ISSN 1870-4069

Hyperparameter Optimization for Transfer Learning in Gastrointestinal Image Classification Using an Evolutionary Algorithm: Proof of Concept

Rogelio García-Aguirre, Luis Torres-Treviño, Alberto Gonzáles-González

Universidad Autónoma de Nuevo León,
Facultad de Ingeniería Mecánica y Eléctrica,
San Nicolás de los Garza, N.L.,
Mexico

`rogelio.garciag@uanl.edu.mx`

Abstract. Endoscopy is the foundation for the diagnosis and treatment of several gastrointestinal ailments. However, it is an operator-dependent procedure. The quality of assessment of endoscopy images relies on the physician's experience, ability, and conditions. A system capable of automatically evaluate endoscopy images and classify different gastrointestinal findings within them is an alternative to enhance the diagnosis quality. Convolutional neural networks (CNN) show great promise to this end. Research on this topic focuses on the generation of new complex architectures. Several publications have stated their concerns about the capability of these state-of-the-art schemes to be deployed in a clinical setting. Mainly due to the uncertainty to employ them in real-time because of the computing power these algorithms need. In this study, we performed hyperparameter optimization during the transfer learning and fine-tuning process using off-the-shelf CNN (more likely to operate in real-time) for the image classification task. To this purpose, we use an evolutionary algorithm. We provide preliminary results to this method, proving that this approach may reach classification performance competitive with the novel deep learning structures while maintaining low complexity in the architecture.

Keywords: Hyperparameter optimization, evolutionary algorithm, medical images, gastrointestinal tract, deep learning.

1 Introduction

In recent years, research on automatic medical image classification has gained significant importance. The implementation of Artificial Intelligence (AI) in medicine has been successful in image-intensive specialties, such as radiology, pathology, ophthalmology, and cardiology [22]. Several publications have reported the current state and expectations of such tools in the area of gastroenterology, in particular for endoscopy [1,3,6,8,23]. Endoscopy is the foundation for the diagnosis and treatment of diseases of the gastrointestinal (GI) tract. This procedure is operator-dependent, which generates substantial interobserver variation in the detection and assessment of GI findings [19]. So, the detection of GI lesions essentially relies on the expertise of the physician [28].

An automated system capable of classifying different GI findings would help to reduce the variation in endoscopists' performance [19]. The development of computer-aided diagnosis (CAD) systems with this purpose is currently an open challenge since the feasibility, effectiveness, and safety of CAD for upper gastrointestinal endoscopy in clinical practice remain unknown [23]. There are different approaches for classifying endoscopic images. Deep learning (DL) techniques usually outperform strategies that use hand-crafted features [23,28]. Consequently, convolutional neural networks (CNN) are the most employed method nowadays.

The majority of research focuses on proposing new architectures or combining existing frameworks to enhance classification/detection performance. However, the systems must allow operating in real-time to achieve the final goal, which is assistance in real-time during endoscopy [1,8,16,23]. Numerous state-of-the-art architectures run too slow to be implemented in a clinical setting [19]. Therefore, optimizing the performance of off-the-shelf architectures (that may allow real-time operation) is a possible solution to this.

We utilize less complex deep architectures to classify endoscopic images of the KVASIR dataset [24]. We implement transfer learning in different off-the-shelf CNN. Formerly, we aim to optimize hyperparameters to improve the classification performance of the algorithms. Borgli et al. [5] presented a similar scheme, yet, they carried out Bayesian optimization. In contrast, we propose to use an evolutionary algorithm for this purpose.

2 Related Work

Several studies are using AI to analyze endoscopic images. A great deal of these focuses on a specific GI finding, such as polyp detection and segmentation (e.g., [26]), gastric cancer detection and diagnosis (e.g., [20]), diagnosis and detection of Helicobacter Pylori infection (e.g., [30]), among others. The publication in 2017 of the KVASIR dataset [24], 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.

