ImmunoInformatics 3–4 (2021) 100008

Contents lists available at ScienceDirect

ImmunoInformatics

journal homepage: www.elsevier.com/locate/immuno

Deep learning for the detection of microsatellite instability from histology images in colorectal cancer: A systematic literature review

Amelie Echle[a], Narmin Ghaffari Laleh[a], Peter L. Schrammen[a], Nicholas P. West[b], Christian Trautwein[a], Titus J. Brinker[c], Stephen B. Gruber[d], Roman D. Buelow[e], Peter Boor[e], Heike I. Grabsch[b,f], Philip Quirke[b], Jakob N. Kather[a,b,g,*]

[a] Department of Medicine III, University Hospital RWTH Aachen, Pauwelsstr. 30, Aachen 52074, Germany
[b] Pathology & Data Analytics, Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, United Kingdom
[c] Digital Biomarkers for Oncology Group, National Center for Tumor Diseases (NCT), German Cancer Research Center (DKFZ), Heidelberg, Germany
[d] Center for Precision Medicine and Department of Medical Oncology, City of Hope National Medical Center, Duarte, CA, United States
[e] Institute of Pathology, University Hospital RWTH Aachen, Aachen, Germany
[f] Department of Pathology, GROW School for Oncology and Developmental Biology, Maastricht University Medical Center+, Maastricht, Netherlands
[g] Medical Oncology, National Center for Tumor Diseases (NCT), University Hospital Heidelberg, Heidelberg, Germany

A B S T R A C T

Microsatellite instability (MSI) or deficient mismatch repair (dMMR) is a clinically important genetic feature affecting 10–15% of colorectal cancer (CRC) patients. Patients with metastatic MSI/dMMR CRC are eligible for therapy with immune checkpoint inhibitors, making MSI/dMMR the most important immuno-oncological biomarker in CRC. Gold standard tests for detection of MSI/dMMR in CRC are based on wet laboratory tests such as immunohistochemistry (IHC) or DNA extraction with subsequent polymerase chain reaction (PCR). However, since 2019, advances in Deep Learning (DL), an Artificial Intelligence (AI) technology, have enabled the prediction of MSI/dMMR directly from digitized routine haematoxylin and eosin (H&E) histopathology slides with high accuracy. In addition to the initial proof-of-concept publication in 2019, twelve subsequent studies have refined, improved, and further validated this approach. At this moment, MSI/dMMR prediction using Deep Learning has become a widely used benchmark task for academic studies in the field of computational pathology. Beyond academic use, this assay has attracted commercial interest from companies with the possibility of approval as a diagnostic device in the near future. In this review, we summarize and quantitatively compare the existing evidence on Deep-Learning-based detection of MSI/dMMR in CRC and discuss the need for further improvement and potential for integration into routine pathological workflows. Ultimately, this DL-based method could facilitate the identification of patients eligible for treatment with immune checkpoint inhibitors by pre-screening or replacement of current methods.

Introduction

The application of AI technology in the field of histopathology has developed rapidly over the last three years. Supervised analysis of histopathological images by Deep Learning methods moved from studies reporting technical advances to real-world applications with a clinical focus [1,2]. Important factors contributing to this development are (1) the increasing use of whole slide scanners in pathology departments generating an enormous number of histopathological images and (2) the evolving quality and availability of algorithms, computing power, and storage media [3]. In addition, advances in new technologies are often driven by the availability of specific benchmark tasks. A benchmark task is a clearly defined problem on which new algorithms can be tested and results can be compared to previous approaches. In the context of computational pathology, prediction of microsatellite instability (MSI) or mismatch repair deficiency (dMMR) from hematoxylin and eosin (H&E) stained histology slides of colorectal cancer (CRC) has become such a benchmark task and multiple studies have specifically addressed this question, suggesting new algorithms and comparing performance to the initial studies [4–17]. In this systematic literature review, we aimed to provide a critical review of studies performing MSI/dMMR prediction on H&E-stained CRC tissue sections with a brief comparison to DL based MSI/dMMR prediction in other cancer types. Differences in methodology and results as well as limitations and possible future work are discussed.

Background: microsatellite instability as a key biomarker to select patients for immunotherapy

Microsatellites are short tandem repeat non-coding DNA sequences that are widely distributed throughout the human genome [18]. These sequences are vulnerable to spontaneous replication errors. In healthy cells, replication errors are constantly corrected by the so-called Mismatch Repair (MMR) system which consists of two major protein com-

* Corresponding author.
E-mail address: jkather@ukaachen.de (J.N. Kather).

https://doi.org/10.1016/j.immuno.2021.100008
Received 25 June 2021; Received in revised form 8 December 2021; Accepted 10 December 2021

plexes. During CRC carcinogenesis, the MMR system can be compromised (become 'deficient') due to the loss of expression of its proteins [19]. Tumor cells with a deficient MMR system are unable to correct replication errors which consequently leads to many mutations in coding DNA sequences. One of the reasons why Microsatellite instability (MSI) or deficient mismatch repair (dMMR) detection are widely studied is its immediate clinical relevance. Approximately 10–15% of colorectal cancer (CRC) display dMMR at protein level [20], and presence of dMMR is almost completely overlapping with presence of MSI at DNA level [21]. Medical guidelines recommend that all CRC patients should undergo screening for dMMR/MSI for a number of reasons [21,22]. First, dMMR/MSI in combination with other molecular alterations can be an indicator for the presence of Lynch syndrome, the most common cause of hereditary CRC [21]. Second, dMMR/MSI in intermediate stage CRC (pT3–4, N0–2) has been associated with a reduced response to Fluorouracil-based chemotherapy as well as lower incidence of locoregional metastases [23]. Finally, the presence of dMMR/MSI is suggestive of potential efficacy of cancer immunotherapy with immune checkpoint inhibitors [24,25] as most dMMR/MSI tumors are highly immunogenic [26]. In metastatic CRC, dMMR/MSI is currently the only biomarker which renders patients eligible for treatment with immune checkpoint inhibitors, as approved by the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA). The EMA approval of immune checkpoint inhibitors in dMMR/MSI CRC includes the anti-PD1 antibody Pembrolizumab in first-line palliative therapy [27] and the combination of anti-CTLA4 antibody Ipilimumab with the anti-PD1 antibody Nivolumab in second-line palliative therapy [28]. According to the FDA approval for anti-PD1 antibodies, dMMR/MSI is not only a biomarker for first-line immunotherapy in metastatic or locally advanced unresectable CRC but also for any unresectable or metastatic solid tumor that has progressed following initial treatment [25,29]. The FDA approval of immune checkpoint inhibitors in dMMR/MSI patients currently allows the use of two diagnostic methods to detect this genetic alteration: (1) detection of loss of expression of one of the MMR proteins by immunohistochemistry (IHC) and/or (2) detection of the presence of MSI using extracted tumor DNA and a polymerase chain reaction (PCR) panel or next-generation sequencing (NGS) [30,31]. Recently, a proof-of-concept study suggested that dMMR/MSI can also be detected directly from routine pathology slides by using a Deep Learning-based diagnostic method [4]. It is currently envisaged that the DL-based assay could be used as a pre-screening tool reducing the number (and costs) of wet laboratory tests or as definitive diagnostic test [4,32].

Histopathological features as predictors of MSI status on H&E slides

The rationale why Deep Learning could detect dMMR/MSI from routine histology slides is that there are known morphological patterns of dMMR/MSI CRC. In other words, humans can see patterns which are associated with dMMR/MSI and so Deep Learning can also detect such patterns (Fig. 1). As early as the 1990s, a relationship between MSI genotype and morphological tumor phenotype was demonstrated by several studies. [31,33-36] Lists of clinico-pathological features enriched in patients with Lynch syndrome, so-called Amsterdam and Bethesda criteria, were formally published in 1991 and 1996 respectively. [31,36] The aim of these criteria was the identification of patients with high likelihood of hereditary CRC, provide genetic counseling to the index patient and relatives and include patients and relatives in appropriate surveillance programs.

