



Kudo's Classification for Colon Polyps Assessment Using a Deep Learning Approach

Sebastian Patino-Barrientos ^{1,†}, Daniel Sierra-Sosa ^{2,*,†}, Begonya Garcia-Zapirain ³, Cristian Castillo-Olea ³ and Adel Elmaghraby ²

- ¹ Apolo Scientific Computing Center, Universidad EAFIT, Medellín 50035, Colombia; spatino6@eafit.edu.co
- ² Department of Computer Science and Engineering, University of Louisville, KY 40292, USA; adel@louisville.edu
- ³ eVida Research Group, University of Deusto, 48007 Bilbao, Spain; mbgarciazapi@deusto.es (B.G.-Z.); cristian.castillo@deusto.es (C.C.-O.)
- * Correspondence: desier01@louisville.edu
- + These authors contributed equally to this work.

Received: 26 November 2019; Accepted: 4 January 2020; Published: 10 January 2020



Abstract: Colorectal cancer (CRC) is the second leading cause of cancer death in the world. This disease could begin as a non-cancerous polyp in the colon, when not treated in a timely manner, these polyps could induce cancer, and in turn, death. We propose a deep learning model for classifying colon polyps based on the Kudo's classification schema, using basic colonoscopy equipment. We train a deep convolutional model with a private dataset from the University of Deusto with and without using a VGG model as a feature extractor, and compared the results. We obtained 83% of accuracy and 83% of F1-score after fine tuning our model with the VGG filter. These results show that deep learning algorithms are useful to develop computer-aided tools for early CRC detection, and suggest combining it with a polyp segmentation model for its use by specialists.

Keywords: colon cancer; deep learning; image processing; medical dataset; VGG

1. Introduction

Colorectal cancer (CRC) is one of the most lethal cancer types in the world, is the second leading cause of cancer death in the world and ranks third in incidence with over 881,000 deaths and 1.8 million cases that were estimated to occur in 2018 [1]. This disease begins as a non-cancerous growth of glandular tissue on the colon's or rectum's inner lining known as a adenomatous polyp. Some of these polyps can be seen as neoplasm, if not treated over time, they can cause cancer and death in the worst case. Consequently, there is evidence that is possible to prevent CRC if these polyps are detected and removed in an early stage of the disease [2,3].

Currently, colonoscopy is one of the most reliable methods for screening and diagnosing of CRC [4]. In order to reduce false negatives, the medical personnel must remove all detected polyps during colonoscopy as there is no a perfect method to decide if a polyp is benign or malignant. Because colonoscopy can still leave polyps undetected, several machine learning techniques have been made to provided computer-aided polyp detection systems [5].

In order to reduce costs and avoid unnecessary removal of benign polyps, several automated classification systems have been made. These systems help endoscopists to make an appropriate optical diagnosis of detected polyps. Some automated classification systems currently exists, but they are based on advanced imaging modalities [6]. One technique used to determine the treatment of colon polyps is mucosal surface analysis. Also called pit-pattern analysis, it is categorized using Kudo's classification schema [7]. Those classified as Type I or II are often not removed because they are usually



non malignant. Type III and IV have a low risk factor of being carcinogenic. Type V has a 56% of being malignant [8,9].

Multiple techniques originated from the area of computer vision have been used in the area of computer assisted polyp classification. Most of these are based on hand-crafting feature extraction methods to train a more basic classifier such as support vector machine (SVM) or a k-nearest neighbor (k-NN) classifier. Some known methods used are wavelet based methods [10,11], texture analysis methods based on computing the local fractal dimension [11] and energy maps obtained from the intensity of valleys [12].

Some authors have already experimented using deep learning for polyp detection and classification. In [13] they trained a large, deep convolutional model to classify between "adenomatous polyp-positive" and "adenomatous polyp-negative" classes. Using 1225 polyp images extracted from various digital sources they obtained an accuracy of 80% with a recall of 90% and a precision of 78%. In [14] the authors used transfer learning to obtain a feature extractor for a SVM that classifies between hyperplasia and adenoma polyps. They obtain an accuracy of 85.9% with 87.6% recall and 87.3% precision, however as this method uses support vector machine for classification, training when new data is available consume significant resources. The benefits of using learned features compared to hand-crafted methods is their ability to generalize better to multiple situations and to unfamiliar datasets. This has been proved in other areas such as pedestrian gender recognition [15].

