

Review

Deep Learning Approaches to Colorectal Cancer Diagnosis: A Review

Lakpa Dorje Tamang  and Byung Wook Kim * 

Department of Information and Communication Engineering, Changwon National University,
Changwon 51140, Korea; l.tamang9@changwon.ac.kr

* Correspondence: bwkim@changwon.ac.kr

Abstract: Unprecedented breakthroughs in the development of graphical processing systems have led to great potential for deep learning (DL) algorithms in analyzing visual anatomy from high-resolution medical images. Recently, in digital pathology, the use of DL technologies has drawn a substantial amount of attention for use in the effective diagnosis of various cancer types, especially colorectal cancer (CRC), which is regarded as one of the dominant causes of cancer-related deaths worldwide. This review provides an in-depth perspective on recently published research articles on DL-based CRC diagnosis and prognosis. Overall, we provide a retrospective synopsis of simple image-processing-based and machine learning (ML)-based computer-aided diagnosis (CAD) systems, followed by a comprehensive appraisal of use cases with different types of state-of-the-art DL algorithms for detecting malignancies. We first list multiple standardized and publicly available CRC datasets from two imaging types: colonoscopy and histopathology. Secondly, we categorize the studies based on the different types of CRC detected (tumor tissue, microsatellite instability, and polyps), and we assess the data preprocessing steps and the adopted DL architectures before presenting the optimum diagnostic results. CRC diagnosis with DL algorithms is still in the preclinical phase, and therefore, we point out some open issues and provide some insights into the practicability and development of robust diagnostic systems in future health care and oncology.



Citation: Tamang, L.D.; Kim, B.W. Deep Learning Approaches to Colorectal Cancer Diagnosis: A Review. *Appl. Sci.* **2021**, *11*, 10982. <https://doi.org/10.3390/app112210982>

Received: 12 October 2021
Accepted: 16 November 2021
Published: 19 November 2021

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



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: colorectal cancer; digital pathology; computer-aided diagnosis; deep learning

1. Introduction

Global cancer statistics from 2018 show that the incidence of colorectal cancer (CRC) ranks highest after lung cancer and breast cancer, and worldwide, it accounts for approximately 10% of the total annual cancer patients among both men and women [1]. Although people aged 65 years and above are the most prevalent victims of this disease, the risk in younger patients is also significant, with the highest risk due to heredity (35%) followed by other factors such as obesity, bad nutritional habits, and smoking [2]. These rates show no trend toward decline, but rather are expected to increase by more than 60% in the next decade, with more than two million new diagnoses and over a million deaths by the next decade [3]. In this regard, there is a need to develop an optimal diagnosis strategy for the early and precise detection of CRC patients.

With routine screening being an important step for the reduction in mortality rates of this disease, colonoscopy (an endoscopic method) is considered a primary and straightforward clinical diagnosis method of choice for CRC [4]. Aside from this method, medical imaging techniques such as CT colonography, a complementary imaging method for polyp detection in CRC, and the histological evaluation of hematoxylin and eosin (H&E) slides remain indispensable approaches to subtle inspections for CRC. While manual observations of these imaging modalities by individual pathologists have been pursued relentlessly, recently, they have been modeled as traditional and unsophisticated approaches that are highly labor-intensive and time-consuming. Besides, inter-observer variation can be significant during pathological diagnosis, resulting in the biased analysis of typing and grading

of cancer tumors [5]. Therefore, a more standardized and automated technique based on computer-aided diagnosis (CAD) has gained a lot of interest and demand lately.

Many CAD systems have been utilized in mainstream radiology to assist physicians, from improved chest X-rays and mammography applications in the 1960s to enabling the early diagnosis of cancers in the 2000s. Considering the medical and economic burdens caused by the prognosis and treatment of CRC-related diseases, researchers have been focusing on developing CAD systems for use in the early and effective diagnosis of CRC. The development of CAD systems for CRC can be dated from the conventional models that require complex a priori knowledge of mathematics [6–8] to advanced machine learning (ML)-based systems [9–12] that can perform beyond human levels of accuracy.

Although cancer diagnosis with deep learning (DL) has been a very popular subject of interest in the medical imaging domain, comprehensive literature reviews covering various aspects of CRC diagnosis and prognosis using state-of-the-art DL schemes are still limited. The existing studies lack surveys based on various types of available standard CRC imaging datasets. In addition, in a short period of time, there has been adequate novel research and findings from DL-based CRC diagnoses. A proper review of these state-of-the-art findings in terms of adapted data preprocessing strategies is required, and a methodology needs to be developed to facilitate upcoming researchers and scholars in this field. Therefore, this review paper intends to fill this gap in four ways. First, it provides a brief retrospective overview of conventional CAD systems based on simple image processing and ML-based approaches to CRC diagnosis. Secondly, we identify and list some of the publicly available imaging datasets collected and archived from various independent sources, which are standardized for DL-based CRC diagnosis. Thirdly, we systematically categorize and highlight the latest studies on DL-based detection, diagnosis, and prognosis regarding different types of CRC, including tumors, microsatellite instability (MSI), and polyps. Lastly, we outline some open issues observed in this area of research and speculate on future studies regarding the optimization of diagnosis accuracy, such that it is practical and suitable for use in the clinical domain.

To sum up the organization of this paper, the next section provides a brief understanding of the CAD approaches based on simple image processing techniques and ML-based techniques. Section III presents a detailed discussion of recently published research into CRC diagnosis by organizing it into different categories, each highlighting a key contribution to data preprocessing, model architectures, and the optimal results obtained. Finally, a discussion and conclusion are presented in Sections 4 and 5, respectively.

2. Simple Image-Processing-Based and Machine-Learning-Based CAD Approaches

Owing to synthetic approaches based on image processing techniques, conventional CAD systems have been used in diagnosing CRC for a few decades. Table 1 lists a brief overview of the conventional studies regarding simple image processing techniques and ML-based CAD systems that have been researched for use in diagnosing CRC. CAD systems based on simple image processing methods rely on mathematical models explicitly defined by human rules for processing images from one modality to another, and they require the case-by-case tuning of model parameters for optimal performance. Specifically, diagnoses are mainly based on feature engineering methods where feature extraction is carried out either from vessel structural or textural analysis of the image patches using a local binary pattern (LBP) [13]. Solely relying on simple image processing algorithms, these CAD systems [6–8] can classify tumor regions or identify various traits of malignant tissues in CRC. Although these systems have been a part of digital pathology for the clinical diagnosis of CRC, they are application-specific and are considered heuristic approaches that require strong domain expertise, which depend upon the unique characteristics of the imaging type involved.

Table 1. Overview of CAD systems based on simple image processing techniques and ML-based techniques.

Technology	Publication (Year)	Objective	Methods Used	Image Types	Maximum Results
Based on image processing techniques	[6] (2011)	To develop a CAD system for the classification of colorectal polyps.	ROI selection, blood vessel segmentation, and feature extraction and classification.	Endoscopic images	Sensitivity = 95.0% Accuracy = 86.8% Specificity = 87.8%
	[7] (2015)	To provide the fully automated and instant classification of colorectal polyps for routine colonoscopy.	Image acquisition, adaptive thresholding, nucleus-like spot labeling, removal of artifacts, and feature extraction.	Endoscopic images	Sensitivity = 92.0% Accuracy = 89.2% Specificity = 79.5%
	[8] (2017)	To evaluate the effectiveness of software developed for endoscopic diagnosis of colorectal lesions using magnifying narrow band imaging (M-NBI)	ROI extraction, grayscale conversion, binarization using moving average methods, morphological operations, and feature extraction.	Endoscopic images	Sensitivity = 83.9% Accuracy = 82.8% Specificity = 82.6%
Based on machine learning techniques	[9] (2013)	To classify NBI images of colorectal tumors into three types (A, B, and C3) based on NBI magnification findings.	Feature extraction using bag of visual words, and classification with support vector machine (SVM) classifiers.	Colonoscopy, NBI samples	Recognition rate = 95.44%
	[10] (2016)	To predict histological diagnoses of colorectal lesions depicted on narrow band imaging samples.	Feature extraction using densely sampled scale-invariant feature transform (SIFT) in a bag-of-features framework; classification with SVM.	Endoscopic images	Sensitivity = 93.0% Specificity = 93.3% Positive prediction value (PPV) = 93.0% Negative prediction value (NPV) = 93.3%
	[11] (2017)	To investigate the feasibility of computer algorithms to identify early Barrett's esophagus (BE) neoplasia on ex vivo volumetric laser endomicroscopy (VLE) images.	Preprocessing of input images, feature extraction with gray-level co-occurrence matrices, local binary patterns, wavelet transforms, and histogram of oriented gradients; classification with SVM, decision trees, k-nearest neighbors, linear regression, and logistic regression.	Volumetric laser endomicroscopy images	Sensitivity = 90% Specificity = 93% AUC = 0.95
	[12] (2019)	To develop a CAD system based on linked color imaging (LCI) to predict the histological results of polyps by analyzing colors of the lesions.	Preprocessing of input images, converting images from RGB space to HLS space and concatenating both images to obtain a 6-D vector for each pixel; classification using a Gaussian mixture model (GMM).	Linked color imaging samples	Sensitivity = 83.3% Accuracy = 78.4% Specificity = 70.1% PPV = 82.6% NPV = 71.2%

