applied sciences

MDPI

Article

A CNN CADx System for Multimodal Classification of Colorectal Polyps Combining WL, BLI, and LCI Modalities

Roger Fonollà [1,*], Quirine E. W. van der Zander [2,3], Ramon M. Schreuder [4], Ad A. M. Masclee [2,5], Erik J. Schoon [4], Fons van der Sommen [1] and Peter H. N. de With [1]

1 Department of Electrical Engineering, Video Coding and Architectures (VCA), Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands; fvdsommen@tue.nl (F.v.d.S.); P.H.N.de.With@tue.nl (P.H.N.d.W.)
2 Division of Gastroenterology and Hepatology, Maastricht University Medical Center, 6229 HX Maastricht, The Netherlands; q.vanderzander@maastrichtuniversity.nl (Q.E.W.v.d.Z.); a.masclee@mumc.nl (A.A.M.M.)
3 School for Oncology and Developmental Biology (GROW), Maastricht University, 6229 ER Maastricht, The Netherlands
4 Department of Gastroenterology and Hepatology, Catharina Hospital, 5623 EJ Eindhoven, The Netherlands; ramonmichel.schreuder@catharinaziekenhuis.nl (R.M.S.); erik.schoon@catharinaziekenhuis.nl (E.J.S.)
5 School of Nutrition & Translational Research in Metabolism (NUTRIM), Maastricht University, 6229 ER Maastricht, The Netherlands
* Correspondence: r.fonolla.navarro@tue.nl

Received: 29 April 2020; Accepted: 20 July 2020; Published: 22 July 2020

Abstract: Colorectal polyps are critical indicators of colorectal cancer (CRC). Blue Laser Imaging and Linked Color Imaging are two modalities that allow improved visualization of the colon. In conjunction with the Blue Laser Imaging (BLI) Adenoma Serrated International Classification (BASIC) classification, endoscopists are capable of distinguishing benign and pre-malignant polyps. Despite these advancements, this classification still prevails a high misclassification rate for pre-malignant colorectal polyps. This work proposes a computer aided diagnosis (CADx) system that exploits the additional information contained in two novel imaging modalities, enabling more informative decision-making during colonoscopy. We train and benchmark six commonly used CNN architectures and compare the results with 19 endoscopists that employed the standard clinical classification model (BASIC). The proposed CADx system for classifying colorectal polyps achieves an area under the curve (AUC) of 0.97. Furthermore, we incorporate visual explanatory information together with a probability score, jointly computed from White Light, Blue Laser Imaging, and Linked Color Imaging. Our CADx system for automatic polyp malignancy classification facilitates future advances towards patient safety and may reduce time-consuming and costly histology assessment.

Keywords: blue light imaging; linked color imaging; colorectal polyp classification; artificial intelligence; deep learning; CADx; CNN

1. Introduction

Colorectal cancer (CRC) is the fourth cause of cancer-related death worldwide, with the highest incident rates in developed countries [1,2]. An early diagnosis of CRC can prevent spreading throughout the colon and avoid further complications. Colorectal polyps (CRP) are precursor lesions and indicators of colorectal cancer. There are roughly two classes of CRPs: (1) non-neoplastic CRPs,

which include the hyperplastic polyps (HP); and (2) pre-malignant CRPs, comprising adenomas (ADs) and the sessile serrated adenomas (SSAs). HPs are considered benign polyps, whilst ADs and SSAs are capable of developing into CRC when kept untreated [3]. In consequence, not all CRPs are removed, as different types of colorectal polyps lead to a different risk of progression into CRC. However, even for experienced and trained endoscopists, it is hard to differentiate between benign and pre-malignant CRPs based on their visual appearance [4]. Therefore, current medical protocols dictate that all detected colorectal polyps should be resected to undergo histological evaluation. However, this protocol has two considerable drawbacks, since (1) unnecessary removal of benign polyps exposes the patient to additional risks of polypectomy-related complications and (2) histological examination of all resected polyps leads to significantly increased costs. In order to minimize both the risk and cost, the strategy of resect-and-discard has been proposed for diminutive ($\leq$5 mm) adenomatous polyps [5–7] and the diagnose-and-leave strategy for diminutive hyperplastic polyps [8,9]. In spite of the aforementioned strategies, visual differentiation of CRPs is an ongoing challenge in the clinical endoscopy routine. White light endoscopy (WL) is the most common technique for detecting lesions in the intestinal tract, but it falls behind when enhancing the visualization of the vessels and the mucosa. Compared to WL, chromoendoscopy techniques [10] are capable of achieving high-contrast results, but they require the injection of chemical dyes into the body. Similar visual effects can be achieved with the use of in-vivo optical filters, like Narrow-Band Imaging (NBI) [11,12]. An alternative to NBI is the four-LED Multi Light Technology (Fujifilm Co.), based on the combination of four types of light as source emitters: blue-violet, blue, green, and red. Blue Light Imaging (BLI) and Linked Color Imaging (LCI) are two of the observation modes of the four-LED Multi Light Technology that allow enhanced visualization of hemoglobin. BLI intensifies the blue light emission in the range of 410 nm, which enhances the visualization of the vessels and the mucosa. Alternatively, LCI accentuates color contrast by decreasing the intensity of blue-light and emphasizing red-light, providing better delineation and detection of lesions and inflammation [13,14].

The increasing development of new acquisition systems and the growing need of in-vivo discrimination between different types of CRPs introduces new challenges onto image interpretability in clinical practice. Each new imaging mode comes with its own learning curve and reveals potential clues that enable in-vivo histology assessment. Computer-aided detection (CADe) and diagnosis (CADx) systems facilitate fast and objective decision-making during clinical assessment, offering medical experts additional information for medical diagnosis. Furthermore, while providing the endoscopist with additional evidence to support the diagnosis, it can also shorten the learning curve for such new modalities by pointing out the relevant visual clues for treatment decisions.