2.1 Dataset

Machine learning (ML) and DL schemes need datasets to be developed, validated, tested, and compared. With this in mind, Pogolerov et al. [24] created and published the KVASIR dataset. The original version of this dataset consists of 8000 images from inside the GI tract. This dataset contains anatomical landmarks, pathological findings, procedures, and normal findings. In concrete, the images belong to one of the following classes: z-line, pylorus, cecum, esophagitis, polyps, ulcerative colitis, dyed lifted polyps, dyed resection margin, normal colon mucosa, and stool. Each class has 1000 images of it. Hence, the dataset is balanced.

2.2 Evaluation Metrics

There exist standard evaluation metrics used to assess classification algorithms. We take the performance measures from [16]. Since it is the most complete and recent paper describing and comparing the performance of different methods for automatic endoscopic image classification. The evaluation metrics are as follows: recall (REC), specificity (SPEC), accuracy (ACC), precision (PREC), Matthews correlation coefficient (MCC), and F_1 value (F1):

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

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

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

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

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

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

In the above, *TP*, *TN*, *FP* and, *FN* stand for *true positive*, *true negative*, *false positive* and, *false negative*, respectively.

2.3 Classification algorithms

To compare and measure the performance of the proposed method, in this investigation, we only consider studies that use one version of the KVASIR dataset for training, validation, and testing. Table 1 shows the studies that use some version of the KVASIR dataset and have the best classification performance.

During the last years, different paradigms for the classification of endoscopic images came to be tested. CNN architectures are the current most effective approach. Regarding classification performance, hybrid architectures have gained notice. Chang et al. [7] denoted the importance of applying an adequate data augmentation technique. In [7], they developed an algorithm to automatically

Table 1. Studies with the best evaluation metrics for endoscopic image classification using some version of the KVASIR dataset. Metrics as presented in [16].

Autor	Architecture	MCC	ACC	F1
Chang et al., 2019 [7]	ResNet34 [10] + SE-ReNext [14] + Attention-inception-v3 [27]	0.9520	0.9946	0.9569
Harzing et al., 2019 [9]	MobileNetV2 [25]	0.9490	0.9936	0.9105
Luo et al., 2019 [21]	10 CNN + LightGBM [17]	0.9480	0.9941	0.9533
Hoang et al., 2019 [13]	ResNet-101 [10] + Faster R-CNN	0.9406	0.9933	0.9464
Hoang et al., 2018 [12]	ResNet-101 [10] + Faster R-CNN	0.9398	0.9932	0.9342
Thambawita et al., 2018 [29]	ResNet-152 [10] + DenseNet-161 [15] + MP	0.9397	0.9932	0.9297
Hicks et al., 2018 [11]	DenseNet-169 [15]	0.9325	0.9924	0.9236

select the data augmentation technique based on the F1 value of a rapid training in 20 groups of randomly selected test samples. For the classification of endoscopic images, they developed a CNN architecture consisting of a reduced version of residual neural network (ResNet34 [10] combined with SE-ReNext [14] and Attention -inception-v3 [27]. The addition of the attention blocks aimed for these to learn the differences between classes. This study introduced multi-epoch fusion, which consists of using the average of the weights of the last 5 training epochs to improve the model generalization and avoid parametric overfitting.

Chang et al. [7] designed their architecture to carry out multi-label classification. So, they realized a threshold selection of belongings to each label. They tested different threshold combinations for each label and selected the one that had the best performance. This work was the best evaluated for classification task in the Biomedia ACM MM Grand Challenge 2019 [16], where they reached an MCC of 0.9520.

Harzig et al. [9], used 2 CNNs for image classification, and although they used data augmentation techniques, their results were affected by the imbalance in the training samples [9]. In this study, the authors focused on making a fast classification, and not only accurate. Consequently, they used smaller CNNs that can run even on mobile devices. With MobileNetV2 [25] they achieved an MCC of 0.95974 in the KVASIR database [9] with an inference time that suggests that this algorithm could be implemented in real-time.

