Automated gastrointestinal abnormalities detection from endoscopic images

Premananth Gowtham*, Mahesan Niranjan+, Anantharajah Kaneswaran*
*Dept. of Computer Engineering, Faculty of Engineering, University of Jaffna, Srilanka
+Dept. of Electronics and Computer Science, University of Southampton, UK
gowtham@eng.jfn.ac.lk, mn@ecs.soton.ac.uk, kanesh@eng.jfn.ac.lk

***Abstract*—Impressive high performance reported in the use of machine learning on computer vision problems is often due to the availability of very large datasets with which deep neural networks can be trained. With inference from medical images, however, this is not the case and available data is often only a small fraction in size in comparison to benchmark natural scene recognition problems. To circumvent this problem, transfer learning is often applied, where a model trained on a large natural image corpus is adapted, or pre-trained, to model the medical problem. In this work, we consider transfer learning applied to a specific medical diagnostics problem, that of abnormality detection in the gastrointestinal tract of a human body using images obtained during endoscopy. We carry out a search over several image recognition architectures and adapt pre-trained models to the endoscopy problem. Using the benchmark KVASIR dataset, we show that transfer learning is effective in outperforming previously reported results, at an accuracy of 98.5±0.27.**

***Index Terms*—Endoscopy, Gastrointestinal abnormalities, Transfer Learning**

I. Introduction

Gastrointestinal abnormalities are very common in the world nowadays that every year millions of people get affected by various gastrointestinal abnormalities and a considerable amount of people die due to these diseases [1]. As most of these diseases can be cured, the lives lost due to these abnormalities can be saved with the proper diagnosis of these abnormalities. When considering the diagnosis procedures used for gastrointestinal abnormalities, a lot of different diagnostic procedures are being practiced around the globe. Out of those procedures, endoscopic procedures are the latest state-of-the-art procedures carried out for the diagnosis of most gastrointestinal abnormalities. Although the endoscopic procedures are most preferred all around the globe for the diagnosis of gastrointestinal abnormalities, as the diagnosis is carried out manually the diagnosis requires a large amount of time from even a trained physician. In this paper, a system is proposed that can diagnose multiple gastrointestinal abnormalities automatically from endoscopic images to assist the physicians in the diagnosis by reducing the time taken by a considerable amount compared to manual diagnosis.

In recent years there has been a significant development in deep convolution neural networks which have created a revival in image recognition and segmentation-based researches worldwide [2], [3]. Various techniques have been employed in the use of deep convolution neural networks for medical image classification. The most commonly used techniques are training of the CNNs from the scratch [4] and the use of Transfer Learning in CNNs [5]. And especially for biomedical image-based datasets as the number of images present in biomedical image datasets are comparatively lower than normal datasets due to the patient privacy and confidentiality issues in the health sector, Transfer learning in convolution neural networks can be selected to be used for the classification tasks as transfer learning can utilize the available amount of images for training and produce better classification results even with such low number of images for training. In this paper, both of the above-mentioned deep CNN techniques are explored for the automated gastrointestinal abnormalities detection from endoscopic images by deploying various CNN architectures differing in width and depth and deploying pre-trained ImageNet based CNN architectures to show how Transfer Learning can be useful in biomedical image classification-based problems in the case of a limited amount of images available for training the CNN.

II. Related work

Computer-aided abnormality detection for gastrointestinal abnormalities has been an active research field for nearly a decade now. Different researches have been done on different data sets containing different sets of images of endoscopy and wireless capsule endoscopy using different methods to predict various gastrointestinal abnormalities and diseases. Jia et al. [6] used a deep convolution neural network architecture for feature extraction and have used the extracted features to be classified using a Support Vector Machine classifier for gastrointestinal bleeding detection. The research done by Coelho et al. [7] is focused on red lesion detection and segmentation from capsule endoscopy videos. For this, they have used a unique convolution neural network architecture called U-Net. The U-Net neural network architecture was introduced by O.Ronneberger et al. [3] for biomedical image segmentation. But in this research Coelho et al [7] were able to segment the regions with red lesions rather than only identifying them. In the research done by Shvets et al. [4], the authors have focused on finding a suitable method for the detection and segmentation of Angiodysplasia using deep convolutional neural networks from wireless-capsule endoscopic images. For detection and segmentation purposes they have used various convolutional neural network architectures. For this, they have

978-1-6654-2637-4/21/$31.00 ©2021 IEEE.

used U-Net network architecture, TernausNet-11 network architecture, and AlbuNet-34 network architecture to train and test wireless-capsule endoscopic images for Angiodysplasia segmentation. And from the experiments, the authors were able to find out that the AlbuNet-34 network was able to produce the highest IOU (intersection over Union) percentage for the segmentation among the three network architectures used.