With the technological advancements in the field of artificial intelligence (AI), a computer can mimic cognitive functions to solve real-world problems by learning all by itself. Within AI, an ML technique that allows computers to learn from real-world data without being explicitly programmed has been extensively applied to the medical imaging domain. Particularly for the clinical diagnosis of CRC patients, several research works based on ML approaches [9–12] have been conducted. The ML-based techniques utilize the handcrafted feature extraction of predefined morphological features that rely on the shape, color, and textural information of the image data. These types of features are usually extracted using different types of procedures, such as LBP, wavelet transform, gray-level co-occurrence matrices, scale invariant feature transform (SIFT), histogram of oriented gradients (HoG), etc. The extracted features are then subjected to ML-based classification algorithms that include, but are not limited to, the support vector machine, k-nearest

neighbors, logistic regression, decision trees, m-Gaussian mixture models, etc. Although these techniques have been introduced as part of the medical diagnoses of CRC patients, they have certain limitations due to their infirm feature extraction procedures. Moreover, ML algorithms cannot produce unbiased representations for the large amounts of data, which makes them highly susceptible to overfitting and errors.

3. Deep Learning-Based Studies for CRC Diagnosis

Deep learning [14] is the branch of ML in the AI paradigm that identifies trends and patterns in the data without the need for human intervention or any feature engineering methods. The DL method makes use of multiple hidden layers to extract an abstract representation of the input at each layer that is appropriate to perform a specific task. DL models have been regarded as superior to ML-based techniques in the presence of large amounts of data and have been popular in multiple disciplines [15,16], including the diagnosis and prognosis of cancer in digital pathology. In Figure 1, a procedural diagram demonstrating the working mechanisms of ML and DL-based CAD systems screening for CRC patients is displayed. ML relies on handcrafted feature extraction before passing features for classification, whereas DL concurrently extracts and classifies features through multiple hidden layers and activation functions. This makes the DL technique suitable for learning more task-specific representations of large-scale image datasets; thus, they are frequently preferred over ML in solving medical imaging classification or in tumor detection problems.

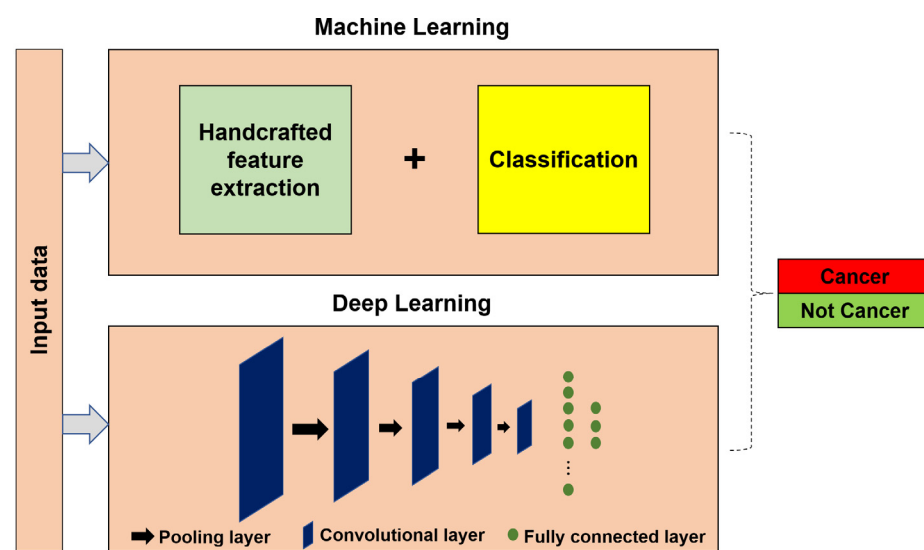


Figure 1. Block diagram showing the fundamental difference between working mechanisms of ML and DL algorithms for classifying images as cancer or non-cancer subtypes.

3.1. Datasets

Data constitute an inevitable part of a DL algorithm, from which the model can learn concealed information or underlying statistics available within them. Data can be in any form such as numbers, audio, images, and videos. Dataset preparation is a long process that includes collection, analysis and treatment, exploration, training, and testing. The data from which the model is trained must be relevant to the specific problem and must resemble real-world data as much as possible. To train a DL model, a large amount of data with significant standard deviation is required. With more data, better accuracy from a DL algorithm can be acquired as the model learns an abundant number of variations and recognizes invariant features and discrete instances of the input samples.

In medical imaging, data are acquired for several purposes, including but not limited to disease diagnosis, therapy planning, intraoperative navigation, and biomedical research [17]. Unlike normal imaging modalities, medical image data are hard to acquire

due to privacy and confidentiality considerations. Besides that, requirements for accurate imaging, specific contrast, minimal artifacts, and a sufficient signal-to-noise ratio make it hard to obtain the optimum image quality required for clinical practice. In cancer diagnosis, particularly in the CRC domain, the analysis of endoscopy/colonoscopy image samples, as shown in Figure 2 (top row), has been popular in the past. Capturing colonoscopy images is considered an invasive procedure where a tiny tube is inserted along the entire length of the colon to provide an interior view of cross-sectional areas. Histopathology imaging, as shown in Figure 2 (bottom row), on the other hand, is a less invasive procedure that provides a more comprehensive view of the disease, and it preserves the underlying tissue architecture. Due to a lack of computational resources and the high cost of digital imaging equipment, this image modality has been overlooked in the past. However, thanks to the high-end computational resources recently developed, spatial analysis of histopathology imagery has been considered the backbone of most automated image analysis techniques and remains the undisputed best way to diagnose vast numbers of diseases, including all cancer types [18]. In digital pathology, histological images are stained with H&E to view cellular and tissue structural details. These H&E-stained slides are utilized to confirm the presence or absence of disease, for disease grading, and for measuring disease progression in CRC.

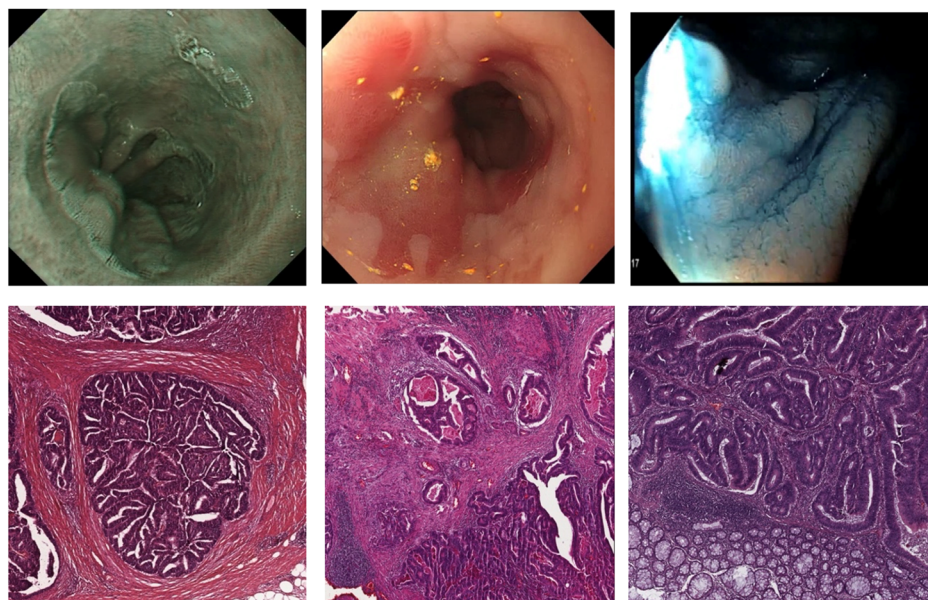


Figure 2. Sample images from a CRC dataset: (top row), endoscopy images, and (bottom row), histological whole slide images (WSI).

To this end, different types of CRC datasets belonging to either colonoscopy or histological imaging have been introduced. These images are preprocessed by applying several techniques before passing them to DL algorithms for specific tasks, such as detection, segmentation, and classification. Table 2 lists some of the popular datasets used in multiple studies based on developing DL-based CAD systems. These datasets provide comprehensive imagery of CRC tissue and tumors and entail disease-specific characteristics that are annotated by experienced pathologists.

Table 2. Some of the publicly available CRC imaging datasets.

