Towards an interpretable model for automatic classification of endoscopy images

Towards an interpretable model for automatic classification of endoscopy images

Rogelio García-Aguirre[1], Luis Torres-Treviño[1], Eva María Navarro-López[2,3], José Alberto González-González[4]

[1] Universidad Autónoma de Nuevo León, Facultad de Ingeniería Mecánica y Eléctrica, Ave. Universidad S/N, Cd. Universitaria, San Nicolás de los Garza, N.L. CP 66455 México.
[2] School of Environment, Education and Development, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom
[3] School of Engineering, Computing and Mathematical Sciences, Faculty of Science and Engineering, University of Wolverhampton, Alan Turing Building, Wulfruna Street, Wolverhampton WV1 1LY, United Kingdom
[4] Servicio de Gastroenterología, Facultad de Medicina, Hospital Universitario "Dr. José E. González", Universidad Autónoma de Nuevo León, Monterrey, N.L. CP 64700 México

Abstract. Deep learning strategies have become the mainstream for computer-assisted diagnosis tools development since they outperform other machine learning techniques. However, these systems can not reach their full potential since the lack of understanding of their operation and questionable generalizability provokes mistrust from the users, limiting their application. In this paper, we generate a Convolutional Neural Network (CNN) using a genetic algorithm for hyperparameter optimization. Our CNN has state-of-the-art classification performance, delivering higher evaluation metrics than other recent papers that use AI models to classify images from the same dataset. We provide visual explanations of the classifications made by our model implementing Grad-CAM and analyze the behavior of our model on misclassifications using this technique.

Keywords: Interpretability · Convolutional Neural Networks · Endoscopy images

1 Introduction

We are living the third artificial intelligence (AI) boom [5,21]. Areas such as computer vision (CV) and natural language processing (NLP) have undergone considerable progress due to the deep learning (DL) schemes developed during the last decade. Referring to CV, Deep Neural Networks (DNN) have exceeded human performance in many applications [2], including medical image classification [1]. Nevertheless, DNNs are far from perfect. Numerous studies have stated their concerns about the relevance of the existing DL models in real-world applications. The focal limitations found for these systems are their interpretability

scarcity [2,9,16,20] (often referred to as the "black box" condition [11,21,25]) and questionable generalizability [2,20,26,28].

In medicine, image-intensive specialties have benefitted from AI systems [15]. For the particular matter of endoscopy, numerous publications describe the current state of the applications and expectations of AI for this field [1,3,4,7,17]. Nevertheless, DL systems are usually unsuitable for clinical application due to the previously stated limitations of these systems (interpretability scarcity and questionable generalizability). As DL models become more and more present for critical applications (such as medicine), model interpretability has been suggested as a solution for the black box condition. Nevertheless, many papers lack a definition of interpretability [13]. For that matter, Lipton [13] stated that *interpretability is not a monolithic concept but reflects several distinct ideas.*

We take the following definition: the interpretability of an AI system refers to the possibility for a human to understand the relation between the system's predictions and the information used to make those predictions [22]. In that sense, for AI applications in medical image classification, the interpretability purpose is not to understand every part of an AI system but to have enough information for the assigned work [22]. Hence, the radiology field requires task-specific interpretability solutions with clinically oriented validations [22]. Following that idea, Reyes et al. [22] concluded that *saliency maps can be integrated easily into the radiology workflow because they work at the voxel level; hence, these visualization maps can be fused or merged with patient images and computer-generated results.*

In this work, we aim to explore the capabilities of the current AI tools to develop a system for automatic endoscopy image classification with the potential of having a clinical application. To that end, we construct an optimized Convolutional Neural Network (CNN) using a framework [6] based on genetic algorithms that perform hyperparameter optimization to increase the CNN classification performance as the model generalizability. Then, as an interpretability method, we apply the Grad-CAM technique [23] to the network to generate heat map-like images that aid the visualization of the relevant zones in the input images for the classification using the optimized CNN. We use the KVASIR dataset of endoscopy images to develop the optimized CNN and test the interpretability method. The classification performance of our optimized CNN is state-of-the-art, and the visualizations using the Grad-CAM technique can locate the regions relevant for the correct classifications.