In 2003, *Greenson* et al. were the first to investigate morphological phenotypic markers of MSI in CRC in a population-based study. By analyzing 528 CRCs they identified tumor infiltrating lymphocytes, poor differentiation, right-sided location, mucinous differentiation, Crohn's like inflammatory reaction in the periphery of the tumor and a lack of so-called dirty necrosis as independent predictors of microsatellite instability. [35] Example histology images of MSI/dMMR and MSS/pMMR tumors are shown in Fig. 1. Six years later, *Greenson* et al. expanded their analysis by increasing the sample size and confirming tumor-infiltrating lymphocytes, Crohn's-like inflammatory reaction, lack of dirty necrosis within the tumor lumen, and mucinous differentiation as independent histological predictors of MSI presence together with the clinical characteristics of age <50 years and a right-sided location of the CRC. When combining all these features within a MSI scoring system, a pathologist looking at the histological slide was able to predict MSI status in 1649 CRC with an area under the receiver operating characteristic curve (AUROC) of 0.85 [34].

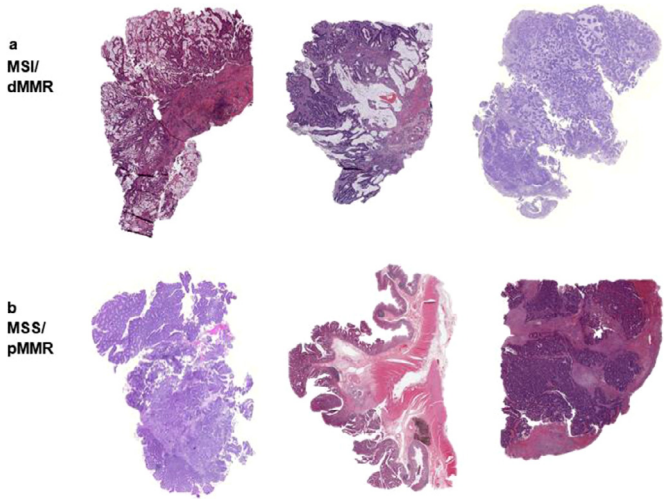

Fig. 1. Example histology slides for MSI/dMMR and MSS/pMMR colorectal tumors. (a) MSI/dMMR tumors display some typical morphological patterns such as a high number of tumor infiltrating lymphocytes, poor differentiation or the presence of mucin, which are known indicators of MSI/dMMR [34]. However, quantification of these features by human experts is not sufficiently accurate for a definitive diagnosis of MSI/dMMR in clinical routine. (b) MSS/pMMR tumors. This collection of histological images also highlights the variability in terms of color, size and presence of different non-tumor tissue types on histopathology slides.

In 2007, *Jenkins* et al. systematically quantified the association between CRC histological features in H&E slides, clinical parameters, and presence of MSI in a large population-based study of CRC patients. Their study confirmed clinical and pathological features such as age at diagnosis younger than 50 years, right-sided tumor location, presence of tumor-infiltrating lymphocytes, mucinous, (focal) signet ring cell differentiation, or undifferentiated histology and a Crohn's-like inflammatory reaction as predictors for the presence of MSI [33]. This approach was validated in an independent patient cohort and obtained an AUROC of 0.89 [33].

In a more recent study in 2021, a much lower performance for pathologist-based MSI/dMMR detection was reported: *Yamashita* et al. evaluated the unassisted prediction of pathologists which reached an average AUROC of 0.605 with a low inter-rater agreement for the unassisted prediction of MSI from whole slide images (WSI) [10]. However, in contrast to the two above mentioned pathologist-based studies, pathologists in this study were blinded to clinical information such as patient age or tumor location.

These H&E based study results from human observers suggest that modern Deep Learning based methods should be able to predict MSI/dMMR status from analyzing H&E tissue sections.

Methods

To identify studies using Deep Learning-based methodology for the prediction of MSI or dMMR status in CRC histology images, we searched

the Medline database using different queries including studies published between January 2017 to August 2021. This literature research was performed by AE and JNK independently.

To be eligible for inclusion in the current review, studies had to be original research papers that used Machine Learning methods to investigate the detection of MSI/dMMR on CRC histology images and had to be published in English. We did not include studies that used other imaging modalities than digitized histopathological images like Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) scans. The initial search terms were "(deep learning) AND (microsatellite instability)", which resulted in 18 studies including 6 reporting original methods for MSI prediction in CRC (based on review of title and abstract). Next, we used the terms "(machine learning) AND (microsatellite instability)" resulting in 35 studies, "(artificial intelligence) AND (microsatellite instability)" resulting in 26 studies and last "(convolutional neural networks) AND (microsatellite instability)" which led to one result. However, none of search terms listed in the previous sentence resulted in additional studies of original methods for dMMR/MSI detection in CRC being found. Most search results were either review articles or studies investigating images from other sources than histology such as from MRI or CT scans, other tumor types, other prediction targets or did not cover image related research but the analysis of RNA sequencing data. To include preprints or publications which are not yet listed in the Medline database, we queried the Google Scholar Database with the terms "deep learning" and "microsatellite instability" and "colorectal cancer", which yielded 577 results. After manual review of titles, article previews and abstracts, we identified 7 additional studies reporting original methods.

Thus, in total, we identified 13 studies published between July 2019 to April 2021, listed in Table 1. To compare the performance of study results, we used the AUROC values as this was the most commonly reported endpoint. As one of the studies did not report AUROC values, we report the F-score, a statistical measure of a test's accuracy for a binary classifier.

Deep learning for MSI/DMMR detection in colorectal cancer

Common methods

The very first step in all Deep Learning (DL) studies is the collection of scanned whole slide images from H&E-stained tumor tissue sections and corresponding clinical information such as MSI/dMMR status as determined by a gold standard method. One important and widely used data source is the publicly available Cancer Genome Atlas (TCGA), from which almost all recent histopathological Deep Learning studies obtained data [37].

Before a histopathological image can be used to train a DL network, several preparatory steps are necessary. The H&E-stained tissue section needs to be digitized using a whole slide scanner, a process which may differ between institutions with respect to scanner model, scanning quality or magnification factor. As soon as images are digitally available for computational analysis, they need to undergo preprocessing steps that may vary slightly between different studies [38]. First, a decision needs to be made whether all available slides from one patient will be used for the DL-based study or whether a particular slide will be selected based on predefined criteria. Second, a decision needs to be made whether the whole tissue on a given slide or only tumor tissue will be used for training purposes. If aiming for the latter, manual or automated tumor annotations will be necessary. Due to the potential large data size of a single scanned WSI and currently still technically limited input size for training neural networks, smaller image patches (tiles) need to be created. After creating image tiles of the whole tissue or from annotated regions of interest, image tiles can be color-normalized to account for variability of staining hue and intensity. Preprocessing methods between studies are summarized in Table 1.

Table 1
Comparison of DL studies for MSI detection in CRC.