Type	Name	Release Date	Total Number of Image Samples	Average Dimensions (in Pixels)
Endoscopy images	The KVASIR dataset [19]	2017	1000	720 × 576 to 1920 × 1072
	The Nerthus dataset [20]	2017	5525	750 × 576
	The Hyper-Kvasir dataset [21]	2019	110,079	224 × 224
	CVC-Colon DB [22]	2017	356	500 × 574 to 1080 × 1920
	CVC-Clinic DB [23]	2012	612 images	384 × 288
	Endoscopy disease detection (EDD) and segmentation [24]	2020	386 images	400 × 400
Histology images	Kather texture dataset [25]	2016	5000 patches	150 × 150
	Colorectal adenocarcinoma gland (CRAG) [26]	2019	38 WSI	1512 × 1516
	Colorectal nuclear segmentation and phenotypes (CoNSeP) [27]	2018	41 WSI	1000 × 1000
	CRC-TIA [28]	2017	139 WSI	1792 × 1792
	Histological images for tumor detection in gastrointestinal cancer [28]	2019	11,977 patches	512 × 512
	Pathology AI platform (PAIP) [29]	2019	118 WSI	29,879 × 23,066
	Warwick-QU dataset [30]	2016	166 images	775 × 522
	UniToPatho [31]	2021	9536 patches	224 × 224

3.2. Tumor Tissue Detection and Classification

Tumors are complex structures in a human organ that comprise multiple and distinct types of tissue. They can be interpreted as abnormal tissue composed of multiple types of cells or a matrix of cells. In CRC, the architecture of a tumor is varied, along with its development, and is the major factor in patient prognosis [25]. Therefore, an automated and highly quantitative analysis of tumor tissue is indispensable for the clinical diagnosis of CRC. Automatic analysis of these tissue regions can be helpful in quantifying their extent, in the grading of tumors, and to investigate a biological hypothesis based on tissue morphology.

Figure 3 shows the different types of tumor tissues obtained from H&E-stained histological slides that are relevant to CRC. These tissue types, when evaluated by pathologists, are visually classified into one of eight different categories (tumor, stroma, complex, lympho, debris, mucosa, adipose, and empty). A DL-based CAD system can automatically classify these tumor regions if provided with adequate amounts of data and if trained with optimal network hyper-parameters.

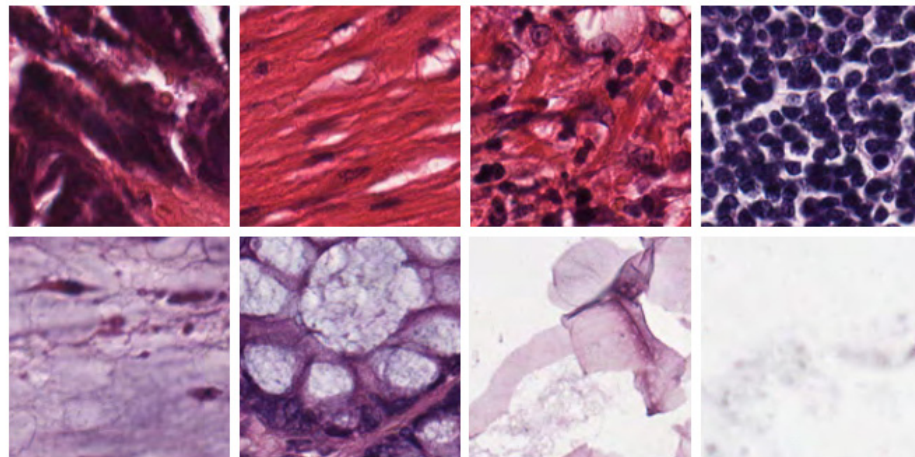


Figure 3. Sample images from a CRC dataset: (**top row**), endoscopy images, and (**bottom row**), histological whole slide images (WSI).

Multiple studies based on DL have been conducted to this end in order to accurately classify CRC tumor regions. Ponzio et al. [32] proposed a CNN framework to distinctly classify adenocarcinomas (a type of tumor) from healthy tissues and benign lesions. As a preprocessing step, they created a total of 13,500 image patches with dimensions of 1089×1096 at $40\times$ magnification of the original whole slide images (WSIs) from H&E-stained images. Subsequently, to compensate for the color inconsistencies, sets of whole image patches were normalized based on mean and standard deviation. These preprocessed datasets were fed into a CNN model with thirteen convolutional layers, five max pooling layers, and three fully connected layers (FCLs) to univocally classify them into one of three tissue subtypes, namely adenocarcinoma, tubulovillous adenoma, and healthy tissue. An initial classification accuracy of around 90% was obtained, which was optimized to 96% by using a transfer learning strategy. Another study [33] developed DL-based automated analysis of CRC image samples with the objective of improving the prognostic stratification of patients. In this study, the original H&E-stained WSIs were split into uniform-dimension tiles of 224×224 pixels, after which, VGG16 [34] (a popular pretrained CNN model) was used to extract intermediate features from the image patches. The extracted 4096 bin feature vector was classified into several tumor types by using a combination of long short-term memory (LSTM) [35] with one of three classifiers: a support vector machine (SVM), logistic regression, or naïve Bayes. LSTM is a type of recurrent neural network (RNN) that is well suited for classifying, processing, and making predictions on the time series data and is famous for its capability of learning the long-term temporal dependencies of input data. The model's performance was assessed with different accuracy metrics, where an area under the curve (AUC) value of 0.69, a hazard ratio of 2.3, and a 95% confidence interval (CI) were achieved. Similar to the previous study, Yue et al. [36] also used a well-known VGG16 framework with some notable changes to its architecture, where classification was carried out with a voting classifier and an SVM classifier. In this study, data preprocessing was applied to the H&E slides before passing them on for feature extraction. The steps included chromatic normalization of the image patches at 224×224 pixels and data augmentation to increase the number of samples for better generalization of the network. The patch level accuracy and F1-score were found to be 70% and 0.67, respectively, while a cluster level experiment significantly outperformed the former with a staggering accuracy of 100% and a unit F1-score.

In DL, the accuracy of a model is significantly dependent upon the type of feature extractor and the classification procedures [37]. Therefore, multiple studies considered using a variety of popular CNN models or designed a model from scratch with the optimal tuning of hyper-parameters. To make use of multiple pretrained models and to evaluate their performance, Kather et al. [38] investigated whether the existing pretrained CNN

models could extract the prognosticators directly from H&E-stained tissue slides. Human cancer tissue slides from multiple patient cohorts (NCT biobank at <http://dx.doi.org/10.5281/zenodo.1214456> (accessed on 18 July 2021), a DACHS study at <http://dx.doi.org/10.5281/zenodo.1214456> (accessed on 18 July 2021), and a TCGA cohort at <http://cancer.digitalarchive.net> (accessed on 18 July 2021) (NCT: National Center for Tumor Diseases, DACHS: Darmkrebs Chancen der Verhütung durch Screening, and TCGA: The Cancer Genome Atlas) were used as training and testing datasets. For data preprocessing, they created several non-overlapping image patches, each at 224×224 pixels, and normalized them with the Macenko method [39]. Five pretrained models (VGG19 [34], AlexNet [40], SqueezeNet v1.1 [41], GoogLeNet [42], and ResNet-50 [43]) were used for feature extraction, while classification was carried out by replacing the classification head with a new fully connected layer. Among them, the best classification accuracy was achieved by VGG19, which was trained on a full set of 100,000 images and tested with an external test set of more than 7000 images, while the least accurate model was SqueezeNet, with a classification accuracy of less than 50%.

A study in [44] segmented and filtered the background area of tumors by using Otsu's thresholding [45] and labeled the tumor area with a self-developed annotation tool before passing it to the CNN model for feature extraction and classification. They built a new model with the combination of DeepLab v2 [46], and ResNet-34 [43] and compared the model's performance with analyses of experienced pathologists. They found that their DL model for the diagnosis of adenoma in CRC was quite similar to the results from the pathologists, where a slide-level accuracy of over 90% and an AUC of 0.92 were obtained. Choi et al. [47] used an approach similar to the one in [44], where data preprocessing was carried out by discarding the unnecessary black regions in the endoscopic image samples via filtering. A transfer learning approach was used where the pretrained weights of various DL models, such as Inception-v3 [48], ResNet-50 [43], and DenseNet-161 [49], were used with 10-fold cross-validation. They evaluated their performance in terms of accuracy, recall, and precision, where they obtained respective values of 92.48%, 99.7%, and 99.2%. Similar studies [50–52] based on the tumor tissue detection type are listed in Table 3.

Table 3. Recently published articles on CRC tumor tissue detection and classification using DL techniques.

Publication (Year)	Objective	Method Used	Image Types	Optimum Results
[50] (2020)	To develop a biomarker of the patient outcome by analyzing the scanned H&E-stained slides with DL models	Data preprocessing: Multiple non-overlapping image tiles were selected with $10 \times$ and $40 \times$ magnification of WSI images. Model: Feature extraction and classification used a DoMore v1 network comprising a MobileNetV2 [53] representation network, a noisy-AND pooling function, and an FCL.	Hematoxylin and eosin-stained slides	Hazard ratio = 3.04 Sensitivity = 52% Accuracy = 76% Specificity = 78% PPV = 19% NPV = 94%
[52] (2021)	To develop a DL system for predicting disease-specific survival for stage II and III CRC patients	Data preprocessing: Constructing a tumor probability heat maps with $20 \times$ magnification of original slides, they generated binary ROI masks for each tumor probability output, applying denoising and dilation with a circular filter. Model: A CNN with depth-wise separable convolution layers such as MobileNet [54], plus hyperparameter tuning via random grid search.	Hematoxylin and eosin-stained slides	AUC = 0.70

Table 3. Cont.

