Review Paper:

Assessing the Current Status of Artificial Intelligence Applications for Predicting Microsatellite Instability in Colon Cancer

Alsughayyir Jawaher
Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, King Saud University, Riyadh 12372, SAUDI ARABIA
jalsughayyir@ksu.edu.sa

Abstract

Effective treatment decisions for colorectal cancer (CRC) depend on the histological classification and microsatellite instability (MSI) status of the patient's biopsy. In recent years, artificial intelligence (AI) has emerged as a valuable tool in the diagnostic process, offering efficiency, reducing the need for extensive manpower and maintaining accuracy. This review explores recent advancements in AI technology and its effectiveness in identifying prognostic biomarkers related to CRC and aims to inform clinicians and gastroenterologists about novel patient management strategies. A narrative non-systematic review of existing literature on using AI for detecting deficient mismatch repair (dMMR)/MSI in CRC diagnosis was performed. Searches were conducted in the PubMed database using a combination of keywords such as colorectal cancer diagnosis, artificial intelligence and deep learning, focusing on publications from 2019 onward.

The reviewed articles exhibited varying outcomes, with each utilizing the CNN model under differing conditions like cohort types and sizes and convolution or filter numbers, highlighting specific strengths and limitations for each model. AI-driven predictive analytics offered researchers superior insights into genomics and proteomics data, elevating patient characterization precision and streamlining pathology workflows.

Keywords: Convolutional neural networks, cancer pathology, machine learning, predictive value of tests, DNA mismatch repair.

Introduction

Through a journey that may last for years, patients and their families will have to deal with the inevitable emotional pain and financial strains associated with cancer. Colorectal cancer (CRC), which includes both colon and rectal cancer, is identified worldwide as the third most common type of cancer and the second most cancer-related death with approximately 1.93 million new cases reported in 2022[11]. The incidence of colon cancer is relatively similar between genders for the 40–59 age categories but the incidence increases slightly in males aged 60 and above[11]. Although rectal cancer exhibits a higher prevalence among males, right-sided colon cancer, a more aggressive form of colon cancer, has a greater incidence in females[1]. Globally, the incidence of CRC is highest in the regions of Australia/New Zealand and Europe, with a rate of 40.6 per 100,000 and a mortality rate of 20.2 per 100,000 for males[54].

Conversely, the lowest incidence rates of the same disease are observed in various African regions and Southern Asia, where the rate is 4.4 per 100,000 with a mortality rate of 2.5 per 100,000 females[54]. The differences in the global rates of CRC reflect the multifaceted impact of various influences such as country's Human Development Index levels (HDI) and genetic factors. Prediction models by the same study anticipated that by 2040, there will be 3.2 million new cases and 1.6 million fatalities due to CRC, with the majority occurring in Nations with high HDI. CRC not only detrimentally affects the quality of life of those diagnosed but also incurs substantial economic burdens. These include expenses associated with early disease stages such as screening, diagnosis, imaging and surgical interventions, as well as expenses related to later stages of the disease (e.g. hospitalization, medication and durable medical equipment).

The estimated global economic cost of CRC is $2.8 trillion internationally, accounting for 10.9% of the global economic cost of cancer[15]. Hence, promoting screening programs designed for early colon cancer detection is crucial to reduce the financial burden associated with therapeutic interventions and to enhance the efficacy of treatment modalities. The risk of developing colon polyps increases with advancing age, male gender, high-fat low-fiber diet, excess alcohol intake, tobacco use and family history of colon cancer[33]. In addition, patients with Crohn's disease and long-standing ulcerative colitis have an increased risk of developing CRC[78]. Such risk factors often lead to the silent development of the disease, leading to nearly half of the patients aged 45-50 years being diagnosed at an advanced stage of the disease—stage III or IV, a point at which the prognosis becomes less favorable[67,79].

Microsatellite instability (MSI) and DNA mismatch repair (MMR) are critical factors in the diagnosis and treatment of colon cancer. MSI is a condition referring to the buildup of insertion or deletion mutations at microsatellite repeat sequences within cancer cells, caused by a functional failure in one or more key DNA MMR proteins which are crucial for correcting DNA replication errors. Identifying MSI status can help to predict prognosis and response to certain

therapies such as immunotherapy[31,84]. Moreover, MMR deficiency, often associated with Lynch syndrome, can guide genetic counseling and testing in affected families; thereby enhancing personalized treatment strategies and improving patient outcomes. Traditionally, hematoxylin and eosin (H&E) staining of tissue biopsies or molecular methods is utilized for determining the dMMR/MSI status.

However, these techniques are generally labor-intensive and costly. Consequently, there is a pressing need for the development of innovative and accurate methods for both diagnosis and prognosis. This is because these markers are essential for determining the most effective treatment options for individual CRC patients. In this regard, advancements have been made to refine the methodologies aimed at improving diagnostic and prognostic protocols by integrating artificial intelligence (AI) into pathology. The aim of this review is to outline the utility of AI in applications related to colon cancer diagnosis and prognosis. Specifically, this review concentrates on the prediction models developed within the preceding five years, utilizing data from routinely stained H and E histopathology slides to determine dMMR/MSI status. Additionally, limitations and future directions for improvement are also discussed.

A narrative non-systematic review of the published literature on AI technology utilization for dMMR/MSI detection within the context of colorectal cancer diagnosis was conducted. This review involved searching PubMed databases with different combinations of keywords and phrases used to narrow down and to identify pertinent sources. These keywords included, but were not limited to, colorectal cancer diagnosis, artificial intelligence, deep learning and MSI. The review encompasses only material published since 2019, with non-English publications excluded.

CRC diagnosis, treatment and prognosis

The development of CRC usually starts with the formation of a neoplastic polyp invading the muscularis mucosa and into the submucosa. While most polyps are benign, a histological examination at this early stage is necessary to determine whether the polyp has the potential to develop into a malignant tumor[52]. If the tumor is found to be malignant, it can be staged based on the tumor-node-metastasis (TNM) histopathological criteria established by the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC)[7]. TNM staging relies on three histopathological criteria: the primary tumor size (T), the number of involved regional lymph nodes (N) and the presence of distant metastases (M)[77].

In addition to the endoscopic biopsy, number of diagnostic imaging procedures can be performed such as preoperative ultrasound, computed tomography scans and magnetic resonance imaging scans to estimate the involvement of the rectum wall and local lymph node metastases[7,8]. The TNM system has similar survival rates for both rectal and colon cancers, supporting the use of the same staging system for both these diseases[3]. The staging element "T" is crucial for prognosis, as research indicates that patients diagnosed with T4, N0 tumors exhibit lower survival rates compared to those with T1-2, N1-2[10,16,30]. Similarly, the regional lymph node classification "N" is crucial for determining the disease metastasis. The AJCC and The College of American Pathologists (CAP) recommend the examination of a minimum of 12 lymph nodes[22,23,74]. Nonetheless, a definitive consensus on the minimum number of lymph nodes required for the accurate staging of stage II cancer is lacking and pathologists are encouraged to retrieve as many lymph nodes as possible for staging colon cancers.

All newly identified cases of colon cancer must undergo screening for genetic mutations, along with performing a comprehensive colonoscopy and establishing baseline levels of carcinoembryonic antigen (CEA). Genetic mutation in MMR genes is characterized by mutations in *MLH1*, *MSH2*, *MSH6*, or *PMS2* which result in DNA replication error. Such mutations can be identified by immunohistochemistry (IHC) to provide information about the expressed MMR proteins. Determining MSI status, an MMR byproduct, is performed by molecular techniques[58]. The high concordance rate of dMMR and MSI, exceeding 95%, makes them nearly interchangeable[84].

A normal IHC test implies that all four MMR proteins are normally expressed and is often reported as proficient MMR (pMMR). The absence of expression of one or more of the four DNA MMR proteins is often reported as positive IHC or deficient MMR (dMMR). In addition to blood biomarkers, patients diagnosed with invasive colon cancer necessitate an initial computed tomography (CT) scan of the chest and abdominopelvic regions[8].