An interesting idea to combine different CNNs in a single model is that proposed by Luo et al. [21], they individually tested some state-of-the-art CNNs and selected the 10 with the best classification results for the KVASIR database. Subsequently, they trained 10 sub-models for each of the selected CNNs using cross-validation with the training data. Then, they used the output as a vector of probabilities of membership to each class of the trained submodels as a set of characteristics to train ML systems for classification. Their best MCC was

0.948035, and they obtained it with a LightGBM [17] classifier.

Hoang et al. [13] proposed and applied a data augmentation technique, which consists of cropping the region of interest for classification and adding this region of the image to others in the same database. These authors implemented a residual neural network in conjunction with a Faster R-CNN. The goal of using these 2 neural networks working together is that ResNet CNN carries out the classification work and, the detection network serves to reiterate the class. With this methodology, the authors achieved an MCC of up to 0.9406 for classification in the Biomedia ACM MM Grand Challenge 2019 test database [16,**?**].

In [29], Thambawita et al. studied different pre-trained models and combinations of these. They concluded that the combination of ResNet-152 and DenseNet-161 to extract image features, with a multi-layer perceptron for the classification lead to the best performance. With this approach, they got an MCC of 0.9397 in the 2018 Medico Classification task.

Hicks et al. [11] conjectured that pre-training the models with a medical dataset could enhance the models' performance. However, they discovered that vast and diverse datasets were better to pre-train, even if they were not similar to the final dataset. These authors reached an MCC of 0.9325 in the 2018 Medico Classification task using DenseNet-169.

2.4 Hyperparameter Optimization

In the previous subsection, we presented several studies concerning endoscopic image classification. In concrete, all of these works use a version of the KVASIR dataset. It is important to denote that every one of the presented studies used transfer learning to adapt the employed model to the target domain. During transfer learning, there are some hyperparameters to tune. These are capable of enhancing or worsen the model's performance. Despite the importance of the hyperparameter, these are usually manually tuned.

The only precedent that currently exists in the literature regarding the automatic tuning of hyperparameters to optimize the classification performance of endoscopic images is the research of Borgli et al. [5]. In this study, they used a Bayesian optimization approach achieving an improvement of up to 10% in terms of accuracy with other works that used the same CNNs for classification in the KVASIR database by adjusting the hyperparameters manually.

Borgli et al. [5] considered 4 hyperparameters: The pre-trained model, the gradient descent optimizing function, the learning rate and, the delimiting layer. The pre-trained model refers to the kind of architecture that is used to classify the images. They used KERAS to train the models, so the architectures and gradient descent optimizing function that they used during the optimization process were the ones available in this API. The architectures prospects were: Xception, VGG16, VGG19, ResNet50, InceptionV3, InceptionRes- NetV2, DenseNet121, DenseNet169, and DenseNet201. The gradient descent optimizing function prospects were SGD, RMSprop, Adagrad, Adadelta, Adam, Adamax, and Nadam. The learning rate was set in a continuous value between 1 and 10^{-4}. Finally, the delimiting layer refers to the number of layers that are trained in the model. This value was set between 0 and the number of layers in the selected model.

3 Methods and Implementation

3.1 Evolutionary Algorithm

Evolutive algorithms use the paradigm of evolution proposed by Darwin, in which the fundamental law is the principle of variation and selection. This principle of changing each generation (through reproduction) is the main component of the evolutionary strategies [4]. *Evolutionary algorithms are based on the collective learning process within a population of individuals, each of which represents a search point in the space of potential solutions to a given problem.* Back, 1993 [2].

In evolutionary algorithms, several individuals explore the solution space of the environment at random points. Then, the best-evaluated individuals pass their genes (information) to the next generation. Evaluation is the procedure of assessing how well the solutions fit the established goals. The genes of the selected individuals are preserved and mixed in new individuals with recombination mechanisms. Also, it is a good practice to consider a mutation factor during this procedure. The mutation is the introduction of random information, which introduces innovation into the population [2]. Fig. 1 shows a general overview of an evolutionary algorithm.