Although gastroenterologists are well-trained in identifying suspicious areas in the colon, there has been a rapid increase in the development of CADe systems to help localize CRPs from colonoscopy images. The introduction of deep learning in medical imaging and the increased availability of large datasets, both public and private, have helped to shape CADe systems towards real-time detection. Several studies have proposed end-to-end CNNs to tackle the problem of colorectal polyp detection, which aims to differentiate between normal bowel tissue and polyps. Such work has been explored on several datasets of WL and NBI [15–19], achieving specificity and sensitivity higher than 90% and successfully locating polyps on video frames. Given the potential of CNN, the next logical step is to observe its capabilities in a real clinical environment. In the study of Urban et al. [20], a CADe system was trained with 8641 WL and NBI polyp images. The CNN was tested on an external dataset with a variety of polyp sizes and achieved an accuracy of 96.1%, with an area under the curve (AUC) of 0.99. Additionally, 3 experts reviewed a set of colonoscopy video frames with and without assistance of the CNN and an increase in polyp detection was observed after employing the CADe system. A similar study was conducted by Wang et al. [21], where an open nonblinded trial was performed on 1058 patients, of which 536 were analyzed with colonoscopy and 522 were subjected to colonoscopy with a CADe system. The system achieved specificity of 95.9% and an AUC of 0.98 on different polyp sizes and types. The study observed an increased performance

on polyp detection when the experts were assisted by a CADe system. Overall, in recent years, systems for polyp detection have achieved impressive results and been shown to be an effective tool on assisting medical experts. Assuming a high detection rate, the following issue presented to clinicians is to visually identify a polyp for being benign or pre-malignant, which imposes an extra challenge that several studies have tried to overcome. In early studies, CRPs were classified based on local features from blood vessels using NBI images [22–24], or exploiting a combination of chromoendoscopy, WL, and NBI [25]. In Scheeve et al. [26], handcrafted features were used to predict the histology of polyps using Support Vector Machines (SVMs) and clinical classification models. The development of CNNs also had a great impact on colorectal polyp classification. Initial work employed classifiers using features extracted from broadly used CNNs [27–30], such as AlexNet, ResNet, or InceptionNet. Next, a more traditional classifier such as SVMs was required to classify CRPs between healthy and malignant. The early success pushed the development of several CADx systems to classify polyps using different classification schemes. In Konami et al. [31], high-accuracy results were obtained (sensitivity, 93.0%; specificity, 93.3%) using SVMs on a dataset of 118 lesions obtained with NBI magnifying colonoscopy. The study developed a CADx system for the Hiroshima classification. In a similar fashion in Mori et al. [32], a CADx was developed for CRP classification from images, obtained with NBI and stained endocytoscopy. The study employed SVM which required a three-step process to perform the polyp prediction. As the availability of medical data increased, alternative approaches using more recent deep learning frameworks allowed the design of end-to-end predictions. In Chen et al. [33], a framework was developed to classify diminutive polyps using magnified NBI. The magnified modalities allow for detailed imaging of lesions but require a high-level of precise movement of the endoscope, which makes it a less desirable technique in clinical usage. More recently, in Byrne et al. [34], a deep learning framework was developed to classify unaltered video frames of nonmagnified NBI to classify polyps using the NICE classification. One main arising issue is that the NICE classification does not incorporate SSAs polyps, which causes doctors to consider them as dangerous as ADs. Despite this aspect, the study presents a great method to accurately classify polyps during real-time colonoscopy. The current CADx studies have focused their potential around NBI, whereas BLI and LCI have not yet seen significant developments with the capabilities of artificial intelligence.

In our previous work [35], we collected a dataset of 203 patients with WL, BLI, and LCI. The limited dataset constrained the classification of our data. Therefore, we have explored and extracted features from a pretrained ResNet50 with the aim to classify the dataset of polyps between benign and pre-malignant. We have trained the features by combining single SVMs for each modality and evaluated the results via Leave-One-Patient-Out Cross-Validation (LOPOCV). Our study was finalized by incorporating the information of WL, BLI, and LCI and combining the posterior probabilities of the trained model.

In this work, we present a new study which builds further on our previous research in the following points.

1. Our dataset is improved with 458 new patients resulting in a total of 2919 images obtained from three different hospitals.
2. We perform a benchmark with mostly used state-of-the-art colorectal polyp deep learning architectures, in order to train an end-to-end CNN, evaluated with a test set of 60 patients obtained with WL, BLI, and LCI.
3. We build a CADx system to classify CRPs between benign and pre-malignant and we compare our results with the knowledge and expertise of 19 endoscopists (13 novices and 6 experts).
4. We present a probability score to the endoscopists, which is computed from the average prediction of WL, BLI, and LCI.
5. Our developed CADx systems provides explainable visual data from the CNN to contribute to smooth decision-making.

Our study concludes with showcasing how our CADx system could perform in clinical routine and how the outcomes can offer benefits to the endoscopists during real-time colonoscopy.

2. Materials and Methods

2.1. Patient Inclusion and Data Acquisition

The data collection was carried out in a prospective fashion, according to a predefined image acquisition protocol, in the Maastricht University Medical Center+ (MUMC+) and the Catharina Hospital Eindhoven (CZE), both in the Netherlands, and the Queen Alexandra Hospital in Portsmouth, United Kingdom. The training dataset consisted of a total of 468 patients, with 2319 pre-malignant polyps and 420 benign, with CRPs of all sizes, from diminutive polyps to large polyps. The dataset includes polyps acquired in WL, BLI, LCI, and I-Scan (HDWL; Mode 1, 2, and 3) modalities. Using I-Scan data adds robustness to the algorithm, as all modes have almost similar visual properties as the three other modalities. The test set was restricted to 60 patients to match with the exact same patients analyzed by endoscopists (further explained in Section 2.6). All the test set was acquired from CZE, with 45 patients identified as pre-malignant polyps and 15 benign. For each test patient, a single image of the polyp was acquired at different time steps with three different modalities—WL, BLI, and LCI—adding to a total of 180 polyp images. Figure 1 summarizes the data collection described above. All collected data was made fully anonymous prior to the study.

