at 57% [70].

Focusing on endometrial cancer, a recent study available as a preprint generated CNNs that had three branches of an InceptionResNet architecture, each analyzing tiles at a different resolution [71]. An optional fully connected layer incorporating clinical features was also evaluated as a fourth branch. This structure, termed Panoptes, allowed

the model to take into account both tissue-level and cellular-level structures, as would a human pathologist using a microscope. MSI classification was one of several tasks that the CNNs were trained to do. While the complex architecture showed strong performance in predicting many histological and molecular features, MSI was best predicted by the existing InceptionResnetV1 architecture, with an AUC of 0.827 (Table 2), which outperformed Kather's previously described ResNet-18 architecture (AUC 0.75). The inclusion of clinical data did not seem to improve the model's performance: when the age and BMI of the patient were added into the model, its performance did not significantly improve [71]. Predicted MSI was correlated with certain histological features, including intratumoral and peritumoral lymphocytic infiltrates.

The strongest-performing model for MSI prediction was developed by Echle et al. by training a CNN on a large cohort of H&E-stained CRC samples from the MSIDETECT consortium, which is comprised of whole slide images from TCGA, DACHS, the United Kingdom-based Quick and Simple and Reliable trial (QUASAR), and the Netherlands Cohort Study (NLCS) [72]. A modified version of the CNN ShuffleNet that was pre-trained on ImageNet was trained on whole slide images from MSIDETECT with known MSI or dMMR status and externally validated on a separate population-based cohort, Yorkshire Cancer Research Bowel Cancer Improvement Programme (YCR-BCIP). For each slide, tumor tissue was manually outlined and the slide was divided into smaller tiles. The patient-level prediction of MSI/dMMR was based on the average tile-level prediction for each patient. The CNN was first trained and tested on individual sub-cohorts. As in earlier-described studies [43,68,70], when a CNN trained on a single sub-cohort was tested on another sub-cohort, performance usually suffered. A positive correlation between the size of the training cohort and the performance of the model was noted. The CNN was then trained on increasing numbers of patients randomly selected from the MSIDETECT cohort. The model showed better performance with greater numbers of patients up until about 5000 patients, after which performance plateaued. After training with samples from 5500 patients, the model attained an AUC of 0.92 when tested on a separate set of patients from MSIDETECT. When tested on the external validation cohort (YCR-BCIP), the model attained a similarly impressive AUC of 0.95. Additionally, when slides were subjected to color normalization, the specificity at given levels of sensitivity increased and a slight improvement in AUC to 0.96 was demonstrated [72]. Though these results are encouraging, it is worth noting that the samples used to train and test this model were derived mostly from European patients. Further validation with more diverse cohorts and prospective studies will be necessary before this model can be applied in a broad clinical context.

Subgroup analysis did reveal some variation in the model's performance for certain tumor characteristics. While the performance was consistent for tumors at stages I-III (AUCs 0.91–0.93), the AUC for stage IV tumors was lower (0.83). The authors do not discuss potential explanations for this discrepancy, but there was a similar reduction in AUC for tumors with high histologic grade (AUC for high grade tumors was 0.83). The relatively low prevalence of MSI/dMMR in stage 4 colorectal cancers would have decreased the number of available images from this subgroup available for training, as would the fact that stage 4 tumors are more likely to come from biopsy specimens than complete resection samples. This lower performance for stage 4 tumors is unfortunate given that ICI therapy is currently primarily used in late-stage colorectal cancer, reducing the model's potential utility for guiding treatment decisions. Additionally, the model predicted MSI more effectively for colon cancer (AUC 0.91) than for rectal cancer (AUC 0.83). Performance did not vary significantly by tumor molecular characteristics (e.g., mutation status) [72].

As noted above, a previous study demonstrated that the performance of ResNet-18 in classifying MSI status plateaued with a quantity of tissue that can be obtained by needle biopsy [43]. However, Echle et al. found a significant decrease in AUC when the CNN trained on surgical specimens was tested on YCR-BCIP biopsy specimens as compared to YCR-BCP surgical specimens (0.78 vs. 0.96). Though size of specimen may be a factor here, artifacts from specimen acquisition and the fact that samples were derived only from

luminal tumor tissue may also affect performance. When the authors performed a 3-fold cross-validated experiment using YCR-BCIP biopsy specimens to both train and test, the AUC improved to 0.89 [72]. However, the model was not tested on samples from sites of metastasis, which are commonly biopsied in the clinical setting. Thus, machine learning models may be effective in classifying the MSI status of biopsy specimens, but will likely perform best when trained on similarly derived specimens.

Taken together, these studies demonstrate that multiple CNNs and machine learning techniques are being evaluated for MSI prediction from histology. There is no clear consensus regarding the optimal network architecture. The use of large and diverse datasets for training may overcome some of the limitations of models whose classification accuracy for MSI status is worse when applied to datasets with differing characteristics, which could be the case when applying these methods across different health systems, regions and populations. With continued experimentation, improvement, and validation of existing models, the use of machine learning to predict MSI may reach a level of accuracy sufficient for clinical application in the future.

6. Predicting Response to Immunotherapy with Deep Learning

While MSI status is currently used to determine a CRC patient's eligibility for ICIs, it is far from a perfect predictor of the efficacy of these treatments. Only 30–50% of CRC patients with MSI respond to ICIs. There is also a subset of microsatellite stable CRC that responds to ICI [32,80,81], demonstrating that ICIs could have a role in the treatment of early-stage proficient MMR tumors. Patients who receive these treatments are at risk for immune-related adverse events including thyroid dysfunction, hepatitis, colitis, pneumonitis and others [82]; there is increasing evidence of an association between response to therapy and the development of immune-related adverse events [83,84]. Thus, alternative methods of determining eligibility and predicting the efficacy and toxicity of ICIs are needed.

Despite the potential for the prediction of MSI from histology discussed above, there is little published research using this method to predict ICI response, and to our knowledge, there are no published results concerning ICI response prediction in CRC. Machine learning can be used to predict ICI response in other tumors from various types of input data, including H&E staining, which may lay the groundwork for similar research in CRC. One study available in abstract form predicted ICI response in melanoma from pre-treatment H&E slides in patients who were treated with first-line ICI therapy [85]. A CNN was trained to classify slides into responders and non-responders and into those who experienced severe adverse events and those who experienced none. The model performed modestly well in predicting ICI response, despite training on slides from only 124 patients. The model was much less effective at predicting adverse events, and research incorporating immunologic biomarkers into the algorithm is ongoing [85]. A similar study on non-small cell lung cancer (NSCLC) samples used the spatial arrangement of TILs as detected by computer algorithms to train a machine learning classifier to predict response to nivolumab, achieving an AUC of 0.64 on an external validation cohort [86].