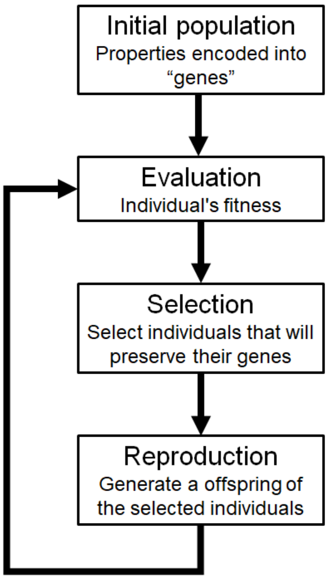

Fig. 1. General evolutionary algorithm.

We considered the whole population for the selection process. We use tournament selection, in which the algorithm takes n individuals randomly from the population, compares them, and selects the individual with the best evaluation. The algorithm repeats this process until it reaches the desired total number of selected individuals. All the individuals have the same chance to participate in the tournament. Nevertheless, the individuals with the highest evaluations are more likely to win the match and preserve their genes.

There are different mechanisms for the generation of new individuals. Overall, this procedure consists of the information (genes) combination of the selected individuals and a mutation factor to introduce novel information into the population. There exist different combination mechanisms, such as recombination (generate a new individual mixing up the parents' genes), raw combination (generate new genes by blending the parents' genes), etc. The combination mechanisms could be static or dynamic. The same happens with the mutation factor. Fig. 2 shows an example of the different combination techniques and the mutation procedure.

During this work, we use a static combination method without recombination. We use the arithmetic mean of the parents' genes as the combination method. We considered a mutation factor as an independent variable for every gen for every individual. Table 2 shows a pseudocode representation of the evolutionary algorithms that we use during the experiments presented in this study.

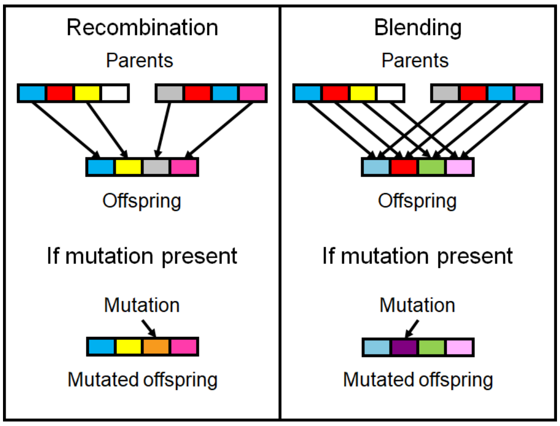

Fig. 2. Generation mechanisms of evolutive algorithms.

3.2 Data

The KVASIR dataset [24] has 8000 labeled images, 1000 for each class. We split these images into three subsets for training, validation, and testing, each with 4000, 2400, and 1600 images, respectively.

Table 2. Pseudocode representation of the evolutive algorithm.

```
1) Random generation of initial population of size $M$.
2) While condition of conclusion is not satisfied.
3)     Evaluation of the population.
4)     Selection of the $N$ individuals for the crossover.
5)     Generation of a new population of size $M$ with the crossover of the $N$
       selected individuals.
6) End.
```

We used data augmentation techniques. In every training iteration, the training images went through a transformation step, where they are randomly rotated at an angle between 1° and 355°. Also, a horizontal flip, vertical flip, and brightness adjustment are applied to every image, with a 50% probability for every transformation.

3.3 Experiments 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. Pytorch 1.7.1 was used to obtain the pretrained CNNs architectures and gradient descent optimization for training, which was Adam algorithm [18].

We included four hyperparameters in the optimization algorithm during the transfer learning: kind of CNN architecture, learning rate, delimiting layer, and training epochs. The gene representing the kind of CNN architecture takes discrete values, one for each architecture available, which were: AlexNet, ResNet-18, ResNet-34, Resnet50, SqueezeNet-1.1, DenseNet-121, DenseNet-169, MobileNet-v2, ShuffleNet-v2-x0.5, and ResNext-50-32x4d. The other three genes take continuous values. For the learning rate, we established bounds between 10^{-4} and 10^{-6}. The bounds of the delimiting layer depended on the selected architecture. The lower bound represents that the last half of the layers are fine-tuned, and the upper bound represents that the training only affects the classification layers. For the training epochs gen, we set the bounds between 5 and 15.