Publication (Year)	Objective	Method Used	Image Types	Optimum Results
[54] (2021)	To develop a CNN-based CAD system for predicting the pathological histology of colorectal adenomas	Data preprocessing: Standardized input images excluded unnecessary black areas by cropping them to 480×480 pixels. Model: Multiple CNN models (ResNet-50 [43], Inception-v3 [48], and DenseNet-161 [49]) with a new classification head consisting of a single FCL.	Endoscopic images	Sensitivity = 77.25% Specificity = 92.42% PPV = 77.16% NPV = 92.58%

3.3. MSI Detection

Microsatellites, which are also known as short tandem repeats (STRs), are tiny repeating stretches of DNA that are scattered across the entire genome region, accounting for approximately 3% of the whole region [55]. The MSI phenotype is one of the molecular changes that occurs in CRC, and it is also observed in different types of cancer, such as adrenocortical, rectal, colon, stomach, and endometrial tumors, and breast and prostate cancer [56]. MSI can also be referred to as a hyper-mutable phenotype that is an outcome of deficient mismatch repair (dMMR). In Figure 4, we can see MSI patches, indicated by yellow arrows, that show activations around the potential patterns of infiltrating immune cells. The identification of MSI status in CRC patients is crucial, because it helps to determine the presence of related diseases such as lynch syndrome, a highly penetrant hereditary cancer syndrome accounting for one-third of patients with MSI. Therefore, a less labor-intensive and broadly accessible MSI testing tool based on DL approaches has been studied lately. These CAD systems contribute an automated screening tool to triage patients when making clinical decisions, so as to identify differential treatment responses.

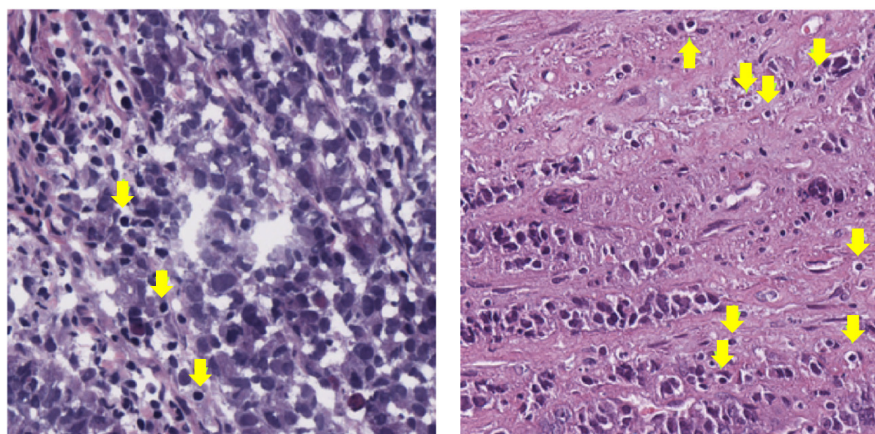


Figure 4. MSI cells in H&E-stained image patches (indicated by yellow arrows).

The authors in [57] introduced adversarial MSI-based assessment (AMIBA), a modality to diagnose microsatellite instability directly from histopathological images. Histological image data with a clinically determined MSI status (MSI-H, MSI-L, and MSS) were obtained from TCGA available at <https://portal.gdc.cancer.gov/> (accessed on 21 July 2021), where high (H), low (L), and microsatellite stable (MSS) each represents highest, lowest, and no presence of microsatellites. For data preprocessing, the image slides were clipped into non-overlapping image patches of 1000×1000 pixels, which were obtained at $20 \times$ magnification from the original slide. Furthermore, patches with more than half of the area empty were not included in the process, which generated a total of 620,833 patches to train the specific DL architecture. Multiple state-of-the-art DL architectures were used,

including ResNet-18 [43], AlexNet [40], and VGG-19 [34], where weights were initialized with parameters pretrained on an ImageNet dataset [58]. By using the Adam optimization algorithm with a learning rate of 0.0001, the authors obtained patch-level and slide-level accuracies of 91.7% and 98.3%, respectively.

Considering ways to facilitate universal MSI screening, the research in [59] studied how deep residual learning can predict the MSI status directly from H&E-stained histological slides. Multiple datasets regarding MSI status were collected from large patient cohorts in TCGA, which were manually annotated and classified to represent one tumor tissue and two non-tumor tissues (dense and loose tissue). The image slides were preprocessed to create 11,977 unique image tiles, each with a 256 μm edge length. Furthermore, to convert all the images into a reference color, a color normalization technique based on the Macenko method was used. The authors conducted initial experiments with multiple convolutional layers, from which ResNet-18 was selected as the optimum model due to its noteworthy advantages, such as a short training time, better classification performance, less risk of overfitting, and comparatively fewer training parameters. All models were trained on an ImageNet dataset, and only the weights of the last 10 layers were fine-tuned, while the rest of them were frozen. By using the Adam optimizer [60] and L2 regularization with multiple learning rates $\{10^{-5}, 10^{-6}\}$, they obtained an area under the curve (AUC) of 0.99. The CI for both true MSI and MSS tiles was found to be 95%.

Another study conducted in 2020 by Lee et al. [61] developed a two-stage DL-based classification pipeline for predicting MSI status in CRC patients. In the two-stage process, the first stage was responsible for segmenting the tumor area into two types of tissue (MSI-H and MSI-L). The latter stage then classified the tissue types into their corresponding class. H&E-stained histological WSIs annotated by professional pathologists were obtained from a pathology AI platform (PAIP) at <http://wisepaip.org/paip> (accessed on 21 July 2021) and were preprocessed before being used as input to the DL architecture. During preprocessing, the WSIs were cropped to magnifications of $20 \times$ and $10 \times$ to obtain image patches of 224×224 pixels before converting the RGB images to the CIE $L \times a \times b$ color space. Other preprocessing methods, such as foreground mask extraction with Otsu's thresholding, were used to segment individual patches. Two DL models were adapted in this research: the feature pyramid network (FPN) [62] and inception ResNet-V2 [43], one for each stage in the classification pipeline. Multiple optimization algorithms such as Adam and RMS prop were used, with each being trained on one of two learning rate schedulers (step decay and cosine annealing) with a learning rate of 10^{-4} . The optimum precision, recall, and F1-score were found to be 0.93, 0.93, and 0.94, respectively.

Similarly, to develop a DL system for detecting CRC tumor specimens with MSI, Echle et al. [63] collected H&E-stained slides from 8836 CRC tumors from the MSI-DETECT consortium (<https://jnkather.github.io/msidetect/> (accessed on 22 July 2021)). All specimens belonged to a large cohort of patients from Germany, the Netherlands, the United Kingdom, and the United States, where each specimen with MSI was identified via genetic analysis. These data were preprocessed by tessellating the slides into individual square tiles of 256 μm edge lengths followed by color normalization with the Macenko method before passing them to a ShuffleNet model [64] for classification. The whole model was trained on Nvidia RTX6000 graphical processing unit (GPU) hardware with the Adam optimizer, L2 regularization, and a learning rate of 5×10^{-5} . The classification results were evaluated based on several performance metrics, where the optimal values of area under the receiver operating characteristics (AUROC), area under the precision recall curve (AUPRC), sensitivity, and specificity were recorded as 0.96, 0.9, 99%, and 98%, respectively.

Recently, several other DL-based studies for MSI detection and/or classification were conducted, which are listed in Table 4.

Table 4. Some recently published articles on MSI detection in CRC by using DL techniques.

Publication (Year)	Objective	Method Used	Image Types	Optimum Results
[65] (2020)	To predict RNA-Seq profiles from WSI without expert annotation and to identify tumors containing MSI	Data preprocessing: Divided WSI into squares tiles of (224 × 224 pixels); segmented white background using Otsu's thresholding. Model: Multi-layer perceptron applied to every tile with 5-fold cross-validation.	Histology images	AUC-ROC = 0.83
[66] (2020)	To investigate the potential of DL-based CAD systems for automated prediction of MSI from H&E-stained WSIs	Data preprocessing: Discarded the non-tissue containing white background by thresholding; partitioned the WSI into non-overlapping tiles at 256 × 256 pixels; color normalized with Macenko method. Model: Based on MobileNetV2 architecture (pretrained on an ImageNet dataset) with two sequential components: tissue type classifier, and MSI classifier, 4-fold cross-validation.	Hematoxylin and eosin-stained slides	Sensitivity = 43.1% Specificity = 94.9% NPV = 89.9% AUROC = 0.964 CI = 95%
[67] (2020)	To develop an ensemble multiple-instance DL model for predicting MSI status based on histopathology images	Data preprocessing: WSI subsequently tiled into patches of 512 × 512 pixels; ROI containing carcinoma in the WSIs manually annotated by experienced pathologists; tumor cells occupying less than 80% of ROI discarded; interfering factors such as creases, bleeding, necrosis, and blurred areas excluded; applied data augmentation and image normalization. Model: CNN model based on ResNet-18 with binary cross entropy loss function; traditional ML-based classifiers such as gradient boosting and naïve Bayes used.	Hematoxylin and eosin-stained slides	ROCAUC = 0.8848 CI = 95% <i>p</i> -value < 0.001