2 Related work

2.1 Classification of endoscopy images

Several studies are using AI to analyze endoscopic images. A great deal of these focuses on a specific gastrointestinal finding, such as polyp detection and segmentation (e.g., [24]), gastric cancer detection and diagnosis (e.g., [14]), diagnosis and detection of Helicobacter Pylori infection (e.g., [29]), among others. The

publication in 2017 of the KVASIR dataset [19], consisting of 8000 images of different GI findings in images of upper endoscopy, made possible the development of a new generation of algorithms for endoscopic image classification. These studies aim to achieve a general classification of the different GI findings that can appear during endoscopy instead of concentrating on a particular suffering or symptom. For a detailed review of papers using AI to classify images of the gastrointestinal tract, refer to Jha et.al [10].

2.2 Interpretability

Interpretability is a critical research topic for AI due to the rise of DL approaches during the last years [22]. There are different kinds of interpretability methods, and this area is continually growing. Nevertheless, numerous interpretability methods have not yet reached the radiology AI systems [22].

In this paper, we focus on providing visual explanations (often refer as saliency maps), which is the typical form of explainability in medical image analysis [27]. Saliency maps are often gradient-based techniques [27], which foundation is the assumption that the magnitude of the gradients correlates with the contribution of voxels to a model's prediction [22]. For an overview of methods for interpretability of DL for medical image analysis, refer to van der Velden et.al. [27].

3 Methods and implementation

3.1 Dataset

For our approach development and evaluation, we used the KVASIR dataset [19], which consists of 8000 images of the gastrointestinal tract insides. This dataset includes anatomical landmarks, pathological findings, procedures, and normal findings. All the images in this dataset belong to one of the following classes: dyed lifted polyps, dyed resection margins, esophagitis, normal cecum, normal pylorus, normal z-line, polyps, and ulcerative colitis. We divided the dataset into three partitions: training, validation, and test. Each partition has 4800, 2000, and 1200 images, respectively.

3.2 Optimized CNN

We performed hyperparameter optimization on off-the-shelf CNNs based on the genetic algorithm presented in [6]. This approach automatically generates CNNs for image classification. The algorithm performs in parallel the training of the CNNs using gradient-based optimization and hyperparameter optimization with a genetic algorithm. The two optimization procedures have different optimization targets and use distinct partitions of the dataset to promote the generalizability of the resulting models. The hyperparameters of the resulted CNN are listed in table 1.

Table 1. Hyperparameters of the resulting CNN using the approach described in [6].

Resulted hyperparameters							
a_γ	b_γ	c_γ	a_δ	b_δ	c_δ	Architecture	Loss function Optimizer
0.0867	6.1762	0.5883	0.5547	2.9659	0.6659	ResNext-50 32x4d	Logit penalty AdaMax

The hyperparameters a_γ, b_γ, and c_γ control the magnitude of the learning rate as a function of the training epoch and the total number of epochs. Similarly, a_δ, b_δ, and c_δ control the trainable layers of the CNN during training. The CNN architecture is the base CNN model. The loss function and optimizer are the hyperparameters used for the training using gradient-based optimization.

3.3 Grad-CAM

We used the Grad-CAM technique [23] to produce *visual explanations* of the classifications made by our CNN. This method is based on the fact that deeper layers in CNNs specialize in higher-level visual features and that convolutional layers maintain the spatial information of the input data [23].

Grad-CAM uses the gradients of the logit (of the class that is desired to know the Grad-CAM) with respect to the activations of the last convolutional layer [23]. Then, these gradients are global-average-pooled over the height and width [23]. These values function as weights that denote the importance of the feature maps in the last layer with respect to the given logit [23]. Then the linear combination of the weighted activations of the last layer is calculated. Finally, these values pass though a ReLU activation function to eliminate the negative values [23]. This is because the Grad-CAM focuses on elucidating the regions in the image that evoke a positive value for the given class's logit, and negative values of the weighted activations are assumed to represent regions in the input data that promote a positive value for the logits of other classes [23].

