Polyp malignancy classification with CNN features based on Blue Laser and Linked Color Imaging

Polyp malignancy classification with CNN features based on Blue Laser and Linked Color Imaging

Roger Fonolla Navarro Fons van der Sommen Peter H.N. de With
Department of Electrical Engineering, Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
{r.fonolla.navarro, fvdsommen, P.H.N.de.With}@tue.nl

Ramon M. Schreuder Erik J. Schoon
Department of Gastroenterology and Hepatology, Catharina Ziekenhuis Eindhoven
P.O. Box 1350, 5602ZA Eindhoven, The Netherlands

{ramonmichel.schreuder, erik.schoon}@catharinaziekenhuis.nl

Abstract

In-vivo classification of benign and pre-malignant polyps is a laborious task that requires histophatology confirmation. In an effort to improve the quality of clinical diagnosis, medical experts have come up with visual models with only limited success. In this paper, a classification approach is proposed to differentiate between polyp malignancy, using features extracted from the Global Average Pooling (GAP) layer of a pre-trained Convolutional Neural Network (CNNs) . Two recently developed endoscopic modalities are used to improve the pipeline prediction: Blue Laser Imaging (BLI) and Linked Color Imaging (LCI). Furthermore, a new strategy of per-class data augmentation is adopted to tackle the differences of unbalanced class distribution. The results are compared with a more general approach, showing how artificial examples can improve results on highly unbalanced problems. For the same reason, the combined features for each patient are extracted and trained using several machine learning classifiers without CNNs. Moreover to speed up computation, a recent GPU based Support Vector Machine (SVM) scheme is employed to substantially decrease the overload during training time. The presented methodology shows the feasibility of using the LCI and BLI techniques for automatic polyp malignancy classification and facilitates future advances to limit the need for time-consuming and costly histopathological assessment.

Keywords— *Polyp classification, Blue Laser Imaging, BLI, Linked Color Imaging, LCI, CNN, Data Augmentation, SVM, GPU Support Vector Machine.*

1. Introduction

Colorectal cancer (CRC) is the second most diagnosed cancer in women and the third in men worldwide, with more than half of its incidence rates arising in developed countries [3]. An early diagnosis of CRC can prevent spreading throughout the colon and avoid further complications. In the initial stages of the cancer, abnormal colorectal tissue or *polyp* can be classified in three different groups, hyperplastic polyps (HPs), adenomas (ADs) and sessile serrated adenomas (SSAs). HPs are considered benigns polyps, whilst ADs and SSAs are identified as pre-maligannt polyps, capable of developing to CRC when kept untreatead [8]. Visual differentiation of benign and pre-malignant polyps is an on-going challenge in a clinical endoscopy routine, often dictated by the expertise of the clinician. White light endoscopy (WLE) is the most common technique to assess lesions in the intestinal tract, but it falls behind when enhancing the visualization of vessels and the mucosa. Compared to WLE, chromoendoscopy techniques [6] are capable of achieving high visual 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 Narrrow-Band Imaging (NBI) [2] [20]. Alternatively, LED-based techniques like Blue Laser Imaging (BLI) and Linked Color Imaging (LCI) [13] exploit the absorption rate of the hemoglobin in the range of the blue-violet light to achieve comparable enhanced results. The advance on less invasive modalities could potentially avoid the use of chemical stains, while still providing the same amount of visual information. The increasing need for early diagnosis of CRC introduces the need of in-vivo differentiation of benign and pre-malignant polyps. Computer aided diag-