3.4. Polyp Detection

The beginning phase of most CRCs is stimulated with a growth of tissues on the inner lining of the colon. These abnormal growths of tissue from the mucous membrane, developing over a period, are called polyps, and are often considered a precursor to CRC. Figure 5a–c, respectively, show a polyp image from a DL-based CAD system extracted from a CRC patient, the corresponding annotations, and the detected polyps. Colonic polyps, especially with a large size and in large numbers, are more likely to be cancerous, and if not treated early, they could develop into colon cancer. CRC polyps can be categorized as neoplastic and non-neoplastic. The former is non-cancerous, while the latter can develop into cancer and can be further sub-categorized into adenomas and serrated polyps. In clinical practice, the detection of polyps is usually accomplished via colonoscopy, which is an expensive, manual, and time-consuming procedure. Frequent reviews of colonoscopy data are required, because 20% of polyps are likely to be missed during a single review. This is extremely labor intensive, and the lack of a thorough inspection of the data might result in the missed detection of polyps. Taking this into account, an automated and non-invasive

procedure based on CAD has been reliable and is considered more robust for accurate detection of polyps. Specifically, DL-based segmentation and classification algorithms have been indispensably applied recently to enable the routine detection of polyps in CRC diagnosis. In this regard, in [68], multiple studies related to colon cancer analysis were collected under the field of colon cancer and deep learning; then, they were categorized into five categories that are detection, classification, segmentation, survival prediction, and inflammatory bowel diseases. In [69], the current systematic review on colorectal cancer detection and localization, and difficulties of a fair comparison and the reproducibility of those methods were addressed.

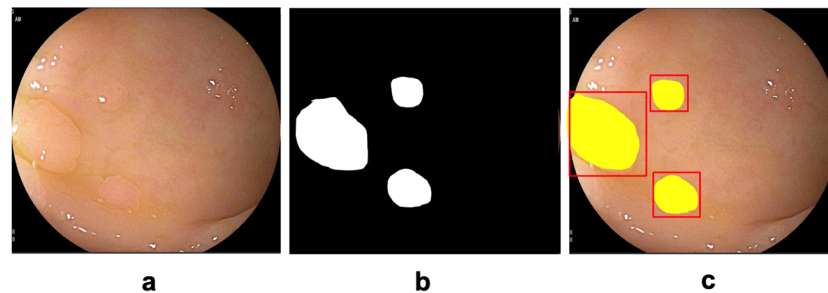


Figure 5. Polyp image samples. (a) original image, (b) corresponding annotation, and (c) detected polyps in the yellow regions.

A conference paper was published by Godkhindi et al. [70] in 2017 with the objective being the automatic detection of polyps via CT colonography using DL techniques. To achieve this, a CT colonography image dataset was collected from the Cancer Imaging Archive (TCIA), available at <https://www.cancerimagingarchive.net/> (accessed on 23 July 2021), containing ground truth information for segmentation purposes. In the data preprocessing steps, the authors discarded the air gap-filled regions in the colon images by using thresholding and filtering techniques. Furthermore, to label each block in the image, binary region of interest (ROI) segmentation was performed. A CNN model with three convolutional layers, three max pooling layers, and a single FCL was designed and trained using 10-fold cross-validation, which obtained classification accuracy, sensitivity, and specificity levels of 88.56%, 88.77%, and 87.35%, respectively.

Another similar study [71] applied state-of-the-art DL algorithms to each colonoscopy frame of a gastrointestinal image analysis (GIANA) (dataset available at <https://giana.grand-challenge.org/> (accessed on 23 July 2021)) that consisted of 18 videos collected from endoscopic results of multiple patients. For data preprocessing, the black edges of the endoscopy image frames were removed, and the images were resized to model-specific dimensions of 284×265 pixels before being passed for data augmentation with horizontal and vertical flips and a blur filter. This approach used ResNet-50 as a fully convolutional neural network (FCNN) to extract descriptive characteristics from the input image. The extracted features were then subjected to a faster RCNN [72] model with two FCLs, each operating as a regression and a classification layer. After extensive experiments and evaluations, their model achieved a precision value of 80.31%, recall of 75.37%, accuracy of 71.99%, and specificity of 65.70%.

Different levels of diagnostic accuracy can be observed by adapting different strategies for data preprocessing and by using DL algorithms. The optimal solution for medical image diagnostics is, however, not obtained through minimal trials and testing, but is considered via continuous and long-running research. In this context, several studies tried to overcome the shortcomings in previous research or developed a completely novel scheme for CRC diagnosis. Considering this, to enhance detection accuracy obtained by reference studies, Lee et al. [73] developed and validated a robust DL algorithm for use in the detection of colorectal polyps. In that study, the authors collected endoscopy data samples from the Asan Medical Center, Korea, between May 2017 and February 2018. The whole training dataset contained 8075 images from 185 colonoscopy videos of 103 patients.

For validation and testing, different sets of data samples were collected from the same place. These datasets were preprocessed by storing them at a fixed resolution of 475×420 pixels before labeling the location and dimension of each polyp in the image with bounding boxes. This study used a one-shot classification model of every object present in the image using YOLO v2 [74] without using an attention mechanism. The classification model was a fine-tuned Darknet19 model provided at <https://pjreddie.com/darknet/> (accessed on 26 July 2021), which was pretrained on an ImageNet dataset. By creating B bounding boxes with a confidence score for the class probability of each box, the model was able to secure a sensitivity level of 96.7% with a false positive rate (FPR) of 6.3%.

Similarly, in 2020, Poudel et al. [75] developed a classification model for use in the identification of adenomas, Crohn's disease, ulcerative colitis, and normal images by using endoscopic image samples from CRC patients. They used two datasets: the first was provided by Gill Hospital in Korea with a total of 3515 images, and the second was a publicly available KVASIR dataset [19] with 4000 endoscopy samples. Each dataset was normalized to model-specific inputs and was subjected to augmentation, which included flipping, scaling, rotating, zoom, contrast normalization, and shearing. A transfer learning approach was used with a ResNet-50 architecture as a baseline model, initialized with pretrained weights from ImageNet. An efficient dilation technique [76] was adopted to preserve the spatial information of the final layers in the network by using dilated convolution layers in ascending and descending order. The original ResNet-50 model was also modified by using DropBlock regularization [77] at deeper layers to make it robust towards noise and artifacts. With extensive experiments using both datasets, the optimal values for precision, recall, and F1-score were found to be 0.932, 0.928, and 0.93, respectively.

Another study in [78] created an endoscopic dataset from different sources and annotated the ground truths by collaborating with experienced gastroenterologists. Due to the severe differences in the existing datasets in terms of image resolution and color temperature (possibly due to different imaging equipment setups), the authors built a new dataset to serve as a benchmark to train and evaluate the DL models for polyp detection and classification. The new dataset included multiple publicly available endoscopic datasets as well as some independently collected from the University of Kansas Medical Center. Due to the extreme imbalance among the total number of image frames in each dataset, an adaptive sampling rate was utilized to homogenize the representativeness of each polyp by extracting important frames from the video. In total, 116 training, 17 validations, and 22 testing sets were generated, each comprising of 28,773, 4254, and 4872 frames, respectively. By using these datasets, eight of the most popular state-of-the-art object detection models were evaluated. These included Faster RCNN [72], YOLOv3 [79], SSD [80], RetinaNet [81], DetNet [82], RefineDet [83], YOLOv4 [84], and ATSS [85]. Using these frameworks, three different types of experiments were conducted: first, frame-based one-class polyps detection, second, two-class polyps detection, and the third, sequence-based two class polyp detection. For the two frame-based detection experiments, the performance was measured by regular object detection metrics, while for the sequence-based detection, regular object detection was applied to each frame. Finally, the voting procedure was applied to pick the mostly predicted polyps. For both the frame-based and sequence-based detection methods, RefineDet performed very well with an F1 score of 88.6, and 86.3, respectively. Other similar studies published recently on polyp detection using endoscopy images samples are listed in Table 5.

Table 5. Some recently published articles on polyp detection in CRC by using DL techniques.