A variety of biomarkers have been evaluated for predicting response to immunotherapy, many of which can be predicted from histology leveraging deep learning. One such biomarker is TMB, which is associated with specific CNN-derived pathological signatures [68]. In most cancers, including CRC, TMB is associated with improved overall survival after treatment with ICIs [87,88]. This association is attributed to the heightened immune response elicited by the multitude of tumor neoantigens [89–91]. However, tumors with low TMB can respond to ICIs, as such tumors may harbor one or more highly immunogenic mutations. This was demonstrated in a case study of a patient with pembrolizumab-responsive proficient MMR metastatic CRC, who was found to have T-cell responses to at least one neoantigen expressed within their tumor [92]. In addition, a recent study of neoadjuvant ICI treatment showed that there was no difference in pretreatment TMB between early-stage proficient MMR (pMMR) CRCs that responded and those that did not [81]. MMR deficient CRC is substantially more immunogenic than unselected MMR

proficient CRC [10,32], in part due to the high TMB including frameshift insertion-deletion mutations [34].

T cell infiltration in the tumor microenvironment has also been studied as a potential biomarker for ICI response. A recent study of early-stage CRCs including both proficient and deficient MMR tumors showed that neoadjuvant therapy with a combination of nivolumab and ipilimumab elicited a pathological response in all dMMR tumors and 27% of proficient MMR tumors [81]. Proficient MMR tumors that responded to this treatment could be predicted by the presence of TILs that co-expressed CD8 and PD-1 on pre-treatment biopsies. No other biomarkers were found to differ significantly between the proficient MMR tumors that responded and those that did not [81]. Increasing density of CLR was observed after treatment. Other factors likely play a role in the response of proficient MMR tumors to ICIs. For example, high expression of IL-17 has been suggested to abrogate the ICI-responsiveness of such tumors, even in the presence of TILs expressing CD8 and PD-1 [93]. CNNs have been used to predict the spatial organization and subtypes of T cells within the tumor microenvironment of CRC and other tumors [69,94,95]. These features have already been used to predict ICI response in NSCLC with some success [86].

CNNs have also successfully predicted PD-L1 expression [59], another biomarker that has been assessed as a potential predictor of ICI response. However, while mechanistically compelling and potentially predictive of ICI response in NSCLC, PD-L1 does not seem to be useful on its own in determining which CRC will respond to ICI [96]. PD-L1 expression was not associated with progression-free or overall survival in CRC patients treated with ICIs, and there was no significant difference between responders and non-responders in PD-L1 expression in pre-treatment samples from patients with dMMR metastatic CRC [32,97]

Other researchers have used different types of clinical data to train machine learning models to predict ICI response. For example, a machine learning method called ImmuCellAI was developed to predict the relative abundance of various types of T cells in pre-treatment samples of melanomas from gene expression data. They then developed a separate model to predict immunotherapy response based on these results, achieving an AUC of 0.80–0.91 [98]. The successful implementation of a similar machine learning technique involving ICI response prediction based on the expression of immune-related genes was also reported in NSCLC and triple-negative breast cancer [99,100].

Radiomics-based machine learning has also been used to predict response to immunotherapy in melanoma and NSCLC based on a defined set of features extracted from pre-treatment CT imaging of primary and metastatic tumors from patients treated with anti-PD-1 therapy [101]. The model produced a radiomic biomarker score for each lesion evaluated, from which anti-PD-1 response was predicted. By combining the predictions from each of an individual patient's lesions, a patient level prediction of anti-PD-1 response could be made. The model achieved significant performance for both tumor types (AUC 0.76 for NSCLC and 0.77 for melanoma) [101]. A similar study used deep learning to develop a TMB radiomic biomarker that can divide NSCLC tumors into high- and low-TMB groups (AUC 0.81) and divide ICI-treated NSCLC patients into high and low risk groups with significantly different overall and progression-free survival [50]. Another model can predict the transcriptomic-based abundance of CD8 T-cells, and response to immunotherapy, from radiologic data. The resultant biomarker was found to be positively associated with response to treatment with anti-PD-1 and anti-PD-L1 therapies [102]. Fluorodeoxyglucose (FDG)-positron emission tomography (PET) scans have also been used to train deep learning networks to determine a biomarker quantifying CD8+ T cell activity against the tumor that can differentiate between those patients more likely to respond to immunotherapy and those who are less likely to respond [103].

Specific somatic mutations may affect tumor response to ICIs. Previous success in predicting genomic data from histology [58,60,71] suggests that it would be possible to perform for other genes, but validation for individual genes would be required. For example, mutations in the DNA polymerase epsilon (*POLE*) gene, which codes for an enzyme involved in DNA proofreading during replication, can lead to a very high mutational burden

without MSI [104]. The predicted number of neoantigens produced by affected CRCs can significantly exceed that of CRCs with MSI [105]. CRCs harboring *POLE* mutations have quantities of TILs similar to those found in dMMR CRCs [106], demonstrating an adaptive host response to these tumors. At least one case report describes a robust response to pembrolizumab treatment in a patient with a metastatic, treatment-refractory microsatellite stable CRC with a confirmed *POLE* mutation [107]. Clinical trials are ongoing to determine the extent of the benefit of ICI treatment in CRC with *POLE* mutations [108]. However, the impact of these studies will be limited, as *POLE* mutations are only found in about 1–2% of CRCs [106].

While multiple machine learning methods have been used to predict ICI efficacy in tumors other than CRC, and several biomarkers associated with ICI efficacy in CRC have been identified, data are lacking regarding machine learning for the prediction of ICI efficacy in CRC. Optimal ICI response prediction in CRC and other cancers will likely require larger datasets and the integration of multiple types of biomarkers incorporating genetic, immunologic, and other data. Leveraging machine learning to predict ICI response in CRC is an appealing goal given the lack of a sufficiently accurate predictive biomarker. The prediction of ICI response from ubiquitously available clinical data, such as H&E slides and radiographic studies, could greatly improve access to these therapies.

7. Future Directions

Table 3 summarizes the potential advantages, current limitations, and suggestions for future development of machine learning for MSI classification from digital pathology. The potential to predict multiomic data from a universally available clinical specimen has substantial advantages that rely on the ability to achieve excellent classification with CNNs. Data that are not routinely collected (e.g., transcriptomics) can be predicted utilizing relatively limited resources if a digital pathology infrastructure already exists. With the increasing use of omics data for clinical decision making for cancer treatment [109], many more institutions and clinicians could have access to this information. Expanding this technology to mobile phones, as with CNNs discussed above, could allow even greater accessibility [67], but the scalability of these methods to settings with limited resources remains to be demonstrated.

Predicting MSI and other molecular features from H&E slides is an attractive goal given the success of CNNs in similar tasks and initial encouraging results. The focus of the published literature has been on H&E stained slides. It is possible that performance can be further improved by using additional histological stains. The previous performance of models created by human pathologists may also point to the attainability of this goal. The most important hurdle will be to demonstrate, through rigorous clinical trials, that utilizing machine learning on clinical samples is superior or non-inferior to the standard of care, which is itself rapidly evolving and non-uniform. For example, next generation sequencing is increasingly performed on tumor specimens, permitting the identification of MSI as part of a broader, clinically relevant, molecular characterization [110,111]. With decreasing sequencing costs and the need to detect certain mutations clinically (e.g., in *KRAS* and *BRAF)*, accurate genomic predictions in addition to MSI classification may be required from histology-based machine learning methods. Blood based tests have shown good accuracy for predicting MSI from the primary tumor [112] and radiomics have also been proposed for the prediction of MSI status [113–115].

Another major challenge is the generalizability of these methods, which, contrary to molecular methods such as PCR or immunohistochemistry, are often not robust to differing patient or tissue characteristics. The reduction in performance seen in several studies when a trained CNN was applied to new datasets may be a barrier to the widespread implementation of these methods (Table 3).