To determine the optimal therapy for each patient, the dMMR/MSI status should be determined for all newly diagnosed patients to guide the disease diagnosis and to determine the treatment plan. Based on the most recent National Comprehensive Cancer Network (NCCN) guidelines, patients with early stages (UICC stage I) and stage II with MSI-H are universally accepted to be treated with surgery accompanied by lymph node resection. Similarly, low-risk stage II colon cancer patients (with MSS or proficient MMR (pMMR)) are also treated with surgery and lymph node resection and can be observed without adjuvant therapy, or considered for capecitabine or 5-FU/leucovorin (LV).

On the other hand, stage II patients (MSS/pMMR) at high risk for systemic recurrence and displaying poor prognostic features are considered for 6 months of adjuvant chemotherapy with 5-FU/LV, capecitabine, or FOLFOX, or 3 months of CAPEOX[6,17]. Low-risk (T1-3, N1) patients with stage III of disease are recommended to have adjuvant treatment of CAPEOX[29], or 3 to 6 months of FOLFOX among other treatment options[25]. Patients with high-risk (T4,

N1-2 or any T, N2) and stage III of disease, are recommended for 6 months of FOLFOX, or 3 to 6 months of CAPEOX, among other treatment options.

Patients with metastatic CRC should have tumor genotying for RAS and BRAF mutations. This is especially important to determine the optimal systemic therapy[7]. Stage IV and recurrent CRC are treated by surgery, chemotherapy and target therapy with monoclonal antibodies targeting growth factors (vascular endothelial growth factor receptors (VEGFR) and epidermal growth factor receptors (EGFR). The specific order and timing of these treatments vary among patients and the choices of adjuvant therapies and management protocols have been extensively reviewed elsewhere and are beyond the scope of this review[7,53].

In general, the prognosis of CRC can be influenced by a number of factors, reflecting the substantial variability in the 5-year survival rate. Key determinants include the patient's age, dietary habits, timing of therapy, tumor localization, body mass index and comorbidities among other factors. The pathological stage is considered as the most significant prognostic indicator. The 5-year overall survival rate for colon cancer is estimated to be 85% for stage I, 70-80% for stage II, 35-65% for stage III and 5% for stage IV. These survival rates translate to 90% for localized tumors, 73% for regional tumors, 13% for distant tumors and an overall 63% for all cases.

The risk of CRC recurrence following curative surgery is estimated to range between 30 and 40% with 40-50% of recurrences occurring within the first few years after the initial resection surgery[28]. The probability of CRC recurrence has been associated with number of risk factors including age at diagnosis, tumor location and stage and number of dissected lymph nodes, pre- and post-operative CEA serum levels and the type of the performed resection procedure[57,65,83].

Based on the insights obtained from the histological examination and imaging, disease management and prognosis can be determined. Although TNM staging has long been considered a universal staging system, it exhibits a notable variance in prognostic accuracy among patients with the same stage, indicating its limitation in predicting outcomes accurately[45]. In response to this limitation, a number of novel tools are being investigated to enhance the prognostic efficacy of CRC. These tools seek to optimize the stratification of patients by distinguishing individuals at low-risk and high-risk of cancer recurrence. Such advancement holds the potential to offer valuable guidance in determining the most suitable treatment regimen, particularly for patients anticipated to derive substantial benefits.

AI in pathology

AI refers to the ability of machines to carry out tasks in a way that mimics human intelligence in terms of problem-solving and pattern recognition among many other tasks. In the field of pathology, AI can analyze histopathological images and differentiate cancerous cells from non-cancerous ones[31]. Unlike static rule-based systems, AI algorithms can modify their outputs based on new data or feedback, making them more flexible and adaptable. Machine learning models can analyze large datasets, including patient history, genetic information and treatment outcomes, to forecast the likely trajectory of the disease[62]. This enables oncologists to tailor personalized treatment plans for each patient, optimizing the chances of recovery and enhancing the quality of care[31]. This process reduces the likelihood of human error and improves diagnostic precision, enabling early detection and better patient outcomes.

AI encompasses a variety of subfields, each contributing distinct methodologies and clinical applications to the field of pathology. Machine learning (ML) refers to the ability of a system to use algorithms to learn from the provided data and subsequently make predictions or solve problems based on this data[37]. In CRC pathology, ML can be used to analyze large datasets of histopathological images to identify patterns correlated with cancerous tissue, thus enhancing diagnostic accuracy[31]. Deep Learning (DL), is a branch of ML that uses layers of neural networks and multiple layers to process raw data without manual feature extraction[70]. For instance, images paired with corresponding class labels (e.g. H&E stained images with benign or malignant cells) are introduced to the system as the training set and upon the introduction of new input data, the system can distinguish and classify the images even without any pre-existing assumptions.

Neural networks (NNs) refer to layers of nodes: an input layer, one or more hidden layers and an output layer. Each node functions as an artificial neuron connected to the next node and each has a specific weight and threshold value. A node activates and transmits data to the following layer if its output surpasses the threshold value; otherwise, it does not pass any data[37]. NNs are trained using data to enhance their accuracy over time and the weights are adjusted during the learning process to minimize error and improve the model's accuracy. NNs are effective in processing data, recognizing patterns and predictions. Once the algorithms are fine-tuned, these networks become important tools by enabling rapid classification and clustering of data.

Deep Neural Networks (DNNs) features utilize multiple layers of nodes that are densely interconnected, which facilitates global feature learning. The term "deep" in deep learning signifies the number of layers within a neural network. When a neural network consists of more than three layers, including input and output layers, it qualifies as a deep learning algorithm[37]. DNNs are commonly used across a range of machine-learning tasks[71]. Convolutional Neural Networks (CNNs) use mathematical operations known as convolution in one or more of their layers. The concept of CNN architecture is illustrated in figure 1.