The capsule endoscopic or endoscopic images obtained from patients have a high chance of possibility for the prevalence of various gastrointestinal abnormalities rather than a selected abnormality, therefore the models produced from the researches focusing on single gastrointestinal abnormalities prediction or segmentation [3], [4], [6], [7] won't be that much of use in a complete diagnostic sense in the real-world scenario. Because then the images had to be run through different systems or models to predict various gastrointestinal abnormalities. But Sekuboyina et al. [8] has proposed a deep convolution neural network to detect multiple gastrointestinal abnormalities from wireless capsule endoscopy images. In the research done by S. Yu et al. [9], the authors have tried to classify multiple gastrointestinal organ regions from the wireless-capsule endoscopic images. For the detection purpose, they have used a convolution neural network structure for the feature extraction and an extreme learning machine for the organ classifier to increase the classification accuracy. By using this method, the authors were able to obtain higher classification accuracy than the previous state-of-the-art procedures including convolution neural network with Support Vector Machine classifiers.

Another research done by K. Pogorelov et al. [10] also focuses on multi classes for gastrointestinal diseases detection. One of the important features of this research is that this research was done on a dataset called KVASIR which is made of endoscopy video images. And they used various classification methods like convolution neural networks, global features-based random forests, and transfer learning to detect gastrointestinal diseases from the images to find out which method suits best for the classification and detection purposes.

Another research was done by K. Pogorelov et al. [11] also focuses on multi classes of gastrointestinal diseases detection in the KVASIR dataset. But in this research, the authors have used two different methods for the detection of the multiple classes to compare and find out which method is most suitable for the detection of multiple classes from the KVASIR dataset. They have used global feature extraction and convolution neural networks for the detection of multiple gastrointestinal abnormalities from the dataset and through the experiments done they were able to find out that the global feature extraction method fared better than the convolution neural network method on the KVASIR dataset. And also in this research, they have tried to implement both these methods in the EIR system which is a real-time based multiple gastrointestinal abnormalities detection system based on wireless capsule endoscopy procedure.

When considering about deep convolution neural networks which are the latest solution for many complex image processing based problems, various CNN architectures [2], [12], [13], [14] have been introduced and used for various different image classification and segmentation tasks.

Although deep convolution neural network architectures have a very high potential to be used in complex image classification and segmentation problems, their potential seemed to be not fully utilized in the gastrointestinal abnormalities detection in the researches [10], [11] based on the KVASIR dataset.

A. Our contribution

Our work is driven based on the researches [10], [11] in which the authors have used the KVASIR dataset which contains images of multiple gastrointestinal abnormalities as well as multiple normal gastrointestinal landmarks as well. Our main goal was to improve on the classification accuracy obtained in those researches which will help to create a better real-time gastrointestinal abnormalities detection system that can detect multiple gastrointestinal abnormalities with lesser error rates so that it can be employed in the real world medical diagnosis applications to help the physicians. In the researches [10], [11] the state-of-the-art accuracy of classification they were able to achieve was by using global features extracted from those endoscopic images. The accuracy of the system they were able to achieve was limited by the number of images available in the dataset. But our work focuses on increasing the classification accuracy with the available images by the use of transfer learning in deep convolution neural networks. Rather than randomly assigning weights for the deep convolution neural networks, pre-trained weights obtained through transfer learning from a natural image dataset have been assigned for the deep convolution neural networks. This eased up the training process of the deep convolution neural networks and resulted in better classification results in terms of accuracy, precision, and specificity when compared to the previous methods that were trained on this particular dataset.