4 Results

4.1 Experimental settings

The experiments were carried out using the following hardware specifications: AMD Ryzen 5 3400G CPU, one NVIDIA GeForce GTX 1660 Ti GPU, 16 GB RAM, and 476 GB system memory. All the algorithms were implemented in Python 3.8.5, using the environment Spyder 4.1.5 and Pytorch 1.7.1 for the CNN modules and gradient-based optimization algorithms.

4.2 Optimized CNN classification performance

The Optimized CNN resulted of the implementation of the method described in Section 3.2 is a ResNext-50 32x4d. We evaluated the classification performance of the optimized CNN using the standar evaluation metrics: recall (REC),

specificity ($SPEC$), accuracy (ACC), precision ($PREC$), Matthews correlation coefficient (MCC), and F_1 value ($F1$).

$$REC = \frac{TP}{TP + FN} \tag{1}$$

$$SPEC = \frac{TN}{TN + FP} \tag{2}$$

$$PREC = \frac{TP}{TP + FP} \tag{3}$$

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{\rho}} \tag{5}$$

with $\rho = (TP + FN)(TN + FP)(TP + FP)(TN + FN)$

$$F1 = 2 \times \frac{PREC \times REC}{PREC + REC} \tag{6}$$

In the above, TP, TN, FP and, FN stand for *true positive*, *true negative*, *false positive* and, *false negative*, respectively. Table 2 shows the evaluation metrics of the optmized CNN using the three partitions of the dataset. Figure 1 shows the confusion matrix of the optimized CNN using the test partition. Figure 2 shows the evaluation metrics of the optimized CNN using the test partition as recommended in [26] to facilitate the visualization of the CNN classification performance.

Table 2. Evaluation metrics of the optimized CNN using the test set.

Data partition	ACC	REC	SPEC	PREC	F1	MCC
Train	0.9962	0.9850	0.9978	0.9850	0.9850	0.9828
Validation	0.9853	0.9415	0.9916	0.9415	0.9415	0.9331
Test	0.9860	0.9441	0.9920	0.9441	0.9442	0.9362

4.3 Visual explanations using Grad-CAM

Using the Grad-CAM technique and image processing, we generated heat map-like images to aid the visualization of the zones in the input images that evoke the CNN classification output. Figure 3 shows an example of the heat map-like images for a correctly classified image per class in the dataset.

We can also construct these heat map-like images to analyze the misclassification made by the optimized CNN. Figure 4 shows examples of misclassified images and compares the heat map-like of the input images for the output logitt of the wrong prediction and the output logit of the true class.

True class \ Predicted class	0	1	2	3	4	5	6	7
0	147	3	0	0	0	0	0	0
1	7	143	0	0	0	0	0	0
2	0	0	128	0	1	21	0	0
3	0	0	0	147	0	0	3	0
4	0	0	1	0	149	0	0	0
5	0	0	18	0	0	132	0	0
6	0	0	0	4	3	0	141	2
7	0	0	0	2	2	0	0	146

Fig. 1. Confusion matrix of the optimized CNN using the test partition. The classes' codes are: 0-dyed lifted polyps, 1-dyed resection margins, 2-esophagitis, 3-normal cecum, 4-normal pylorus, 5-normal z-line, 6-polyps, and 7-ulcerative colitis.

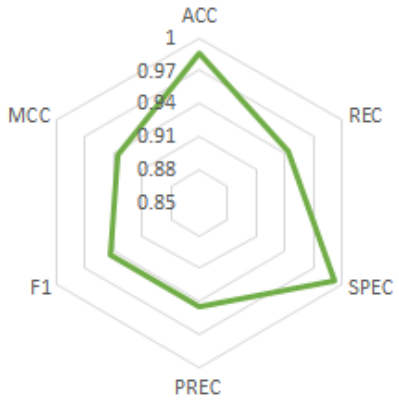

Fig. 2. Evaluation metrics of the optimized CNN using the test partition.