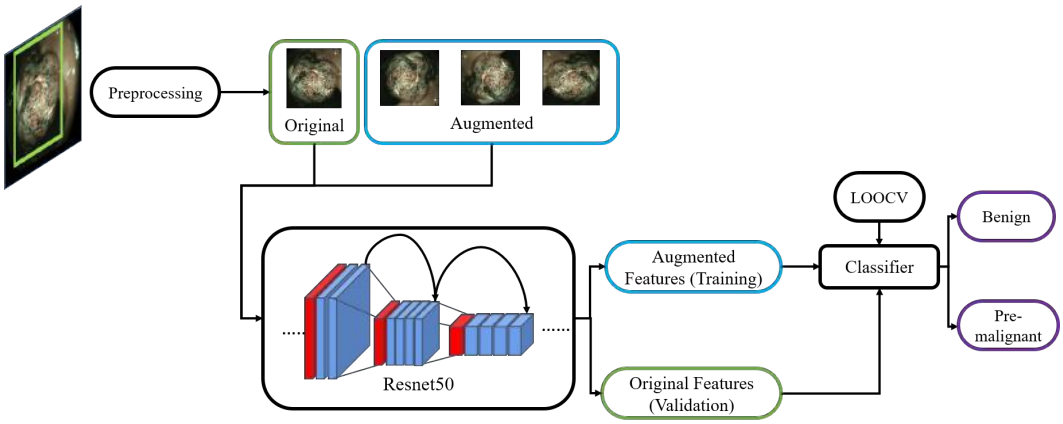

Figure 1: Pipeline of the proposed method. An ROI covering the polyp is selected from a full-size image. After an initial preprocessing, the proposed data augmentation pipeline is applied. Then, the features from the last layer of Resnet50 are extracted and used to train several classifiers. Finally, LOPOCV is used to evaluate those classifiers and predict the malignancy of the test images.

nosis (CAD) systems offer a side-opinion to help on clinical decisions, and potentially prevent costly resections and hystopathological analysis. In the past years, CAD systems have taken advantage of machine learning methods and visual descriptors or features to automatically classify several pathologies in the gastrointestinal tract [10]. Previous studies on polyp classification extracted local features from blood vessels using NBI images [4] [19] [5], while others achieved similar results by combining chromoendoscopy, WLE and NBI [17]. In recent years Convolutional Neural Networks (CNNs) have proven to be strong feature extractors for a wide variety of domains, including medical image analysis [18]. In the study of Zhang *et al*. [23] several features of a CNN from non-medical databases were used to detect and classify colorectal polyps in three different classes (This approach is further referred to as CNN features). Similar work was achieved by Murata *et al*. [12], where CNN features were combined with multiple SVMs and a voting system to classify polyps in between benign and pre-malignant using dye imaging, WLE and NBI.

In this work, a pipeline for polyp malignancy classification is designed, using residual CNN features from a Resnet50 architecture [7]. The scheme uses features obtained from a prospectively acquired dataset of polyps. Several machine learning classifiers are used to learn the CNN features. The different models are then evaluated using Leave-One Patient-Out Cross-Validation (LOPOCV). Moreover, an upsampling approach for image augmentation is adopted to tackle the unbalancing between benign and pre-malignant classes. In Sec. 2 the dataset and the proposed method are explained. In Sec. 3 several classifiers are compared and the results are presented. Finally in Sec. 4 the findings are discussed as well as future approaches.

2. Methods

2.1. Data description

The data collection was carried out in a prospective fashion, according to a pre-defined image acquisition protocol, in the Catharina (Ziekenhuis) Hospital Eindhoven (CZE). A total of 115 patients were included in this study. After histopathology examination, 95 patients were found with pre-malignant polyps and 20 with definitively benign polyps. For each patient, a single image of the polyp was acquired at different time steps with three different modalities, WLE, BLI and LCI. All collected data was fully anonymized prior to the study.

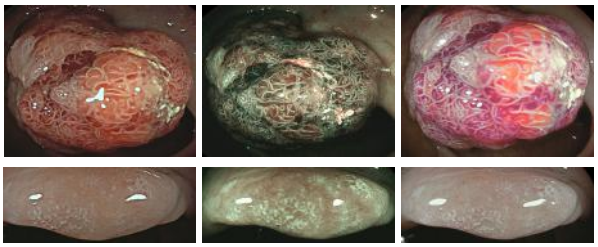

Figure 2: (Top) Example of a pre-malignant polyp from the CZE dataset. (Bottom) Example of a benign polyp. From left to right, a visualization of the three image acquisition modalities used in the study; White Light Endoscopy, Blue Light Imaging and Linked Color Imaging, respectively.