Paper	Date of publication	# patients	Performance	Methods			
				Tumor detection	Color norm.	DL model	other aspects
Human performance							
Jenkins et al. [33]	July 2007	1098	AUC = 0.89	N/A	N/A	N/A	MSI scoring system (pathology WSIs + clinical parameters)
Greenson et al. [34]	January 2009	1649	AUC = 0.85	N/A	N/A	N/A	
Yamashita et al. [10]	January 2021	40	AUC = 0.605	N/A	N/A	N/A	Pathology WSI
External validation							
Kather et al. [4]	July 2019	738	AUC = 0.84	auto	yes	ResNet18	N/A
Kather et al. [5]	July 2020	805	AUC = 0.89	manual	yes	Shufflenet	N/A
Cao et al. [6]	September 2020	121	AUC = 0.649	manual	yes	ResNet18	new method to generate slide-level predictions
Echle et al. [7]	October 2020	8836	AUC = 0.96	manual	yes	Shufflenet	N/A
Bilal et al. [8]	January 2021	~300	AUC = 0.98	auto (*training*), manual (*ext. validation*)	yes	ResNet34	analysis of cellular composition
Yamashita et al. [10]	January 2021	579	AUC = 0.78	auto	yes	MobileNetV2	Tissue classifier
Yamashita et al. [9]	February 2021	200	AUC = 0.876	no	yes	MobileNetV2	Medically-irrelevant random style augmentation
Lee et al. [16]	April 2021	484	AUC = 0.787	auto	yes	Inception V3	additional analysis of metastasized CRC
Internal testing							
Schmauch et al. [12]	August 2020	465	AUC = 0.82	no	no	ResNet18	N/A
Zhu et al. [13]	October 2020	360	AUC = 0.81	manual	yes	ResNet18	feature extraction highlights the importance of texture and color
Ke et al. [14]	December 2020	360	AUC = 0.802	no	not reported	mislabel-aware classifier	
Lee et al. [15]	December 2020	57	F-score = 0.83	manual	yes	Inception-Resnet-V2	multiclass classification method (normal tissue vs. MSI vs. MSS)
Schirris et al. [17]	July 2021	360	AUC = 9.903	manual	yes	SimCLR + VarMIL	pre-trained feature extractor and Multiple Instance Learning with a feature variability module

Norm. Normalization, N/A not applicable, AUC Area under the curve, WSI whole-slide image, # number of.

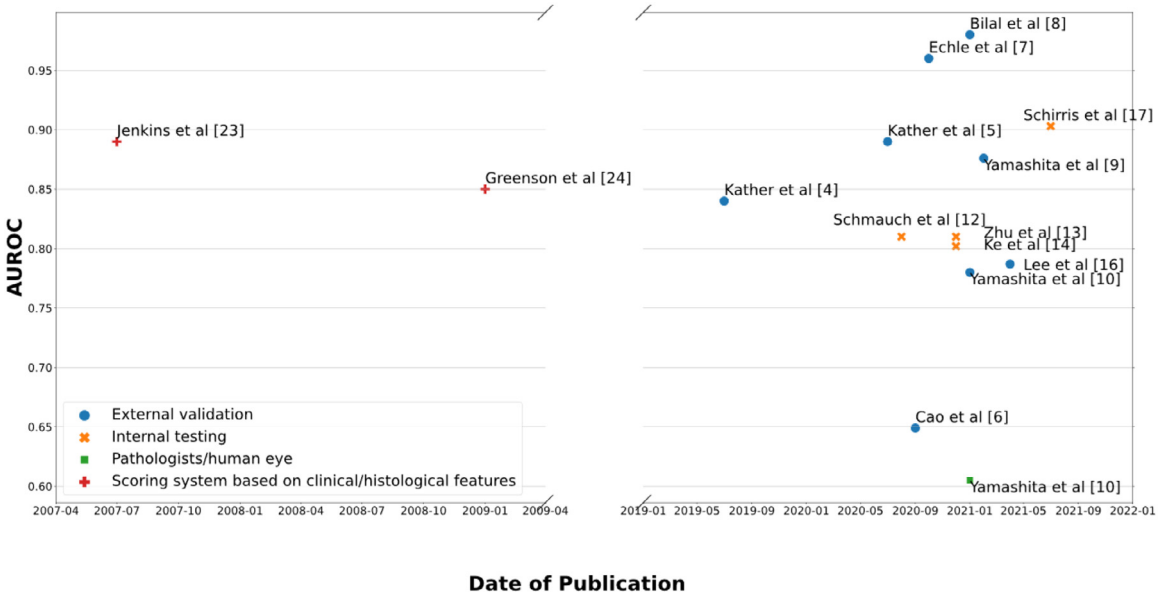

Fig. 2. Overview of Deep Learning studies for MSI/dMMR detection in colorectal cancer histology. Comparison of performances of Deep Learning studies that performed an external validation or only internal testing and studies using an MSI/dMMR scoring system. AUROC, area under the receiver-operator curve.

The first studies using deep learning for MSI/DMMR prediction (2019 and 2020)

The first fully automated, end-to-end Deep-Learning-based detection of MSI/dMMR status in CRC was published in July 2019 [4]. *Kather* et al. presented a supervised transfer learning approach in which a classifier based on a pretrained resnet18 network was trained on CRC H&E images from TCGA and evaluated on a second independent subset of CRCs from the DACHS trial, a large-scale multi-institutional cohort from southwest Germany. [39,40] This led to a reasonable performance with an AUROC of 0.84 within the TCGA cohort as well as in the DACHS cohort [4]. Building on these initial results, a slightly different technical pipeline using ShuffleNet, e.g. as a different neural network model, together with manually annotated tumor regions yielded a higher external prediction performance of MSI/dMMR status in DACHS with an AUROC of 0.89 [5].

Validation by multiple research groups (2020)

After these two initial studies, six subsequent studies published in 2020 confirmed the ability of DL to predict MSI/dMMR status in CRC H&E-stained digitized tissue sections (Fig. 2) by further improving the existing methods as well as by developing new technical approaches:

In August 2020, *Schmauch* et al. presented a method to infer gene expression profiles from CRC H&E images of the TCGA database using DL [12]. Compared to previous studies they investigated a new approach by first using a subset of patients to train a model to predict RNA-sequencing data and second applying transfer learning to this pre-trained model to train the MSI/dMMR status classifier. This method of DL-based MSI/dMMR status prediction resulted in an AUROC of 0.81 which was higher than the reported AUROC of 0.71 for prediction of MSI/dMMR status in the same subset of the TCGA data without using the pre-trained gene expression model. These results suggest that pre-training a model on transcriptomic representation might help the model to detect biologically relevant patterns and could therefore be useful to improve prediction performances of DL models. Furthermore, in an additional experiment, they performed a direct prediction of MSI/dMMR status from WSI without using a pre-trained gene expression model in the full TCGA CRC dataset and reported an AUROC of 0.82 [12]. This study was primarily driven by a commercial entity (Owkin, Inc.; 12 Rue Martel, 75,010 Paris, France) demonstrating the commercial interest in using DL to predict the presence of MSI/dMMR. Notably, another commercial entity (Tempus, Inc., 600 W Chicago Ave. Ste 510, Chicago, IL 60,654, USA), has filed a US Patent application #20,190,347,557 (retrieved from Google Patents on 1st April 2021) for a DL based system to predict MSI/dMMR status from histopathology images. These two publicly available pieces of information demonstrate the early commercial interest in the use of this technology.

Also in 2020, *Cao et al.* developed an MSI/DMMR status using the TCGA CRC image dataset in a similar way as previously described by *Kather* et al. (based on a pretrained resnet18 network). This approach resulted in a comparable AUROC of 0.885 in the TCGA dataset, but performance dropped to an AUROC of 0.649 in an external validation set. By using 10% of cases from the external validation cohort for fine-tuning the model, prediction performance increased to an AUROC of 0.85 for the external validation set [6]. This study demonstrated the challenges and difficulties of training a DL-based MSI/dMMR status classifier that is robust and shows a similar performance in unseen data of external validation cohorts. In addition, *Cao* et al. compared methods to generate the needed prediction on WSI level from the predictions of single tiles resulting in a new model for this task that combines different methods of Multiple Instance Learning [6].