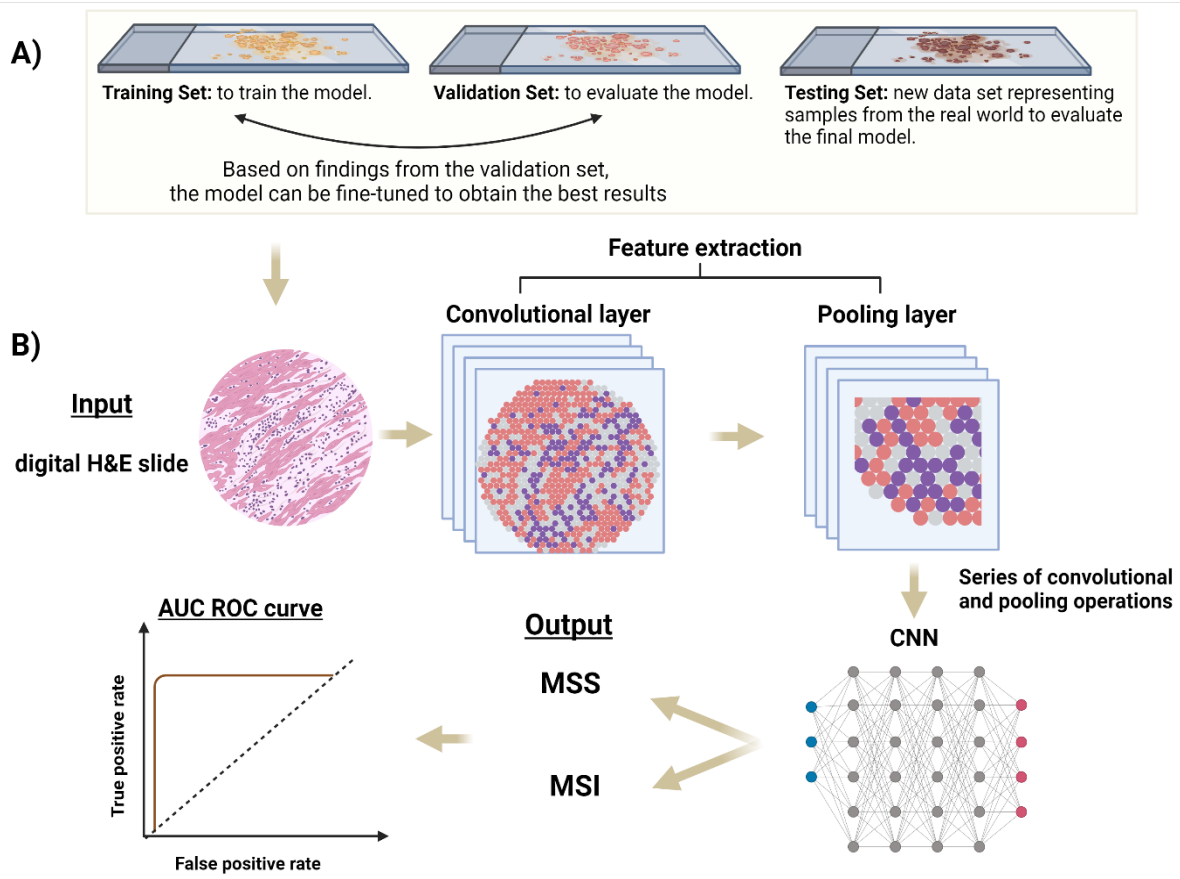

Figure 1: An overview of convolutional neural networks (CNNs) process in digital pathology. A) The training dataset is used for learning and to fit the initial parameters of the classifier. The validation dataset is used to adjust the parameters of the classifier. The test dataset consists of held-out data that provides an unbiased evaluation of the final prediction model. B) H and E scanned images are derived from the original tissue blocks. The feature extraction process utilizes a convolutional neural network (CNN), which is an intricate network of interconnected processes organized in layers, tasked with extracting higher-level features from the images. Pooling layers summarize the features extracted by convolution layers and reduce the dimensions of the images. Following a series of convolution and pooling layers, the classification layer predicts the output by depicting AUC-ROC value.

However, unlike CNNs, DNNs do not have specialized layers for capturing spatial hierarchies, making them more versatile but potentially less efficient for tasks specifically related to images[61]. The ability of CNNs to automatically identify features from images makes them well-suited for image classification in the field of pathology especially when the original WSI cannot be used for AI and requires processing prior analysis[61].

In CNN, the convolution process involves the addition of a filter over the input data to produce a feature map, capturing spatial hierarchies in the data[62]. CNNs are structured into two primary components: the feature extraction layer and the classification layer, each serving distinct purposes. The feature extraction layer integrates convolution and pooling layers. The convolution layer generates newly modified versions of images by applying various types of filters, such as th identifying edges and enhancing contrast. The pooling layer condenses the features of images produced by the convolution layer[62]. Repeating the preceding process enables the segregation of significant attributes from the input data[56]. Following a series of multiple cycles of convolution and pooling, the classification layer utilizes the extracted features to classify the input data into predefined categories, thereby forming the network's prediction. The predicted result is then compared against a reference standard provided in the data[31]. If discrepancies are identified, the filters are adjusted to enhance prediction accuracy[31]. This makes CNNs highly efficient for tasks involving image recognition and video analysis.

Inception-V3, ResNET, ShuffleNet, MSInet, MIL, WisMSI and MSIntuit are among the CNN architectures widely used for image classification tasks, each providing their unique advantages[2,26,38,82]. For instance, Inception-V3 is most suitable for high-performance settings due to its depth and accuracy[50], while ShuffleNET is superior in efficiency and speed[60]. MSInet is specialized in multi-scale feature extraction and is mostly used in clinical oncology imaging[80]. By operating these AI techniques, the field of pathology is advancing toward more accurate and efficient diagnostic protocols.

To maximize the performance of deep learning models, selecting large and diverse training and validation data sets is crucial[62]. The training dataset consists of a set of pathology images that the AI model uses to learn and identify patterns associated with different pathological conditions. Through this process, the model adjusts its algorithms to improve accuracy and performance. The validation dataset is used to evaluate the model's performance and to ensure that the results are not confined to the initial training dataset, thereby confirming its generalizability[63]. Finally, the testing dataset includes held-out data to evaluate how well the model performs in real-world scenarios outside the training data. This approach helps in fine-tuning the model, preventing overfitting and ensuring that it can accurately interpret new and unseen pathology slides[55].

The evaluation of these models is based on their Area Under the Receiver Operating Characteristic curve (AU-ROC, or AUC) values, which serve as an essential metric in assessing the performance of CNNs. The AUC quantitatively measures the model's capacity to differentiate between classes, thereby offering a comprehensive indication of its classification accuracy. The plot represents the relationship between the true positive rate and the false positive rate across various classification thresholds. A high AUC (close to 1) indicates that the model is effective in distinguishing between the two classes and its predictions are reliable.

Conversely, a low AUC (close to 0) suggests poor performance. An AUC around 0.5 implies that the model is essentially making random guesses, demonstrating no ability to separate the classes and indicating that the model is not learning any meaningful patterns from the data[12,32]. The scientific community continues to eagerly compete in the development of models that achieve improved AUC values, thereby advancing the ability to accurately determine MSI/dMMR status in clinical practice.

Predicting dMMR/MSI status by Deep Learning approaches

Immune checkpoint inhibitors (ICI) have been regarded as a successful treatment strategy for various solid tumors. ICI acts by counteracting tumor-mediated immune suppression, creating a proinflammatory microenvironment that can facilitate the destruction of cancer cells. Among the first FDA-approved ICI therapeutics were nivolumab (anti-PD1)[75], atezolizumab (anti-PDL1) and ipilimumab (anti-CTLA-4)[4,47]. These approvals were due to the observation that tumors with high levels of MSI (MSI-H), regardless of their site, are resistant to chemotherapy and sensitive to ICI[8,45,48]. Despite the NCCN recommendation for dMMR/MSI screening, not all patients can have the test[18,73].

This limitation arises from the restricted availability of these tests mostly in tertiary hospitals and their high cost[18,41]. The fact that only a small percentage of CRC patients (~15%) have MSI-H, AI technology enables the prediction of therapeutic efficacy by forecasting patient responses to treatment[49,81]. This predictive capability can mitigate the risk of patients experiencing toxicity without therapeutic benefit. In the past few years, a significant body of research has emerged focusing on AI-based MSI prediction models derived from whole slide images (WSIs). These models, particularly within the context of CRC, have yielded promising outcomes in the field (Table 1).

The reviewed articles demonstrated variable degrees of outcomes as each study used the CNN model in different settings, including variations in the type and size of the cohorts, as well as the number of convolutions or filters employed. The first fully automated, end-to-end deep learning system designed to detect dMMR/MSI status in CRC was developed by Kather et al[44] in 2019 with a performance accuracy of 0.84. Since then, a number of groups have developed novel CNN models with diverse architectures and specialized functionalities. For instance, the randomized controlled study performed by Jiang et al[40] tested five sets of machine learning models for their abilities to determine MMR/MSI status.

A total of 2,279 patients between the training and test groups were enrolled and twelve clinicopathological features were incorporated into the development of the predictive models. Among the five predictive models, the AUC value of the machine-learning random forest (RF) model outperformed the logistic regression (LR) method in identifying dMMR/MSI. The RF model in the same study exhibited an AUC of 0.85, indicating superior performance compared to the other models. The study also conducted a comprehensive analysis of the strengths and limitations inherent in each model.

The support vector machine (SVM) model exhibited the capability to perform both linear and nonlinear classification and regression tasks. However, it encountered significant challenges when processing complex and extensive datasets. Conversely, RF demonstrated its ability to recognize and identify dMMR and pMMR samples.