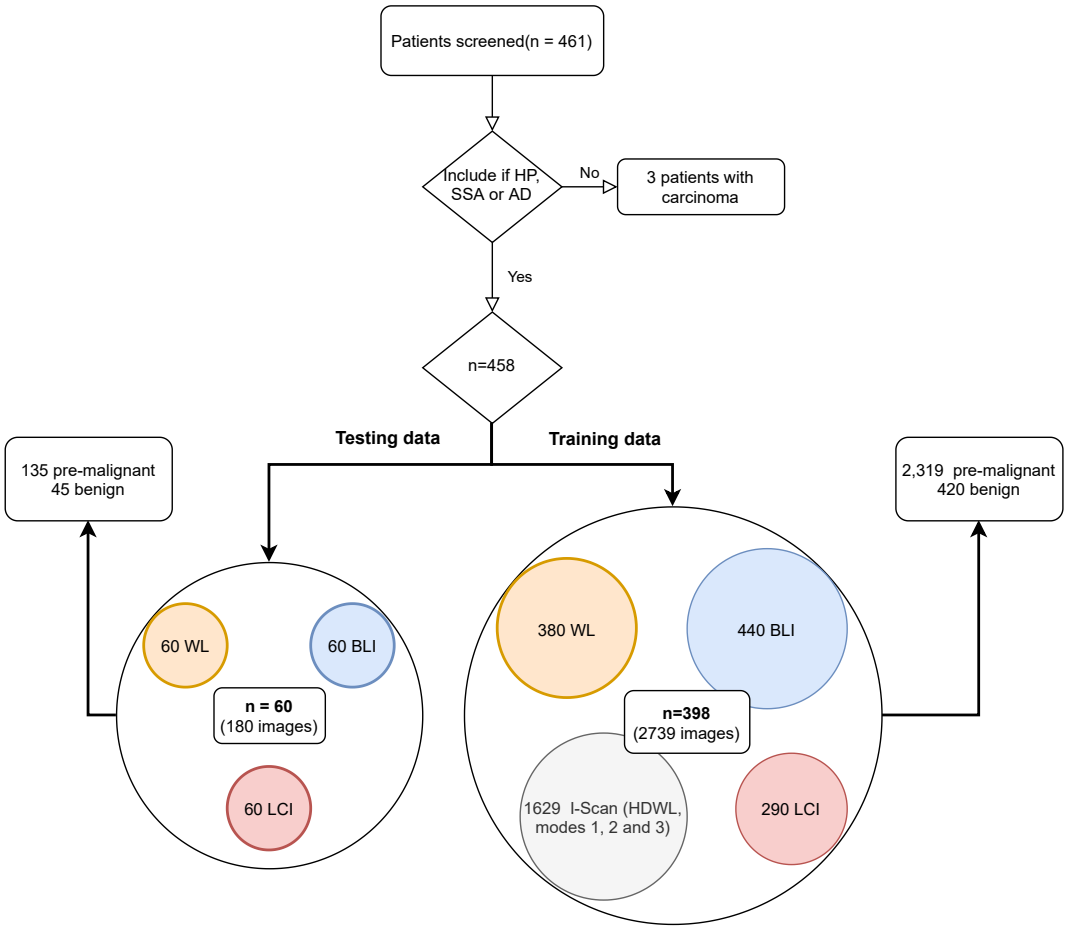

Figure 1. Data collection diagram. Patients with carcinoma were excluded. The dataset contained representatives of several modalities, White Light (WL) and High Definition White Light (HDWL), Blue Laser Imaging (BLI), Linked Color Imaging (LCI).

2.2. Data Preprocessing

In order to obtain optimal classification, the central region of the image was automatically selected as the ROI. The cropped region ensures a coverage of the polyp area, as well as its surrounding texture. Successively, the dataset was normalized by subtracting the mean and by dividing the standard deviation of the pretrained ImageNet data. For the last step, each input image was resized to 229×229 pixels in the RGB color space. To increase the generalization of the network, data augmentation was used to enhance the model capabilities for our classification task. In this study, the training images are augmented by a combination of flipping, shifting and $\pm 90^{\circ}$ rotation, contrast enhancement, blurring, and zooming.

2.3. Network Architectures

We performed a benchmark with different architectures and based on the results, we selected EfficientNet [36] as the main architecture for our system. This family of models achieves state-of-the-art accuracy on the ImageNet dataset by employing a simple, yet powerful concept where the network models are not only scaled in depth, but also in width and resolution. To achieve such a scheme, the authors propose a compound coefficient (Φ) that uniformly scales the network along the three dimensions. The coefficient controls the distribution of resources available for scaling the model under the constraint of a maximum operation growth of 2^{Φ} FLOPS. For our CADx, we employed the variant B4, which has a total of 19 million parameters. The B4 variant was the preferred option for two main reasons; one, it achieved a higher performance against state-of-the-art polyp classification architecture while reducing the number of parameters; and second, it allowed the best memory performance on our CADx setup. Additionally, several other commonly-used architectures were considered for the development of our CADx system, therefore we trained all alternatives and compared their performance with EfficientNet. The architectures were selected based on the most common networks employed in state-of-the-art polyp detection and classification studies. For this reason, we selected the following networks: VGG16, ResNet50, ResNet101, Xception, and InceptionResNet.

2.4. Training

All the networks were initialized with ImageNet weights and trained from scratch with Stochastic Gradient Descent (SGD) using a momentum of 0.9. For all the networks except EfficientNet, we used a batch size of 32, and for the latter a batch size of 8, due to the memory restrictions on our single GPU. We chose to use an exponential learning rate, with hard restarts at every two epochs, ranging from 1×10^{-2} to 4×10^{-3}—except for VGG16, where the learning rate ranged from 1×10^{-3} to 1×10^{-4}. The results of each architecture can be found in Table 1. Finally, the model was trained for 100 epochs or until convergence on the validation set, using a single TitanXp GPU. As input, the network received a single image of any of the modalities present in the training set (WL, BLI, and LCI), which allowed for shared features between all modalities. To ensure that each class is representative, during training, an independent image generator was created for the benign and the pre-malignant class. During inference time, we divided the results in WL, LCI, and BLI and obtained the posterior probabilities to observe the classification to the final prediction.