Table 3. Advantages, limitations, and future directions for MSI classification from digital pathology using machine learning.

	Advantages	Limitations	Future Directions
Classification accuracy	- Rapidly improving accuracy of several CNN architectures	- Insufficient accuracy to obviate universal MSI screening in resource-rich settings currently - Decreased accuracy when applied to biopsy samples - Lacking data on post treatment samples (e.g., rectal cancer)	- Determine optimal CNN architecture/s for distinct clinical scenarios - Leverage machine learning to prioritize molecular testing in settings with limited resources
Generalizability	Excellent performance on well curated cohorts that are similar to training data	Performance not robust to differing patient and tissue characteristics	Increase availability of datasets for global and local model refinement
Accessibility	Potential to expand access to settings without pathology experts or molecular testing, including via cellular devices	- Digital pathology not yet widely available - No published studies conducted in settings with reduced access to healthcare	Design dedicated CNNs for settings with reduced access to healthcare
Clinical endpoint prediction	Very good classification of MSI	No direct prediction of clinical endpoints	Shift from surrogate marker classification to clinical endpoint prediction
Patient selection for immunotherapy		Decreased accuracy of MSI classification in metastatic disease, where immunotherapy is approved	- Directly predict response to immunotherapy, regardless of MSI status - Improve MSI prediction in metastatic disease
Identification of Lynch Syndrome		Inability to distinguish between somatic and germline etiology of MSI, such that confirmatory testing is required	- Predict somatic vs. germline etiology of MSI
Comparison with next generation sequencing (NGS)	Rapid and cost-effective after initial investment	Cannot currently reliably detect *KRAS* or *BRAF* mutations, tumor mutational burden and other clinically actionable alterations	- Attempt to directly predict response to therapy based on histology - Improve classification of other molecular alterations - Integrate genomic and histologic data for optimal prediction
Cost effectiveness	Long term savings on molecular assays	Initial investment required in hardware and software for digital pathology	Expand use of and access to digital pathology

Table 3 outlines the advantages, limitations, and future directions as they relate to detecting MSI from histology using machine learning. CNN: convolutional neural network.

The few peer-reviewed data that are available suggest that the accuracy of the current machine learning algorithms for the prediction of MSI may not yet be sufficient to guide clinical care in high-resource settings. However, as methods continue to improve and more training datasets become available, it is plausible that CNNs will be able to predict MSI status more accurately. Predictions for differing populations, primary cancer sites and tissue preparation methods (including true biopsies) are some of the challenges that exist. The accuracy of prediction for post-treatment pathology slides has not been explored and may be relevant for rectal cancer patients undergoing neoadjuvant therapy. Until the performance of CNNs improves, emphasis could be placed on achieving a near-perfect sensitivity for the detection of MSI, tolerating a certain number of false positives. MSI/dMMR assays could be avoided for most samples, but confirmation of CNN-predicted MSI would be required. Since MSI is a biomarker, the major potential for machine learning to improve upon MSI testing is if it were able to predict clinically relevant genomic features, such as *MLH1* hypermethylation or germline dMMR mutations, or clinically relevant outcomes, such as response to chemotherapy and immunotherapy (Table 3). The Immunoscore already uses digital pathology for prognostication [10,57] but machine learning could be utilized to improve predictions and to identify subsets of patients with microsatellite stable CRC that could benefit from immunotherapy [81].

Another limitation of the current research is that, to our knowledge, all CNN models trained to identify MSI status in CRC have been trained on surgical samples derived from resection of the primary tumor. Under current guidelines, immunotherapy is most commonly used to treat patients with stage IV tumors, who often have tissue samples available only from biopsies of metastatic sites. If machine learning models are not able to accurately predict MSI based on such samples, one of the most promising applications of MSI prediction from histological samples would be restricted to a much smaller segment of potential beneficiaries. Thus, future research should work to optimize machine learning algorithms for the prediction of MSI from biopsy samples of distant metastases.

Lastly, to accelerate the acceptance of CNNs as clinical tools and inform other areas of research, further insights into the features that drive CNN classification are needed. Without understanding what features CNNs are using to classify images, there exists a risk of introducing bias and error.

Funding: This work received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data used for this review is available by accessing the citations below.

Acknowledgments: A.M. would like to thank Itay Maoz for helpful discussions regarding machine learning and artificial intelligence. All the authors would like to thank the Boston University Medical Center Internal Medicine Residency Program for supporting research during residency.

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

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef] [PubMed]
2. Araghi, M.; Soerjomataram, I.; Jenkins, M.; Brierley, J.; Morris, E.; Bray, F.; Arnold, M. Global trends in colorectal cancer mortality: Projections to the year 2035. *Int. J. Cancer* **2019**, *144*, 2992–3000. [CrossRef] [PubMed]
3. Siegel, R.L.; Miller, K.D.; Sauer, A.G.; Fedewa, S.A.; Butterly, L.F.; Anderson, J.C.; Cercek, A.; Smith, R.A.; Jemal, A. Colorectal cancer statistics, 2020. *CA Cancer J. Clin.* **2020**, *70*, 145–164. [CrossRef] [PubMed]
4. Lal, N.; White, B.S.; Goussous, G.; Pickles, O.; Mason, M.J.; Beggs, A.D.; Taniere, P.; Willcox, B.E.; Guinney, J.; Middleton, G. KRAS Mutation and Consensus Molecular Subtypes 2 and 3 Are Independently Associated with Reduced Immune Infiltration and Reactivity in Colorectal Cancer. *Clin. Cancer Res.* **2018**, *24*, 224–233. [CrossRef] [PubMed]