B. Description of the dataset

The KVASIR dataset version-1 [10], [11] has been used for all training and evaluation purposes in this study. The KVASIR dataset version-1 consists of 4,000 images manually annotated by physicians which consist of 8 classes of images such as gastrointestinal abnormalities, normal anatomical landmarks, and endoscopic procedures done in the gastrointestinal tract. Among the 8 classes, 3 are anatomical landmarks which are normal-cecum, normal pylorus, and normal-z-line, 3 of them are gastrointestinal abnormalities which are esophagitis, polyps, and ulcerative-colitis and the rest 2 are of endoscopic procedures namely dyed lifted-polyps and dyed-resection-margins. An example for all of those classes of images is shown in Fig. 1. All the images in the dataset have a resolution ranging from 720x576 up to 1920x1072 because they were taken from various endoscopic procedures which used various endoscopic apparatus.

III. Methodology

In this research, as the images in the KVASIR dataset are of different sizes the pre-processing of images was needed

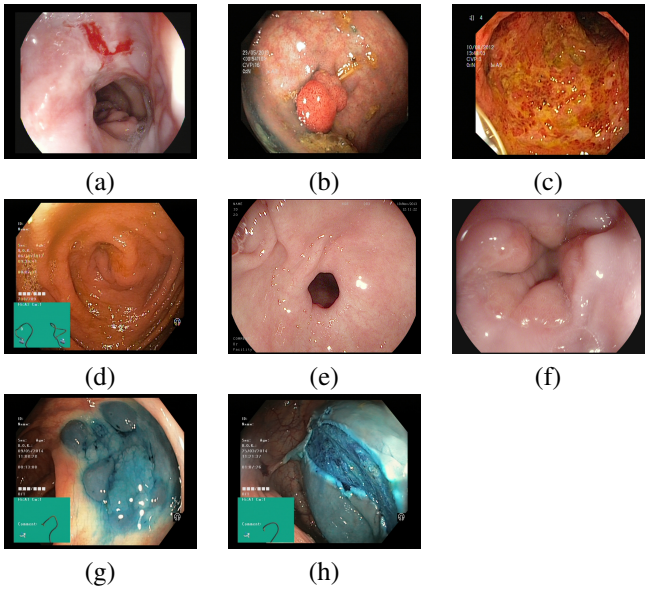


Fig. 1. Images of all 8 classes in KVASIR dataset (a) esophagitis (b) polyps (c) ulcerative colitis (d) normal-cecum (e) normal-pylorus (f) normal-z-line (g) dyed lifted polyps (h) dyed resection margins

at first to feed them into deep convolution neural networks. As a pre-processing step, all the images in the dataset were resized to 224x224 images because that was the input size of images in the Imagenet dataset as the pre-trained weights learned using the Imagenet dataset [2] were used for transfer learning. After that, the resized images were transformed using random Horizontal Flip to make sure that the orientation of the image doesn't affect classification results.

Then various convolution neural networks including Alexnet [2], Resnet [12], Densenet [13], and Squeezenet [14] were used for the classification of multiple classes of gastrointestinal abnormalities available in the dataset [10]. These CNNs were both trained from scratch by randomly assigning weights and also trained using transfer learning of weights obtained by training the models using the Imagenet dataset [2]. Although the ImageNet dataset is not a medical dataset, CNNs comprehensively trained on large datasets like this one was used to transfer weights to small datasets based classification models as they made the models more effective. Because these deep CNNs possess millions of free parameters and they require a significant number of images to train them and that liberty can't be obtained with these biomedical datasets due to their low number of images.

These various deep convolution neural network architectures were used for the classification of multiple classes available in the dataset to find out which deep CNN architecture produces the better detection model for gastrointestinal abnormalities. All the CNN models were trained on the Pytorch framework for 50 epochs with an SGD optimizer with a learning rate of 0.001 and a momentum of 0.9. The learning rate was decreased by a factor of 10 after every 10 epochs. NVIDIA GEFORCE GTX 1050 GPUs were used to train the CNNs.

Resnet [12] and Densenet [13] CNN models have multiple versions of them depending on the depth of the network architectures. In this study, multiple versions of both networks were used to find out which configuration produces the best model for classification as these were the better performing models when compared to all the models that were used in this research when it came to transfer learning.