The maximization of the validation accuracy was the optimization target for the evolutionary algorithm. During the generation process, the genes had a probability of crossover (with arithmetic mean) of 50%, except the CNN architecture, which automatically inherited the gene of the best-evaluated parent. The learning rate, delimiting layer, and training epochs genes had a mutation probability of 15%, and the CNN architecture a probability of 25%. The population during the optimization consisted of 20 individuals. The evolutionary algorithm had 5 generations in total since the experiment goal was to prove that this kind of optimization is effective for this task.

4 Results

The best-evaluated individual of the initial population (0-generation) had a validation accuracy of 0.8995, and the best-evaluated individual of the fifth generation has a validation accuracy of 0.9821. Table 3 shows the characteristics of the best-evaluated individual of each generation in detail.

Table 3. Best evaluated individuals per generation.

Generation	Learning rate	Architecture	Layers pretrained	Epochs Epochs	Validation Acc	Validation MCC	Validation F1
0	8.888^{-4}	ShuffleNet-v2-x0.5	0.9341	11	0.9718	0.8709	0.8870
1	1.892^{-4}	ResNext-50-32x4d	0.6615	12	0.9786	0.9023	0.9145
2	1.892^{-4}	ResNet-18	0.9587	12	0.9769	0.8942	0.9075
3	2.209^{-4}	ResNext-50-32x4d	0.6950	11	0.9783	0.9006	0.9130
4	3.231^{-4}	MobileNet-v2	0.8870	11	0.9804	0.9103	0.9215
5	2.306^{-5}	ResNet-50	0.9101	11	0.9821	0.9183	0.9285

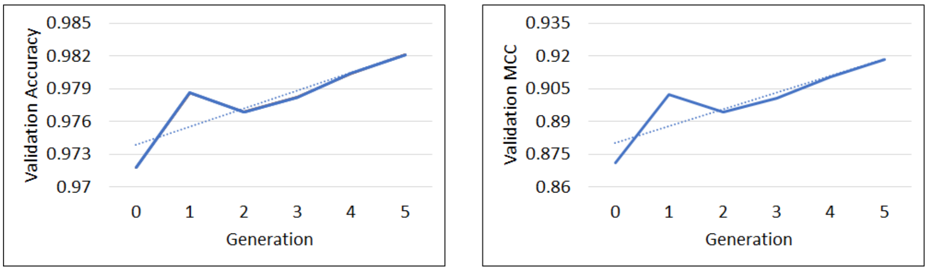

Fig. 3. Validation accuracy (left) and MCC (right) of the best-evaluated individuals per generation.

5 Conclusions

The experiment results imply that an evolutionary strategy can improve the accuracy of an endoscopy image classification algorithm. The results presented are evidence that this kind of optimization paradigm can lead to classification performance comparable to that of the best-evaluated architectures since the validation accuracy, MCC, and F1 value reached during the experiment are similar to those presented in table 1. Nevertheless, the objective of the experiment was to demonstrate that optimization of hyperparameter using a genetic algorithm is capable of improving the classification performance of off-the-shelf architectures.

The results presented are enough to prove it, considering that we got a rising of 0.0474% in the validation MCC in 5 optimization steps using only 20 individuals and basic generation mechanisms.

An inconvenience of using evolutionary strategies is that this kind of optimization algorithms usually needs several individuals. In computationally expensive tasks (such as this), it can take too long to reach the optimal solution. A possible solution to this issue is using surrogate models.

During the experiment, we only considered four genes (CNN architecture, learning rate, delimiting layer, and training epochs), and we set the bounds of the searching space based on the literature. In the future, we can extend the number of genes by including other hyperparameters, and we can refine the bounds of the searching space by carrying out a characterization of the solution space of the optimization problem.