5. Becht, E.; De Reyniès, A.; Giraldo, N.A.; Pilati, C.; Buttard, B.; Lacroix, L.; Selves, J.; Sautès-Fridman, C.; Laurent-Puig, P.; Fridman, W.H. Immune and Stromal Classification of Colorectal Cancer Is Associated with Molecular Subtypes and Relevant for Precision Immunotherapy. *Clin. Cancer Res.* **2016**, *22*, 4057–4066. [CrossRef] [PubMed]
6. Guinney, J.; Dienstmann, R.; Wang, X.; De Reyniès, A.; Schlicker, A.; Soneson, C.; Marisa, L.; Roepman, P.; Nyamundanda, G.; Angelino, P.; et al. The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **2015**, *21*, 1350–1356. [CrossRef]
7. Seppälä, T.T.; Bohm, J.; Friman, M.; Lahtinen, L.; Väyrynen, V.M.J.; Liipo, T.K.E.; Ristimäki, A.P.; Kairaluoma, M.V.J.; Kellokumpu, I.H.; Kuopio, T.H.I.; et al. Combination of microsatellite instability and BRAF mutation status for subtyping colorectal cancer. *Br. J. Cancer* **2015**, *112*, 1966–1975. [CrossRef]
8. Sjoquist, K.M.; Renfro, L.A.; Simes, R.J.; Tebbutt, N.C.; Clarke, S.; Seymour, M.T.; Adams, R.; Maughan, T.S.; Saltz, L.; Goldberg, R.M.; et al. Personalizing Survival Predictions in Advanced Colorectal Cancer: The ARCAD Nomogram Project. *J. Natl. Cancer Inst.* **2018**, *110*, 638–648. [CrossRef]
9. Hynes, S.O.; Coleman, H.G.; Kelly, P.J.; Irwin, S.; O'Neill, R.F.; Gray, R.T.; Mcgready, C.; Dunne, P.D.; McQuaid, S.; James, J.A.; et al. Back to the future: Routine morphological assessment of the tumour microenvironment is prognostic in stage II/III colon cancer in a large population-based study. *Histopathology* **2017**, *71*, 12–26. [CrossRef]
10. Mlecnik, B.; Bindea, G.; Angell, H.K.; Maby, P.; Angelova, M.; Tougeron, D.; Church, S.E.; Lafontaine, L.; Fischer, M.; Fredriksen, T.; et al. Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability. *Immunity* **2016**, *44*, 698–711. [CrossRef]
11. Rozek, L.S.; Schmit, S.L.; Greenson, J.K.; Tomsho, L.P.; Rennert, H.S.; Rennert, G.; Gruber, S.B. Tumor-Infiltrating Lymphocytes, Crohn's-Like Lymphoid Reaction, and Survival from Colorectal Cancer. *J. Natl. Cancer Inst.* **2016**, *108*. [CrossRef] [PubMed]
12. Li, G.-M. Mechanisms and functions of DNA mismatch repair. *Cell Res.* **2007**, *18*, 85–98. [CrossRef] [PubMed]
13. Vilar, E.; Gruber, S.B. Microsatellite instability in colorectal cancer—The stable evidence. *Nat. Rev. Clin. Oncol.* **2010**, *7*, 153–162. [CrossRef] [PubMed]
14. Cerretelli, G.; Ager, A.; Arends, M.J.; Frayling, I.M. Molecular pathology of Lynch syndrome. *J. Pathol.* **2020**, *250*, 518–531. [CrossRef]
15. Boland, P.M.; Yurgelun, M.B.; Boland, C.R. Recent progress in Lynch syndrome and other familial colorectal cancer syndromes. *CA Cancer J. Clin.* **2018**, *68*, 217–231. [CrossRef] [PubMed]
16. Sinicrope, F.A.; Rego, R.L.; Halling, K.C.; Foster, N.; Sargent, D.J.; La Plant, B.; French, A.J.; Laurie, J.A.; Goldberg, R.M.; Thibodeau, S.N.; et al. Prognostic Impact of Microsatellite Instability and DNA Ploidy in Human Colon Carcinoma Patients. *Gastroenterology* **2006**, *131*, 729–737. [CrossRef]
17. Samowitz, W.S.; Curtin, K.; Ma, K.N.; Schaffer, D.; Coleman, L.W.; Leppert, M.; Slattery, M.L. Microsatellite instability in sporadic colon cancer is associated with an improved prognosis at the population level. *Cancer Epidemiol. Biomark. Prev.* **2001**, *10*, 917–923.
18. Maoz, A.; Dennis, M.; Greenson, J.K. The Crohn's-Like Lymphoid Reaction to Colorectal Cancer-Tertiary Lymphoid Structures with Immunologic and Potentially Therapeutic Relevance in Colorectal Cancer. *Front. Immunol.* **2019**, *10*, 1884. [CrossRef]
19. Overman, M.J.; McDermott, R.; Leach, J.L.; Lonardi, S.; Lenz, H.-J.; Morse, M.A.; Desai, J.; Hill, A.; Axelson, M.; Moss, R.A.; et al. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): An open-label, multicentre, phase 2 study. *Lancet Oncol.* **2017**, *18*, 1182–1191. [CrossRef]
20. Ribic, C.M.; Sargent, D.J.; Moore, M.J.; Thibodeau, S.N.; French, A.J.; Goldberg, R.M.; Hamilton, S.R.; Laurent-Puig, P.; Gryfe, R.; Shepherd, L.E.; et al. Tumor Microsatellite-Instability Status as a Predictor of Benefit from Fluorouracil-Based Adjuvant Chemotherapy for Colon Cancer. *N. Engl. J. Med.* **2003**, *349*, 247–257. [CrossRef]
21. De Vos tot Nederveen Cappel, W.H.; Meulenbeld, H.J.; Kleibeuker, J.H.; Nagengast, F.M.; Menko, F.H.; Griffioen, G.; Cats, A.; Morreau, H.; Gelderblom, H.; Vasen, H.F.A. Survival after adjuvant 5-FU treatment for stage III colon cancer in hereditary nonpolyposis colorectal cancer. *Int. J. Cancer* **2004**, *109*, 468–471. [CrossRef] [PubMed]
22. Sargent, D.J.; Marsoni, S.; Monges, G.; Thibodeau, S.N.; Labianca, R.; Hamilton, S.R.; French, A.J.; Kabat, B.; Foster, N.R.; Torri, V.; et al. Defective Mismatch Repair as a Predictive Marker for Lack of Efficacy of Fluorouracil-Based Adjuvant Therapy in Colon Cancer. *J. Clin. Oncol.* **2010**, *28*, 3219–3226. [CrossRef] [PubMed]
23. Webber, E.M.; Kauffman, T.L.; O'Connor, E.; Goddard, K.A. Systematic review of the predictive effect of MSI status in colorectal cancer patients undergoing 5FU-based chemotherapy. *BMC Cancer* **2015**, *15*, 1–12. [CrossRef] [PubMed]
24. Sepulveda, A.R.; Hamilton, S.R.; Allegra, C.J.; Grody, W.; Cushman-Vokoun, A.M.; Funkhouser, W.K.; Kopetz, S.E.; Lieu, C.; Lindor, N.M.; Minsky, B.D.; et al. Molecular Biomarkers for the Evaluation of Colorectal Cancer: Guideline From the American Society for Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and American Society of Clinical Oncology. *J. Mol. Diagn.* **2017**, *19*, 187–225. [CrossRef]
25. Shaikh, T.; Handorf, E.A.; Meyer, J.E.; Hall, M.J.; Esnaola, N.F. Mismatch Repair Deficiency Testing in Patients with Colorectal Cancer and Nonadherence to Testing Guidelines in Young Adults. *JAMA Oncol.* **2018**, *4*, e173580. [CrossRef]
26. Greenson, J.K.; Huang, S.-C.; Herron, C.; Moreno, V.; Bonner, J.D.; Tomsho, L.P.; Ben-Izhak, O.; Cohen, H.I.; Trougouboff, P.; Bejhar, J.; et al. Pathologic Predictors of Microsatellite Instability in Colorectal Cancer. *Am. J. Surg. Pathol.* **2009**, *33*, 126–133. [CrossRef]
27. Jenkins, M.A.; Hayashi, S.; O'Shea, A.-M.; Burgart, L.J.; Smyrk, T.C.; Shimizu, D.; Waring, P.M.; Ruszkiewicz, A.R.; Pollett, A.F.; Redston, M.; et al. Pathology Features in Bethesda Guidelines Predict Colorectal Cancer Microsatellite Instability: A Population-Based Study. *Gastroenterology* **2007**, *133*, 48–56. [CrossRef]