Densenet model proposed in [13] deploys a convolution neural network model which connects each layer in the model to every other layer in a feed-forward fashion. For each layer, the feature maps of all of the previous layers in the model are used as inputs and the feature maps of that layer are used as inputs to all the subsequent layers. Densenet model produced state-of-the-art performances across several competitive datasets such as CIFAR-10 [16], CIFAR-100 [16], SVHN dataset [17] and Imagenet dataset [2].

Resnet model proposed in [12] deploys a model architecture that uses residual connections between layers in the model to ease the training of the networks that are substantially deeper than those used previously. The model explicitly reformulates the layers as learning residual functions with reference to layer inputs, instead of learning unreferenced functions. This model was used on the Imagenet dataset [2] and produced an error of 3.57% and won 1st place on the ILSVRC 2015 classification task. The model architecture of Resnet is mainly inspired by the philosophy of VGG nets [15] but with shortcut connections that convert the VGG net-styled model into the residual Network model. The identity shortcuts (1x1 convolutions) are used in the Resnet model for the shortcut connections. The architecture of the Resnet-34 model which is the proposed model for Transfer-learning based Gastrointestinal abnormalities detection on the KVASIR dataset is shown in Fig. 2.

IV. Evaluation and Discussion

In this section, the performances of various CNN models trained on the KVASIR dataset [10] are evaluated and compared for the classification of gastrointestinal abnormalities. The models were both trained from scratch and also trained using transfer learning of weights from the Imagenet dataset trained models. For evaluation purposes, 5-fold cross-validation was used. It was done by splitting the dataset into 5 equal folds and alternatively using 4 of them for training and 1 of them for testing. The classification using various CNN architectures was evaluated and compared using multiple performance metrics which are used in researches [10], [11] to compare this study's findings with those researches. The used performance metrics are Precision, Recall, Specificity, Accuracy, Matthew's correlation coefficient, and F1 score.

TABLE I
Comparison of the performance various CNN architectures with and without transfer learning on KVASIR dataset

Model	Accuracy without Transfer learning	Accuracy with Transfer learning
Resnet-18	91.06±0.09	98.31±00.10
Resnet-34	90.63±0.14	98.5±00.27
Densenet121	89.12±0.12	98.19±00.26
Alexnet	91.19±0.28	96.79±00.19
Squeezenet	90.55±0.19	97.12±00.25

Table I compares the performances between the best performing models which were trained with and without transfer learning. All the models were trained and evaluated under