Zhu et al. repeated MSI/dMMR status prediction in the CRC TCGA dataset using the already preprocessed images provided by *Kather et al.* [41] achieving the same AUROC of 0.84. In addition, they performed a feature-level analysis and identified texture characteristics and color features of H&E images as important for the prediction of MSI/dMMR status as they contributed the most to the MSI/dMMR predictions [13].

The strategies provide interpretability in two aspects. On the one hand, the image-level interpretability is achieved by generating localization heat maps of important regions based on the deep learning network; on the other hand, the feature-level interpretability is attained through feature importance and pathological feature interaction analysis. More interestingly, both from the image-level and feature-level interpretability, color features and texture characteristics are shown to contribute the most to the MSI predictions [13].

In October 2020, results from a study with the largest and most heterogeneous CRC patient cohort so far was published by *Echle et al.* [7]. Images from more than 8500 CRC patients collected within the academic MSIDETECT consortium (www.msidetect.eu) from different

countries and institutions were used for the development of a DL based MSI/dMMR status classifier. This classifier obtained an AUROC of 0.96 in an additional external validation cohort [7], which was the best performing DL classifier for MSI/dMMR status prediction from H&E images at that point. The sensitivity of this classifier is similar to that of the gold standard immunohistochemical and PCR based tests [30]. The most relevant difference in this approach compared to the previous ones was the number of histopathological images available for the DL study, which appears to be the main reason for the good performance as this study demonstrated an increasing MSI/dMMR status prediction performance with increasing number of patients [7].

At the end of 2020, results from two smaller studies using only one patient cohort were published. One of them by *Ke* et al. yielded an AUROC of 0.802 using the CRC TCGA data[14]. Their study focused on the potential problem of mislabeling image tiles in the data. Mislabeling, or noisy labels, possibly occurs when all image tiles inherit the label of the WSI, which is the method used in all previous studies. Their aim was to develop a robust and mislabel-aware classifier by cleaning the data before training the DL network. Therefore, the authors built a model to mark image tiles as high or low fidelity and pathologists tagged the most representative samples for noise-robust training [14]. However, this approach seems to rely primarily on human performance and with an AUROC of 0.802 did not improve MSI prediction performance in the TCGA CRC dataset when compared to previous studies.

Lee et al. used a dataset of just 45 patients from the "Pathology Artificial Intelligence Platform (PAIP) - Challenge" [42]. This study proposed a two-stage classification method. First, the regions with tumor tissue are detected and second, the MSI/dMMR status DL model trains on the detected tumor tissue, a method that was also used by *Kather* et al. in 2019. The authors reported a good performance by this two-stage classification model but did not report the AUROC, limiting the quantitative comparison to other studies. The F-score was 0.83 in this case [15].

Recent advances resulting in further performance boost (2021)

Contributions to the field continued at the beginning of 2021 with a publication by *Yamashita* et al. using a proprietary cohort from Stanford University for training, reaching a high intra-cohort AUROC of 0.93. However, performance of the MSI/dMMR status prediction models dropped to an AUROC of 0.78 for external validation in the TCGA CRC dataset. This study thus confirmed the challenges of training an MSI/dMMR status classifier which maintains its good performance in unseen data of external validation cohorts. Whilst MSI/dMMR status prediction in CRC histology images was similar to previous studies, a new two-step method was introduced: first a tissue-type classifier was used to select tumor-epithelial and mucin containing tiles which were then used for training of the MS/dMMR status DL classifier in a second step [10]. This more complex methodical pipeline is an example for ever-growing complexity of DL workflows on the MSI/dMMR status benchmark task. While hand-crafting complex image processing pipelines can improve performance, such an approach might also have weaknesses when compared to simple, off-the-shelf approaches: the more processing steps any pipeline has, the more potential breakpoints are present - if there are more steps in an image processing pipeline this could make the analysis more error-prone and more difficult to reproduce.

Also, in 2021, *Bilal et al.* presented the highest classification performance for MSI/dMMR status prediction in CRC so far. They used CRC TCGA data to train a DL system which led to an intra-cohort AUROC of 0.86. When applying the MSI/dMMR status classifier to the small PAIP challenge dataset of 47 patients as external validation cohort, the DL classifier yielded an AUROC of 0.98 [8]. Their DL framework involved three models, first a tumor detection model, second the MSI/dMMR prediction model and third a segmentation model to analyze cellular composition of image parts ranked highly predictive of MSI/dMMR status by the DL classifier. Furthermore, they used DL for the detection of other molecular alterations such as CpG island methylator phenotype and chromosomal instability as well as for BRAF and TP53 mutation prediction [8]. Certainly, such high performance for MSI/dMMR status prediction in CRC histopathological images is very encouraging but needs to be interpreted with caution as the validation cohort comprises only 47 images which were manually chosen by the PAIP challenge committee. Therefore, the validation cohort might potentially not be representative of real-world CRC patient cohorts.

In a study from February 2021, *Yamashita* et al. used the same Stanford and TCGA dataset as in their above-mentioned study to investigate the improvement of a DL classifier when trained on H&E images that underwent medically-irrelevant style augmentation, a method that was not investigated in any of the previous studies. This means that the style of a random artistic painting replaces the style of the histopathology image (including texture, color, and contrast) with an uninformative style while preserving global object shapes [9]. This led to an AUROC of 0.876 in the TCGA cohort used as an external validation set showing a pronounced improvement to their findings as well as those of other groups in the TCGA CRC dataset and also outperforming other normalization and augmentation methods in a systematic comparison [9].

In April 2021, *Lee* et al. developed a DL based image analysis pipeline that first uses a classifier to exclude artifacts, second a tumor detection network and in the last step a MSI/dMMR status prediction network. They used the TCGA CRC dataset for training and validated their classifier in a cohort from Seoul reaching an AUROC of 0.78. Compared to the previous study, a new aspect investigated in this work is the prediction of MSI on distant metastasis of CRC. Applying the TCGA based primary CRC classifier to metastatic CRC in the liver or lung led to an initial AUROC of only 0.484. However, retraining and testing a DL network within the metastasized tumors only yielded much better results with an AUROC of 0.801 [16].

In the most recent study from July 2021, *Schirris* et al. used the TCGA-based dataset provided by *Kather et al. [41]* to use a two-step approach for MSI/dMMR detection. First, a simple framework for contrastive learning of visual representations (SimCLR, [43]) was used to pre-train a feature extractor. Second, a Multiple Instance Learning (MIL) approach was extended by a feature variability module to generate the MSI/dMMR prediction. By that combination of methods, the authors achieved an AUC of 0.903 within the TCGA dataset and outperformed previous approaches for that dataset [17].

All methodological approaches and detailed results of all mentioned studies are summarized in Table 1 and Fig. 2. While the study inclusion for this article ended in April 2021, it can be assumed that further studies will increase the performance even more: in non-medical domains, image classification accuracy on complex tasks has continuously increased from 2012 to 2021 [44]. By applying these improved non-medical DL models to medical tasks, the performance on medical benchmark tasks, such as MSI/dMMR status prediction in CRC, will conceivably increase in the future. In the next section, we will point out limitations and potential routes for further development.