28. Greenson, J.K.; Bonner, J.D.; Ben-Yzhak, O.; Cohen, H.I.; Miselevich, I.; Resnick, M.B.; Trougouboff, P.; Tomsho, L.D.; Kim, E.; Low, M.; et al. Phenotype of Microsatellite Unstable Colorectal Carcinomas: Well-Differentiated and Focally Mucinous Tumors and the Absence of Dirty Necrosis Correlate With Microsatellite Instability. *Am. J. Surg. Pathol.* **2003**, *27*, 563–570. [CrossRef]
29. Walsh, M.D.; Cummings, M.C.; Buchanan, D.D.; Dambacher, W.M.; Arnold, S.; McKeone, D.; Byrnes, R.; Barker, M.A.; Leggett, B.A.; Gattas, M.; et al. Molecular, Pathologic, and Clinical Features of Early-Onset Endometrial Cancer: Identifying Presumptive Lynch Syndrome Patients. *Clin. Cancer Res.* **2008**, *14*, 1692–1700. [CrossRef]
30. Alexander, J.; Watanabe, T.; Wu, T.-T.; Rashid, A.; Li, S.; Hamilton, S.R. Histopathological Identification of Colon Cancer with Microsatellite Instability. *Am. J. Pathol.* **2001**, *158*, 527–535. [CrossRef]
31. Buckowitz, A.; Knaebel, H.-P.; Benner, A.; Bläker, H.; Gebert, J.; Kienle, P.; Doeberitz, M.V.K.; Kloor, M. Microsatellite instability in colorectal cancer is associated with local lymphocyte infiltration and low frequency of distant metastases. *Br. J. Cancer* **2005**, *92*, 1746–1753. [CrossRef] [PubMed]
32. Le, D.T.; Uram, J.N.; Wang, H.; Bartlett, B.R.; Kemberling, H.; Eyring, A.D.; Skora, A.D.; Luber, B.S.; Azad, N.S.; Laheru, D.; et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N. Engl. J. Med.* **2015**, *372*, 2509–2520. [CrossRef] [PubMed]
33. Williams, D.S.; Bird, M.J.; Jorissen, R.N.; Yu, Y.L.; Walker, F.; Zhang, H.H.; Nice, E.C.; Burgess, A.W. Nonsense Mediated Decay Resistant Mutations Are a Source of Expressed Mutant Proteins in Colon Cancer Cell Lines with Microsatellite Instability. *PLoS ONE* **2010**, *5*, e16012. [CrossRef] [PubMed]
34. Willis, J.A.; Reyes-Uribe, L.; Chang, K.; Lipkin, S.M.; Vilar, E. Immune Activation in Mismatch Repair–Deficient Carcinogenesis: More Than Just Mutational Rate. *Clin. Cancer Res.* **2020**, *26*, 11–17. [CrossRef] [PubMed]
35. Umar, A.; Boland, C.R.; Terdiman, J.P.; Syngal, S.; De La Chapelle, A.; Rüschoff, J.; Fishel, R.; Lindor, N.M.; Burgart, L.J.; Hamelin, R.; et al. Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability. *J. Natl. Cancer Inst.* **2004**, *96*, 261–268. [CrossRef]
36. Hyde, A.; Fontaine, D.; Stuckless, S.; Green, R.; Pollett, A.; Simms, M.; Sipahimalani, P.; Parfrey, P.; Younghusband, B. A Histology-Based Model for Predicting Microsatellite Instability in Colorectal Cancers. *Am. J. Surg. Pathol.* **2010**, *34*, 1820–1829. [CrossRef]
37. Fujiyoshi, K.; Yamaguchi, T.; Kakuta, M.; Takahashi, A.; Arai, Y.; Yamada, M.; Yamamoto, G.; Ohde, S.; Takao, M.; Horiguchi, S.-I.; et al. Predictive model for high-frequency microsatellite instability in colorectal cancer patients over 50 years of age. *Cancer Med.* **2017**, *6*, 1255–1263. [CrossRef]
38. Roman, R.; Verdu, M.; Calvo, M.; Vidal, A.; Sanjuan, X.; Jimeno, M.; Salas, A.; Autonell, J.; Trias, I.; González, M.; et al. Microsatellite instability of the colorectal carcinoma can be predicted in the conventional pathologic examination. A prospective multicentric study and the statistical analysis of 615 cases consolidate our previously proposed logistic regression model. *Virchows Arch.* **2010**, *456*, 533–541. [CrossRef]
39. Bessa, X.; Alenda, C.; Paya, A.; Álvarez, C.; Iglesias, M.; Seoane, A.; Dedeu, J.M.; Abulí, A.; Ilzarbe, L.; Navarro, G.; et al. Validation Microsatellite Path Score in a Population-Based Cohort of Patients with Colorectal Cancer. *J. Clin. Oncol.* **2011**, *29*, 3374–3380. [CrossRef]
40. Brazowski, E.; Rozen, P.; Pel, S.; Samuel, Z.; Solar, I.; Rosner, G. Can a gastrointestinal pathologist identify microsatellite instability in colorectal cancer with reproducibility and a high degree of specificity? *Fam. Cancer* **2012**, *11*, 249–257. [CrossRef]
41. Bera, K.; Schalper, K.A.; Rimm, D.L.; Velcheti, V.; Madabhushi, A. Artificial intelligence in digital pathology—New tools for diagnosis and precision oncology. *Nat. Rev. Clin. Oncol.* **2019**, *16*, 703–715. [CrossRef] [PubMed]
42. Acs, B.; Rantalainen, M.; Hartman, J. Artificial intelligence as the next step towards precision pathology. *J. Intern. Med.* **2020**, *288*, 62–81. [CrossRef] [PubMed]
43. 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] [PubMed]
44. Rashidi, H.H.; Tran, N.K.; Betts, E.V.; Howell, L.P.; Green, R. Artificial Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods. *Acad. Pathol.* **2019**, *6*, 2374289519873088. [CrossRef] [PubMed]
45. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
46. Mobadersany, P.; Yousefi, S.; Amgad, M.; Gutman, D.A.; Barnholtz-Sloan, J.S.; Vega, J.E.V.; Brat, D.J.; Cooper, L.A. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, e2970–e2979. [CrossRef]
47. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
48. Wolf, L.; Hassner, T.; Maoz, I. Face recognition in unconstrained videos with matched background similarity. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 529–534.
49. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
50. 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.
51. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1–9.