Publication (Year)	Objective	Method Used	Image Types	Optimum Results
[86] (2021)	To detect polyps in real time through localization and segmentation approaches	<p>Data preprocessing: Normalized images by subtracting mean of the image by its standard deviation; annotated with bounding boxes; resized to fixed 512×512 pixels; applied data augmentation such as horizontal, vertical flipping, random rotation, random scale, and random cropping.</p> <p>Model: Used multiple models: feature extraction using EfficientDet [87] (EfficientNet [88] as a backbone architecture) with a bidirectional feature pyramid network (FPN), and a shared class/box prediction network; Faster R-CNN as detector network with region proposal network (RPN) as the proposal network, YOLOv3 [79] with multi-class logistic loss modeled with regularizers such as objectness prediction scores, and YOLOv4 [84] with on-the-fly data augmentation such as mosaic, and cut-mix.</p>	Endoscopy images	<p>Average precision = 0.8000 Mean intersection over union (IoU) = 0.8100 Detection speed = 180 frames per second (fps), dice coefficient = 0.8206, segmentation average speed = 182.36 fps.</p>
[89] (2021)	To use a DL approach to accomplish multitasks, such as colorectal image classification and polyp image segmentation	<p>Data preprocessing: Removed unclear and unclear colorectal images with data filtering; labeled images into one of three classes: normal tissue, polyp, and tumor; split the dataset into train, validation, and testing set with the ration of 2:1:2; applied data augmentation; resized to $440 \times 440 \times 3$ to maintain dimensional uniformity.</p> <p>Model: Used self-paced regularization method to assign different sample weights for different training samples; VGG19 pretrained on ImageNet dataset for feature extraction; replaced FCL with GAP layer for classification of colorectal images; used U-Net-based [90] automatic polyp-region segmentation.</p>	Endoscopy images	Accuracy = 96.0%

4. Discussion

Multiple DL algorithms discussed in aforementioned sections have achieved highly reliable results in accurately detecting different types of tumors, MSI cells, and colorectal polyps. The evaluations of these models have been conducted through several validation tests and are designed to perform domain specific tasks such as segmenting tumorous from non-tumorous tissue or classifying cancerous cells from the healthy ones. For tumor tissue classification tasks, the EfficientNet [88] model was shown to display superior performance, while U-Net [90], and YOLO architectures [74,79,84] showed high precision in solving polyp segmentation and detection tasks, respectively. Using DL methods, the clinical inspection of CRC-related patients is quickly performed with high diagnostic accuracy. With that being said, the accuracy of DL algorithms can vary significantly and is dependent upon the amount of data with which the model is trained. Especially in the medical imaging sector, the availability of a publicly usable, large-scale standard dataset for

conducting experiments is considered rare, relative to other fields, for example, in natural images (ImageNet). In most scenarios, to complement the scarcity of vast amounts of data, techniques such as data augmentation are widely practiced. Ranging from traditional augmentation techniques such as flipping, shifting, and rotation, novel techniques such as generative adversarial networks (GANs) [91] and style transfer techniques [92] have also been extensively used to create and add synthetic instances to increase the data samples, guaranteeing higher efficiency in the DL models. In order to address the complications created by limited data, other techniques, such as transfer learning, can mitigate the model's dependency on training data sample size by using pretrained weights of other large-scale datasets to initialize the model hyperparameters. Because real-world medical imaging data are hard to acquire, data augmentation and synthetic imaging techniques can be helpful to enhance the accuracy of DL models in diagnosing CRC. Similarly, a clear difference in the accuracy of the DL model can be perceived, depending upon the use of preprocessed and non-preprocessed datasets to train any model. For a quantitative evaluation of the DL model, data preprocessing such as ROI extraction, color normalization, thresholding, etc., must be incorporated. In addition, data cleaning/preprocessing eliminates a portion of low-quality data or outliers, such as image pairs with suboptimal registration.

Apart from that, image annotation such as tumor tissue labeling in CRC is considered a highly sophisticated and time-consuming task, and thus, there is a need for highly skilled and experienced pathologists to prepare high-quality datasets for training and testing. Not only limited to image annotations, the requirement for medical associates and pathologists is also extremely important in order to consider the values and preferences of the patients, the medical judgments, the interventional procedures, policy making, and other tasks that cannot be accomplished by computer programs alone. Therefore, the need for pathologists remains essential for medical practices diagnosing not only CRC but also other cancer variants.

Current DL models exist in various forms and architectures, and frequent optimizations of those models are being released to ensure highly accurate results in CRC diagnosis. However, only an abundant number of experiments and user-based experiences can guarantee the reliability of those models for clinical purposes. Therefore, it is necessary to apply several DL algorithms to identify and detect each type of CRC malignancy and compare them to find the optimal diagnosis procedure. Moreover, improving the existing theoretical foundation of the DL on the basis of the type of experimental data must be taken into consideration to quantify the performance of multiple DL-based CRC detection modalities. Such improvements must address the data-specific assessment of any algorithm, its computational complexity, and the hyperparameter tuning strategies [93]. The fact that currently incorporated models can be biased towards non-CRC datasets cannot be overlooked, and thus, specific criteria should be validated for CRC-specific DL models in order to obtain intuitive insights into their optimization characteristics and certainties. CRC diagnosis and prognosis with DL technologies are almost ready to be commercialized for practical use cases in clinical settings. By exploring several other opportunities regarding data preparation and model architectures, there is still room for improvement in the accuracy of those models that are still in the suboptimal phase.

5. Conclusions

DL has expanded rapidly over the past few years in the field of oncology, especially for screening and the diagnosis of CRC-related diseases. Putting this into perspective, in this paper, we reviewed the publicly available CRC imaging datasets and recently published research works that focused on detecting different types of CRC, including tumor detection, MSI detection, and polyp detection. Furthermore, we also outlined some issues regarding the scarcity of data and preprocessing strategies and provided insights into developing problem-specific DL architectures to diagnose CRC patients in real time to enable their commercialization for clinical practice. Through extensive research, and development of medical application-oriented DL models, and by collaborating with

experienced pathologists in collecting high-quality annotated datasets, we believe that the reliable and automated screening of one of the most fatal cancer subtypes will be possible in the near future.

Author Contributions: Conceptualization, L.D.T. and B.W.K.; methodology, L.D.T. and B.W.K.; formal analysis, L.D.T. and B.W.K.; investigation, L.D.T.; writing—original draft preparation, L.D.T. and B.W.K.; writing—review and editing, L.D.T. and B.W.K.; supervision, B.W.K.; project administration, B.W.K.; funding acquisition, B.W.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the project for Industry-Academic Cooperation Based Platform R&D funded Korea Ministry of SMEs and Startups in 2020 (S3014213), and the National Research Foundation of Korea (NRF) grant funded by the Korean government (NRF-2019R1A2C4069822).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegal, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef] [PubMed]
2. Ballinger, A.B.; Anggiansah, C. Colorectal Cancer. *BMJ* **2007**, *335*, 715–718. [CrossRef]
3. Arnold, M.; Sierra, M.S.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Patterns and Trends in Colorectal Cancer Incidence and Mortality. *GUT* **2017**, *66*, 683–691. [CrossRef]
4. Geboes, K.; Mourin, A.J. Endoscopy and Histopathology. Available online: <https://www.intechopen.com/chapters/44215> (accessed on 8 July 2021).
5. Rabe, K.; Snir, O.L.; Bossuyt, V.; Harigopal, M.; Celli, R.; Reisenbichler, E.S. Interobserver Variability in Breast Carcinoma Grading Results in Prognostic Stage Differences. *Hum. Pathol.* **2019**, *94*, 51–57. [CrossRef]
6. Gross, S.; Trautwein, C.; Behrens, A.; Winograd, R.; Palm, S.; Lutz, H.H.; Sokhan, R.S.; Hecker, H.; Aach, T.; Tischendorf, J.J.W. Computer-based Classification of Small Colorectal Polyps by Using Narrow-band Imaging with Optical Magnification. *Gastro. Endosc.* **2011**, *74*, 1354–1362. [CrossRef]
7. Mori, Y.; Kudo, S.E.; Wakamura, K.; Misawa, M.; Ogawa, Y.; Kutsukawa, M.; Kudo, T.; Hayashi, T.; Miyachi, H.; Ishida, F.; et al. Novel Computer-aided Diagnostic System for Colorectal Lesions by Using Endocytoscopy (with videos). *Gastro. Endosc.* **2014**, *81*, 621–629. [CrossRef] [PubMed]
8. Tamai, N.; Saito, Y.; Sakamoto, T.; Nakajima, T.; Matsuda, T.; Sumiyama, K.; Tajiri, H.; Koyama, R.; Kido, S. Effectiveness of computer-aided diagnosis of colorectal lesions using novel software for magnifying narrow-band imaging: A pilot study. *Endosc. Int. Open* **2017**, *5*, E690–E694. [CrossRef] [PubMed]
9. Tamaki, T.; Yoshimuta, J.; Kawakami, M.; Raytchev, B.; Kaneda, K.; Yoshida, S.; Takemura, Y.; Onji, K.; Miyako, R.; Tanaka, S. Computer-aided colorectal tumor classification in NBI endoscopy using local features. *Med. Image Anal.* **2013**, *17*, 78–100. [CrossRef]
10. Kominami, Y.; Yoshida, S.; Tanaka, S.; Sanomura, Y.; Hirakawa, T.; Raytchev, B.; Tamaki, T.; Koide, T.; Kaneda, K.; Chayama, K. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. *Gastro. Endosc.* **2016**, *83*, 643–651. [CrossRef] [PubMed]
11. Swager, A.F.; Sommen, F.V.D.; Klomp, S.R.; Zinger, S.; Meijer, S.L.; Schoon, E.J.; Bergman, J.J.G.H.M.; With, P.H.; Curvers, W.L. Computer-aided detection of early Barrett’s neoplasia using volumetric laser endomicroscopy. *Gastro. Endosc.* **2017**, *86*, 839–846. [CrossRef]
12. Min, M.; Su, S.; He, W.; Bi, Y.; Ma, Z.; Liu, Y. Computer-aided diagnosis of colorectal polyps using linked color imaging colonoscopy to predict histology. *Sci. Rep.* **2019**, *9*, 2881. [CrossRef] [PubMed]
13. Linder, N.; Konsti, J.; Turkki, R.; Rahtu, E.; Lundin, M.; Nordling, S.; Haglund, C.; Ahonen, T.; Pietikainen, M.; Lundin, J. Identification of tumor epithelium and stroma in tissue microarrays using texture analysis. *Diagno. Pathol.* **2012**, *7*, 22. [CrossRef]
14. Gonzalez, R.C. Deep convolutional neural networks [Lecture Notes]. *IEEE Sig. Proc. Mag.* **2018**, *35*, 79–87. [CrossRef]
15. Tamang, L.D.; Kim, B.W. Deep D2C-Net: Deep learning-based display-to-camera communications. *Opt. Express* **2021**, *29*, 11494–11511. [CrossRef]
16. Fang, L.; Monroe, F.; Novak, S.W.; Kirk, L.; Schiavon, C.R.; Yu, S.B.; Zhang, T.; Wu, M.; Kastner, K.; Latif, A.A.; et al. Deep learning-based point scanning super resolution imaging. *Nat. Methods* **2021**, *18*, 406–416. [CrossRef]
17. Kohli, A.D.; Summers, R.M.; Geis, J.R. Medical imaging data and datasets in the era of machine learning. *J. Dig. Imaging* **2017**, *30*, 392–399. [CrossRef]
18. Gurcan, M.N.; Boucheron, L.; Can, A.; Madabhushi, A.; Rajpoot, N.; Yener, B. Histopathological image analysis: A review. *IEEE Rev. Biomed. Eng.* **2010**, *2*, 147–171. [CrossRef]