Visualization methods and limitations

Deep learning detects histological patterns with known link to MSI/DMMR status

More than a decade ago, Greenson et al. presented and validated visual features in H&E-stained tissue sections from CRC which were correlated with MSI/dMMR status [34]. An important plausibility check of studies using Deep Learning to predict MSI/dMMR status is whether the automatic approach learns these known visual features or whether the unbiased nature of training Deep Learning networks on image data allows the networks to identify additional, previously unknown visual features. Various methods have been proposed to provide explainability and visualization of such classifiers [45–47]. Applying visualization methods also raises the awareness towards potential biases such as batch

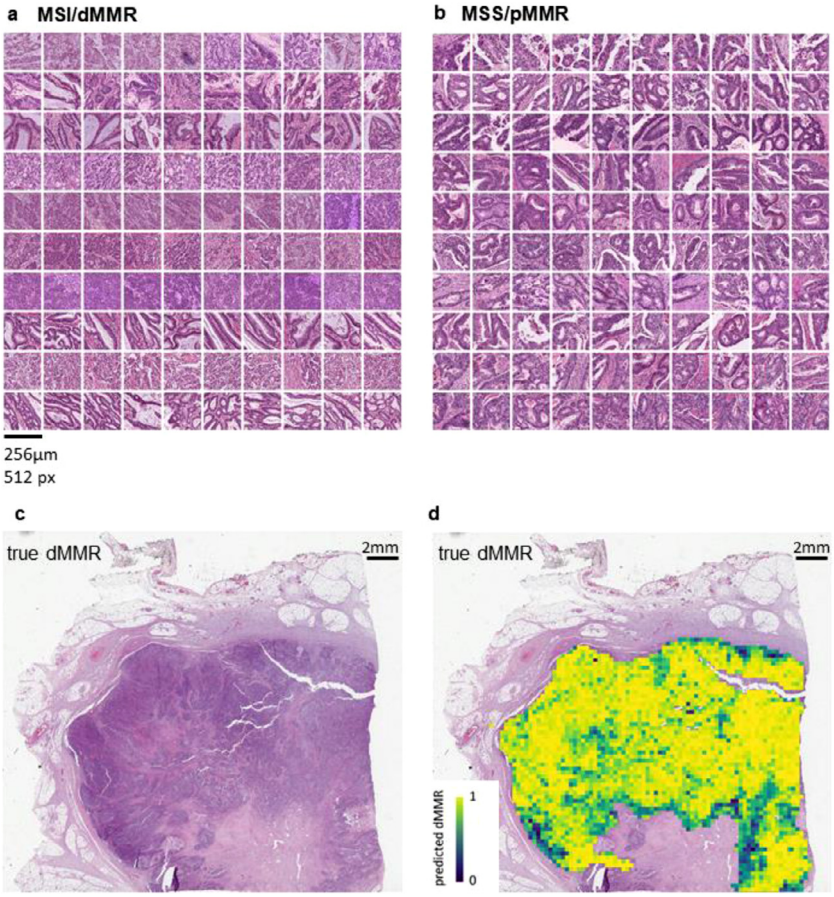

Fig. 3. Visualization approaches for feature detection in histopathology. (a-b) The ten image tiles (columns) with the highest prediction score (as predicted by a DL model) in the ten highest-scoring patients (rows) for microsatellite instable / deficient mismatch repair (MSI/dMMR) and microsatellite stable / proficient mismatch repair (MSS/pMMR). These tiles are based on predictions from a Deep Learning model of a previously published work [7]. (c) An original whole slide image of CRC from a patient with deficient MMR, (d) Corresponding visualization heatmap of tile-level scores of MSI/dMMR status.

effects due to different staining or scanning techniques which can arise when samples are derived from different institutions [48]. One widely-used method is to visualize and analyze image parts which are classified as highly predictive by the Deep Learning network (Fig. 3a-b). In this visualization of "highly predictive image tiles", known histopathological predictors of MSI/dMMR status such as lymphocytic infiltration and poor tumor differentiation have been shown to be present in MSI/dMMR tiles[7,8] as for example shown in Fig. 3a. Although these image tiles are only a few out of several million tiles processed by such models, and the size of these tiles is much smaller than tissue regions which are usually assessed by pathologists, the ability of deep learning models to rank tiles and let the user interactively explore the content of these tiles provides a potentially useful tool for pathology research in the context of MSI/dMMR and beyond. Related to such a collection of highly scoring image tiles, tile-level predictions have been visualized as heat maps by multiple studies [4,5,12,47]. These landscapes can illustrate the spatial context of highly scoring tiles, potentially providing additional possibilities to discover visual features (Fig. 3d).

The importance of common standards for deep learning classifiers

Because of the known presence of histopathological features and the rapidly evolving DL techniques as well as the growing interest in their possible clinical utility, it is not surprising that more and more research groups started working on MSI/dMMR status prediction from scanned histopathological slides since 2019. However, studies so far vary markedly regarding the number of patients and cohorts. Moreover, not all classifiers were evaluated in external cohorts, which is considered helpful to prove generalizability to unseen data as needed in a possible diagnostic setting. In particular, all studies listed in Table 1 used slightly different methods pipelines. Therefore, no conclusion can be made from existing studies which DL model or preprocessing pipeline is universally optimal. Recently *Kleppe* et al. posited that predefined analysis protocols for Deep Learning studies are needed to prevent selection biases [49]. Generally, consented quality attributes will be crucial for the further development of this research field and should be taken into account by every person contributing to the evolution. Minimal criteria for artificial-intelligence-based diagnostic accuracy studies are currently being developed by the STARD diagnostic guidelines group [50]. Furthermore, MI-CLAIM checklist may serve as a potential documentation standard [51]. However, these checklists are not yet widely adopted in the field and none of the above-mentioned studies explicitly declared adherence to all relevant guidelines. As suggested by those guidelines most of the authors give public access to their trained models at the time of publication, which is a very positive trend allowing other research groups to reproduce experiment results and directly compare their own work. For example, the publicly available image data and programming codes from the first DL model for MSI prediction by *Kather* et al. from 2019[4] were re-used by three of the studies presented in this review. [6,9,13]

Future directions - how will the field evolve?

Technological improvements

Over time, due to the increased computational power of computers, it became possible to train deep learning models with a larger amount of available open-source datasets [52–54] and use transfer learning and achieve relatively high performance even in smaller patient cohorts with less than 300 patients. Based on the focus of non-medical fields on developing new deep learning models, development of newer model architectures and more efficient classifiers are expected to continue. As an ex-

ample of recently developed models, multi-instance learning, a weakly supervised approach, and completely new artificial intelligence models such as vision transformers [55] were proposed, but have not been applied to the problem of MSI/dMMR status prediction yet. In other benchmark tasks, these models can result in better performances and higher explainability of the trained models [56,57]. Therefore, application of these new approaches to the task of MSI/dMMR status prediction in large cohorts is still awaited. The increasing accessibility of Deep Learning technologies - for example, through initiatives like FastAI (www.fast.ai) decreases the obstacles for biological and medical researchers to enter the field of computational pathology. This increasing accessibility of such easily usable programming libraries could conceivably further boost innovation and thus improve the performance of MSI/dMMR status classifiers. Ultimately, but on a longer time scale, further improvements in computer hardware - in particular, graphics processing units with larger memory - will conceivably enable larger models to be trained on histopathological gigapixel images, thereby facilitating approaches that work directly on whole slide images without extensive image preprocessing.