52. Bizzego, A.; Bussola, N.; Chierici, M.; Maggio, V.; Francescatto, M.; Cima, L.; Cristoforetti, M.; Jurman, G.; Furlanello, C. Evaluating reproducibility of AI algorithms in digital pathology with DAPPER. *PLoS Comput. Biol.* **2019**, *15*, e1006269. [CrossRef]
53. Nagpal, K.; Foote, D.; Liu, Y.; Chen, P.-H.C.; Wulczyn, E.; Tan, F.; Olson, N.; Smith, G.; Mohtashamian, A.; Wren, J.H.; et al. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *NPJ Digit. Med.* **2019**, *2*, 1–10. [CrossRef]
54. Bychkov, D.; Linder, N.; Turkki, R.; Nordling, S.; Kovanen, P.E.; Verrill, C.; Walliander, M.; Lundin, M.; Haglund, C.; Lundin, J. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci. Rep.* **2018**, *8*, 1–11. [CrossRef]
55. Kather, J.N.; Krisam, J.; Charoentong, P.; Luedde, T.; Herpel, E.; Weis, C.-A.; Gaiser, T.; Marx, A.; Valous, N.A.; Ferber, D.; et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS Med.* **2019**, *16*, e1002730. [CrossRef] [PubMed]
56. Skrede, O.-J.; De Raedt, S.; Kleppe, A.; Hveem, T.S.; Liestøl, K.; Maddison, J.; Askautrud, H.A.; Pradhan, M.; Nesheim, J.A.; Albregtsen, F.; et al. Deep learning for prediction of colorectal cancer outcome: A discovery and validation study. *Lancet* **2020**, *395*, 350–360. [CrossRef]
57. Pagès, F.; Mlecnik, B.; Marliot, F.; Bindea, G.; Ou, F.-S.; Bifulco, C.; Lugli, A.; Zlobec, I.; Rau, T.T.; Berger, M.D.; et al. International validation of the consensus Immunoscore for the classification of colon cancer: A prognostic and accuracy study. *Lancet* **2018**, *391*, 2128–2139. [CrossRef]
58. Coudray, N.; Ocampo, P.S.; Sakellaropoulos, T.; Narula, N.; Snuderl, M.; Fenyo, 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] [PubMed]
59. Yip, S.S.F.; Sha, L.; Osinski, B.L.; Ho, I.Y.; Tan, T.L.; Willis, C.; Weiss, H.; Beaubier, N.; Mahon, B.M.; Taxter, T.J. Multi-field-of-view deep learning model predicts nonsmall cell lung cancer programmed death-ligand 1 status from whole-slide hematoxylin and eosin images. *J. Pathol. Inform.* **2019**, *10*, 24. [CrossRef] [PubMed]
60. Couture, H.D.; Williams, L.A.; Geradts, J.; Nyante, S.J.; Butler, E.N.; Marron, J.S.; Perou, C.M.; Troester, M.A.; Niethammer, M. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. *NPJ Breast Cancer* **2018**, *4*, 1–8. [CrossRef] [PubMed]
61. Sirinukunwattana, K.; Domingo, E.; Richman, S.D.; Redmond, K.L.; Blake, A.; Verrill, C.; Leedham, S.J.; Chatzipli, A.; Hardy, C.; Whalley, C.M.; et al. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. *Gut* **2020**. [CrossRef]
62. Steu, S.; Baucamp, M.; Von Dach, G.; Bawohl, M.; Dettwiler, S.; Storz, M.; Moch, H.; Schraml, P. A procedure for tissue freezing and processing applicable to both intra-operative frozen section diagnosis and tissue banking in surgical pathology. *Virchows Arch.* **2008**, *452*, 305–312. [CrossRef]
63. Sprung, R.W.; Brock, J.W.C.; Tanksley, J.P.; Li, M.; Washington, M.K.; Slebos, R.J.C.; Liebler, D.C. Equivalence of Protein Inventories Obtained from Formalin-fixed Paraffin-embedded and Frozen Tissue in Multidimensional Liquid Chromatography-Tandem Mass Spectrometry Shotgun Proteomic Analysis. *Mol. Cell. Proteom.* **2009**, *8*, 1988–1998. [CrossRef]
64. Yeh, Y.-C.; Nitadori, J.-I.; Kadota, K.; Yoshizawa, A.; Rekhtman, N.; Moreira, A.L.; Sima, C.S.; Rusch, V.W.; Adusumilli, P.S.; Travis, W.D. Using frozen section to identify histological patterns in stage I lung adenocarcinoma of $\leq$3 cm: Accuracy and interobserver agreement. *Histopathology* **2015**, *66*, 922–938. [CrossRef]
65. Ratnavelu, N.D.; Brown, A.P.; Mallett, S.; Scholten, R.J.; Patel, A.; Founta, C.; Galaal, K.; Cross, P.; Naik, R. Intraoperative frozen section analysis for the diagnosis of early stage ovarian cancer in suspicious pelvic masses. *Cochrane Database Syst. Rev.* **2016**, *2016*, 010360. [CrossRef]
66. Mantel, H.T.; Westerkamp, A.C.; Sieders, E.; Peeters, P.M.J.G.; De Jong, K.P.; Boer, M.T.; de Kleine, R.H.; Gouw, A.S.H.; Porte, R.J. Intraoperative frozen section analysis of the proximal bile ducts in hilar cholangiocarcinoma is of limited value. *Cancer Med.* **2016**, *5*, 1373–1380. [CrossRef]
67. Kather, J.N.; Heij, L.R.; Grabsch, H.I.; Loeffler, C.; Echle, A.; Muti, H.S.; Krause, J.; Niehues, J.M.; Sommer, K.A.J.; Bankhead, P.; et al. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat. Rev. Cancer* **2020**, *1*, 789–799. [CrossRef]
68. Cao, R.; Yang, F.; Ma, S.-C.; Liu, L.; Zhao, Y.; Li, Y.; Wu, D.-H.; Wang, T.; Lu, W.-J.; Cai, W.-J.; et al. Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in Colorectal Cancer. *Theranostics* **2020**, *10*, 11080–11091. [CrossRef] [PubMed]
69. Schmauch, B.; Romagnoni, A.; Pronier, E.; Saillard, C.; Maillé, P.; Calderaro, J.; Kamoun, A.; Sefta, M.; Toldo, S.; Zaslavskiy, M.; et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nat. Commun.* **2020**, *11*, 1–15. [CrossRef] [PubMed]
70. Zhang, R.; Osinski, B.L.; Taxter, T.J.; Perera, J.; Lau, D.J.; Khan, A.A. Adversarial deep learning for microsatellite instability prediction from histopathology slides. In Proceedings of the 1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands, 4–6 July 2018.
71. Hong, R.; Liu, W.; DeLair, D.; Razavian, N.; Fenyö, D. Predicting Endometrial Cancer Subtypes and Molecular Features from Histopathology Images Using Multi-resolution Deep Learning Models. *bioRxiv* **2020**. [CrossRef]
72. Echle, A.; Grabsch, H.I.; Quirke, P.; van den Brandt, P.A.; West, N.P.; Hutchins, G.G.; Heij, L.R.; Tan, X.; Richman, S.D.; Krause, J.; et al. Clinical-Grade Detection of Microsatellite Instability in Colorectal Tumors by Deep Learning. *Gastroenterology* **2020**, *159*, 1406–1416.e11. [CrossRef]