19. Pogorelov, K.; Randel, K.R.; Griwodz, C.; Eskeland, S.L.; Lange, T.; Johansen, D.; Spampinato, C.; Nguyen, D.T.D.; Lux, M.; Schmidt, P.T.; et al. Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In Proceedings of the 8th ACM on Multimedia Systems Conference (MMSYS), Taipei, Taiwan, 20–23 June 2017; pp. 164–169.
20. Pogorelov, K.; Randel, K.R.; Griwodz, C.; Eskeland, S.L.; Lange, T.; Johansen, D.; Spampinato, C.; Nguyen, D.T.D.; Lux, M.; Schmidt, P.T.; et al. Nerthus: A bowel preparation quality video dataset. In Proceedings of the 8th ACM on Multimedia Systems Conference, Taipei, Taiwan, 20–23 June 2017; pp. 170–174.
21. Borgli, H.; Thambawita, V.; Smedsrud, P.H.; Hicks, S.; Jha, D.; Eskeland, S.L.; Randel, K.R.; Pogorelov, K.; Lux, M.; Nguyen, D.T.D. *HyperKvasir*, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Sci. Data* **2020**, *7*, 28. [[CrossRef](#)]
22. Sanchez, F.Z.; Bernal, J.; Montes, C.S.; Miguel, C.R.; Esparrach, G.F. Bright spot regions segmentation and classification for specular highlights detection in colonoscopy videos. *Mach. Vis. Appl.* **2017**, *28*, 917–936. [[CrossRef](#)]
23. Bernal, J.; Sanchez, F.J.; Esparrach, G.F.; Gil, D.; Rodriguez, C.; Vilarino, F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Comput. Med. Imaging Graph.* **2015**, *43*, 99–111. [[CrossRef](#)] [[PubMed](#)]
24. Ali, S.; Ghatwary, N.; Braden, B.; Lamarque, D.; Bailey, A.; Realdon, S.; Cannizzaro, R.; Rittscher, J.; Daul, C.D.; East, J. Endoscopy disease detection challenge. *arXiv* **2003**, arXiv:2003.03376.
25. Kather, J.K.; Weis, C.A.; Bianconi, F.; Melchers, S.M.; Schad, L.R.; Gaiser, T.; Marx, A.; Zollner, F.G. Multi-class texture analysis in colorectal cancer histology. *Sci. Rep.* **2016**, *6*, 27988. [[CrossRef](#)] [[PubMed](#)]
26. Graham, S.; Chen, H.; Gamper, J.; Dou, J.; Heng, P.; Snead, D.; Tsang, Y.W.; Rajpoot, N. MILD-net: Minimal information loss dilated network for gland instance segmentation in colon histology images. *Med. Image Anal.* **2019**, *52*, 199–211. [[CrossRef](#)] [[PubMed](#)]
27. Graham, S.; Vu, Q.D.; Raza, S.E.A.; Azam, A.; Tsang, Y.W.; Kwak, J.T.; Rajpoot, N. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *arXiv* **2019**, arXiv:1812.06499. [[CrossRef](#)]
28. Shaban, M.; Awan, R.; Fraz, M.M.; Azam, A.; Snead, D.; Rajpoot, A.M. Context aware convolutional neural network for grading of colorectal cancer histology images. *IEEE Trans. Med. Imaging* **2020**, *39*, 2395–2405. [[CrossRef](#)] [[PubMed](#)]
29. Kim, Y.J.; Jang, H.; Lee, K.; Park, S.; Min, S.G.; Hong, C.; Park, J.H.; Lee, K.; Kim, J.; Hong, W.; et al. PAIP 2019: Liver cancer segmentation challenge. *Med. Image Anal.* **2021**, *67*, 101854. [[CrossRef](#)]
30. Sirinukunwattana, K.; Snead, D.R.J.; Rajpoot, N.M. A Stochastic Polygons Model for Glandular Structures in Colon Histology Images. *IEEE Trans. Med. Imaging* **2015**, *34*, 2366–2378. [[CrossRef](#)]
31. Barbano, C.A.; Perlo, D.; Tartaglione, E.; Fiandrotti, A.; Bertero, L.; Cassoni, P.; Grangetto, M. UniToPatho, a labeled histopathological dataset for colorectal polyps classification and adenoma dysplasia grading. *arXiv* **2021**, arXiv:2101.09991.
32. Ponzio, F.; Macii, E.; Ficarram, E.; Cataldo, S.D. Colorectal cancer classification using deep convolutional networks. In Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies-BIOIMAGING, Madeira, Portugal, 19–21 January 2018; pp. 58–66.
33. Bychkov, D.; Linder, N.; Turkki, R.; Nordling, S.; Kovanen, P.E.; Verill, C.; Walliander, M.; Lundin, M.; Haglund, C.; Lundin, J. Deep learning-based tissue analysis predicts outcome in colorectal cancer. *Sci. Rep.* **2018**, *8*, 3395. [[CrossRef](#)]
34. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
35. Pulver, A.; Lyu, S. LSTM with working memory. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 845–851.
36. Yuw, X.; Dimitriou, N.; Arandjelovic, O. Colorectal cancer outcome prediction from H&E whole slide images using machine learning and automatically inferred phenotype profiles. *arXiv* **2019**, arXiv:1902.03582.
37. Jogin, M.; Madhulika, M.S.; Divya, G.D.; Meghana, R.K.; Apoorva, S. Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning. In Proceedings of the 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 18–19 May 2018; pp. 2319–2323.
38. Kather, J.N.; Krisam, J.; Charoentong, P.; Luedde, T.; Herpel, E.; Weis, C.A.; Gaiser, T.; Marx, A.; Valous, N.A.; Ferber, D.; et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS Med.* **2019**, *16*, e1002730. [[CrossRef](#)]
39. Macenko, M.; Niethammer, M.; Marron, J.S.; Borland, D.; Woosely, J.T.; Guan, X.; Schmitt, C.; Thomas, N.E. A method for normalizing histology slides for quantitative analysis. In Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, MA, USA, 28 June–1 July 2009; pp. 1107–1110.
40. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
41. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and <0.5 MB model size. *arXiv* **2021**, arXiv:1602.07360.
42. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
43. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