Other biomarkers, tumor types other than CRC, additional imaging modalities

Beyond the detection of MSI/dMMR status in CRC, an increasing number of studies used similar methods for other molecular markers, in different tumor types and with additional imaging modalities. A detailed review of such approaches can be found elsewhere [38]. However, it is of relevance to note that MSI/dMMR status prediction has also been applied to gastric and endometrial cancer, in which the clinical utility of MSI/dMMR status prediction is comparable to CRC. [58,59]. However, for those tumor types the performance for detection of MSI/dMMR is not as high as in CRC [60]. A study from April 2020 presented a decent performance with an AUROC of 0.73 in the TCGA cohort of endometrial carcinoma [61]. In gastric cancer, the number of DL studies for MSI/dMMR status prediction is higher than for endometrial cancer, which might be due to the larger patient numbers and therefore more easily accessible patient cohorts. DL used for MSI/dMMR status prediction in gastric cancer yielded AUROCs between 0.66 and 0.879[4,12–14] which is lower than the performance of DL classifiers in CRC. Besides MSI/dMMR status, other molecular alterations are relevant for diagnosis, classification, or treatment of CRC such as mutations in *BRAF, KRAS, NRAS* or *TP53*. Several studies have shown that these genetic alterations are also to a certain degree detectable by DL from routine H&E histology images [5,8,62]. Additionally, *Wang* et al. showed that DL can predict tumor mutational burden as a quantitative biomarker from histology slides [11]. High tumor mutational burden (TMB) has been correlated with MSI/dMMR status in CRC[23], but future studies need to identify which visual features are specific to high TMB and which are specific to MSI/dMMR status. Similar to histopathology, other imaging modalities have been used to infer genetic information in CRC. It is remarkable that the prediction of MSI/dMMR status in colorectal cancer by DL is not limited to histological information but works also directly from Magnetic Resonance Imaging and computed tomography scans with AUROCs up to 0.811 as shown by two small-scale studies [63,64] However large-scale validation in radiological imaging modalities is currently not available.

Collecting large and diverse training cohort in computational pathology

It is known that, up to some point, a DL network performs better the more training data is available and this aspect also applies to MSI/dMMR status prediction in CRC [7]. As computational pathology approaches rely on sensitive medical patient data, a sufficient number of training images is not always easy to collect. Until now, different strategies have evolved to address this challenge. Collaboration between different institutions, countries and disciplines is essential in the field of DL in medicine. The establishment of academic consortia across countries [65] can enable the collection of large and heterogeneous patient cohorts as well as their interdisciplinary exchange. Yet, transferring images and patient data from involved sites to the institution performing the DL study comes with different challenges regarding infrastructure, data protection, and ethical regulations, possibly slowing down the research progress. Therefore, another strategy to collaborate on DL studies with increasing patient numbers is the so-called federated learning approach, which relies on decentralized training of multiple small models which are subsequently merged. This federated learning approach is mainly promoted by Owkin, Inc [66]. while other commercial entities such as Paige (11 Times Square, Fl 37 New York, NY, 10,036, USA) have acquired commercial usage rights for a large number of archived patient samples [2]. Based on public information, other companies including Tempus Inc. (600 W Chicago Ave. Ste 510, Chicago, IL 60,654, USA) and PathAI Inc. (120 Brookline Ave. Boston, Massachusetts 02,215, USA) are active in this field of research, but have not launched any clinically approved products yet. The newest strategy to collect a sufficient amount of data is the use of generative adversarial networks that can be trained to create synthetic histological images based on real image data [67]. Recently it was shown that these synthetic images even contained sufficient information about MSI/dMMR status to enable the development of a Deep Learning classifier purely on synthetic images [68]. Thus, this technique holds the potential to create patient cohorts that do not face ethical regulations or privacy issues as the images do not belong to an actual patient. Together, it can be expected from these approaches that they will enable researchers to access larger amounts of data in the future, potentially giving rise to computational models with higher performance

Adaption and advancement towards MSI/DMMR status prediction on diagnostic biopsy samples

All presented DL studies focused on the prediction of MSI/dMMR status from histology slides from tissue obtained by surgery. However, before patients undergo surgery, all patients usually have a diagnostic endoscopic biopsy [69]. Consequently, the biopsy tissue is available at an earlier point in the patient pathway and can be the only available tumor tissue sample in patients having a complete pathological response after neoadjuvant treatment [70]. Therefore it is of particular importance that DL based MSI/dMMR status prediction systems are also applicable to digitized endoscopic biopsy samples of CRC. This task raises additional challenges as biopsy samples are much smaller, can contain technical artifacts and are usually fragmented. Furthermore, the biopsy-derived tissue represents the luminal portions of the tumor only. *Echle* et al. reported that a DL system trained on images from tumor resection specimens performs considerably worse when used to predict MSI/dMMR status from endoscopic biopsy tissue [7]. However, training and testing a DL system within biopsy samples improved performance markedly leading to the hypothesis that DL based MSI/dMMR status prediction on biopsy samples most likely needs a DL classifier trained on the same sample type. For future research in this area, key challenges will be to collect a sufficient number of biopsy data to investigate whether a robust, well-performing MSI/dMMR status DL-based detection system can be developed. This raises the additional question whether a minimum amount/region with cancer and a larger patient sample might be required for such a DL system to work well.

Implementation in routine workflows - how can this be achieved?

Today, routine pathology workflows across the world are still predominantly based on examining physical glass slides under a microscope. However, it is expected that routine workflows will be ultimately digital, relying on digitized whole slide images evaluated by human observers with or without the aid of computer-based algorithms [71]. Currently, workflows in radiology departments are almost ubiquitously digi-

tal, with images being stored in a picture archiving and communication system (PACS) and expert observers interactively working with these images on computer workstations. Once the glass-slide based workflow in pathology departments has shifted to a similar, PACS-based approach [3], automatic computer-based image analysis methods could be more easily embedded in routine workflows. A pathology PACS could conceivably automatically trigger deep-learning-based testing of whole slide images in the background, potentially improving patient selection for molecular tests or directly providing definitive subtyping [72]. One important challenge that needs to be tackled before computer-based algorithms can be integrated in clinical workflows is the prospective validation of those methods in randomized patient trials to prove robustness and generate further evidence in real-world clinical settings which is also an important step towards regulatory approval. In addition to prospective validation, complex diagnostic procedures or devices need to be fine-tuned to local data and infrastructure. This is the case for diverse devices and methods such as computer tomographs, linear accelerators for radiation therapy and immunohistochemistry assays in pathology. In principle, such computer-based tests could be implemented in clinical routine by using commercially developed diagnostic algorithms or in-house in-vitro diagnostic test approaches. In this light, the developments around MSI/dMMR status prediction directly from routine histopathology images could serve as a blueprint for other biomarkers and provide a clear incentive for further digitization efforts of routine workflows in pathology.

Conclusions

Microsatellite instability (MSI))/deficient mismatch repair (dMMR) is a clinically important genetic trait that affects a substantial portion of colorectal cancer (CRC) patients. It is currently the only clinically approved biomarker that allows CRC patients to be treated with immune checkpoint inhibitors in the USA and Europe. Since 2019, advances in Deep Learning (DL) have made it possible to predict MSI/dMMR status directly from digitized routine hematoxylin and eosin (H&E) histopathology slides with high accuracy, as shown by 13 published studies to date. This is clinically relevant because DL could facilitate screening of CRC patients for dMMR/MSI. In the context of computational pathology, prediction of MSI/dMMR status with DL is currently one of the most widely studied problems. It has become a de-facto benchmarking task on which new DL algorithms are routinely tested. The broad interest of academia and industry suggests that DL-based assays for MSI detection could come onto the market in the next few years, potentially making MSI/dMMR status prediction in CRC one of the first clinically implemented DL algorithms for molecular subtyping of cancer.