2.5. Explainable CADx System

During the assessment of CRPs, the CADx system provides the endoscopist with a quantitative measure of how likely the observed CRP is to be benign or pre-malignant. Although a probability measure might be sufficient for agreement between the system and the endoscopist, there is a likelihood that a visual inspection of the polyps is required for further confirmation. Gradient-weighted Class Activation Mapping (Grad-CAM) [37] is an effective method that allows the visualization of the decision region of a CNN. Through the average product of the feature maps and a class-activation function, we are able to produce a visual map and add explainability to the endoscopist's observations.

Table 1. Comparison of our computer aided diagnosis (CADx) results with the group of endoscopists. The first block presents the results of the experts and novices based on their intuition. The second block shows the results of the endoscopists after the post-training stage on the BLI Adenoma Serrated International Classification (BASIC) classification. The third block includes the CADx results of WL, BLI, and LCI combined modalities, evaluated with common state-of-the art architectures in polyp classification.

		Accuracy (%)	Specificity (%)	Sensitivity (%)	AUC
Intuition (±SD)					
	Novices (n = 13)	66.7 ± 8.4	93.2 ± 4.0	46.2 ± 16.7	N.A
	Experts (n = 6)	79.5 ± 6.6	95.6 ± 4.9	50.0 ± 16.2	N.A
	Total (n = 19)	70.7 ± 9.8	93.9 ± 4.3	47.4 ± 16.2	N.A
BASIC (±SD)					
	Novices (n = 13)	66.5 ± 9.3	92.1 ± 6.2	55.4 ± 14.7	N.A
	Experts (n = 6)	81.7 ± 4.2	94.1 ± 1.8	61.1 ± 5.0	N.A
	Total (n = 19)	71.3 ± 10.7	92.7 ± 5.2	57.2 ± 12.6	N.A
CADx					
	VGG16	85.0	84.4	86.7	0.90
	ResNet50V2	83.3	93.3	80.0	0.92
	ResNet101V2	88.3	93.3	73.3	0.94
	Xception	85.0	84.4	86.7	0.93
	InceptionResNetV2	85.0	86.7	80.0	0.94
	EfficientNetB4	95.0	93.3	95.6	0.97

2.6. Clinical Benchmark

To benchmark the performance of the proposed algorithm, a prospective, endoscopist-blinded, noninterventional study was conducted at the Maastricht University Medical Center+ (MUMC+) and Catharina Hospital Eindhoven (CZE). The study was in accordance with the declaration of Helsinki as well as the General Data Protection Regulation. A total of 19 endoscopists optically diagnosed 60 colonoscopy images containing a single polyp acquired in WL, BLI, and LCI modalities (later referred to as test data). Two person groups were derived from the medical professionals. The first group consisted of six expert endoscopists from the international BLI-expert group, who were knowledgeable in using BLI and BLI Adenoma Serrated International Classification (BASIC) [13,38] (Table 2) and brought an experience of more than 2000 colonoscopies. The second group consisted of thirteen Dutch novices with limited colonoscopy experience (<400 colonoscopies) and without prior experience in using BLI or BASIC.

The benchmark study was divided in two stages. For both stages, only WL and BLI were present as modalities. In the first stage, the experts and novices were instructed to classify the set of 60 polyps between HP, SSA, or AD based on their intuition and expertise, with a time limit of 30 seconds. After a washout period of four weeks, the medical group was further trained in using BLI and BASIC. Following the instruction period, the same endoscopists were asked to classify the exact same dataset of polyps based on the BASIC classification [13]. Moreover, each endoscopist had to report the level of confidence for each CRP.

Table 2. The BASIC classification comprises a list of visual BLI features, employed by endoscopists to classify hyperplastic, sessile serrated, and adenomatous polyps. A more in-depth clinical analysis can be found in the study of Subramanian et al. [38].

		Hyperplastic	Adenoma	Sessile Serrated
Surface				
	Presence of mucus	No	No	Yes
	Regular/irregular	Regular	Regular or irregular	Regular or irregular
	Pseudodepression	No	Yes	No
	Depression	No	No	No
Pit Pattern				
	Featureless	Yes	No	No
	Type	Round	Not round	Round pits with/without dark spots
	Distribution	Homogenous	Homogenous or heteregenous without focal loss	Homogenous or heteregenous
Vessels				
	Present?	Yes/No	Yes	Yes/No
	Type if present	Lacy	Pericryptal	Pericryptal

3. Evaluation and Results

The test set was evaluated using a single network that was simultaneously trained with all acquisition modalities, which allowed the CNN to learn shared features across all domains. For the single and combined modality, we computed the accuracy and the area under the curve (AUC). Additionally, sensitivity, defined as the rate of correct pre-malignant polyps classified as such; and specificity, defined as the correct rate of benign polyps classified as benign, were calculated as well. The performance of the CADx system and all associated architectures was evaluated and compared to outcomes of the experts and novices.

The best CADx system achieves an accuracy of 95.0% with specificity of 93.3%, sensitivity of 95.6%, and AUC of 0.97. For each input image, the CADx system computes the region for decision-making via Grad-CAM and the malignancy prediction. The average prediction of the three modalities offers the endoscopists with the best possible diagnosis result. The end-to-end framework described above is depicted in Figure 2.

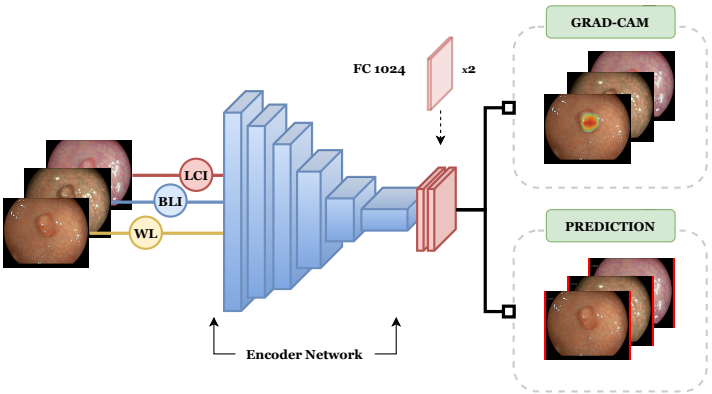

Figure 2. End-to-end framework of the proposed CADx system. The training batch contains independent images of each modality (WL, BLI, and LCI) and at least a representative of each polyp class. The output of the system is a prediction for each single image.