44. Song, Z.; Yu, C.; Zou, S.; Wang, W.; Huang, Y.; Ding, X.; Liu, J.; Shao, L.; Yuan, J.; Gou, X.; et al. Automatic deep learning-based colorectal adenoma detection system and its similarities with pathologists. *BMJ Open* **2020**, *10*, e036423. [[CrossRef](#)]
45. Liu, D.; Yu, J. Otsu Method and K-means. In Proceedings of the Ninth International Conference on Hybrid Intelligent Systems, Shenyang, China, 12–14 August 2009; pp. 344–349.
46. Chen, L.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)]
47. Choi, K.; Choi, S.J.; Kim, E.S. Computer-Aided Diagnosis for Colorectal Cancer using Deep Learning with Visual Explanations. In Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, Canada, 20–24 July 2020; pp. 1156–1159.
48. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
49. Huang, G.; Liu, Z.; Maaten, L.V.D.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
50. Skrede, O.J.; Raedt, S.D.; Kleppe, A.; Hveem, T.S.; Liestol, K.; Maddison, J.; Askautrud, H.A.; Pradhan, M.; Nesheim, J.A.; Albrechtsen, A. Deep learning for prediction of colorectal cancer outcome: A discovery and validation study. *Lancet* **2020**, *395*, 350–360. [[CrossRef](#)]
51. Wulczyn, E.; Steiner, D.F.; Moran, M.; Plass, M.; Reihls, R.; Tan, F.; Flament-Auvigne, I.; Brown, T.; Regitnig, P.; Chen, P.C.; et al. Interpretable survival prediction for colorectal cancer using deep learning. *NPJ Digit. Med.* **2021**, *4*, 71. [[CrossRef](#)] [[PubMed](#)]
52. Choi, S.J.; Kim, E.S.; Choi, K. Prediction of the histology of colorectal neoplasm in white light colonoscopic images using deep learning algorithms. *Sci. Rep.* **2021**, *11*, 5311. [[CrossRef](#)] [[PubMed](#)]
53. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
54. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
55. Nojadeh, J.N.; Sharif, S.B.; Sakhinia, E. Microsatellite instability in colorectal cancer. *EXCLI J.* **2018**, *17*, 159–168. [[CrossRef](#)]
56. Li, K.; Luo, H.; Huang, L.; Luo, H.; Zhu, X. Microsatellite instability: A review of what the oncologist should know. *Cancer Cell International*. **2020**, *20*, 16. [[CrossRef](#)]
57. Zhang, W.; Yin, H.; Huang, Z.; Zhao, J.; Zheng, H.; He, D.; Li, M.; Tan, W.; Tian, S.; Song, B. Development and validation of MRI-based deep learning models for prediction of microsatellite instability in rectal cancer. *Cancer Med.* **2021**, *10*, 4164–4173. [[CrossRef](#)] [[PubMed](#)]
58. Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; Li, F.F. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
59. Kather, J.N.; Pearson, A.T.; Halama, N.; Jager, D.; Krause, J.; Loosen, S.H.; Marx, A.; Boor, P.; Tacke, F.; Neumann, P.U.; et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* **2019**, *25*, 1054–1056. [[CrossRef](#)]
60. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
61. 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. *Appl. Sci.* **2020**, *11*, 254. [[CrossRef](#)]
62. Lin, T.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
63. Echle, A.; Grabsch, H.K.; Quirke, P.; Brandt, P.A.; West, N.P.; Hutchins, G.G.A.; Heij, L.R.; Tan, X.; Richman, S.D.; Krause, J.; et al. Clinical-Grade Detection of Microsatellite Instability in Colorectal Tumors by Deep Learning. *Gastroenterology* **2020**, *159*, 1406–1416. [[CrossRef](#)] [[PubMed](#)]
64. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856.
65. Schmauch, B.; Romagnoni, A.; Pronier, E.; Saillard, C.; Maille, P.; Calderaro, J.; Kamoun, A.; Sefta, M.; Toldo, S.; Zaslavskiy, M.; et al. A deep learning model to predict RNA-Seq expression of tumors from whole slide images. *Nat. Commun.* **2020**, *11*, 3877. [[CrossRef](#)] [[PubMed](#)]
66. Yamashita, R.; Long, J.; Longrace, T.; Peng, L.; Berry, G.; Martin, B.; Higgins, J.; Rubin, D.L.; Shen, J. Deep learning model for the prediction of microsatellite instability in colorectal cancer: A diagnostic study. *Lancet Oncol.* **2021**, *22*, 132–141. [[CrossRef](#)]
67. Cao, R.; Yang, F.; Ma, S.; Liu, L.; Zhao, Y.; Li, Y.; Wu, D.; Wang, T.; Lu, W.; Cai, W.; et al. Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in colorectal cancer. *Theranostics* **2020**, *10*, 11080–11091. [[CrossRef](#)]
68. Pacal, I.; Karaboga, D.; Basturk, A.; Akay, B.; Nalbantoglu, U. A comprehensive review of deep learning in colon cancer. *Comput. Biol. Med.* **2020**, *126*, 104003. [[CrossRef](#)]

69. Sanchez-Peralta, L.F.; Bote-Curiel, L.; Picon, A.; Sanchez-Margallo, F.M.; Pagador, J.B. Deep learning to find colorectal polyps in colonoscopy: A systematic literature review. *Artif. Intell. Med.* **2020**, *108*, 101923. [[CrossRef](#)] [[PubMed](#)]
70. Godkhindi, A.M.; Gowda, R.M. Automated detection of polyps in CT colonography images using deep learning algorithms in colon cancer diagnosis. In Proceedings of the International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India, 1–2 August 2017; pp. 1722–1728.
71. Duran-Lopez, L.; Luna-Perejon, F.; Amaya-Rodriguez, I.; Civit-Masot, J.; Civit-Balcells, A.; Vincente-Diaz, S.; Linares-Barranco, A. Polyp detection in gastrointestinal images using faster regional convolutional neural network. In Proceedings of the 14th International Joint Conference on Computer Vision, Imaging, and Computer Graphics Theory and Applications, Prague, Czech Republic, 25–27 February 2019; pp. 626–631.
72. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **2015**, arXiv:1506.01497. [[CrossRef](#)]
73. Lee, J.Y.; Jeong, J.; Song, E.M.; Ha, C.; Lee, H.J.; Koo, J.E.; Yang, D.H.; Kim, N.; Byeon, J. Real-time detection of colon polyps during colonoscopy using deep learning: Systematic validation with four independent datasets. *Sci. Rep.* **2020**, *10*, 8379. [[CrossRef](#)]
74. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525.
75. Poudel, S.; Kim, Y.J.; Vo, D.M.; Lee, S. Colorectal Disease Classification Using Efficiently Scaled Dilation in Convolutional Neural Network. *IEEE Access* **2020**, *8*, 99227–99238. [[CrossRef](#)]
76. Li, Y.; Zhang, X.; Chen, D. CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1091–1100.
77. Ghiasi, G.; Lin, T.Y.; Le, Q.V. DropBlock: A regularization method for convolutional networks. *arXiv* **2018**, arXiv:1810.12890.
78. Li, K.; Fathan, I.F.; Patel, K.; Zhang, T.; Zhong, C.; Bansal, A.; Rastogi, A.; Wang, J.S.; Wang, G. Colonoscopy polyp detection and classification: Dataset creation and comparative evaluations. *PLoS ONE* **2021**, *16*, e0255809. [[CrossRef](#)]
79. Redmon, J.; Farhadi, A. YOLOv3: AN incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
80. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; Berg, A.C. SSD: Single Shot MultiBox Detector. *arXiv* **2015**, arXiv:1512.02325.
81. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for object detection. *arXiv* **2017**, arXiv:1708.02002.
82. Li, Z.; Peng, C.; Yu, G.; Zhang, X.; Deng, Y.; Sun, J. DetNet: A backbone network for object detection. *arXiv* **2018**, arXiv:1804.06215v2.
83. Zhang, S.; Wen, L.; Bian, X.; Lei, Z.; Li, S.Z. Single shot refinement neural network for object detection. *arXiv* **2017**, arXiv:1711.06897.
84. Bochkovskiy, A.; Wang, C.; Liao, H.M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
85. Zhang, S.; Chi, C.; Yao, Y.; Lei, Z.; Li, S.Z. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. *arXiv* **2020**, arXiv:1912.02424.
86. Jha, D.; Ali, S.; Tomar, N.K.; Johansen, H.D.; Johansen, D.D.; Rittscher, J.; Riegler, M.A.; Halvorsen, P. Real time polyp detection, localization, and segmentation in colonoscopy using deep learning. *arXiv* **2020**, arXiv:2011.07631. [[CrossRef](#)]
87. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10778–10787.
88. Tan, M.; Le, Q.V. EfficientNet: Rethinking model scaling for convolutional neural networks. *arXiv* **2019**, arXiv:1905.11946.
89. Yao, Y.; Gou, S.; Tian, R.; Zhang, X.; He, S. Automated classification and segmentation in colorectal images based on self-placed transfer network. *BioMed. Res. Int.* **2021**, *2*, 1–7. [[CrossRef](#)]
90. Ronnerberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. *arXiv* **2015**, arXiv:1505.04597.
91. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *arXiv* **2014**, arXiv:1406.2661. [[CrossRef](#)]
92. Gatys, L.A.; Ecker, A.S.; Bethge, M. A neural algorithm of artistic style. *arXiv* **2015**, arXiv:1508.06576. [[CrossRef](#)]
93. Zhou, S.K.; Greenspan, H.; Davatzikos, C.; Duncan, J.S.; Ginneken, B.; Madabhushi, A.; Prince, J.L.; Rueckert, D.; Summers, R.M. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *arXiv* **2021**, arXiv:2008.09104. [[CrossRef](#)]