5 Discussion and conclusions

The classification performance of our optimized CNN is state-of-the-art. Table 3 shows the evaluation metrics of other recent papers using AI models to classify images of the same dataset.

The work presented by Hicks et.al [8] also used a CNN to classify the KVASIR dataset images and the Grad-CAM technique for visualization. The main difference between that paper and ours is the method used to develop the CNN. Hicks et.al [8] used a VGG-19 with no reported methodology to choose the other hyperparameters. Instead, we performed hyperparameter optimization based on genetic algorithms with the specific aim of improving the model classification performance and generalizability [6], taking into account that high classification

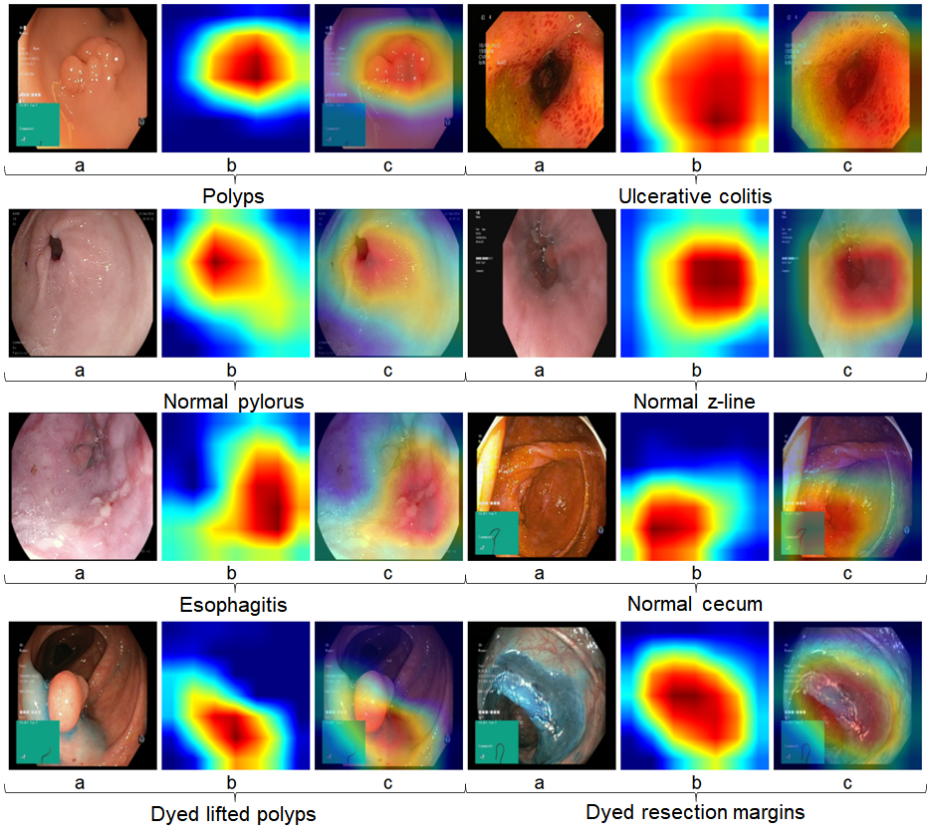

Fig. 3. Examples of every class correctly classified using the optimized CNN and its corresponding Grad-CAM: "a" correspons to the original image, "b" is the Grad-CAM rezied to match "a", and "c" is the superimpossed Grad-CAM "b" over "a".

performance is always desired, and the lack of generalization is currently one focal limitation for the adoption of these kind of systems into clinical practice.

Hicks et.al [8] used the visualizations generated with the Grad-CAM to find properties in the input images that were evoking misclassifications. Then, with this information, they used a preprocessing designed to correct the CNN behavior on the misclassified images, achieving a significant improvement in the CNN classification performance, proving that the visualizations generated using Grad-CAM can help understand both the areas in an image that evoke a specific classification and posible misbehaviors of the model. Table 4 provides a comparison of the highest evaluation metrics achieved by Hicks et.al. [8] and ours.