4. Discussion

In clinical practice, endoscopists must perform visual inspection of all detected colorectal polyps. Experience and expertise are factors that currently dictate decision-making during real-time endoscopy. In this study, we have evaluated a CADx system against expert and novice endoscopists. Firstly, in Table 1, a noticeable difference between novices and experts is observed for both intuition and the BASIC classification. Experts showed a higher diagnostic accuracy compared to novices (79.5% vs. 66.7%, $p = 0.005$) and (81.7% vs. 66.5%, $p = 0.002$), respectively. Both groups improved the diagnosis of pre-malignant polyps during the second round of assessments, which positively reflects on the training with the BASIC classification received after the first stage of the study. Following the first point, our CADx system showed an overall better performance than both novices and experts in all of the trained architectures. All our trained models contained information of three different hospitals with WL, BLI, and LCI modalities, with a total of 2739 polyp images. Compared to the endoscopists, our training stage utilized the benefits of deep learning, which allowed us to build a robust diagnostic tool. Our final CADx system (EfficientNetB4) correctly assessed a total of 57 out of 60 polyps. The CADx system erroneously classified only three polyps: two adenomas and one hyperplastic polyp.

Table 3 presents the individual contribution of each modality to the overall test set. In this study, the performance of the endoscopists was evaluated on the knowledge of WL and BLI. Moreover, the BASIC classification only takes into account the latter modality, hence, comparison with all combined modalities may lead to unreliable one-to-one comparison. Solely observing BLI, our CADx system outperforms experts and novices for both intuition and BASIC. In contrast, WL and LCI do not outperform both groups on classifying enough HP polyps. On the one hand, WL offers limited visualization and enhancement, which makes it difficult to identify small benign polyps; while on the other hand, LCI is not the most common modality in current clinical practice. This is reflected in our training dataset, which contained far more WL and BLI images than LCI data. Lastly, when comparing each single modality with the combined results, we did not find a statistically significant improvement compared to combining modalities. This might indicate that for a further study, the sample size for the test set should be increased.

Table 3. Individual modality results of our CADx system based on EfficienNetB4. The combined results are obtained after averaging the CNN output probabilities for each modality. The results are presented together with confidence intervals at 95%. AUC—Area under the curve.

	Accuracy (%)	Specificity (%)	Sensitivity (%)	AUC
WL	86.7 (78.1–95.3)	86.7 (78.1–95.3)	86.7 (78.1–95.3)	0.94 (0.86–0.99)
BLI	93.3 (87.0–99.6)	93.3 (87.0–99.6)	93.3 (87.0–99.6)	0.96 (0.88–1.00)
LCI	90.0 (82.4–97.6)	80.0 (69.9–90.1)	93.3 (83.2–100)	0.89 (0.79–0.97)
Combined	95.0 (89.5–100)	93.3 (87.0–99.6)	95.6 (89.2–100)	0.97 (0.92–1.00)

In Figure 3, we showcase the strength of our CADx system, where three predictions are performed—one for each modality—to obtain a final prediction based on the three observations. In the work of Murata et al. [30], the authors proposed a voting system from the predictions of each combination. Although the voting system would be suitable, we prefer to adopt the same methodology as our previous study [35], where the posterior probabilities acquired for each modality are subsequently averaged to compute one unique probability per polyp. In addition, Grad-CAM provides the endoscopists with visual information on the decision region. This has two benefits, one is to provide the endoscopists with visual feedback on the CADx judgment, and the second is to offer an alternative method for tracking polyps (either benign or pre-malignant) during real-time video endoscopy. Our CADx system excels in cases where not all modalities are predicted as correct, such as in the example of Figure 4, where a pre-malignant polyp is incorrectly predicted in the WL modality. Furthermore, a visual informative inspection is supported by the results of Grad-CAM

region, which focuses on the intestinal tract instead of the CRP. Therefore, the combined information of BLI and LCI allows for a correct decision of malignancy, facilitating endoscopists with a better diagnosis during colonoscopy.