73. Gatalica, Z.; Snyder, C.; Maney, T.; Ghazalpour, A.; Holterman, D.A.; Xiao, N.; Overberg, P.; Rose, I.; Basu, G.D.; Vranic, S.; et al. Programmed Cell Death 1 (PD-1) and Its Ligand (PD-L1) in Common Cancers and Their Correlation with Molecular Cancer Type. *Cancer Epidemiol. Biomark. Prev.* **2014**, *23*, 2965–2970. [CrossRef]
74. Wang, H.; Wang, X.; Xu, L.; Zhang, J.; Cao, H. Analysis of the transcriptomic features of microsatellite instability subtype colon cancer. *BMC Cancer* **2019**, *19*, 605. [CrossRef]
75. 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.
76. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, MA, USA, 18–23 June 2018; pp. 6848–6856.
77. Quellec, G.; Cazuguel, G.; Cochener, B.; Lamard, M. Multiple-Instance Learning for Medical Image and Video Analysis. *IEEE Rev. Biomed. Eng.* **2017**, *10*, 213–234. [CrossRef] [PubMed]
78. Wang, T.; Lu, W.; Yang, F.; Liu, L.; Dong, Z.; Tang, W.; Chang, J.; Huan, W.; Huang, K.; Yao, J. Microsatellite Instability Prediction of Uterine Corpus Endometrial Carcinoma Based on H&E Histology Whole-Slide Imaging. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 1289–1292. [CrossRef]
79. Khan, A.A. Generalizable and Interpretable Deep Learning Framework for Predicting MSI from Histopathology Slide Images. U.S. Patent 20190347557, 14 November 2019.
80. Hellmann, M.D.; Kim, T.-W.; Lee, C.; Goh, B.-C.; Miller, W.; Oh, D.-Y.; Jamal, R.; Chee, C.-E.; Chow, L.; Gainor, J.; et al. Phase Ib study of atezolizumab combined with cobimetinib in patients with solid tumors. *Ann. Oncol.* **2019**, *30*, 1134–1142. [CrossRef]
81. Chalabi, M.; Fanchi, L.F.; Dijkstra, K.K.; Berg, J.G.V.D.; Aalbers, A.G.; Sikorska, K.; Lopez-Yurda, M.; Grootscholten, C.; Beets, G.L.; Snaebjornsson, P.; et al. Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. *Nat. Med.* **2020**, *26*, 566–576. [CrossRef]
82. Le, D.T.; Kim, T.W.; Van Cutsem, E.; Geva, R.; Jäger, D.; Hara, H.; Burge, M.; O'Neil, B.; Kavan, P.; Yoshino, T.; et al. Phase II Open-Label Study of Pembrolizumab in Treatment-Refractory, Microsatellite Instability–High/Mismatch Repair–Deficient Metastatic Colorectal Cancer: KEYNOTE-164. *J. Clin. Oncol.* **2020**, *38*, 11–19. [CrossRef] [PubMed]
83. Postow, M.; Sidlow, R.; Hellmann, M.D. Immune-Related Adverse Events Associated with Immune Checkpoint Blockade. *N. Engl. J. Med.* **2018**, *378*, 158–168. [CrossRef]
84. Das, S.; Johnson, D.B. Immune-related adverse events and anti-tumor efficacy of immune checkpoint inhibitors. *J. Immunother. Cancer* **2019**, *7*, 1–11. [CrossRef]
85. Johannet, P.; Coudray, N.; Jour, G.; Donnelly, D.M.; Bajaj, S.; Moran, U.; Dawood, Z.; Nomikou, S.; Kim, R.H.; Pavlick, A.C.; et al. Using machine learning algorithms to predict response and toxicity to immune checkpoint inhibitors (ICIs) in melanoma patients. *J. Clin. Oncol.* **2019**, *37*, 2581. [CrossRef]
86. Barrera, C.V.; Velu, P.; Bera, K.; Wang, X.; Prasanna, P.; Khunger, M.; Khunger, A.; Velcheti, V.; Romero, E.; Madabhushi, A. Computer-extracted features relating to spatial arrangement of tumor infiltrating lymphocytes to predict response to nivolumab in non-small cell lung cancer (NSCLC). *J. Clin. Oncol.* **2018**, *35*, 12115. [CrossRef]
87. Samstein, R.M.; Lee, C.; Shoushtari, A.N.; Hellmann, M.D.; Shen, R.; Janjigian, Y.Y.; Barron, D.A.; Zehir, A.; Jordan, E.J.; Omuro, A.; et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* **2019**, *51*, 202–206. [CrossRef]
88. Schrock, A.; Ouyang, C.; Sandhu, J.; Sokol, E.; Jin, D.; Ross, J.; Miller, V.; Lim, D.; Amanam, I.; Chao, J.; et al. Tumor mutational burden is predictive of response to immune checkpoint inhibitors in MSI-high metastatic colorectal cancer. *Ann. Oncol.* **2019**, *30*, 1096–1103. [CrossRef]
89. Mandal, R.; Samstein, R.M.; Lee, K.-W.; Havel, J.J.; Wang, H.; Krishna, C.; Sabio, E.Y.; Makarov, V.; Kuo, F.; Blecua, P.; et al. Genetic diversity of tumors with mismatch repair deficiency influences anti–PD-1 immunotherapy response. *Science* **2019**, *364*, 485–491. [CrossRef] [PubMed]
90. Liu, L.; Bai, X.; Wang, J.; Tang, X.-R.; Wu, D.-H.; Du, S.; Du, X.-J.; Zhang, Y.-W.; Zhu, H.-B.; Fang, Y.; et al. Combination of TMB and CNA Stratifies Prognostic and Predictive Responses to Immunotherapy Across Metastatic Cancer. *Clin. Cancer Res.* **2019**, *25*, 7413–7423. [CrossRef] [PubMed]
91. Tran, E.; Ahmadzadeh, M.; Lu, Y.-C.; Gros, A.; Turcotte, S.; Robbins, P.F.; Gartner, J.J.; Zheng, Z.; Li, Y.F.; Ray, S.; et al. Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science* **2015**, *350*, 1387–1390. [CrossRef] [PubMed]
92. Smith, K.N.; Llosa, N.J.; Cottrell, T.R.; Siegel, N.; Fan, H.; Suri, P.; Chan, H.Y.; Guo, H.; Oke, T.; Awan, A.H.; et al. Persistent mutant oncogene specific T cells in two patients benefitting from anti-PD-1. *J. Immunother. Cancer* **2019**, *7*, 40. [CrossRef]
93. Llosa, N.J.; Luber, B.; Tam, A.J.; Smith, K.N.; Siegel, N.; Awan, A.H.; Fan, H.; Oke, T.; Zhang, J.; Domingue, J.; et al. Intratumoral Adaptive Immunosuppression and Type 17 Immunity in Mismatch Repair Proficient Colorectal Tumors. *Clin. Cancer Res.* **2019**, *25*, 5250–5259. [CrossRef]
94. Saltz, J.H.; Gupta, R.; Hou, L.; Kurc, T.; Singh, P.; Nguyen, V.; Samaras, D.; Shroyer, K.R.; Zhao, T.; Batiste, R.; et al. Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. *Cell Rep.* **2018**, *23*, 181–193.e7. [CrossRef]