To provide a computer-aided diagnosis tool for CRC prevention, we propose a deep learning algorithm based on any standard colonoscopy equipment that classifies polyps into high risk and low risk of malignancy based on pit patterns described by Kudo. Those classified as Type V are considered high-risk, and those classified as Type IV or below are considered as low-risk.

2. Methodology

Early detection of cancerous polyps allows for reducing the colon cancer incidence rates. High quality endoscopes with high-magnification or other imaging capabilities makes the classification of polyp Types easier. Nonetheless, these devices are not always available, therefore, it is important to develop methods agnostic to the colonoscopy devices to reduce costs and make these tests more affordable for all regions in the world where the access to special equipment to provide a good results is scarce.

Following Kudo's polyp classification [7], we decided to separate polyps between malignant and nonmalignant. Given that the first two categories are not associated with malignancy and Types III and IV have a low risk of containing carcinoma, we decided to group them as nonmalignant. Type V polyps otherwise has a 56% risk of malignancy, so we can confidently label that type as the only one malignant in our test environment [8]. These polyp types can be seen in Figure 1. Based on this, we can train an algorithm that will help in reduction of incidence ratio with a reduced need for intervention. This way specialists could have an additional tool improving the odds of a patient, and potentially reducing colorectal cancer incidence.



Figure 1. Examples of Kudo's classification scale of Polyps ranging from I through V.

For our training and testing dataset, experts associated to the University of Deusto collected and classified a set of 600 images into five different categories of pit patterns defined by Kudo. These images where collected from 142 patients in the Urduliz hospital and Biodonostia Heatlh Research Institute with the required ethics committee approval. Polyp images were extracted from colonoscopy videos

recorded using different colonoscopy cameras and under multiple lighting conditions. The collected images ended up having one of the following resolutions: 576×768 , 576×1047 , 1024×1280 , 1048×1232 , 1072×1440 , 1080×1350 . To make our system general we included a preprocessing step, downscaling the images to a resolution of 150×150 , it should be noted that this process is not necessary in the deep learning model, but enable the system to work efficiently with less computing capacity. The number of images contained in each Kudos' category is presented in the first row of Table 1, due to the limited amount of images per category data augmentation techniques were applied to the dataset. In our experiments we applied rotation, horizontal and vertical shifting, shear, zoom, and mirroring, to the images leading to the number of records presented in the second row of Table 1.

Table 1. Amount of polyps per category.

	Ι	II	III	IV	V
Original dataset	47	90	183	187	93
With augmentation	282	540	1098	1122	558

After checking the final data count, we merge the first four categories and treated them as non malignant polyps, while leaving Type V polyps as malignant for our experiments. We sampled equal quantities from each of the first four classes to obtain a similar amount of images compared to Type V images in order to have a balanced dataset.

Before the rise of deep learning, researchers used to produce hand crafted transformations in order to feed the images in a model such as SVM. A modern convolutional neural network pipeline can be seen as a full feature extractor and a classifier join together in one model. In the previous model, this separation can be noticed where the convolutional layers get flatted and used in a dense layer.

Our plan for classifying the polyps into malignant and non malignant consisted of an incremental approach based on ideas done in similar works such as [14]. We first built a small deep convolutional model to test if it could learn effective transformations in order to extract useful features for a robust polyp classification. We then compared this network with another one using a pre-trained feature extractor based on the VGG-16 neural network architecture [16]. We tested two situations with this model. First, we just added a dense layer on top of the convolutional feature extractor and trained it, freezing the convolutional part so it didn't change during training. After checking the results, we fine tuned the model unfreezing some convolutional layers from the top of the network, that are supposed to represent high level features. We trained the model with this change and observe the results. To measure our results, we divided our data into three sets. 68% of the data as a training set to optimize the model, 16% of the data as a validation set to select the best hyperparameters and another 16% of the data as a test set to obtain the final measurements.