Funding sources

PB is supported by the German Research Foundation (DFG; SFB/TRR219 Project-ID 322900939, BO3755/13-1 Project-ID 454024652), the German Federal Ministries of Education and Research (BMBF: STOP-FSGS-01GM1901A), Health (DEEP LIVER, ZMVI1-2520DAT111) and Economic Affairs and Energy (EMPAIA to PB) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 101001791). SBG is supported by R01 CA197350 and R01 CA81488 (National Institutes of Health, NIH), and a generous gift from Daniel and Maryann Fong. CT is supported by the German Research Foundation (DFG) (SFB CRC1382, SFB-TRR57). PQ and NW are supported by Yorkshire Cancer Research Program Grants L386 and L394. PQ is a National Institute of Health Senior investigator. JNK is supported by the German Federal Ministry of Health (DEEP LIVER, ZMVI1-2520DAT111) and the Max-Eder-Programme of the German Cancer Aid (Grant #70113864).

Author contributions

A.E. and J.N.K. designed the concept of the paper, performed literature research and analysis. A.E. and J.N.K. wrote the first draft of the paper. N.G.L, P.L.S., S.B.G., N.P.W., S.B.G., C.T., T.J.B., R.D.B., P.B., H.I.G. and P.Q. critically revised the paper and provided additional intellectual input. All authors collectively made the decision to submit for publication.

Declaration of Competing Interest

JNK declares consulting services for Owkin, France and Panakeia, UK. TJB reports owning a company that develops mobile apps, outside the scope of the submitted work (Smart Health Heidelberg GmbH, Handschuhsheimer Landstr. 9/1, 69120 Heidelberg). SBG is co-founder of Brogent International LLC with equity, outside the scope of the submitted work. PQ and NW are supported by Yorkshire Cancer Research program grants L386 and L394. PQ is a National Institute of Health Senior investigator. No other potential conflicts of interest are reported by any of the authors.

References

[1] Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nat Med 2018;24:1559–67.
[2] Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 2019;25:1301–9.
[3] Pantanowitz L, Sharma A, Carter AB, Kurc T, Sussman A, Saltz J. Twenty years of digital pathology: an overview of the road travelled, what is on the horizon, and the emergence of vendor-neutral archives. J Pathol Inform 2018;9:40.
[4] Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med 2019;25:1054–6.
[5] Kather JN, Heij LR, Grabsch HI, Loeffler C, Echle A, Muti HS, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. Nat Cancer 2020. doi:10.1038/s43018-020-0087-6.
[6] Cao R, Yang F, Ma S-C, Liu L, Zhao Y, Li Y, et al. Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in Colorectal Cancer. Theranostics 2020;10:11080–91.
[7] Echle A, Grabsch HI, Quirke P, van den Brandt PA, West NP, Hutchins GGA, et al. Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. Gastroenterology 2020;159:1406–16 e11.
[8] Bilal M., Raza S.E.A., Azam A., Graham S., Ilyas M., Cree I.A., et al. Novel deep learning algorithm predicts the status of molecular pathways and key mutations in colorectal cancer from routine histology images. bioRxiv 2021. doi:10.1101/2021.01.19.21250122.
[9] Yamashita R., Long J., Banda S., Shen J., Rubin D.L. Learning domain-agnostic visual representation for computational pathology using medically-irrelevant style transfer augmentation. arXiv [eessIV] 2021.
[10] Yamashita R, Long J, Longacre T, Peng L, Berry G, Martin B, et al. Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. Lancet Oncol 2021;22:132–41.
[11] Wang L, Jiao Y, Qiao Y, Zeng N, Yu R. A novel approach combined transfer learning and deep learning to predict TMB from histology image. Pattern Recognit Lett 2020;135:244–8.
[12] Schmauch B, Romagnoni A, Pronier E, Saillard C, Maillé P, Calderaro J, et al. A deep learning model to predict RNA-Seq expression of tumors from whole slide images. Nat Commun 2020;11:3877.
[13] Zhu J., Wu W., Zhang Y., Lin S., Jiang Y., Liu R., et al. Computational analysis of pathological image enables interpretable prediction for microsatellite instability. arXiv [statML] 2020.
[14] Ke J, Shen Y, Wright JD, Jing N, Liang X, Shen D. Identifying patch-level MSI from histological images of Colorectal Cancer by a knowledge distillation model. In: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2020. p. 1043–6.
[15] Lee H, Seo J, Lee G, Park J, Yeo D, Hong A. Two-stage classification method for MSI status prediction based on deep learning approach. NATO Adv Sci Inst Ser E Appl Sci 2020;11:254.
[16] Lee SH, Song IH, Jang H-J. Feasibility of deep learning-based fully automated classification of microsatellite instability in tissue slides of colorectal cancer. Int J Cancer 2021. doi:10.1002/ijc.33599.
[17] Schirris Y., Gavves E., Nederlof I., Horlings H.M., Teuwen J. DeepSMILE: Self-supervised heterogeneity-aware multiple instance learning for DNA damage response defect classification directly from H&E whole-slide images. arXiv [eessIV] 2021.
[18] Ellegren H. Microsatellites: simple sequences with complex evolution. Nat Rev Genet 2004;5:435–45.

[19] West NP, Gallop N, Kaye D, Glover A, Young C, Hutchins GGA, et al. Lynch syndrome screening in colorectal cancer: results of a prospective 2-year regional programme validating the NICE diagnostics guidance pathway throughout a 5.2-million population. Histopathology 2021. doi:10.1111/his.14390.

[20] Kawakami H, Zaanan A, Sinicrope FA. Microsatellite instability testing and its role in the management of colorectal cancer. Curr Treat Options Oncol 2015;16:30.

[21] 1 Recommendations 2021 Molecular testing strategies for Lynch syndrome in people with colorectal cancer | guidance | NICE n.d.

[22] Stjepanovic N, Moreira L, Carneiro F, Balaguer F, Cervantes A, Balmaña J, et al. Hereditary gastrointestinal cancers: ESMO clinical practice guidelines for diagnosis, treatment and follow-up†. Ann Oncol 2019;30:1558–71.

[23] Kather JN, Halama N, Jaeger D. Genomics and emerging biomarkers for immunotherapy of colorectal cancer. Semin Cancer Biol 2018;52:189–97.

[24] Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med 2015;372:2509–20.

[25] André T, Shiu K-K, Kim TW, Jensen BV, Jensen LH, Punt C, et al. Pembrolizumab in Microsatellite-Instability-High Advanced Colorectal Cancer. N Engl J Med 2020;383:2207–18.

[26] Mlecnik B, Bindea G, Angell HK, Maby P, Angelova M, Tougeron D, et al. Integrative analyses of colorectal cancer show immunoscore is a stronger predictor of patient survival than microsatellite instability. Immunity 2016;44:698–711.

[27] Anonymous. Keytruda 2018. https://www.ema.europa.eu/en/medicines/human/EPAR/keytruda (accessed September 3, 2021).

[28] Anonymous. Opdivo 2018. https://www.ema.europa.eu/en/medicines/human/EPAR/opdivo (accessed September 3, 2021).

[29] Marcus L, Lemery SJ, Keegan P, Pazdur R. FDA approval summary: pembrolizumab for the treatment of microsatellite instability-high solid tumors. Clin Cancer Res 2019;25:3753–8.

[30] 4 Evidence Molecular testing strategies for Lynch syndrome in people with colorectal cancer | Guidance | NICE n.d. 2021.

[31] Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, et al. Revised Bethesda guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst 2004;96:261–8.

[32] Kacew AJ, Strohbehn GW, Saulsberry L, Laiteerapong N, Cipriani NA, Kather JN, et al. Artificial intelligence can cut costs while maintaining accuracy in colorectal cancer genotyping. Front Oncol 2021;11. doi:10.3389/fonc.2021.630953.