In 2021, Lee et al[49] developed the highest-performing model for predicting MSI at the time with an AUC reaching 0.972. The group trained Inception-V3 on a cohort of image patches from The Cancer Genome Atlas study (TCGA) WSI data set and Saint Mary's Hospital (SMH) to differentiate between normal and cancerous tissues and to determine the MSI status. Artifacts in the WSI were automatically excluded, eliminating the need for human intervention.

As a result, only accurate tissue patches were used to train the normal/tumor classifiers, enabling a clear distinction between tumor and normal regions in a WSI. The model exhibited AUC of 0.89 when tested on TCGA data and an AUC of 0.97 on the SMH dataset, representing the highest performance among the studies reviewed.

Table 1
Recent research focusing on DL-based MSI/dMMR status prediction.

Author/Year	CNN Model	Training Population	Internal Validation	External Validation	Test population AUC (95% CI)	Method of MSI analysis
Kather et al, 2019[44]	ResNet-18 (CRC and other tumors)	TCGA CRC FFPE (n=260 pts)	Random split	DACHS	TCGA CRC FFPE (n=100 pts) AUC 0.77 (0.62-0.87) DACHS FFPE (n=378) AUC 0.84 (0.72-0.92)	TCGA PCR DACHS PCR
		TCGA CRC (n= 269 frozen)			TCGA CRC frozen (n=109 pts) AUC 0.84 (0.73-0.91) DACHS FFPE (n=378) AUC 0.61 (0.50-0.73)	
Pressman et al, 2020[64]	ResNet18	TCGA (n=360 WSIs)	NA	Gangnam Sev.	TCGA AUC: 0.79; Gangnam Sev. (n= 170) AUC: 0.76	NA
Schmauch et al, 2020[69]	HE2RNA with ResNet50	TCGA FFPE (n=465 pts)	Three-fold cross validation	None	TCGA FFPE: 0.82	PCR
Kather et al, 2020[43]	ShuffleNet	TCGA CRC FFPE (n=426 pts)	Three-fold cross-validation	DACHS	DACHS FFPE (n=379 pts) AUC 0.89 (0.88-0.92)	TCGA: PCR DACHS: PCR
Echle et al, 2020[21]	ShuffleNet	MSIDETECT CRC (n=6,406 pts)	Random split	Yes	MSIDETECT AUC: 0.92 (0.90 – 0.93)	DACHS: PCR; TCGA: PCR, QUASAR and NLCS: IHC; YCR-BCIP:IHC
			Three-fold cross-validation	YCR-BCIP	MSIDETECT AUC: 0.92 (0.91-0.93) YCR-BCIP-RESECT (n=771) AUC: 0.96 (0.93-0.98) YCR-BCIP-Biopsy (n=1531 pts) AUC: 0.78 (0.75-0.81)	
		YCR-BCIP-Biopsy (n=1,531 pts)	Three-fold cross-validation	None	YCR-BCIP-Biopsy AUC: 0.89 (0.88-0.91)	
Yamashita et al, 2021[80]	MSInet	Stanford dataset (n=85 pts)	Random split	None	Stanford dataset (n=15 pts) AUC: 0.93 (0.77-1.00)	Stanford dataset: IHC/PCR TCGA:PCR
			Four-fold cross-validation	TCGA	Stanford dataset (n=15 pts) AUC: 0.93 (0.77-1.0) TCGA (n=479 pts) AUC: 0.77 (0.72-0.83)	

Krause et al, 2021[46]	ShuffleNet	TCGA FFPE (n=256 patients)	Random split	None	TCGA FFPE (n=142) AUC: 0.74 (0.68-0.85)	PCR
Lee et al, 2021[49]	Inception-V3	TCGA FFPE (n=470,825 patches); SMH FFPE (n=274 WSI)	10-fold cross-validation	None	TCGA FFPE AUC 0.89 (0.85-0.92) SMH FFPE AUC: 0.97 (0.95-0.98)	TCGA: PCR SMH:PCR/IHC
		TCGA FFPE (n=470,825 patches)		SMH FFPE	TCGA FFPE AUC: 0.86 (0.81-0.90) SMH FFPE AUC: 0.78 (0.74-0.83)	
		TCGA frozen (n=562,837 patches)		None	TCGA Frozen AUC: 0.94 (0.92-0.95)	
Cao et al, 2020[13]	ResNet-18 using the EPLA model	TCGA-15COAD (n=429 frozen)	Random split	Asian-CRC of all stages from Tongshu Biotechnology Co., Ltd. FFPE	TCGA-COAD AUC: 0.88 (0.81-0.95) Asian-CRC AUC: 0.64 (0.60-0.69)	PCR
		TCGA-COAD frozen (90%) and Asian-CRC FFPE (10%)			Asian-CRC AUC: 0.85 (0.75-0.93)	
		TCGA-COAD frozen (30%) and Asian-CRC FFPE (70%)			Asian-CRC AUC: 0.92 (0.88-0.97)	
Bilal et al, 2021[9]	ResNet18 Adapted ResNet34 HoVer-Net	TCGA-CRC-DX (n=499 pts)	Random split	TCGA-CRC-DX	PAIP challenge cohort (n=47 slides): AUC: 0.86 (0.04-0.74)	PCR
Jiang et al, 2022[39]	MIL model	TCGA (n=441 WSI)	Three-fold cross validation	PAIP	AUC 0.88 ± 0.03	IHC
Echle et al, 2022[20]	ResNet18	DACHS (n=2,039 pts). All FFPE, only TCGA had a small number of frozen sections	Leave-one-cohort-out cross-validation	Yes	AUC within-cohort 0.91 (0.88-0.93) AUC deployment 0.89 (0.87 – 0.92)	PCR or IHC
		QUASAR (n=1,774 pts)			AUC within-cohort 0.90 (0.88-0.92) AUC deployment 0.93 (0.91-0.95)	
		TCGA (n=426 pts)			AUC within-cohort 0.79 (0.72-0.85)	

					AUC deployment 0.91 (0.87-0.95)	
		NLCS (n=2,098 pts)			AUC within-cohort 0.85 (0.82-0.87) AUC deployment 0.92 (0.90-0.94)	
		YCR-BCIP (n=805 pts)			AUC within-cohort 0.93 (0.90-0.96) AUC deployment 0.96 (0.94-0.98)	
		DUSSEL (n=196 pts)			AUC within-cohort 0.75 (0.64-0.85) AUC deployment 0.85 (0.74-0.93)	
		MECC (n=683 pts)			AUC within-cohort 0.70 (0.64-0.75) AUC deployment 0.74 (0.69- 0.80)	
		UMM (n=35 pts)			AUC within-cohort 0.98 (0.93-1.00) AUC deployment 0.92 (0.69-1.00)	
		MUNICH (n=287 pts)			AUC within-cohort 0.80 (0.71-0.88) AUC deployment 0.88 (0.80-0.95)	
Ding et al, 2022[19]	-	TCGA-COAD (n=459 WSI images)	NA	TCGA-READ CPTAC-COAD	AUC 83.92 (77.41-87.59) TCGA-COAD AUC: 83.92 (77.41-87.59) TCGA-READ AUC: 61.28 (53.28-67.93)	PCR or IHC
Jiang et al, 2022[39]	MIL	TCGA (n=441) SYSUCC-surgical (n=355) SYSUCC-biopsy (n=341) PAIP (n=78)	3-fold cross validation	Yes	AUC 0.88 (0.85-0.92) AUC 0.84 (0.82-0.86) AUC 0.76 (0.73-0.80) AUC 0.88 (0.85-0.90)	IHC
Schirris et al, 2022[68]	DeepSMile	TCGA-CR (n=360)	Random split	NA	0.82 (0.77-0.86)	PCR
Lou et al, 2022[51]	PPsNET	Shandong Hospitals (n=144)	Random split	NA	0.94	IHC
Guo et al 2023[27]	Swin-T	TCGA-CRC-DX	Intra-cohort four-fold cross-validation	Inter-cohort external validation and	TCGA-CRC-DX AUC 0.91± 0.02 (mean ± SD) (23% improvement over recently published AUC values on the same dataset (0.86 from Bilal[9]) and	IHC