3. Experiments

To feed the images into our machine learning algorithm we did a preprocessing step that consisted of rescaling each channel from RGB images into a $\{0, 1\}$ scale and resizing the images to the 150×150 resolution in order to have a constant size for the model. Additionally, we used image augmentation to improve our results, including rotation, horizontal and vertical shift, shearing, zoom and a horizontal flip as operations. This was conducted for all the experiments presented in this paper.

Additionally, we used a bootstrap undersampling technique in which we randomly selected a subset of the nonmalignant category in the training set and all images from the malignant category in the training set and trained for 10 epochs. After that, we resampled the nonmalignant images without replacement and continued training for 10 more epochs. This process was repeated until all images from the nonmalignant category are sampled.

All experiments were conducted in the deep learning framework Keras [17]. For the small convolutional model we used a sequential architecture based on a series of convolution and max pooling operations. The model was composed by four convolutional layers, each one of them followed by a max pooling layer to make the model translation invariant. After that, we add a dropout layer to reduce overfitting and then add a dense layer. This dense layer is then connected to a sigmoid neuron that will output the probability of the current polyp being malignant. For training we used the RMSprop optimizer with a learning rate of 1×10^{-4} A sketch of the model can be seen in Figure 2.



Figure 2. Proposed CNN architecture as described in text.

Then we used a model with the VGG-16 architecture trained with ImageNet [18] provided by Keras to be our feature extractor. This model received as input a 150×150 image, and passed trough a stack of convolutional layers with receptive filters of 3×3 . We added a dense layer and a sigmoid neuron on top of that to do the prediction. This model was trained using the RMSprop optimizer with a learning rate of 2×10^{-5} .

Finally, we fine tuneed the feature extractor making it part of the whole training pipeline, but freezing most of the top layers of the VGG network. There is a big risk of overfitting when doing this, so we stored the previous model and then tested unfreezing layers one by one. We stayed with only unfreezing three layers to avoid a greater risk of overfit. We trained using the RMSprop optimizer too with a learning rate of 1×10^{-5} .

4. Results and Discussion

Given the necessity of making more accessible these tests around the world, and in order to support endoscopists, a computer-assisted diagnosis tool that doesn't depend on special equipment is the key for solving this problem. In order to reach this solution, we present the results obtained by all three models described in this article. It is important to indicate that we had to reduce our usable data given the unbalanced dataset that we had available for this. Also, separating data between training, validation and testing limits us even more, and given the low amount of data for testing, it is possible that some small bias is present in the evaluation of the models.

We evaluated the models using four metrics: accuracy, precision, recall, and the F1 score. We report the results of these experiments on Table 2, it should be noted that these results are agnostic to the colonoscopy modality, and as we are using a fully connected layer for classification, our results are more general when compared with a SVM classification strategy. As we can see, this incremental approach allows us to obtain increasing gains in every metric. Additionally, these results indicate that models trained using generic images without any medical context whatsoever can still be useful to learn good convolutional transformation for feature extraction and transfer learning. Given that medical datasets are usually too small for training very deep neural networks effectively, this experiment also supports the previously tested hypothesis that CNN features from nonmedical domains can be very effective when used with medical datasets [14].

	Accuracy	Precision	Recall	F1 score
Base CNN Model	0.76	0.83	0.66	0.74
Using VGG as a feature extractor	0.80	0.84	0.73	0.78
Fine tuning	0.83	0.81	0.86	0.83

Table 2. Resulting metrics.

In order to establish a comparison between the proposed model and traditional techniques, we conducted experiments by using both our approach and the classicals. We trained two models using our dataset, a SVM with a very high regularization value (C = 15) because when running without regularization this model proves to be incapable of recognize the classes; and a k-NN with an arbitrary number of neighbours. Both of them used histograms of oriented gradient (HOG) [19] as a feature extractor for the images. The results are presented in Table 3.

Table 3. Resulting metrics for classical models using HOG.