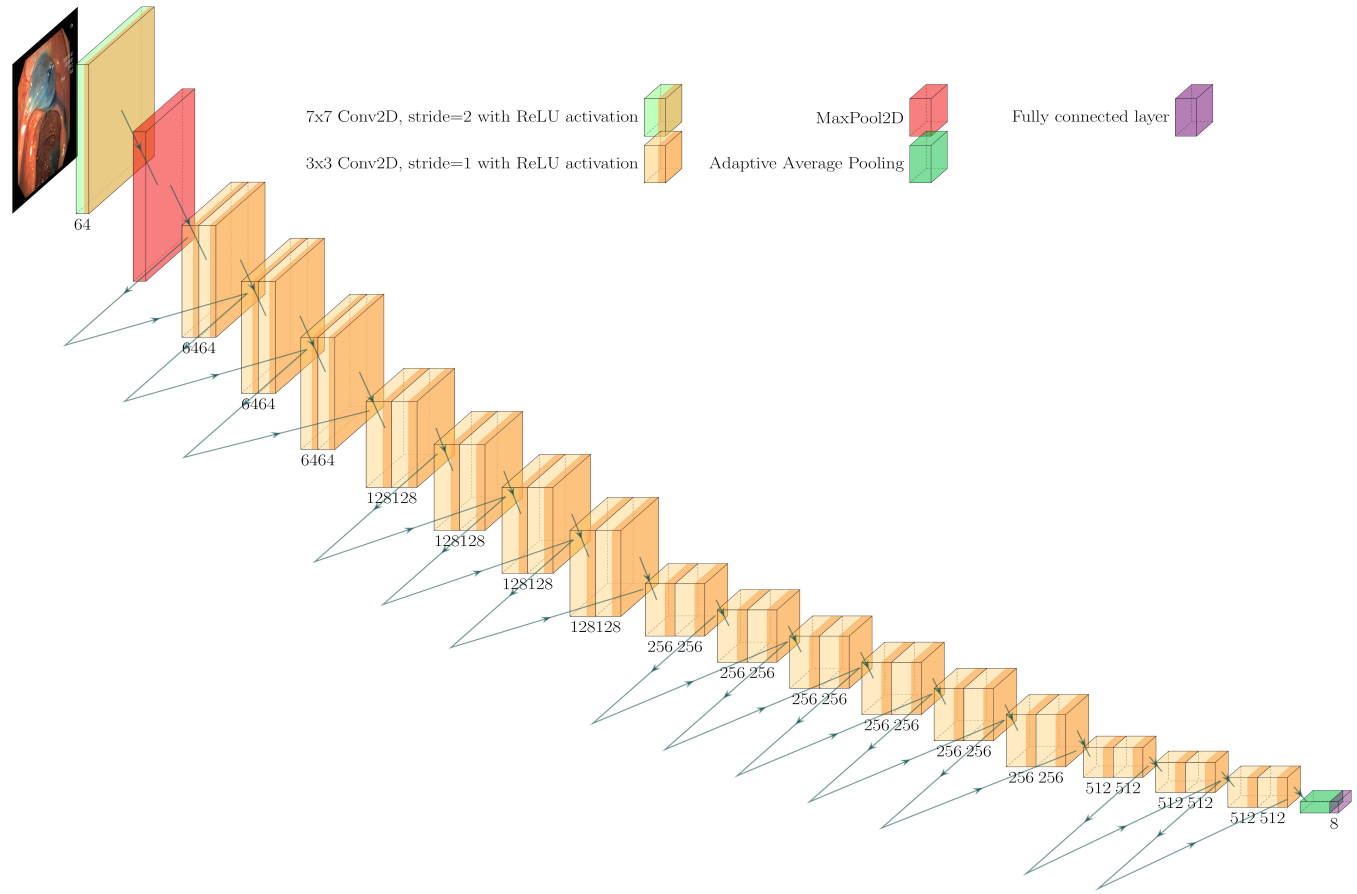


Fig. 2. Resnet-34 [12] model architecture used for gastrointestinal abnormalities detection

similar conditions as mentioned in the methodology. The results tabulated show that all the models performed better when they were trained after transferring weights from the models trained on Imagenet.

Table II gives a detailed description of the top performances produced by the transfer learning of various CNN architectures on KVASIR dataset classification. According to the results summarized in Table II , it can be seen that transfer learning of weights obtained by the training of resnet-34 architecture on Imagenet dataset then transferred and trained on the KVASIR dataset has produced the best classification performance according to all the performance metrics used.

Fig 3 shows the average confusion matrix obtained for the resnet-34 model for KVASIR classification for the 5 cross-validated folds. According to the confusion matrix, it can be observed that the major effect on the wrong classification can be confined to mainly 2 sets of wrong classifications. They are the wrong classifications between dyed and lifted polyps and dyed resection margins and wrong classifications between Esophagitis and Z-line classes.

The reason for the wrong classifications between Esophagitis and z-line is because those are images of the same area of the gastrointestinal tract. After all, Esophagitis is an abnormal condition caused by inflammation in the esophagus of the gastrointestinal tract and z-line is the anatomical landmark that indicates the esophageal and gastric junction. The reason

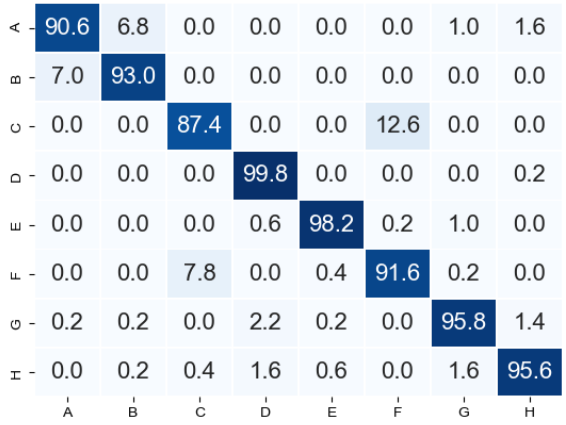


Fig. 3. Average Confusion Matrix of the 5-fold classification using resnet-34 TL on the KVASIR dataset (A) dyed lifted polyps (B) dyed resection margins (C) esophagitis (D) normal-cecum (E) normal-pylorus (F) normal-z-line (G) polyps (H) ulcerative colitis