					0.74 from Kather et al MCO AUC: 0.90 (0.85-0.95)	
Jiang et al, 2023[40]	XGBoost (extreme gradient boosting, SVM (support vector machine), Naïve Bayes (NB), And RF (random forest and logistic regression.	Wuhan Union Hospital (n=2,279 patients)	No	10-time cross validation	XGBoost AUC: 0.80 SVM AUC: 0.81 NB AUC: 0.74 RF AUC: 0.85 LR AUC: 0.78	IHC
Chang et al, 2023[14]	WiseMSI	TSMCC (n=1579) TCGA (n=609)	NA	10-fold cross validation	0.95 (0.94-0.96) 0.63 (0.70-0.73)	PCR
Saillart et al, 2023[66]	MSIntuit	TCGA (n=859) PAIP (n= 47) MPATH (n=600)	NA	NA	TCGA: 0.93 (0.90-0.96) PAIP: 0.97 (0.90-0.99) MPATH-DP200: 0.88 (0.84-0.91) MPATH-UFS: 0.86 (0.83-0.90)	IHC

* AUC, Area Under the Curve; TCGA, The Cancer Genome Atlas study; CRC, Colorectal Cancer; NA: Not Available, WSI, Whole Slide Images; pts, patients; FFPE, Formalin-Fixed Paraffin-Embedded; PAIP: Pathology Artificial Intelligence Platform. DACHS, Darmkrebs: Chancen der Verhütung durch Screening (CRC prevention through screening study abbreviation in German); (Stanford dataset, Stanford University Medical Center (USA) Gangnam sev, Gangnam Severance Hospital (South Korea); MSIDETECT:A consortium composed of 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 Center Improvement Programme; SMH, Saint Mary's Hospital (South Korea). TSMCC: TongShu MSI colorectalcancer; MPATH, medipath. PPsNET: depth refinement network. MIL: multiple instance learning.

The same study trained a second model exclusively on the TCGA dataset and when tested on the SMH dataset, it demonstrated a reduced AUC of 0.78 and AUC of 0.86 on the TCGA cohort. The superior performance of the first model can be attributed to the inclusion of a portion of the SMH dataset in the training phase which improved the ethnic diversity. Overall, these results demonstrated that by automatically removing artifacts and selecting tumor patches with high tumor probability, the DL-based system could screen out a considerable number of tissue slides for their MSI status.

Selecting an appropriate methodological model for CNNs is as crucial as the choice of training and validation cohorts which can impact the performance and generalizability of the model. Ensuring the dataset includes a comprehensive, representative sample of the population is essential to avoid biases and improve the model's generalization. Echle et al[21] addressed this challenge by developing a deep learning-based classifier for determining dMMR/MSI status in two studies. In the initial study performed in 2020, the group trained ShuffleNet on a large dataset comprising 6,406 slides from nine patient cohorts representing various countries and ethnicities.

The model achieved remarkable clinical-grade performance in predicting MSI status, evidenced by an AUC of 0.96 for the external cohort and AUC of 0.92 for the internal cohort, all accomplished without the use of any manual annotations. In the second study performed in 2022, a ResNet18 neural network model was trained on H and E stained slides obtained from 8,343 patients with colorectal tumors across different countries and of different ethnicities (Table 1). The classifiers demonstrated a clinical-grade performance with an AUC of 0.96 without the need for pre-processing, suggesting potential utility for high-throughput, cost-

effective evaluation of colorectal tissue specimens. During the same time period, Bilal and colleagues[9] developed a weakly supervised deep learning architecture incorporating three distinct CNNs to predict MSI status, clinically relevant mutations and molecular pathways.

The study employed ResNet34 and ResNet18 architectures. The algorithms were trained using balanced datasets comprising of tumor and non-tumor tiles. By utilizing the same patient cohort and dividing the dataset into distinct training and testing sets, the authors compared their MSI prediction performance with that of Kather et al[44]. Their iterative sampling method demonstrated superior performance, achieving AUC of 0.90, as opposed to the work published by Kather et al[44] which achieved an AUC of 0.77. Thus, the developed algorithm is considered as a promising tool for the rapid prediction of clinically significant mutations and molecular pathways, such as MSI, facilitating patient stratification for targeted therapies more rapidly than traditional sequencing or IHC methods.

Another novel label-free approach using infrared imaging and AI to classify MSI status in CRC was developed by Gerwert et al[24]. The novelty of the developed model was its ability to classify MSI status based on unstained FFPE sections, thereby leaving the specimen unmodified for subsequent analyses. The study utilized Quantum Cascade Laser Infrared (QCL-IR) microscopes and classified the data using two CNN architectures. The first architecture, a modified U-Net, was utilized to identify regions containing cancer cells whereas the second architecture, VGG-Net, was used to determine MSI status. The model exhibited a promising performance with the validation AUC score reaching 0.9 (Table 1).

In addition to CNN models and cohort factors, several challenging technical issues must be addressed during the development of new DL tools. For instance, processing WSIs using CNNs presents significant challenges in terms of memory and computational processing time. WSIs are extremely high-resolution images, often exceeding gigapixels in size, which demand substantial memory capacity that can overwhelm standard computing resources[42,72]. Additionally, the computational processing time required to analyze such large-scale images through CNNs is relatively long, leading to prolonged training and prediction times that can hinder real-time diagnostic applications.

To address these issues, Tong et al[76] implemented a downscaling methodology where a segmentation-based technique was utilized to derive pixel-level image data through computational analysis. This groundbreaking whole-slide-level dMMR/pMMR deep learning detector, referred to as SPEED, significantly accelerated computational processing, reducing it by a factor of 1,700 while maintaining exceptional performance. In an internal validation set, it achieved an AUC of 0.989, surpassing the results of prior research in the field. These results indicate that artificial intelligence designed for the assessment of dMMR/MSI status presents a significantly faster and more economical alternative to conventional molecular assays. Consequently, the implementation of AI in this context could enhance diagnostic efficiency and reduce associated costs.

Limitations and prospects

Light microscopy remains the gold standard tool for assessing tissue sections in pathology. However, this technique faces several limitations including labor-intensive processes and susceptibility to subjective interpretation. The fact that numerous patients do not undergo MSI testing despite clinical guidelines advocating for its universal testing underscores the urgent need for the development and adoption of new technologies[73]. The design of an effective CNN architecture requires careful consideration of numerous parameters, such as the number of convolution layers, filter sizes and the complexity of the models themselves.

Additionally, balancing the trade-offs between model complexity and computational efficiency is crucial, as overly complex models may lead to overfitting, while simplified models might underperform and achieving the right balance is essential for achieving robust and accurate CNNs[31,61,62].

A common limitation among the reviewed articles was the bias in patient selection where studies used a single cohort (e.g. TCGA) without incorporating additional cohorts. This limitation restricts the diversity of the study population and consequently affects its generalizability[46,69]. The ability to have diverse multi-national cohorts is essential for developing generalizable models that reflect differences between diverse ethnicities. For instance, a model trained with TCGA data exhibited poor performance when applied to a Japanese cohort[44]. Similarly, Cao et al[13] demonstrated low model performance when a model was trained on a TCGA cohort or tested on an Asian cohort; however, incorporating an Asian cohort into the training set significantly improved the model's performance[13].

Additionally, securing an independent external validation set is essential in model evaluation[63] to clearly differentiate between training and validation groups. Thus, according to these observations, it has been hypothesized that morphological variances in colon cancer between Caucasian and Asian patients exist and may hinder the generalizability of deep learning systems.

Based on the data published by Pressman et al[64], the performance of the model on the Gangnam test dataset was observed to be within one standard deviation of its performance on the TCGA test dataset, where the model was trained exclusively on the TCGA data, which included a limited number of Asian patients[64].