2.2. Feature extraction with Resnet50

Residual networks have claimed its prominence since its major breakthrough after obtaining the first place in both Imagenet [15] and COCO competitions [11] in 2015. The

introduction of the residual blocks presented a new approach for feature transfer by defining a refinement module between successive layers, which has led to an improvement in many image recognition tasks [7]. Furthermore, the Global Average Pooling (GAP) contained in the last layer of the architecture allows for a substantial reduction on the number of features, compared to conventional sequential models like AlexNet [9] or VGG [16], being an excellent choice as feature extractor. Therefore, in this work a pre-trained Resnet50 with Imagenet weights was selected as the most desirable architecture in terms of performance and computational cost.

2.3. Preprocessing

In order to achieve optimal classification, a manual selection of the ROI was drawn for each polyp. The cropped region ensured a coverage of the polyp area as well as its surrounding texture. Successively, the dataset was normalized by the mean and standard deviation of the pre-trained Imagenet weights. As final step, each input image was required to be resized by 224×224 color pixels.

Data augmentation was used in this work to enhance the model ability to classify a polyp between benign and pre-malignant. Previous work showcased the increased performance of a classifier when using different degrees of data augmentation [22]. A similar approach was reviewed by Asperty and Mastronardo in [1], where data augmentation was applied to increase the classification results of several gastrointestinal tasks. In this work, a pipeline of chained augmentations was adopted to increase the number of training examples. For each image a combination of *flipping, shifting,* $\pm$ *90°rotation* and *zoom* were applied. Each combination was permutable and the number of operations per combination was chosen randomly for each image. The degrees of rotation were imposed to uniquely be -90°or +90°to avoid interpolation artifacts. For each image a total of 12 combinations were empirically found as a minimum for obtaining the best possible results.

2.4. Training

The dataset used in this study was highly unbalanced, the ratio was found to be about one benign image per four pre-malignant images. Unbalancing between data points can lead to low-confidence intervals for the minority class, more often than not caused by the predominance of the higher class, inaccurately measuring both sensitivity and specificity. In order tackle this problem, a balancing scheme was applied. Other studies that focused on polyp classification did perform a down-sampling strategy [12] [23] to match the highest class with the lowest. In this work a different approach was adopted using *per-class* data augmentation. To analyze the methodology and compare the differences between class differentiation, two artificial datasets were generated. On the one hand, the *unbalanced* dataset was created by applying the same number of augmentation for each class, increasing the number of training examples, but not the ratio-difference between benign and pre-malignant. On the other hand, the *balanced* dataset was generated by applying an increased number of augmentation for the class with smallest number of samples, ensuring a similar amount of artificial images when compared to the largest class.

As mentioned in Section 2.2, features extracted from residual modules benefit from the refinement of previous layers, hence the GAP layer of the Resnet50 architecture was chosen as best feature extractor. For each augmented and original image, the feature vector was computed. The output dimensions of the feature layer was $1 \times 1 \times 2048$. For each training example, the total feature vector had the final dimensions of $N \times 2048$, where N denotes the number of training examples.

To analyze the extracted features, *Naives-Bayes*, *k-Nearest Neighbor*, *Random Forest* and *Support Vector Machine* were implemented using the *scikit-learn* python-package [14]. For each classifier, a *grid search* was used to select the best combination of parameters. The high amount of features to be processed, both for the *unbalanced* and *balanced* dataset, required a large computational task, hence the library *thunder-svm* [21] in combination with the dedicated GPU, were used to substantially reduce the training and testing time of the SVM, thereby working towards a real-time application.

3. Evaluation and Results

Leave-One Patient-Out Cross-Validation (LOPOCV) was employed for validation of the proposed approach. Each patient contained three images with the three different modalities, WLE, BLI and LCI. To ensure unbiased results, the following protocol was applied.

1. Select one patient as test subject and obtain the features from the original images for the WLE, BLI and LCI modalities.
2. Select the remaining 114 patients as training subjects and obtain the features from the augmented images for the WLE, BLI and LCI modalities.
3. Use the features of Step 2 to train the classifier.
4. Use the features of Step 1 to test the classifier.
5. Repeat Step 1 to 4 for all the patients on the dataset.
6. Compute the posterior probabilities for each classifier.
7. Select the *optimal cut-off* value for the classifier.
8. Compute the final predictions for each polyp image.