From the confusion matrix of our optimized CNN (Figure 1), we can observe that the majority of misclassification involve an anatomical landmark (normal z-

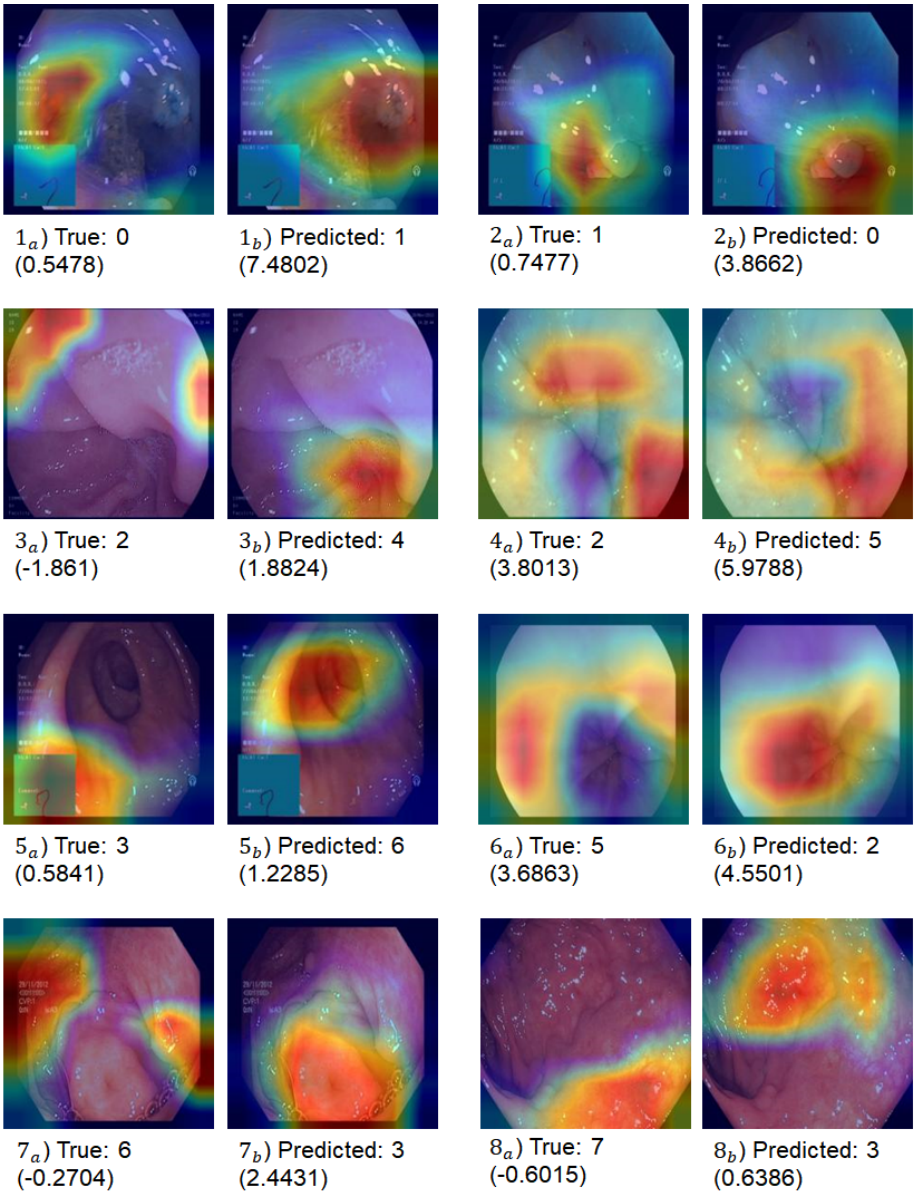

Fig. 4. Examples of misclassifications. The subscript a is for the original images superimposed with the Grad-CAM of the true class, and the subscript b is for the original images superimposed with the Grad-CAM of the predicted class. In parenthesis is the output logit for the given class. The classes' codes are: 0-dyed lifted polyps, 1-dyed resection margins, 2-esophagitis, 3-normal cecum, 4-normal pylorus, 5-normal z-line, 6-polyps, and 7-ulcerative colitis.