[33] Jenkins MA, Hayashi S, O'Shea A-M, Burgart LJ, Smyrk TC, Shimizu D, et al. Pathology features in Bethesda guidelines predict colorectal cancer microsatellite instability: a population-based study. Gastroenterology 2007;133:48–56.

[34] Greenson JK, Huang S-C, Herron C, Moreno V, Bonner JD, Tomsho LP, et al. Pathologic predictors of microsatellite instability in colorectal cancer. Am J Surg Pathol 2009;33:126–33.

[35] Greenson JK, Bonner JD, Ben-Yzhak O, Cohen HI, Miselevich I, Resnick MB, 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;27:563–70.

[36] Lipton LR, Johnson V, Cummings C, Fisher S, Risby P, Eftekhar Sadat AT, et al. Refining the Amsterdam criteria and Bethesda guidelines: testing algorithms for the prediction of mismatch repair mutation status in the familial cancer clinic. J Clin Oncol 2004;22:4934–43.

[37] Cancer Genome Atlas NetworkComprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330–7.

[38] Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: a new generation of clinical biomarkers. Br J Cancer 2020. doi:10.1038/s41416-020-01122-x.

[39] Brenner H, Chang-Claude J, Seiler CM, Hoffmeister M. Interval cancers after negative colonoscopy: population-based case-control study. Gut 2012;61:1576–82.

[40] Carr PR, Weigl K, Edelmann D, Jansen L, Chang-Claude J, Brenner H, et al. Estimation of absolute risk of colorectal cancer based on healthy lifestyle, genetic risk, and colonoscopy status in a population-based study. Gastroenterology 2020;159:129–38 e9.

[41] Kather J.N. Histological images for MSI vs. MSS classification in gastrointestinal cancer, snap-frozen samples. 2019. doi:10.5281/zenodo.2532612.

[42] PAIP2020 - Grand Challenge n.d. https://paip2020.grand-challenge.org/ (accessed March 25, 2021).

[43] Chen T., Kornblith S., Norouzi M., Hinton G. A simple framework for contrastive learning of visual representations. arXiv [csLG] 2020.

[44] Papers with Code - ImageNet Benchmark (Image Classification) n.d. https://paperswithcode.com/sota/image-classification-on-imagenet (accessed April 17, 2021).

[45] Kather JN, Krisam J, Charoentong P, Luedde T, Herpel E, Weis C-A, et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. PLoS Med 2019;16:e1002730.

[46] Diao JA, Wang JK, Chui WF, Mountain V, Gullapally SC, Srinivasan R, et al. Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. Nat Commun 2021;12:1613.

[47] Fu Y, Jung AW, Torne RV, Gonzalez S, Vöhringer H, Shmatko A, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nature Cancer 2020:1–11.

[48] Schmitt M, Maron RC, Hekler A, Stenzinger A, Hauschild A, Weichenthal M, et al. Hidden variables in deep learning digital pathology and their potential to cause batch effects: prediction model study. J Med Internet Res 2021;23:e23436.

[49] Kleppe A, Skrede O-J, De Raedt S, Liestøl K, Kerr DJ, Danielsen HE. Designing deep learning studies in cancer diagnostics. Nat Rev Cancer 2021. doi:10.1038/s41568-020-00327-9.

[50] Sounderajah V, Ashrafian H, Golub RM, Shetty S, De Fauw J, Hooft L, et al. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ Open 2021;11:e047709.

[51] Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, et al. Minimum information about clinical artificial intelligence modeling: the MI–CLAIM checklist. Nat Med 2020;26:1320–4.

[52] Deng J, Dong W, Socher R, Li L, Li Kai, Fei-Fei Li. ImageNet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, ieeexplore.ieee.org; 2009. p. 248–55.

[53] Cohen G, Afshar S, Tapson J, van Schaik A. EMNIST: extending MNIST to handwritten letters. In: 2017 International Joint Conference on Neural Networks (IJCNN), ieeexplore.ieee.org; 2017. p. 2921–6.

[54] Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al.. Microsoft COCO: common objects in context. computer vision - ECCV 2014. Springer International Publishing; 2014. p. 740–55.

[55] Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., et al. An image is worth 16x16 words: transformers for image recognition at scale. arXiv [csCV] 2020.

[56] Das K, Conjeti S, Roy AG, Chatterjee J, Sheet D. Multiple instance learning of deep convolutional neural networks for breast histopathology whole slide classification. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), ieeexplore.ieee.org; 2018. p. 578–81.

[57] Xu Y, Zhu J, Chang E, Tu Z. Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering. In: 2012 IEEE conference on computer vision and pattern recognition, ieeexplore.ieee.org; 2012. p. 964–71.

[58] Brooks RA, Fleming GF, Lastra RR, Lee NK, Moroney JW, Son CH, et al. Current recommendations and recent progress in endometrial cancer. CA Cancer J Clin 2019;69:258–79.

[59] Ratti M, Lampis A, Hahne JC, Passalacqua R, Valeri N. Microsatellite instability in gastric cancer: molecular bases, clinical perspectives, and new treatment approaches. Cell Mol Life Sci 2018;75:4151–62.

[60] Muti H.S., Heij L.R., Keller G., Kohlruss M., Langer R., Dislich B., et al. Development and validation of deep learning classifiers to detect Epstein-Barr virus and microsatellite instability status in gastric cancer: a retrospective multicentre cohort study. The Lancet Digital Health 2021. doi:10.1016/S2589-7500(21)00133-3.

[61] Wang T, Lu W, Yang F, Liu L, Dong Z, Tang W, et al. Microsatellite instability prediction of uterine corpus endometrial carcinoma based on H E histology whole-slide imaging. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI); 2020. p. 1289–92.

[62] Jang H-J, Lee A, Kang J, Song IH, Lee SH. Prediction of clinically actionable genetic alterations from colorectal cancer histopathology images using deep learning. World J Gastroenterol 2020;26:6207–23.

[63] Golia Pernicka JS, Gagniere J, Chakraborty J, Yamashita R, Nardo L, Creasy JM, et al. Radiomics-based prediction of microsatellite instability in colorectal cancer at initial computed tomography evaluation. Abdom Radiol (NY) 2019;44:3755–63.

[64] Zhang W., Huang Z., Zhao J., He D., Li M., Yin H., et al. MRI-based deep learning analysis can predict microsatellite instability in rectal cancer 2020. doi:10.2139/ssrn.3569821.

[65] MSIDETECT n.d. https://jnkather.github.io/msidetect/ (accessed March 25, 2021).

[66] Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digit Med 2020;3:119.

[67] Levine AB, Peng J, Farnell D, Nursey M, Wang Y, Naso JR, et al. Synthesis of diagnostic quality cancer pathology images by generative adversarial networks. J Pathol 2020. doi:10.1002/path.5509.

[68] Krause J, Grabsch HI, Kloor M, Jendrusch M, Echle A, Buelow RD, et al. Deep learning detects genetic alterations in cancer histology generated by adversarial networks. J Pathol 2021. doi:10.1002/path.5638.

[69] Argilés G, Tabernero J, Labianca R, Hochhauser D, Salazar R, Iveson T, et al. Localised colon cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 2020;31:1291–305.

[70] Glynne-Jones R., Wyrwicz L., Tiret E., Brown G., Rodel C., Cervantes A., et al. 2021 Rectal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up n.d. doi:10.1093/annonc/mdx224.

[71] Cui M, Zhang DY. Artificial intelligence and computational pathology. Lab Invest 2021;101:412–22.

[72] Calderaro J., Kather J.N. Artificial intelligence-based pathology for gastrointestinal and hepatobiliary cancers. Gut 2020:gutjnl 2020. doi:10.1136/gutjnl-2020-322880.