	Accuracy	Sensitivity	Specificity	FPR	FNR	AUC
Unbalanced						
Naives Bayes	79.71%	84.56%	56.67%	43.33%	15.44%	**0.71**
k-Nearest Neighbor	75.65%	75.44%	76.67%	23.33%	24.56 %	**0.81**
Random Forest	75.65%	75.79%	75.00%	25.00 %	24.21%	**0.84**
Support Vector Machine	81.45%	81.40%	81.67%	18.33%	18.60%	**0.90**
Balanced						
Naives Bayes	78.55%	85.26%	46.67%	53.33 %	14.74%	**0.67**
k-Nearest Neighbor	74.49%	74.39%	75.00%	25.00%	25.61%	**0.83**
Random Forest	78.55%	78.60%	78.33%	21.67 %	21.40%	**0.85**
Support Vector Machine	81.45%	81.05%	83.33%	16.67 %	18.95%	**0.89**

Table 1: Leave One Out results for polyp malignancy classification using features from GAP-Resnet50.

For each classifier, the accuracy and the area under the curve (AUC) were computed. Moreover, *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. Furthermore, the misclassification rates defined as *false positive rate* (FPR) and *false negative rate* (FNR) respectively, were also computed.

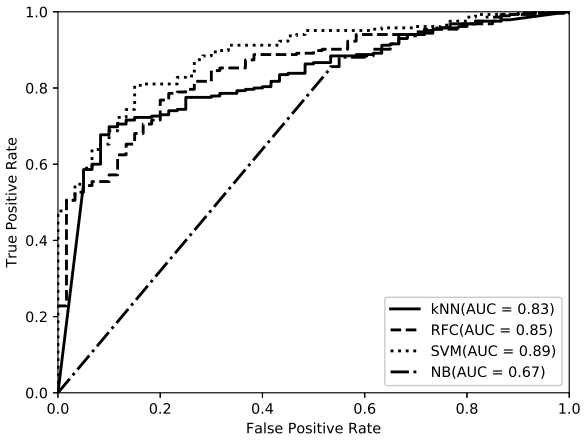

Figure 3: ROC curve for the *balanced* dataset.

The results of LOPOCV for each classifier are presented in Table 1. The best results were achieved using the SVM classifier, for both the unbalanced and balanced dataset. The unbalancing between classes conditioned the selection of a high cut-off value to achieve optimal specificity results. The *per-class* data augmentation strategy achieved small improvements in specificity values, increasing the correct amount of classified benign polyps at the cost of a slight decrease on sensitivity. The observed ROC curve shown in Fig. 3 and the AUC values derived from it, suggest that the proposed models are powerful enough to discriminate between the two different classes, but they still may suffer from deviations towards the higher class, probably explained from the limited amount of data used in the study. Nevertheless, the reported results show promising feasibility for classifying polyp malignancy, when several modalities are combined to achieve an improved prediction of the final class.

4. Conclusions

This paper, has presented an algorithm for polyp malignancy classification. The proposed method is able to classify a polyp between benign and pre-malignant from White Light Endoscopy, and from two recently developed acquisition modalities, Blue Laser Imaging and Linked Color Imaging. Furthermore, an alternative approach to adjust a highly unbalanced dataset is presented, based on separately augmenting the inferior class over the most dominant class, showing that the use of *per-class* data augmentation can substantially improve results. The last layer of a pre-trained Resnet50 with Imagenet is used as feature extractor. Several machine learning classifiers have been evaluated on the CZE dataset. Despite the prominent results, the current processing chain suffers from manual ROI selection. To alleviate this, the use of full-sized images can provide more information towards an end-to-end CNN, decreasing the rate of miss-classification. Moreover, the use of a bigger and more balanced dataset can significantly improve the already feasible results. Overall, the proposed CAD system can provide support to endoscopists with in-vivo polyp classification, possibly avoiding unnecessary resections of benign polyps and avoiding expensive histopathological assessment.