Table 3. Evaluation metrics of recent studies using the KVASIR dataset.

Study	Year	ACC	REC	SPEC	PREC	F1	MCC	FPS
Lafraxo et.al. [12]	2020	0.9680	0.8770	—	0.874	0.876	—	—
Ozturk et.al. [18]	2020	0.9790	0.9232	0.9910	0.9446	0.9264	—	—
Our optimized CNN	2022	0.9860	0.9441	0.9920	0.9441	0.9442	0.9362	35.5

Table 4. Evaluation metrics of studies that use a CNN to classify images of the KVASIR dataset and use a visualization technique for interpretability.

Study	Year	ACC	REC	SPEC	PREC	F1	MCC	FPS
Hicks et.al. [8]	2018	0.9440	0.7980	0.7530	0.9680	0.7780	0.7780	—
Our optimized CNN	2022	0.9860	0.9441	0.9920	0.9441	0.9442	0.9362	35.5

line) and a pathological finding (esophagitis), and the misclassification examples shown in Figure 4 illustrated that the CNN is focusing in different regions of the images for the classification of that two classes. Also, both classes have a positive logit in the examples of misclassification shown in Figure 4. That means that the CNN determines that the image belongs to both classes, but the current operation mode of the CNN is to classify the image only in the class with the highest logit. Since the z-line is in the esophagus and the esophagitis is a pathology of it. It would be interesting to have a gastroenterologist assess if the misclassified images between these two classes, in fact, have both findings (esophagitis and normal z-line), as the positive logits suggest.

In this work, we used the existing techniques in the literature to develop a system for endoscopy image classification with interpretability criteria. Our optimized CNN has state-of-the-art classification performance, delivering higher evaluation metrics than other recent papers that use AI models to classify images from the same dataset. The visualizations constructed using the Grad-CAM provide information on the regions that evoke a given output logit. However, it is important to collaborate with physicians to fully understand the implications of the visualizations. In future work, we can explore other approaches, such as Prototype-based interpretation to provide explanations by example or adopt a holistic approach, combining different forms of explanation.

References

1. Alagappan, M., Brown, J., Mori, Y., Berzin, T.: Artificial intelligence in gastrointestinal endoscopy: The future is almost here. World Journal of Gastrointestinal Endoscopy **10**, 239–249 (10 2018). https://doi.org/10.4253/wjge.v10.i10.239
2. Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L.: Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data **8**(1), 53 (dec 2021). https://doi.org/10.1186/s40537-021-00444-8, https://doi.org/10.1186/s40537-021-