95. Lu, Z.; Xu, S.; Shao, W.; Wu, Y.; Zhang, J.; Han, Z.; Feng, Q.; Huang, K. Deep-Learning–Based Characterization of Tumor-Infiltrating Lymphocytes in Breast Cancers From Histopathology Images and Multiomics Data. *JCO Clin. Cancer Inform.* **2020**, *4*, 480–490. [CrossRef]
96. Huyghe, N.; Baldin, P.; Eynde, M.V.D. Immunotherapy with immune checkpoint inhibitors in colorectal cancer: What is the future beyond deficient mismatch-repair tumours? *Gastroenterol. Rep.* **2019**, *8*, 11–24. [CrossRef]
97. Chakrabarti, S.; Huebner, L.J.; Finnes, H.D.; Muranyi, A.; Clements, J.; Singh, S.; Hubbard, J.M.; McWilliams, R.R.; Shanmugam, K.; Sinicrope, F.A. Intratumoral CD3+ and CD8+ T-Cell Densities in Patients with DNA Mismatch Repair–Deficient Metastatic Colorectal Cancer Receiving Programmed Cell Death-1 Blockade. *JCO Precis. Oncol.* **2019**, *3*, 1–7. [CrossRef]
98. Miao, Y.-R.; Zhang, Q.; Lei, Q.; Luo, M.; Xie, G.; Wang, H.; Guo, A.-Y. ImmuCellAI: A Unique Method for Comprehensive T-Cell Subsets Abundance Prediction and its Application in Cancer Immunotherapy. *Adv. Sci.* **2020**, *7*, 1902880. [CrossRef]
99. Wiesweg, M.; Mairinger, F.; Reis, H.; Goetz, M.; Walter, R.F.H.; Hager, T.; Metzenmacher, M.; Eberhardt, W.; McCutcheon, A.; Köster, J.; et al. Machine learning-based predictors for immune checkpoint inhibitor therapy of non-small-cell lung cancer. *Ann. Oncol.* **2019**, *30*, 655–657. [CrossRef] [PubMed]
100. Bao, X.; Shi, R.; Zhao, T.; Wang, Y.; Anastasov, N.; Rosemann, M.; Fang, W. Integrated analysis of single-cell RNA-seq and bulk RNA-seq unravels tumour heterogeneity plus M2-like tumour-associated macrophage infiltration and aggressiveness in TNBC. *Cancer Immunol. Immunother.* **2020**, 1–14. [CrossRef] [PubMed]
101. Trebeschi, S.; Drago, S.; Birkbak, N.; Kurilova, I.; Călin, A.; Pizzi, A.D.; Lalezari, F.; Lambregts, D.; Rohaan, M.; Parmar, C.; et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann. Oncol.* **2019**, *30*, 998–1004. [CrossRef] [PubMed]
102. Sun, R.; Limkin, E.J.; Vakalopoulou, M.; Dercle, L.; Champiat, S.; Han, S.R.; Verlingue, L.; Brandao, D.; Lancia, A.; Ammari, S.; et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: An imaging biomarker, retrospective multicohort study. *Lancet Oncol.* **2018**, *19*, 1180–1191. [CrossRef]
103. Park, C.; Na, K.J.; Choi, H.; Ock, C.-Y.; Ha, S.; Kim, M.; Park, S.; Keam, B.; Kim, T.M.; Paeng, J.C.; et al. Tumor immune profiles noninvasively estimated by FDG PET with deep learning correlate with immunotherapy response in lung adenocarcinoma. *Theranostics* **2020**, *10*, 10838–10848. [CrossRef]
104. Briggs, S.; Tomlinson, I.P.M. Germline and somatic polymerase ϵ and δ mutations define a new class of hypermutated colorectal and endometrial cancers. *J. Pathol.* **2013**, *230*, 148–153. [CrossRef]
105. Giannakis, M.; Mu, X.J.; Shukla, S.A.; Qian, Z.R.; Cohen, O.; Nishihara, R.; Bahl, S.; Cao, Y.; Amin-Mansour, A.; Yamauchi, M.; et al. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep.* **2016**, *17*, 1206. [CrossRef]
106. Domingo, E.; Freeman-Mills, L.; Rayner, E.; Glaire, M.A.; Briggs, S.; Vermeulen, L.; Fessler, E.; Medema, J.P.; Boot, A.; Morreau, H.; et al. Somatic POLE proofreading domain mutation, immune response, and prognosis in colorectal cancer: A retrospective, pooled biomarker study. *Lancet Gastroenterol. Hepatol.* **2016**, *1*, 207–216. [CrossRef]
107. Gong, J.; Wang, C.; Lee, P.P.; Chu, P.; Fakih, M. Response to PD-1 Blockade in Microsatellite Stable Metastatic Colorectal Cancer Harboring aPOLEMutation. *J. Natl. Compr. Cancer Netw.* **2017**, *15*, 142–147. [CrossRef]
108. Picard, E.; Verschoor, C.P.; Ma, G.W.; Pawelec, G. Relationships Between Immune Landscapes, Genetic Subtypes and Responses to Immunotherapy in Colorectal Cancer. *Front. Immunol.* **2020**, *11*, 369. [CrossRef]
109. Rodon, J.; Soria, J.-C.; Berger, R.; Miller, W.H.; Rubin, E.; Kugel, A.; Tsimberidou, A.; Saintigny, P.; Ackerstein, A.; Braña, I.; et al. Genomic and transcriptomic profiling expands precision cancer medicine: The WINTHER trial. *Nat. Med.* **2019**, *25*, 751–758. [CrossRef] [PubMed]
110. Salipante, S.J.; Scroggins, S.M.; Hampel, H.L.; Turner, E.H.; Pritchard, C.C. Microsatellite Instability Detection by Next Generation Sequencing. *Clin. Chem.* **2014**, *60*, 1192–1199. [CrossRef]
111. Middha, S.; Zhang, L.; Nafa, K.; Jayakumaran, G.; Wong, D.; Kim, H.R.; Sadowska, J.; Berger, M.F.; Delair, D.F.; Shia, J.; et al. Reliable Pan-Cancer Microsatellite Instability Assessment by Using Targeted Next-Generation Sequencing Data. *JCO Precis. Oncol.* **2017**, *2017*, 1–17. [CrossRef]
112. Willis, J.; Lefterova, M.I.; Artyomenko, A.; Kasi, P.M.; Nakamura, Y.; Mody, K.; Catenacci, D.V.; Fakih, M.; Barbacioru, C.; Zhao, J.; et al. Validation of Microsatellite Instability Detection Using a Comprehensive Plasma-Based Genotyping Panel. *Clin. Cancer Res.* **2019**, *25*, 7035–7045. [CrossRef]
113. Fan, S.; Li, X.; Cui, X.; Zheng, L.; Ren, X.; Ma, W.; Ye, Z. Computed Tomography-Based Radiomic Features Could Potentially Predict Microsatellite Instability Status in Stage II Colorectal Cancer: A Preliminary Study. *Acad. Radiol.* **2019**, *26*, 1633–1640. [CrossRef]
114. Pernicka, J.S.G.; Gagniere, J.; Chakraborty, J.; Yamashita, R.; Nardo, L.; Creasy, J.M.; Petkovska, I.; Do, R.R.K.; Bates, D.D.B.; Paroder, V.; et al. Radiomics-based prediction of microsatellite instability in colorectal cancer at initial computed tomography evaluation. *Abdom. Radiol.* **2019**, *44*, 3755–3763. [CrossRef]
115. Wu, J.; Zhang, Q.; Zhao, Y.; Liu, Y.; Chen, A.; Li, X.; Wu, T.; Li, J.; Guo, Y.; Liu, A. Radiomics Analysis of Iodine-Based Material Decomposition Images with Dual-Energy Computed Tomography Imaging for Preoperatively Predicting Microsatellite Instability Status in Colorectal Cancer. *Front. Oncol.* **2019**, *9*, 1250. [CrossRef]