5. Acknowledgment

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

References

[1] A. Asperti and C. Mastronardo. The effectiveness of data augmentation for detection of gastrointestinal diseases from endoscopical images. *CoRR*, abs/1712.03689, 2017. 3

[2] J. E. East, T. Guenther, R. H. Kennedy, and B. P. Saunders. Narrow band imaging avoids potential chromoendoscopy risks. *Gut*, 56(8):1168–1169, 2007. 1

[3] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer*, 136(5):E359–E386, 2015. 1

[4] S. Gross, T. Stehle, A. Behrens, R. Auer, T. Aach, R. Winograd, C. Trautwein, and J. Tischendorf. A comparison of blood vessel features and local binary patterns for colorectal polyp classification. *SPIE Medical Imaging*, 7260:72602Q–72602Q–8, 2009. 2

[5] S. Gross, C. Trautwein, A. Behrens, R. Winograd, S. Palm, H. H. Lutz, R. Schirin-Sokhan, H. Hecker, T. Aach, and J. J. W. Tischendorf. Computer-based classification of small colorectal polyps by using narrow-band imaging with optical magnification. *Gastrointestinal Endoscopy*, 74(6):1354–1359, 2011. 2

[6] O. Har-Noy, L. Katz, T. Avni, R. Battat, T. Bessissow, D. E. Yung, T. Engel, A. Koulaouzidis, R. Eliakim, S. Ben-Horin, and U. Kopylov. Chromoendoscopy, narrow-band imaging or white light endoscopy for neoplasia detection in inflammatory bowel diseases. *Digestive Diseases and Sciences*, 62(11):2982–2990, Nov 2017. 1

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. *Arxiv.Org*, 7(3):171–180, 2015. 2, 3

[8] J. R. Jass. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology*, 50(1):113–130, 2007. 1

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, pages 1–9, 2012. 3

[10] M. Liedlgruber and A. Uhl. Computer-aided decision support systems for endoscopy in the gastrointestinal tract: A review. *IEEE Reviews in Biomedical Engineering*, 4:73–88, 2011. 2

[11] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312, 2014. 2

[12] M. Murata, H. Usami, Y. Iwahori, and W. Aili. Polyp Classification Using Multiple CNN-SVM Classifiers from Endoscope Images. *The Ninth International Conferences on Pervasive Patterns and Applications*, (c):109–112, 2017. 2, 3

[13] H. Nishida, S. Ozawa, K. Shimomura, S. Abe, K. Yoshida, E. Ohashi, and T. Mishima. Development of a New Endoscope platform. Technical Report 62-2017, Fujifilm Research & Development, 2016. 1

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 3

[15] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. 2

[16] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations (ICRL)*, pages 1–14, 2015. 3

[17] T. Tamaki, J. Yoshimuta, M. Kawakami, B. Raytchev, K. Kaneda, S. Yoshida, Y. Takemura, K. Onji, R. Miyaki, and S. Tanaka. Computer-aided colorectal tumor classification in NBI endoscopy: Using local features. *Medical Image Analysis*, 17(1):78–100, 2013. 2

[18] T. Tamaki, J. Yoshimuta, M. Kawakami, B. Raytchev, K. Kaneda, S. Yoshida, Y. Takemura, K. Onji, R. Miyaki, and S. Tanaka. Computer-aided colorectal tumor classification in NBI endoscopy: Using CNN features. 2016. 2

[19] J. J. W. Tischendorf, S. Gross, R. Winograd, H. Hecker, R. Auer, A. Behrens, C. Trautwein, T. Aach, and T. Stehle. Computer-aided classification of colorectal polyps based on vascular patterns: A pilot study. *Endoscopy*, 42(3):203–207, 2010. 2

[20] I. Viovan, M. Tanu, O. Pascu, L. Ciobanu, and A. Tanu. The role of narrow band imaging in colorectal polyp detection. *Bosnian Journal of Basic Medical Sciences*, 17(2):152–158, 2017. 1

[21] Z. Wen, J. Shi, B. He, Q. Li, and J. Chen. ThunderSVM: A fast SVM library on GPUs and CPUs. *To appear in arxiv*, 2018. 3

[22] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell. Understanding Data Augmentation for Classification: When to Warp? *2016 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2016*, 2016. 3

[23] R. Zhang, Y. Zheng, W. C. T. Mak, R. Yu, S. H. Wong, J. Lau, and C. Poon. Automatic Detection and Classification of Colorectal Polyps by Transferring Low-level CNN Features from Non-Medical Domain. *IEEE Journal of Biomedical and Health Informatics*, 21(1):1–1, 2016. 2, 3