00444-8 https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8
3. Berzin, T., Parasa, S., Wallace, M., Gross, S., Repici, A., Sharma, P.: Position statement on priorities for artificial intelligence in gi endoscopy: a report by the asge task force. Gastrointestinal Endoscopy **92** (06 2020). https://doi.org/10.1016/j.gie.2020.06.035
4. Chahal, D., Byrne, M.: A primer on artificial intelligence and its application to endoscopy. Gastrointestinal Endoscopy **92** (05 2020). https://doi.org/10.1016/j.gie.2020.04.074
5. Fujita, H.: AI-based computer-aided diagnosis (AI-CAD): the latest review to read first. Radiological Physics and Technology **13**(1), 6–19 (2020). https://doi.org/10.1007/s12194-019-00552-4, https://doi.org/10.1007/s12194-019-00552-4
6. García-Aguirre, R., Torres-Treviño, L., Navarro-López, E.M., González-González, J.A.: Automatic generation of optimized convolutional neural networks for medical image classification using a genetic algorithm (2022). https://doi.org/10.2139/ssrn.4167905
7. Gross, S., Sharma, P., Pante, A.: Artificial intelligence in endoscopy. Gastrointestinal Endoscopy **91** (12 2019). https://doi.org/10.1016/j.gie.2019.12.018
8. Hicks, S., Riegler, M., Pogorelov, K., Anonsen, K.V., de Lange, T., Johansen, D., Jeppsson, M., Ranheim Randel, K., Losada Eskeland, S., Halvorsen, P.: Dissecting deep neural networks for better medical image classification and classification understanding. In: 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS). pp. 363–368 (2018). https://doi.org/10.1109/CBMS.2018.00070
9. Jha, D., Ali, S., Hicks, S., Thambawita, V., Borgli, H., Smedsrud, P.H., de Lange, T., Pogorelov, K., Wang, X., Harzig, P., Tran, M.T., Meng, W., Hoang, T.H., Dias, D., Ko, T.H., Agrawal, T., Ostroukhova, O., Khan, Z., Atif Tahir, M., Liu, Y., Chang, Y., Kirkerød, M., Johansen, D., Lux, M., Johansen, H.D., Riegler, M.A., Halvorsen, P.: A comprehensive analysis of classification methods in gastrointestinal endoscopy imaging. Medical Image Analysis **70**, 102007 (2021). https://doi.org/https://doi.org/10.1016/j.media.2021.102007, https://www.sciencedirect.com/science/article/pii/S1361841521000530
10. Jha, D., Ali, S., Hicks, S., Thambawita, V., Borgli, H., Smedsrud, P.H., de Lange, T., Pogorelov, K., Wang, X., Harzig, P., Tran, M.T., Meng, W., Hoang, T.H., Dias, D., Ko, T.H., Agrawal, T., Ostroukhova, O., Khan, Z., Atif Tahir, M., Liu, Y., Chang, Y., Kirkerød, M., Johansen, D., Lux, M., Johansen, H.D., Riegler, M.A., Halvorsen, P.: A comprehensive analysis of classification methods in gastrointestinal endoscopy imaging. Medical Image Analysis **70**, 102007 (2021). https://doi.org/https://doi.org/10.1016/j.media.2021.102007, https://www.sciencedirect.com/science/article/pii/S1361841521000530
11. Kochhar, G.S., Carleton, N.M., Thakkar, S.: Assessing perspectives on artificial intelligence applications to gastroenterology. Gastrointestinal Endoscopy **93**(4), 971–975.e2 (2021). https://doi.org/10.1016/j.gie.2020.10.029, https://doi.org/10.1016/j.jns.2019.116544
12. Lafraxo, S., El Ansari, M.: Gastronet: Abnormalities recognition in gastrointestinal tract through endoscopic imagery using deep learning techniques. In: 2020 8th International Conference on Wireless Networks and Mobile Communications (WINCOM). pp. 1–5 (2020). https://doi.org/10.1109/WINCOM50532.2020.9272456
13. Lipton, Z.C.: The mythos of model interpretability (2016). https://doi.org/10.48550/ARXIV.1606.03490, https://arxiv.org/abs/1606.03490