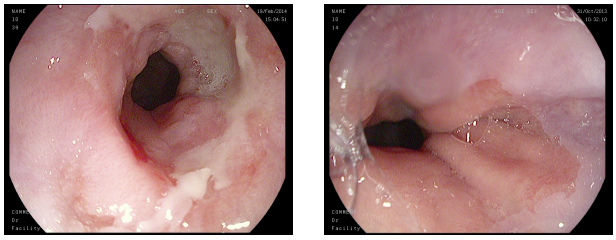
for the wrong classification can be seen in Fig.4 in which examples from both the classes are shown. The inflammation which causes Esophagitis can be very small and the model can miss it and can classify it as a normal-z-line because both

TABLE II
CLASSIFICATION PERFORMANCE IN TERMS OF WEIGHTED AVERAGE OF VARIOUS CNN ARCHITECTURES WITH TRANSFER LEARNING ON KVASIR DATASET

Model	Precision	Recall	Specificity	Accuuracy	MCC	F1-score
Resnet-18	93.35±00.39	93.25±00.40	99.04±00.06	98.31±00.10	92.30±00.45	93.21±00.41
Resnet-34	94.08±01.07	94.00±01.08	99.14±00.15	98.5±00.27	93.16±01.23	93.99±01.08
Densenet121	92.86±01.03	92.78±01.06	98.97±00.15	98.19±00.26	91.76±01.20	92.75±01.06
Alexnet	87.91±00.71	87.15±00.75	98.16±00.10	96.79±00.19	85.54±00.84	87.08±00.76
Squeezenet	88.83±00.97	88.48±01.00	98.35±00.14	97.12±00.25	86.93±01.13	88.43±01.00

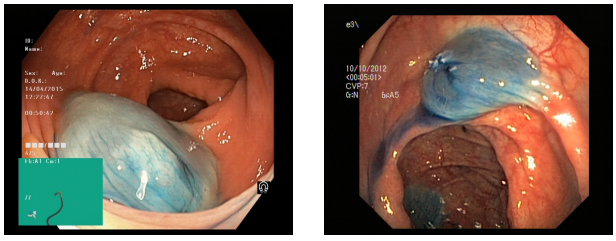
are of the same anatomical regions only with and without the inflammation.

The reason for the wrong classifications between dyed and lifted polyps and dyed resection margins is because both the images are of the same regions before and after a polyp removal process after injecting saline and indigo carmine which gives an indigo color to the region and also some polyps are very small from which the difference between the presence of a polyp and its absence after it is removed also couldn't be seen properly. This can be noticed in Fig. 5 which shows the images from both classes but without much difference in the indigo region which was injected for the polyp removal process. These are the reasons why it causes a slight misunderstanding in the model and the model wrongly classifies them.



(a) (b)

Fig. 4. Images of the 2 confusing classes in KVASIR dataset (a) esophagitis (b) normal-z-line



(a) (b)

Fig. 5. Images of the 2 confusing classes in KVASIR dataset (a) dyed lifted polyps (b) dyed resection margins

Table III and Table IV compares the performance of various researches done on multi-class gastrointestinal abnormalities classification. And it can be seen that the proposed model of transfer learning of weights in resnet-34 CNN architecture has outperformed all the performances of other models used for multi-class gastrointestinal abnormalities classification using 8-class and 6-class KVASIR datasets.