	Accuracy	Precision	Recall	F1 Score
SVM	0.66	0.6	1.0	0.74
k-NN with 15 neighbours	0.66	0.6	1.0	0.74
Proposed approach	0.83	0.81	0.86	0.83

It should be noted that both classical techniques obtained the same results when synthesizing our dataset. This can be interpreted as a train overfit in both cases, proving the generalization capacity that deep convolutional neural networks have. An additional advantage of using the proposed Deep Learning model approach for the colon polyps classification is the possibility of extend these results by creating a framework, aiming to provide a narrative from the obtained results as described in [20].

5. Conclusions

In this paper we show that a small deep CNN model trained from scratch can provide good results for pit pattern classification intended to the colorectal cancer diagnosing. We tested using an incremental approach based on learning a feature extractor as part of the training pipeline, and then using a feature extractor trained from a nonmedical domain. We obtained good results after fine tuning the selected model.

These results show that modern deep learning techniques are already robust enough to provide assistance in computer-aided diagnosis. Given that the consequences of a non-correctly diagnosed polyp can be dying from colon cancer these advancements provide a solution to reduce incidence rate, and in turn, mortality of cancer patients.

This model can be easily extended in future work into a real time algorithm that processes live feed from a magnifying colonoscopy that help provide computer-aided diagnosis of risk. We are currently collecting more records to enhance our dataset, we are confident that results will be improved over time using this additional data and exploring new techniques such as video analysis instead of independent frames. In Figure 3 we show a schema from a computer-aided diagnosis tool for polyp detection and classification based on our work, that is currently being developed.



Figure 3. Diagram of a computer-aided diagnosis tool based on our work on polyp detection and segmentation.

Author Contributions: Conceptualization: D.S.-S., B.G.-Z. and A.E.; methodology: D.S.-S.; investigation: S.P.-B.; Project administration: D.S.-S., B.G.-Z. and A.E.; resources: B.G.-Z., and C.C.-O.; software: S.P.-B.; supervision: D.S.-S.; writing—original draft: S.P.-B.; writing—review and Editing, D.S.-S., S.P.-B. and A.E.; data curation, C.C.-O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basque Government "Aids for health research projects" and the publication fees supported by the Basque Government Department of Education (eVIDA Certified Group IT905-16).

Acknowledgments: The authors would like to thank to Hospital Universitario Donostia (Inés Gil and Luis Bujanda), Hospital Universitario de Cruces (Manuel Zaballa and Ignacio Casado), Hospital Universitario de Basurto (Angel José Calderón and Ana Belen Díaz), Hospital Universitario de Araba (Aitor Orive and Maite Escalante), Hospital de San Eloy (Fidencio Bao and Iñigo Kamiruaga) and Hospital Galdakao (Alain Huerta) health centres and Osakidetza Central (Isabel Idígoras and Isabel Portillo) for their collaboration in the research. The authors also wish to thank the anonymous reviewers for their helpful comments.

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

Abbreviations

The following abbreviations are used in this manuscript:

- CRC Colorectal Cancer
- VGG Visual Geometry Group (Name of the Architecture)
- SVM Support Vector Machine
- k-NN k Nearest Neighbor
- HOG Histograms of Oriented Gradient

References

- 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] [PubMed]
- 2. Shin, Y.; Qadir, H.A.; Aabakken, L.; Bergsland, J.; Balasingham, I. Automatic colon polyp detection using region based deep cnn and post learning approaches. *IEEE Access* **2018**, *6*, 40950–40962. [CrossRef]