The broad applicability of AI-based classifiers remains a significant concern for the applicability of DL models in the clinical setting. Jang et al[34-36] and others[48,49] found that classifiers designed to differentiate various molecular characteristics from tissue images using TCGA datasets exhibited suboptimal performance when applied to Korean datasets.

An effective AI-based analysis for WSI also relies on several factors including the type of sample (frozen or FFPE), the staining quality and the scanning technology employed. Thus, the limited generalizability of these models can be attributed, in part, to different sample preservation and preparation methods. For instance, frozen slides exhibit a distinct morphology compared to FFPE slides due to the freezing process[35].

In comparing their classification performance, patch-level results were robust only when classifiers were trained exclusively on one tissue type. Conversely, when classifiers were trained on two tissue types (i.e. frozen and FFPE tissues), the performance for each tissue modality was suboptimal.

Thus, the technique employed in sample collection and preparation can adversely impact the model performance[49]. This observation is supported by Jang et al[34-36] findings where model performance was reported to vary among the investigated biomarker genes where higher AUCs were achieved for APC and KRAS using frozen WSIs over FFPE WSIs, indicating that frozen WSIs may be more advantageous for molecular testing, despite FFPE methods excelling in cellular morphology preservation. In addition to the type of sample, discrepancies were noted among the studies concerning the techniques used to determine dMMR/MSI status in the original patient samples, with some studies employing PCR and others using IHC.

Considering such differences is important as each test provides different information about the tumor samples. Specifically, IHC provides information about the MMR proteins expressed in the sample, whereas MSI by PCR measures MMR function by detecting changes in DNA that result when major MMR function is lost. Consequently, these differences can have a significant impact on the model validation[61].

Finally, due to memory limitations, even server-grade graphics processing units (GPUs) are incapable of processing high-resolution WSIs. Consequently, the majority of deep learning models have been developed using classification-based approaches, whereby small patches are extracted from WSIs for training purposes rather than utilizing the full-size WSIs[61]. This necessitates the segmentation of a WSI into thousands of smaller patches which compromise the continuity of the image and result in a significant increase in computational data[5]. These issues can be partially solved by using modern WSI scanners designed with an autofocus optics system that choose focal planes to precisely capture the three-dimensional structure of tissue as a two-dimensional digital image[59].

Nevertheless, these scanners might generate digital images with out-of-focus or blurry sections if the AF optics system incorrectly selects focus points at an improper plane relative to the tissue's actual height. Additionally, color normalization can pose challenges. The high cost of infrastructure, such as rapid scanners and large storage spaces, poses a significant challenge[59]. Therefore, standardizing scanner systems, image processing models and virtual microscopy interfaces, is vital for advancing image analysis methods in pathology.

Conclusion

Light microscopy remains the gold standard in pathology, but there is an increasing demand for image analysis tools capable of performing complex analyses with high precision and reproducibility. The cost of infrastructure including rapid scanners, large storage spaces and integration into medical information systems, presents a barrier. Hence, advancing image analysis methods are becoming increasingly crucial for pathologists, driving the need for sophisticated tools to meet these demands. AI-based predictive analytics allowed scientists to analyze complex data sets from genomics, proteomics and entire pathology slides, resulting in a more thorough characterization of patients and improved precision in prognosis. These advancements promise to streamline pathological workflows.

In addition, the employment of AI technology can reduce variability due to subjective assessments, making pathology more consistent. Although AI has not yet replaced traditional diagnostic tools, it shows significant potential in predicting MSI status based on morphological features. This technology also tackles challenges like high costs, labor intensity and subjective interpretation. Given that many patients miss out on testing despite clinical recommendations for universal MSI testing, the study highlights the need for new technologies.

References

1. Abotchie P.N., Vernon S.W. and Du X.L., Gender Differences in Colorectal Cancer Incidence in the United States, 1975–2006, *Journal of Women's Health*, **21(4)**, 393–400 **(2012)**

2. Alzumaili B., Xu B., Saliba M., Abuhashem A., Ganly I., Ghossein R. and Katabi N., Clinicopathologic Characteristics and Prognostic Factors of Primary and Recurrent Pleomorphic Adenoma A Single Institution Retrospective Study of 705 Cases, *American Journal of Surgical Pathology*, **46(6)**, 854–862 **(2022)**

3. Amin M.B., Greene F.L., Edge S.B., Compton C.C., Gershenwald J.E., Brookland R.K., Meyer L., Gress D.M., Byrd D.R. and Winchester D.P., The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging, *CA: A*

Cancer Journal for Clinicians, **67(2)**, 93–99 **(2017)**

4. André T. et al, Pembrolizumab in Microsatellite-Instability–High Advanced Colorectal Cancer, *New England Journal of Medicine*, **383(23)**, 2207–2218 **(2020)**

5. Bahadir C.D., Omar M., Rosenthal J., Marchionni L., Liechty B., Pisapia D.J. and Sabuncu M.R., Artificial intelligence applications in histopathology, *Nature Reviews Electrical Engineering*, **1(2)**, 93–108 **(2024)**

6. Baxter N.N. et al, Adjuvant Therapy for Stage II Colon Cancer: ASCO Guideline Update, *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **40(8)**, 892–910 **(2022)**

7. Benson A.B. et al, Colon Cancer, Version 3.2024, NCCN Clinical Practice Guidelines in Oncology, *Journal of the National Comprehensive Cancer Network*, **22(2D)**, e240029 **(2024)**

8. Benson A.B. et al, Colon Cancer, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology, *Journal of the National Comprehensive Cancer Network*, **19(3)**, 329–359 **(2021)**

9. Bilal M., Raza S.E.A., Azam A., Graham S., Ilyas M., Cree I.A., Snead D., Minhas F. and Rajpoot N.M., Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study, *The Lancet Digital Health*, **3(12)**, e763–e772 **(2021)**

10. Bilimoria K.Y., Palis B., Stewart A.K., Bentrem D.J., Freel A.C., Sigurdson E.R., Talamonti M.S. and Ko C.Y., Impact of Tumor Location on Nodal Evaluation for Colon Cancer, *Diseases of the Colon & Rectum.*, **51(2)**, 154–161 **(2008)**

11. Bray F., Laversanne M., Sung H., Ferlay J., Siegel R.L., Soerjomataram I. and Jemal A., Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA: A Cancer Journal for Clinicians*, **74(3)**, 229–263 **(2024)**

12. Brown I. and Mues C., An experimental comparison of classification algorithms for imbalanced credit scoring data sets, *Expert Systems with Applications*, **39(3)**, 3446–3453 **(2021)**

13. Cao R. et al, Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in Colorectal Cancer, *Theranostics*, **10(24)**, 11080–11091 **(2020)**

14. Chang X., Wang J., Zhang G., Yang M., Xi Y., Xi C., Chen G., Nie X., Meng B. and Quan X., Predicting colorectal cancer microsatellite instability with a self-attention-enabled convolutional neural network, *Cell Reports Medicine*, **4(2)**, 100914 **(2023)**

15. Chen S. et al, Estimates and Projections of the Global Economic Cost of 29 Cancers in 204 Countries and Territories From 2020 to 2050, *JAMA Oncology*, **9(4)**, 465 **(2023)**

16. Chu Q.D., Zhou M., Medeiros K. and Peddi P., Positive surgical margins contribute to the survival paradox between patients with stage IIB/C (T4N0) and stage IIIA (T1-2N1, T1N2a) colon cancer, *Surgery*, **160(5)**, 1333–1343 **(2016)**

17. Compton C.C. et al, Prognostic Factors in Colorectal Cancer, *Archives of Pathology & Laboratory Medicine*, **124(7)**, 979–994 **(2000)**

18. Debniak T., Kurzawski G., Gorski B., Kladny J., Domagala W. and Lubinski J., Value of pedigree/clinical data, immunohistochemistry and microsatellite instability analyses in reducing the cost of determining hMLH1 and hMSH2 gene mutations in patients with colorectal cancer, *European Journal of Cancer*, **36(1)**, 49–54 **(2000)**