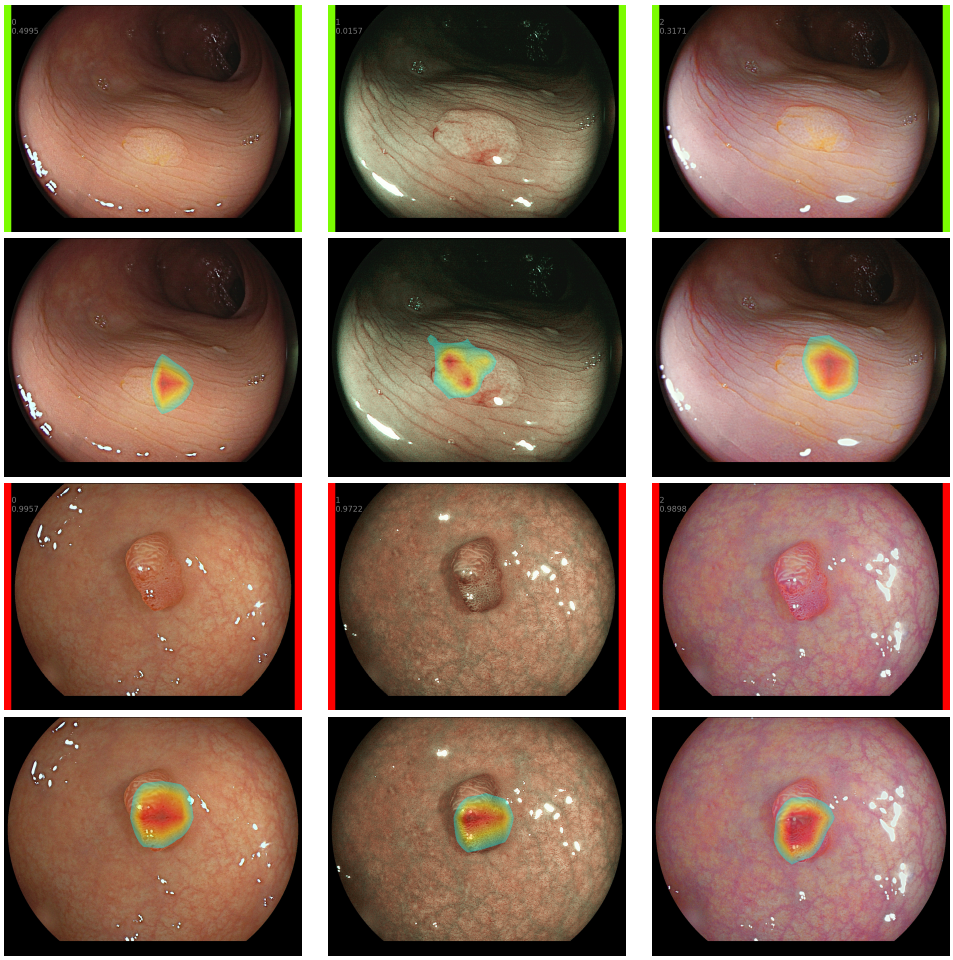

Figure 3. Output generated from the CADx system where an image of a colorectal polyp (CRP) is received from the endoscope. Two outputs are presented, (1) a prediction of whether the polyp is benign (green) or pre-malignant (red), and (2) a GRAD-CAM map that allows the endoscopists to judge the decisions of the CADx. From left to right, the three modalities employed in this study are shown: White Light, Blue Light Imaging, and Linked Color Imaging. In the first row, a benign polyp is predicted from the three modalities with the CADx system decision shown in the following row. In the third and fourth rows, a pre-malignant polyp is predicted from the combined modalities.

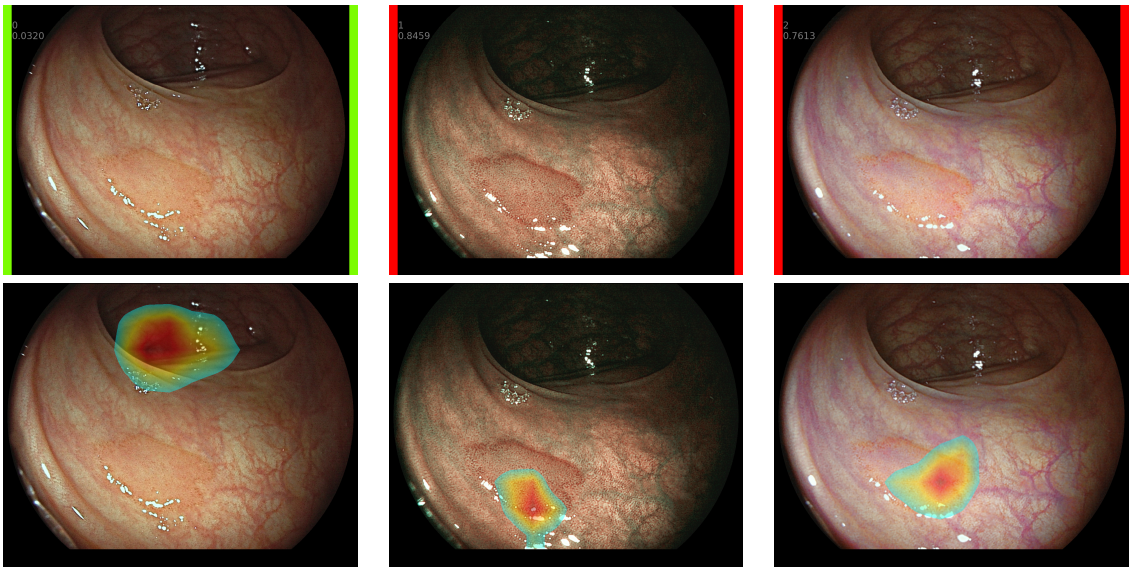

Figure 4. Example of misclassification due to a wrong decision in the CADx for the WL modality. In this case, BLI and LCI correctly identified the pre-malignant polyp and the CADx system is capable pf identifying it as such.

5. Conclusions

In this study, we have developed an end-to-end CADx system with a state-of-the-art deep learning architecture to classify colorectal polyps between benign and pre-malignant, obtained with three different image modalities (WL, BLI, and LCI). We evaluated our framework on an independent test set of 60 patients and compared the results with the diagnosis of 19 endoscopists (6 experts and 13 novices). Our CADx system was trained with a dataset of 2739 images collected from three different hospitals. Anticipating clinical application, we employed EfficientNetB4, which is a state-of-the art architecture for classification. The network demonstrates excellent performance for our CADx system compared to most common networks found in polyp classification literature. Besides its optimal performance, one of the most noticeable downsides we found during training is that the complex computations for scaling and depth limited the batch size for training. We have opened the path for endoscopists towards combining WL, BLI, and LCI to predict polyp histology, and the results have shown that there is potential to enhance the prediction of individual polyps from several modalities, but further studies should conclude the findings with a broader testing set. Moreover, we have experimented with Grad-CAM to offer endoscopists an interpretable answer of the CADx decisions. In further studies, we will investigate our system performance in real-time colonoscopy. In conclusion, we present a CADx system that could be used in routine colonoscopy to classify benign and pre-malignant colorectal polyps. If the CADx system successfully predicts CRP histology, then potentially diminutive hyperplastic polyps could be left in the colon and the suggested 'diagnose-and-leave' strategy could be applied. The same principle can be applied to diminutive adenomatous polyps as well, which then could be resected without performing histological evaluation following the 'resect-and-discard' strategy. Overall, the presented study may allow for improved diagnosis of CRC and decrease the current cost burden of histological examinations.