- Lieberman, D.A.; Weiss, D.G.; Bond, J.H.; Ahnen, D.J.; Garewal, H.; Harford, W.V.; Provenzale, D.; Sontag, S.; Schnell, T.; Durbin, T.E.; et al. Use of Colonoscopy to Screen Asymptomatic Adults for Colorectal Cancer. *N. Engl. J. Med.* 2000, 343, 162–168. [CrossRef] [PubMed]
- 4. Navarro, M.; Nicolas, A.; Ferrandez, A.; Lanas, A. Colorectal cancer population screening programs worldwide in 2016: An update. *World J. Gastroenterol.* **2017**, *23*, 3632. [CrossRef] [PubMed]
- 5. Sanchez-Gonzalez, A.; Garcia-Zapirain, B.; Sierra-Sosa, D.; Elmaghraby, A. Automatized colon polyp segmentation via contour region analysis. *Comput. Biol. Med.* **2018**, *100*, 152–164. [CrossRef] [PubMed]
- Mori, Y.; Kudo, S.e.; Berzin, T.M.; Misawa, M.; Takeda, K. Computer-aided diagnosis for colonoscopy. *Endoscopy* 2017, 49, 813–819. [CrossRef] [PubMed]
- 7. Kudo, S.E.; Tamura, S.; Nakajima, T.; Yamano, H.o.; Kusaka, H.; Watanabe, H. Diagnosis of colorectal tumorous lesions by magnifying endoscopy. *Gastrointest. Endosc.* **1996**, *44*, 8–14. [CrossRef]
- 8. Neilson, L.J.; Rutter, M.D.; Saunders, B.P.; Plumb, A.; Rees, C.J. Assessment and management of the malignant colorectal polyp. *Frontline Gastroenterol.* **2015**, *6*, 117–126. [CrossRef] [PubMed]
- 9. Tanaka, S.; Kaltenbach, T.; Chayama, K.; Soetikno, R. High-magnification colonoscopy (with videos). *Gastrointest. Endosc.* **2006**, *64*, 604–613. [CrossRef] [PubMed]
- 10. Wimmer, G.; Tamaki, T.; Tischendorf, J.J.; Häfner, M.; Yoshida, S.; Tanaka, S.; Uhl, A. Directional wavelet based features for colonic polyp classification. *Med. Image Anal.* **2016**, *31*, 16–36. [CrossRef] [PubMed]
- 11. Häfner, M.; Tamaki, T.; Tanaka, S.; Uhl, A.; Wimmer, G.; Yoshida, S. Local fractal dimension based approaches for colonic polyp classification. *Med. Image Anal.* **2015**, *26*, 92–107. [CrossRef] [PubMed]
- Bernal, J.; Sánchez, F.J.; Fernández-Esparrach, G.; Gil, D.; Rodríguez, C.; Vilariño, F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computer. Med. Imaging Graph.* 2015, 43, 99–111. [CrossRef] [PubMed]
- Li, T.; Cohen, J.; Craig, M.; Tsourides, K.; Mahmud, N.; Berzin, T.M. Mo1979 The Next Endoscopic Frontier: A Novel Computer Vision Program Accurately Identifies Colonoscopic Colorectal Adenomas. *Gastrointest. Endosc.* 2016, 83, AB482. [CrossRef]
- Zhang, R.; Zheng, Y.; Mak, T.W.C.; Yu, R.; Wong, S.H.; Lau, J.Y.; Poon, C.C. Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain. *IEEE J. Biomed. Health Inform.* 2016, *21*, 41–47. [CrossRef] [PubMed]
- Antipov, G.; Berrani, S.A.; Ruchaud, N.; Dugelay, J.L. Learned vs. hand-crafted features for pedestrian gender recognition. In Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 26–30 October 2015; pp. 1263–1266.
- 16. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
- 17. Available online: https://keras.io (accessed on 8 January 2019).
- Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, USA, 22–24 June 2009; pp. 248–255.
- Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of the International Conference on Computer Vision & Pattern Recognition (CVPR'05), IEEE Computer Society, San Diego, CA, USA, 20–26 June 2005; Volume 1, pp. 886–893.
- Holzinger, A.; Kieseberg, P.; Weippl, E.; Tjoa, A.M. Current Advances, Trends and Challenges of Machine Learning and Knowledge Extraction: From Machine Learning to Explainable AI. In *Springer Lecture Notes in Computer Science LNCS*; Springer: Cham, Switzerland, 2018; Volume 11015, pp. 1–8.



© 2020 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 (http://creativecommons.org/licenses/by/4.0/).