19. Ding K., Zhou M., Wang H., Zhang S. and Metaxas D.N., Spatially aware graph neural networks and cross-level molecular profile prediction in colon cancer histopathology: a retrospective multi-cohort study, *The Lancet Digital Health*, **4(11)**, e787–e795 **(2022)**

20. Echle A. et al, Artificial intelligence for detection of microsatellite instability in colorectal cancer—a multicentric analysis of a pre-screening tool for clinical application, *ESMO Open*, **7(2)**, 100400 **(2022)**

21. Echle A. et al, Clinical-Grade Detection of Microsatellite Instability in Colorectal Tumors by Deep Learning, *Gastroenterology*, **159(4)**, 1406-1416.e11 **(2020)**

22. Fridman W.H., Pagès F., Sautès-Fridman C. and Galon J., The immune contexture in human tumours: impact on clinical outcome, *Nature Reviews Cancer*, **12(4)**, 298–306 **(2012)**

23. Galon J. et al, Type, Density and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome, *Science*, **313(5795)**, 1960–1964 **(2006)**

24. Gerwert K. et al, Fast and Label-Free Automated Detection of Microsatellite Status in Early Colon Cancer Using Artificial Intelligence Integrated Infrared Imaging, *Eur J Cancer*, **182**, 122-131 **(2022)**

25. Grothey A. et al, Duration of Adjuvant Chemotherapy for Stage III Colon Cancer, *New England Journal of Medicine*, **378(13)**, 1177–1188 **(2018)**

26. Gündüz E., Alçin Ö.F., Kizilay A. and Piazza C., Radiomics and deep learning approach to the differential diagnosis of parotid gland tumors, *Current Opinion in Otolaryngology and Head and Neck Surgery*, **30(2)**, 107–113 **(2022)**

27. Guo B., Li X., Yang M., Jonnagaddala J., Zhang H. and Xu X.S., Predicting microsatellite instability and key biomarkers in colorectal cancer from H&E-stained images: achieving state-of-the-art predictive performance with fewer data using Swin Transformer, *The Journal of Pathology: Clinical Research*, **9(3)**, 223–235 **(2023)**

28. Guraya S.Y., Pattern, Stage and Time of Recurrent Colorectal Cancer After Curative Surgery, *Clinical Colorectal Cancer*, **18(2)**, e223–e228 **(2019)**

29. Haller D.G., Tabernero J., Maroun J., de Braud F., Price T., Van Cutsem E., Hill M., Gilberg F., Rittweger K. and Schmoll H.J., Capecitabine plus oxaliplatin compared with fluorouracil and folinic acid as adjuvant therapy for stage III colon cancer, *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, **29(11)**, 1465–71 **(2011)**

30. Hari D.M., Leung A.M., Lee J.H., Sim M.S., Vuong B., Chiu C.G. and Bilchik A.J., AJCC Cancer Staging Manual 7th Edition Criteria for Colon Cancer: Do the Complex Modifications Improve Prognostic Assessment?, *Journal of the American College of Surgeons*, **217(2)**, 181–190 **(2013)**

31. Hildebrand L.A., Pierce C.J., Dennis M., Paracha M. and Maoz A., Artificial Intelligence for Histology-Based Detection of Microsatellite Instability and Prediction of Response to Immunotherapy in Colorectal Cancer, *Cancers*, **13(3)**, 391 **(2021)**

32. Hong H., Pradhan B., Xu C. and Tien Bui D., Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines, *CATENA*, **133**, 266–281 **(2015)**

33. Hossain M.S. et al, Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies, *Cancers*, **14(7)**, 1732 **(2022)**

34. Jang H.J., Go J.H., Kim Y. and Lee S.H., Deep Learning for the Pathologic Diagnosis of Hepatocellular Carcinoma, Cholangiocarcinoma and Metastatic Colorectal Cancer, *Cancers*, **15(22)**, 5389 **(2023)**

35. Jang H.J., Lee A., Kang J., Song I.H. and Lee S.H., Prediction of clinically actionable genetic alterations from colorectal cancer histopathology images using deep learning, *World Journal of Gastroenterology*, **26(40)**, 6207–6223 **(2020)**

36. Jang H.J., Lee A., Kang J., Song I.H. and Lee S.H., Prediction of genetic alterations from gastric cancer histopathology images using a fully automated deep learning approach, *World Journal of Gastroenterology*, **27(44)**, 7687–7704 **(2021)**

37. Janiesch C., Zschech P. and Heinrich K., Machine learning and deep learning, *Electronic Markets*, **31(3)**, 685–695 **(2021)**

38. Jiang T. et al, Deep learning-assisted diagnosis of benign and malignant parotid tumors based on ultrasound: a retrospective study, *BMC Cancer*, **24(1)**, 510 **(2024)**

39. Jiang W. et al, Clinical actionability of triaging DNA mismatch repair deficient colorectal cancer from biopsy samples using deep learning, *eBio Medicine*, **81**, 104120 **(2022)**

40. Jiang Z. et al, Development and Interpretation of a Clinicopathological-Based Model for the Identification of Microsatellite Instability in Colorectal Cancer, *Disease Markers*, **2023**, 5178750 **(2023)**

41. Kalloger S.E., Allo G., Mulligan A.M., Pollett A., Aronson M., Gallinger S., Torlakovic E.E. and Clarke B.A., Use of Mismatch Repair Immunohistochemistry and Microsatellite Instability Testing, *American Journal of Surgical Pathology*, **36(4)**, 560–569 **(2012)**

42. Kalra S., Tizhoosh H.R., Choi C., Shah S., Diamandis P., Campbell C.J.V. and Pantanowitz L., Yottixel – An Image Search Engine for Large Archives of Histopathology Whole Slide Images, *Medical Image Analysis*, **65**, 101757 **(2020)**

43. Kather J.N. et al, Pan-cancer image-based detection of clinically actionable genetic alterations, *Nature Cancer*, **1(8)**, 789–799 **(2020)**

44. Kather J.N. et al, Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer, *Nature Medicine*, **25(7)**, 1054–1056 **(2019)**

45. Kirilovsky A., Marliot F., El Sissy C., Haicheur N., Galon J. and Pagès F., Rational bases for the use of the Immunoscore in routine clinical settings as a prognostic and predictive biomarker in cancer patients, *International Immunology*, **28(8)**, 373–82 **(2016)**

46. Krause J. et al, Deep learning detects genetic alterations in cancer histology generated by adversarial networks, *The Journal of Pathology*, **254(1)**, 70–79 **(2021)**

47. Le D.T. et al, PD-1 Blockade in Tumors with Mismatch-Repair Deficiency, *New England Journal of Medicine*, **372(26)**, 2509–2520 **(2015)**

48. Lee S.H., Lee Y. and Jang H., Deep learning captures selective features for discrimination of microsatellite instability from pathologic tissue slides of gastric cancer, *International Journal of Cancer*, **152(2)**, 298–307 **(2023)**

49. Lee S.H., Song I.H. and Jang H., Feasibility of deep learning-based fully automated classification of microsatellite instability in tissue slides of colorectal cancer, *International Journal of Cancer*, **149(3)**, 728–740 **(2021)**

50. León F., Plazas M. and Martínez F., An inception deep architecture to differentiate close-related Gleason prostate cancer scores, In Brieva J., Romero E. and Lepore N., 15th International Symposium on Medical Information Processing and Analysis, SPIE, 66 **(2020)**

51. Lou J., Xu J., Zhang Y., Sun Y., Fang A., Liu J., Mur L.A.J. and Ji B., PPsNet: An improved deep learning model for microsatellite instability high prediction in colorectal cancer from whole slide images, *Computer Methods and Programs in Biomedicine*, **225**, 107095 **(2022)**

52. Mathews A.A., Draganov P.V. and Yang D., Endoscopic management of colorectal polyps: From benign to malignant polyps, *World Journal of Gastrointestinal Endoscopy*, **13(9)**, 356–370 **(2021)**