Author Contributions: Conceptualization, R.F.; methodology, R.F.; software, R.F.; validation, R.F., Q.E.W.v.d.Z., and R.M.S.; formal analysis, Q.E.W.v.d.Z.; investigation, R.F. and Q.E.W.v.d.Z.; resources, Q.E.W.v.d.Z., R.M.S., and E.J.S.; data curation, R.F., Q.E.W.v.d.Z., and R.M.S.; writing—original draft preparation, R.F.; writing—review and editing, F.v.d.S. and P.H.N.d.W.; visualization, R.F.; supervision, F.v.d.S. and P.H.N.d.W.; project administration, A.A.M.M., E.J.S., F.v.d.S., and P.H.N.d.W.; funding acquisition, Q.E.W.v.d.Z., A.A.M.M., E.J.S., F.v.d.S., and P.H.N.d.W. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 721766.

Acknowledgments: We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

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

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef]
2. Ferlay, J.; Colombet, M.; Soerjomataram, I.; Mathers, C.; Parkin, D.M.; Piñeros, M.; Znaor, A.; Bray, F. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer* **2019**, *144*, 1941–1953. [CrossRef] [PubMed]
3. Jass, J.R. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology* **2017**, *50*, 113–130. [CrossRef] [PubMed]
4. Vleugels, J.L.A.; Dijkgraaf, M.G.W.; Hazewinkel, Y.; Wanders, L.K.; Fockens, P.; Dekker, E. Effects of Training and Feedback on Accuracy of Predicting Rectosigmoid Neoplastic Lesions and Selection of Surveillance Intervals by Endoscopists Performing Optical Diagnosis of Diminutive Polyps. *Gastroenterology* **2018**, *154*, 1682–1693.e1681. [CrossRef] [PubMed]
5. Ignjatovic, A.; East, J.E.; Suzuki, N.; Vance, M.; Guenther, T.; Saunders, B.P. Optical diagnosis of small colorectal polyps at routine colonoscopy (Detect InSpect ChAracterise Resect and Discard; DISCARD trial): A prospective cohort study. *Lancet Oncol.* **2009**, *10*, 1171–1178. [CrossRef]
6. Hassan, C.; Pickhardt, P.J.; Rex, D.K. A Resect and Discard Strategy Would Improve Cost-Effectiveness of Colorectal Cancer Screening. *Clin. Gastroenterol. Hepatol.* **2010**, *8*, 865–869.e3. [CrossRef]
7. Tsuji, S.; Takeda, Y.; Tsuji, K.; Yoshida, N.; Takemura, K.; Yamada, S.; Doyama, H. Clinical outcomes of the "resect and discard" strategy using magnifying narrow-band imaging for small (<10 mm) colorectal polyps. *Endosc. Int. Open* **2018**, *6*, E1382–E1389.
8. Neumann, H.; Neumann Sen, H.; Vieth, M.; Bisschops, R.; Thieringer, F.; Rahman, K.F.; Gamstätter, T.; Tontini, G.E.; Galle, P.R. Leaving colorectal polyps in place can be achieved with high accuracy using blue light imaging (BLI). *United Eur. Gastroenterol. J.* **2018**, *6*, 1099–1105. [CrossRef]
9. Kandel, P.; Wallace, M.B. Should We Resect and Discard Low Risk Diminutive Colon Polyps. *Clin. Endosc.* **2019**, *52*, 239. [CrossRef]
10. Har-Noy, O.; Katz, L.; Avni, T.; Battat, R.; Bessissow, T.; Yung, D.E.; Engel, T.; Koulaouzidis, A.; Eliakim, R.; Ben-Horin, S.; et al. Chromoendoscopy, narrow-band imaging or white light endoscopy for neoplasia detection in inflammatory bowel diseases. *Dig. Dis. Sci.* **2017**, *62*, 2982–2990. [CrossRef]
11. East, J.E.; Guenther, T.; Kennedy, R.H.; Saunders, B.P. Narrow band imaging avoids potential chromoendoscopy risks. *Gut* **2007**, *56*, 1168–1169. [PubMed]
12. Vişovan, I.I.; Tanțău, M.; Pascu, O.; Ciobanu, L.; Tanțău, A. The role of narrow band imaging in colorectal polyp detection. *Bosn. J. Basic Med. Sci.* **2017**, *17*, 152. [CrossRef] [PubMed]
13. Bisschops, R.; Hassan, C.; Bhandari, P.; Coron, E.; Neumann, H.; Pech, O.; Correale, L.; Repici, A. BASIC (BLI Adenoma Serrated International Classification) classification for colorectal polyp characterization with blue light imaging. *Endoscopy* **2018**, *50*, 211–220. [CrossRef] [PubMed]
14. Yoshida, N.; Dohi, O.; Inoue, K.; Yasuda, R.; Murakami, T.; Hirose, R.; Inoue, K.; Naito, Y.; Inada, Y.; Ogiso, K.; et al. Blue Laser Imaging, Blue Light Imaging, and Linked Color Imaging for the Detection and Characterization of Colorectal Tumors. *Gut Liver* **2019**, *13*, 140–148. [CrossRef]
15. Yu, L.; Chen, H.; Dou, Q.; Qin, J.; Heng, P.A. Integrating Online and Offline Three-Dimensional Deep Learning for Automated Polyp Detection in Colonoscopy Videos. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 65–75. [CrossRef]
16. Wang, D.; Zhang, N.; Sun, X.; Zhang, P.; Zhang, C.; Cao, Y.; Liu, B. AFP-Net: Realtime Anchor-Free Polyp Detection in Colonoscopy. In Proceedings of the ICTAI 2019, Portland, OR, USA, 4–6 November 2019; pp. 636–643.
17. Zhang, X.; Chen, F.; Yu, T.; An, J.; Huang, Z.; Liu, J.; Hu, W.; Wang, L.; Duan, H.; Si, J. Real-time gastric polyp detection using convolutional neural networks. Real-time gastric polyp detection using convolutional neural networks. *PLoS ONE* **2019**, *14*, e0214133.