14. Luo, H., Xu, G., Li, C., He, L., Luo, L., Wang, Z., Jing, B., Deng, Y., Jin, Y., Li, Y., Tan, W., He, C., Seeruttun, S., Wu, Q., Huang, J., Huang, D.w., Chen, B., Lin, S.b., Xu, R.h.: Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. The Lancet Oncology **20** (10 2019). https://doi.org/10.1016/S1470-2045(19)30637-0
15. Maddox, T., Rumsfeld, J., Payne, P.: Questions for artificial intelligence in health care. JAMA **321** (12 2018). https://doi.org/10.1001/jama.2018.18932
16. Miotto, R., Wang, F., Wang, S., Jiang, X., Dudley, J.T.: Deep learning for healthcare: review, opportunities and challenges. Briefings in bioinformatics **19 6**, 1236–1246 (2018)
17. Mori, Y., Kudo, s.e., Mohmed, H., Misawa, M., Ogata, N., Itoh, H., Oda, M., Mori, K.: Artificial intelligence and upper gastrointestinal endoscopy: Current status and future perspective. Digestive Endoscopy **31** (12 2018). https://doi.org/10.1111/den.13317
18. Öztürk, S., Özkaya, U.: Gastrointestinal tract classification using improved LSTM based CNN. Multimedia Tools and Applications **79**(39-40), 28825–28840 (2020). https://doi.org/10.1007/s11042-020-09468-3
19. Pogorelov, K., Randel, K.R., Griwodz, C., Eskeland, S.L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D.T., Lux, M., Schmidt, P.T., Riegler, M., Halvorsen, P.: Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In: Proceedings of the 8th ACM on Multimedia Systems Conference. pp. 164–169. MMSys'17, ACM, New York, NY, USA (2017). https://doi.org/10.1145/3083187.3083212, http://doi.acm.org/10.1145/3083187.3083212
20. Prevedello, L., Halabi, S., Shih, G., wu, C., Kohli, M., Chokshi, F., Erickson, B., Kalpathy-Cramer, J., Andriole, K., Flanders, A.: Challenges related to artificial intelligence research in medical imaging and the importance of image analysis competitions. Radiology: Artificial Intelligence **1**, e180031 (01 2019). https://doi.org/10.1148/ryai.2019180031
21. Quinn, T.P., Jacobs, S., Senadeera, M., Le, V., Coghlan, S.: The three ghosts of medical ai: Can the black-box present deliver? Artificial Intelligence in Medicine p. 102158 (2021). https://doi.org/https://doi.org/10.1016/j.artmed.2021.102158, https://www.sciencedirect.com/science/article/pii/S0933365721001512
22. Reyes, M., Meier, R., Pereira, S., Silva, C., Dahlweid, MD, P.M., Tengg-Kobligk, H., Summers, R., Wiest, R.: On the interpretability of artificial intelligence in radiology: Challenges and opportunities. Radiology: Artificial Intelligence **2**, e190043 (05 2020). https://doi.org/10.1148/ryai.2020190043
23. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: Visual explanations from deep networks via gradient-based localization. International Journal of Computer Vision **128**(2), 336–359 (oct 2019). https://doi.org/10.1007/s11263-019-01228-7, https://doi.org/10.1007%2Fs11263-019-01228-7
24. Shin, Y., Balasingham, I.: Automatic polyp frame screening using patch based combined feature and dictionary learning. Computerized Medical Imaging and Graphics **69**, 33–42 (08 2018). https://doi.org/10.1016/j.compmedimag.2018.08.001
25. Stead, W.W.: Clinical implications and challenges of artificial intelligence and deep learning. JAMA **320 11**, 1107–1108 (2018)
26. Thambawita, V., Jha, D., Hammer, H.L., Johansen, H.D., Johansen, D., Halvorsen, P., Riegler, M.A.: An extensive study on cross-dataset bias and evaluation metrics interpretation for machine learning applied to gastrointestinal tract

abnormality classification. ACM Trans. Comput. Healthcare **1**(3) (6 2020). https://doi.org/10.1145/3386295, https://doi.org/10.1145/3386295
27. van der Velden, B.H., Kuijf, H.J., Gilhuijs, K.G., Viergever, M.A.: Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Medical Image Analysis **79**, 102470 (jul 2022). https://doi.org/10.1016/j.media.2022.102470, https://doi.org/10.1016%2Fj.media.2022.102470
28. Yao, A., Cheng, D., Pan, I., Kitamura, F.: Deep learning in neuroradiology: A systematic review of current algorithms and approaches for the new wave of imaging technology. Radiology: Artificial Intelligence **2**, e190026 (03 2020). https://doi.org/10.1148/ryai.2020190026
29. Yasuda, T., Hiroyasu, T., Hiwa, S., Okada, Y., Hayashi, S., Nakahata, Y., Yasuda, Y., Omatsu, T., Obora, A., Kojima, T., Ichikawa, H., Yagi, N.: Potential of automatic diagnosis system with linked color imaging for diagnosis of Helicobacter pylori infection. Digestive Endoscopy **32**(3), 373–381 (2020). https://doi.org/10.1111/den.13509