TABLE III
COMPARISON OF VARIOUS GASTROINTESTINAL ABNORMALITIES CLASSIFICATION MODELS TRAINED AND REPORTED ON 8-CLASS KVASIR DATASET

Models	Accuracy
6 CNN [10]	0.914
Inception V3-TL [10]	0.924
6 GF-LMT [10]	0.937
Resnet-34-TL (Proposed Model)	**0.985±0.0027**

TABLE IV
COMPARISON OF VARIOUS GASTROINTESTINAL ABNORMALITIES CLASSIFICATION MODELS TRAINED AND REPORTED ON 6-CLASS KVASIR DATASET

Model	Accuracy
CNN [11]	0.975
6 Global features [11]	0.969
Resnet-34-TL (Proposed Model)	**0.9842±0.0041**

V. CONCLUSION

In this work, we have shown the power of transfer learning for a benchmark medical image classification problem, outperforming previously reported results. Searching through several architectures of deep neural network models, pre-trained on natural scene recognition tasks (ImageNet), we show that the ResNet architecture gives the best performance. While the only previous attempt at transfer learning on this particular dataset which uses the Inception V3-TL model achieves an accuracy of 92.4%, the ResNet model reaches 98.5%. This is a significant margin, suggesting that the choice of architecture is important in transfer learning. A particular difficulty in making comparisons is that previous authors do not report uncertainty in their results. Hence the statistical significance of the differences cannot be formally established. However, analyzing the errors made by the classifiers, we can show interpretable misclassifications of the system.

REFERENCES

[1] M. Avramidou, F. Angst, J. Angst, A. Aeschlimann, W. Rössler, and U.Schnyder, "Epidemiology of gastrointestinal symptoms in young and middleaged Swiss adults: Prevalences and comorbidities in a longitudinal population cohort over 28 years," BMC Gastroenterol., vol. 18, no. 1, pp. 1–10, 2018, doi: 10.1186/s12876-018-0749-3.

[2] B. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, 2012.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9351, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4 28.

[4] A. A. Shvets, V. I. Iglovikov, A. Rakhlin, and A. A. Kalinin, “Angiodysplasia Detection and Localization Using Deep Convolutional Neural Networks,”Proc. - 17th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2018, pp. 612–617,2019, doi: 10.1109/ICMLA.2018.00098.

[5] H. C. Shin et al., “Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning,” IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1285–1298, 2016, doi: 10.1109/TMI.2016.2528162.

[6] X. Jia, S. Member, and M. Q. Meng, “A Deep Convolutional Neural Network for Bleeding Detection in Wireless Capsule Endoscopy Images *,” 2016 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., pp. 639–642, 2016, doi: 10.1109/EMBC.2016.7590783.

[7] P. Coelho, A. Pereira, A. Leite, M. Salgado, and A. Cunha, “A Deep LearningApproach for Red Lesions Detection in Video Capsule Endoscopies,” vol. 1,pp. 267–270, 2014, doi: 10.1016/B978-0-12-396501-1.00009-1.

[8] A. K. Sekuboyina, S. T. Devarakonda, and C. S. Seelamantula, “A convolutional neural network approach for abnormality detection in wireless capsule endoscopy,”2017 IEEE 14th Int. Symp. Biomed. Imaging (ISBI 2017), pp. 1057–1060, 2017, doi: 10.1109/ISBI.2017.7950698

[9] J. S. Yu, J. Chen, Z. Q. Xiang, and Y. X. Zou, “A hybrid convolutional neural networks with extreme learning machine for WCE image classification,” 2015 IEEE Int. Conf. Robot. Biomimetics, IEEE-ROBIO 2015, pp. 1822–1827, 2015, doi: 10.1109/ROBIO.2015.7419037.

[10] K. Pogorelov et al., “Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection,” Proc. 8th ACM Multimed. Syst. Conf. MMSys 2017, 2017, doi: 10.1145/3083187.3083212.

[11] K. Pogorelov et al., “Efficient disease detection in gastrointestinal videos – global features versus neural networks,” pp. 22493–22525, 2017, doi: 10.1007/s11042-017-4989-y.

[12] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.,vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[13] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 2261–2269, 2017, doi: 10.1109/CVPR.2017.243.

[14] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K.Keutzer, “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters42 and ¡0.5MB model size,” pp. 1–13, 2016.

[15] K. Simonyan and A. Zisserman, “Very deep convolutional networks for largescale image recognition,” 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf.Track Proc., pp. 1–14, 2015

[16] A. Krizhevsky, “Learning Multiple Layers of Features from Tiny Images,” 2009.

[17] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, “Reading Digits in Natural Images with Unsupervised Feature Learning,” 2011.