18. Sornapudi, S.; Meng, F.; Yi, S. Region-Based Automated Localization of Colonoscopy and Wireless Capsule Endoscopy Polyps. *Appl. Sci.* **2019**, *9*, 2404. [CrossRef]
19. Qadir, H.A.; Balasingham, I.; Solhusvik, J.; Bergsland, J.; Aabakken, L.; Shin, Y. Improving Automatic Polyp Detection Using CNN by Exploiting Temporal Dependency in Colonoscopy Video. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 180–193. [CrossRef]
20. Urban, G.; Tripathi, P.; Alkayali, T.; Mittal, M.; Jalali, F.; Karnes, W.; Baldi, P. Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. *Gastroenterology* **2018**, *155*, 1069–1078. [CrossRef]
21. Wang, P.; Berzin, T.M.; Glissen-Brown, J.R.; Bharadwaj, S.; Becq, A.; Xiao, X.; Liu, P.; Li, L.; Song, Y.; Zhang, D.; et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: A prospective randomised controlled study. *Gut* **2019**, *68*, 1813–1819. [CrossRef]
22. Gross, S.; Stehle, T.; Behrens, A.; Auer, R.; Aach, T.; Winograd, R.; Trautwein, C.; Tischendorf, J. A comparison of blood vessel features and local binary patterns for colorectal polyp classification. *SPIE Med. Imaging* **2009**, *7260*, 72602Q.
23. Tischendorf, J.J.W.; Gross, S.; Winograd, R.; Hecker, H.; Auer, R.; Behrens, A.; Trautwein, C.; Aach, T.; Stehle, T. Computer-aided classification of colorectal polyps based on vascular patterns: A pilot study. *Endoscopy* **2010**, *42*, 203–207. [CrossRef]
24. Gross, S.; Trautwein, C.; Behrens, A.; Winograd, R.; Palm, S.; Lutz, H.; Schirin-Sokhan, R.; Hecker, H.; Aach, T.; Tischendorf, J. Computer-based classification of small colorectal polyps by using narrow-band imaging with optical magnification. *Gastrointest. Endosc.* **2011**, *74*, 1354–1359. [CrossRef] [PubMed]
25. Tamaki, T.; Yoshimuta, J.; Kawakami, M.; Raytchev, B.; Kaneda, K.; Yoshida, S.; Takemura, Y.; Onji, K.; Miyaki, R.; Tanaka, S. Computer-aided colorectal tumor classification in NBI endoscopy: Using local features. *Med. Image Anal.* **2013**, *17*, 78–100. [CrossRef] [PubMed]
26. Scheeve, T.; Schreuder, R.M.; van der Sommen, F.; IJspeert, J.E.; Dekker, E.; Schoon, E.J. Computer-aided classification of colorectal polyps using blue-light and linked-color imaging. In *Medical Imaging 2019: Computer-Aided Diagnosis*; International Society for Optics and Photonics: Bellingham, WA, USA, 2019; Volume 10950, p. 1095012.
27. Tamaki, T.; Yoshimuta, J.; Kawakami, M.; Raytchev, B.; Kaneda, K.; Yoshida, S.; Takemura, Y.; Onji, K.; Miyaki, R.; Tanaka, S. Computer-aided colorectal tumor classification in NBI endoscopy: Using CNN features. *Med. Image Anal.* **2016**, *17*, 78–100. [CrossRef] [PubMed]
28. Zhang, R.; Zheng, Y.; Mak, W.C.T.; Yu, R.; Wong, S.H.; Lau, J.; Poon, C. Automatic Detection and Classification of Colorectal Polyps by Transferring Low-level CNN Features from Non-Medical Domain. *IEEE J. Biomed. Health Inform.* **2016**, *21*, 41–47. [CrossRef]
29. Ribeiro, E.; Häfner, M.; Wimmer, G.; Tamaki, T.; Tischendorf, J.J.; Yoshida, S.; Tanaka, S.; Uhl, A. Exploring texture transfer learning for colonic polyp classification via convolutional neural networks. In Proceedings of the ISBI 2017, Melbourne, Australia, 18–21 April 2017; pp. 1044–1048.
30. Murata, M.; Usami, H.; Iwahori, Y.; Aili, W. Polyp Classification Using Multiple CNN-SVM Classifiers from Endoscope Images. In Proceedings of the PATTERNS 2017, Athens, Greece, 19–23 February 2017; pp. 109–112.
31. 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. *Gastrointest. Endosc.* **2016**, *83*, 643–649. [CrossRef]
32. Mori, Y.; Kudo, S.E.; Misawa, M.; Saito, Y.; Ikematsu, H.; Hotta, K.; Ohtsuka, K.; Urushibara, F.; Kataoka, S.; Ogawa, Y.; et al. Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study. *Ann. Intern. Med.* **2018**, *169*, 357–366. [CrossRef]
33. Chen, P.J.; Lin, M.C.; Lai, M.J.; Lin, J.C.; Lu, H.H.; Tseng, V.S. Accurate Classification of Diminutive Colorectal Polyps Using Computer-Aided Analysis. *Gastroenterology* **2018**, *154*, 568–575. [CrossRef]
34. Byrne, M.F.; Chapados, N.; Soudan, F.; Oertel, C.; Linares-Pérez, M.; Kelly, R.; Iqbal, N.; Chandelier, F.; Rex, D.K. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut* **2019**, *68*, 94–100. [CrossRef]

35. Fonolla, R.; Van Der Sommen, F.; Schreuder, R.M.; Schoon, E.J.; De With, P.H.N. Multi-modal classification of polyp malignancy using CNN features with balanced class augmentation. In Proceedings of the ISBI 2019, Venice, Italy, 8–11 April 2019; pp. 74–78.
36. Tan, M.; Le, Q. Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning. *PMLR* **2019**, *97*, 6105–6114.
37. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. Comput. Vis.* **2019**, *128*, 336–359. [CrossRef]
38. Subramaniam, S.; Hayee, B.; Aepli, P.; Schoon, E.; Stefanovic, M.; Kandiah, K.; Thayalasekaran, S.; Alkandari, A.; Bassett, P.; Coron, E.; et al. Optical diagnosis of colorectal polyps with Blue Light Imaging using a new international classification. *United Eur. Gastroenterol. J.* **2019**, *7*, 316–325. [CrossRef] [PubMed]