Also, in this work, we set as optimization target the maximization of the validation accuracy. In future research, we can implement multi-objective optimization. For example, we can set as optimization target the maximization of other evaluation metrics, such as the MCC value, along with the minimization of the training time.

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)
2. Back, T., Schwefel, H.P.: An overview of evolutionary algorithms for parameter optimization. Evolutionary Computation 1, 1–23 (03 1993)
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)
4. Beyer, H.G.: The theory of evolution strategies (01 2001)
5. Borgli, R.J., Kvale Stensland, H., Riegler, M.A., Halvorsen, P.: Automatic hyperparameter optimization for transfer learning on medical image datasets using bayesian optimization. In: 2019 13th International Symposium on Medical Information and Communication Technology (ISMICT). pp. 1–6 (2019)
6. Chahal, D., Byrne, M.: A primer on artificial intelligence and its application to endoscopy. Gastrointestinal Endoscopy 92 (05 2020)
7. Chang, Y., Huang, Z., Chen, W., Shen, Q.: Gastrointestinal tract diseases detection with deep attention neural network. In: Proceedings of the 27th ACM International Conference on Multimedia. MM '19, Association for Computing Machinery, New York, NY, USA (2019), https://doi.org/10.1145/3343031.3356061
8. Gross, S., Sharma, P., Pante, A.: Artificial intelligence in endoscopy. Gastrointestinal Endoscopy 91 (12 2019)
9. Harzig, P., Einfalt, M., Lienhart, R.: Automatic disease detection and report generation for gastrointestinal tract examination. In: Proceedings of the 27th ACM International Conference on Multimedia. MM '19, Association for Computing Machinery, New York, NY, USA (2019), https://doi.org/10.1145/3343031.3356066
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015)

11. Hicks, S., Smedsrud, P., Halvorsen, P., Riegler, M.: Deep learning based disease detection using domain specific transfer learning (11 2018)
12. Hoang, T.H., Nguyen, H.D., Nguyen, T.A., Nguyen, V.T., Tran, M.T.: An application of residual network and faster - rcnn for medico: Multimedia task at mediaeval 2018. In: CEUR (ed.) MEDIAEVAL 2018, MediaEval Benchmarking Initiative for Multimedia Evaluation Workshop, 29-31 October 2018, Sophia-Antipolis, France. Sophia-Antipolis (2018), cEUR
13. Hoang, T.H., Nguyen, H.D., Nguyen, V.A., Nguyen, T.A., Nguyen, V.T., Tran, M.T.: Enhancing endoscopic image classification with symptom localization and data augmentation. In: Proceedings of the 27th ACM International Conference on Multimedia. MM '19, Association for Computing Machinery, New York, NY, USA (2019), https://doi.org/10.1145/3343031.3356073
14. Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E.: Squeeze-and-excitation networks (2019)
15. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks (2018)
16. 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://www.sciencedirect.com/science/article/pii/S1361841521000530
17. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y.: Lightgbm: A highly efficient gradient boosting decision tree. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017), https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2017)
19. de Lange, T., Halvorsen, P., Riegler, M.: Methodology to develop machine learning algorithms to improve performance in gastrointestinal endoscopy. World Journal of Gastroenterology 24 (12 2018)
20. 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)
21. Luo, Z., Wang, X., Xu, Z., Li, X., Li, J.: Adaptive ensemble: Solution to the biomedia acm mm grandchallenge 2019. In: Proceedings of the 27th ACM International Conference on Multimedia. MM '19, Association for Computing Machinery, New York, NY, USA (2019), https://doi.org/10.1145/3343031.3356078
22. Maddox, T., Rumsfeld, J., Payne, P.: Questions for artificial intelligence in health care. JAMA 321 (12 2018)
23. 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)
24. 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), http://doi.acm.org/10.1145/3083187.3083212

25. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
26. 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)
27. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision (2015)
28. 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
29. Thambawita, V., Jha, D., Riegler, M., Halvorsen, P., Hammer, H.L., Johansen, H.D., Johansen, D.: The medico-task 2018: Disease detection in the gastrointestinal tract using global features and deep learning (2018)
30. 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)