53. Meyers B.M., Cosby R., Quereshy F. and Jonker D., Adjuvant Chemotherapy for Stage II and III Colon Cancer Following Complete Resection: A Cancer Care Ontario Systematic Review, *Clinical Oncology (Royal College of Radiologists*, **29(7)**, 459–465 **(2017)**

54. Morgan E., Arnold M., Gini A., Lorenzoni V., Cabasag C.J., Laversanne M., Vignat J., Ferlay J., Murphy N. and Bray F., Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN, *Gut.*, **72(2)**, 338–344 **(2023)**

55. Mormont R., Geurts P. and Maree R., Multi-Task Pre-Training of Deep Neural Networks for Digital Pathology, *IEEE Journal of Biomedical and Health Informatics*, **25(2)**, 412–421 **(2021)**

56. Murchan P., Ó'Brien C., O'Connell S., McNevin C.S., Baird A.M., Sheils O., Ó Broin P. and Finn S.P., Deep Learning of Histopathological Features for the Prediction of Tumour Molecular Genetics, *Diagnostics*, **11(8)**, 1406 **(2021)**

57. NCT00002575, C go, identifier, Laparoscopic-Assisted Surgery Compared With Open Surgery in Treating Patients With Colon Cancer **(2016)**

58. Nguyen H. and Duong H., The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy (Review), *Oncology Letters*, **16(1)**, 9–18 **(2018)**

59. Niazi M.K.K., Parwani A.V. and Gurcan M.N., Digital pathology and artificial intelligence, *The Lancet Oncology*, **20(5)**, e253–e261 **(2019)**

60. Panceri S.S., Mutz F., Cardoso V.B., Carneiro R.V., Oliveira-Santos T., Badue C. and de Souza A.F., Detecting Cancerous Tissue in Mammograms Using Deep Neural Networks, 2021 International Joint Conference on Neural Networks (IJCNN), IEEE, 1–8 **(2021)**

61. Park J.H., Kim E.Y., Luchini C., Eccher A., Tizaoui K., Shin J. Il. and Lim B.J., Artificial Intelligence for Predicting Microsatellite Instability Based on Tumor Histomorphology: A Systematic Review, *International Journal of Molecular Sciences*, **23(5)**, 2462 **(2022)**

62. Park S.H., Artificial Intelligence in Medicine: Beginner's Guide, *Journal of the Korean Society of Radiology*, **78(5)**, 301 **(2018)**

63. Park S.H. and Han K., Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction, *Radiology*, **286(3)**, 800–809 **(2018)**

64. Pressman I.S., Xu H., Kang J., Cha Y.J., Lee S.H. and Hwang T.H., Abstract 2100: Deep learning can predict microsatellite instability from histology in colorectal cancer across different ethnic groups, *Cancer Research*, **80(16)**, 2100–2100 **(2020)**

65. Rotermund C., Djinbachian R., Taghiakbari M., Enderle M.D., Eickhoff A. and von Renteln D., Recurrence rates after endoscopic resection of large colorectal polyps: A systematic review and meta-analysis, *World Journal of Gastroenterology*, **28(29)**, 4007–4018 **(2022)**

66. Saillard C. et al, Validation of MSIntuit as an AI-based pre-screening tool for MSI detection from colorectal cancer histology slides, *Nature Communications*, **14(1)**, 6695 **(2023)**

67. Sauter M., Keilholz G., Kranzbühler H., Lombriser N., Prakash M., Vavricka S.R. and Misselwitz B., Presenting symptoms predict local staging of anal cancer: a retrospective analysis of 86 patients, *BMC Gastroenterology*, **16(1)**, 46 **(2016)**

68. Schirris Y., Gavves E., Nederlof I., Horlings H.M. and Teuwen J., DeepSMILE: Contrastive self-supervised pre-training benefits MSI and HRD classification directly from H&E whole-slide images in colorectal and breast cancer, *Medical Image Analysis*, **79**, 102464 **(2022)**

69. Schmauch B. et al, A deep learning model to predict RNA-Seq expression of tumours from whole slide images, *Nature Communications*, **11(1)**, 3877 **(2020)**

70. Serag A., Ion-Margineanu A., Qureshi H., McMillan R., Saint Martin M.J., Diamond J., O'Reilly P. and Hamilton P., Translational AI and Deep Learning in Diagnostic Pathology, *Frontiers in Medicine*, **6**, 185 **(2019)**

71. Shahamatdar S., Saeed-Vafa D., Linsley D., Khalil F., Lovinger K., Li L., McLeod H.T., Ramachandran S. and Serre T., Deceptive learning in histopathology, *Histopathology*, **85(1)**, 116–132 **(2024)**

72. Shen A., Wang F., Paul S., Bhuvanapalli D., Alayof J., Farris A.B., Teodoro G., Brat D.J. and Kong J., An integrative web-based software tool for multi-dimensional pathology whole-slide image analytics, *Physics in Medicine & Biology*, **67(22)**, 224001 **(2022)**

73. Sun B.L., Current Microsatellite Instability Testing in Management of Colorectal Cancer, *Clinical Colorectal Cancer*, **20(1)**, e12–e20 **(2021)**

74. Sun G., Dong X., Tang X., Qu H., Zhang H. and Zhao E., The prognostic value of immunoscore in patients with colorectal cancer: A systematic review and meta-analysis, *Cancer Medicine*, **8(1)**, 182–189 **(2019)**

75. Tang J., Yu J.X., Hubbard-Lucey V.M., Neftelinov S.T., Hodge J.P. and Lin Y., The clinical trial landscape for PD1/PDL1 immune checkpoint inhibitors, *Nature Reviews Drug Discovery*, **17(12)**, 854–855 **(2018)**

76. Tong Z. et al, Development of a whole-slide-level segmentation-based dMMR/pMMR deep learning detector for colorectal cancer, *iScience*, **26(12)**, 108468 **(2023)**

77. Weiser M.R., AJCC 8th Edition, Colorectal Cancer, *Annals of Surgical Oncology*, **25(6)**, 1454–1455 **(2018)**

78. Williams H. and Steinhagen R.M., Historical Perspectives: Malignancy in Crohn's Disease and Ulcerative Colitis, *Clinics in Colon and Rectal Surgery*, **37(01)**, 005–012 **(2024)**

79. Yahagi M., Okabayashi K., Hasegawa H., Tsuruta M. and Kitagawa Y., The Worse Prognosis of Right-Sided Compared with Left-Sided Colon Cancers: a Systematic Review and Meta-analysis, *Journal of Gastrointestinal Surgery*, **20(3)**, 648–655 **(2016)**

80. Yamashita R., Long J., Longacre T., Peng L., Berry G., Martin B., Higgins J., Rubin D.L. and Shen J., Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study, *The Lancet Oncology*, **22(1)**, 132–141 **(2021)**

81. Yanus G.A. et al, The spectrum of Lynch syndrome-associated germ-line mutations in Russia, *European Journal of Medical Genetics*, **63(3)**, 103753 **(2020)**

82. Yu Q. et al, Deep learning-assisted diagnosis of benign and malignant parotid tumors based on contrast-enhanced CT: a multicenter study, *European Radiology*, **33(9)**, 6054–6065 **(2023)**

83. Zare-Bandamiri M., Fararouei M., Zohourinia S., Daneshi N. and Dianatinasab M., Risk Factors Predicting Colorectal Cancer Recurrence Following Initial Treatment: A 5-year Cohort Study,

Asian Pacific Journal of Cancer Prevention, **18(9)**, 2465–2470 **(2017)**

84. Zhao P., Li L., Jiang X. and Li Q., Mismatch repair deficiency/microsatellite instability-high as a predictor for anti-PD-1/PD-L1 immunotherapy efficacy, *Journal of Hematology & Oncology*, **12(1)**, 54 **(2019)**.

(Received 24th September 2024, accepted 28th November 2024)

*****