Interpretability of Deep Neural Networks to diagnose Inflammatory Bowel Disease

José Maurício
a2018056151@isec.pt

Inês Domingues
ines.domingues@isec.pt

Instituto Politécnico de Coimbra,
Instituto Superior de Engenharia,
Rua Pedro Nunes - Quinta da Nora,
3030-199 Coimbra, Portugal

Centro de Investigação do Instituto Português
de Oncologia do Porto (CI-IPOP):
Grupo de Física Médica, Radiobiologia
e Protecção Radiológica

Abstract

The number of patients with inflammatory bowel disease (IBD) has been increasing. The diagnosis is a difficult task for the gastroenterologist performing the endoscopic examination. However, in order to prescribe medical treatment and provide quality of life to the patient, the diagnosis must be quick and accurate. This paper presents a study where the objective is to collect and analyse endoscopic images referring to Crohn's disease and Ulcerative colitis using four deep neural networks. The main focus is on the understanding of the networks, offering through this paper a comparative study of five interpretability models.The obtained results demonstrate that it is possible to automate the process of diagnosing patients with IBD using deep networks for processing images collected during an endoscopic examination. Thus, we can develop tools that, with the aid of interpretability models, assist medical specialists in diagnosing the disease by understanding the specific region of the mucosa the network considered when making a decision.

1 Introduction

Inflammatory bowel disease (IBD) is characterized as a disease of unknown cause that results from the interaction between genetic and environmental factors, triggering an immune response that causes digestive disorders and inflammation in the gastrointestinal tract. These types of diseases are divided into Ulcerative colitis and Crohn's disease [11]. It is estimated that in Portugal it already affects 7,000 to 15,000 people, with 2.9 cases per 100,000 inhabitants per year with Ulcerative colitis and 2.4 cases per 100,000 inhabitants per year with Crohn's disease [4].

The methodology proposed in the present work allows the classification of images through deep neural networks to predict the type of inflammatory bowel disease present in that particular image. In addition, it is also suggested the use of interpretability algorithms of the networks so that the medical specialist can evaluate the region of the mucosal that was considered to make the decision. Towards this end, this study demonstrates a comparison between five interpretability algorithms, namely, Grad-CAM, LIME, SHAP values, RISE and Occlusion sensitivity.

This paper is organized into three sections: section 2 presents the methodology developed during this work and the dataset that was used for the classification; section 3 collects the quantitative results and shows the interpretability of the neural networks; and section 4 summarises the conclusions and directions for future work.

2 Methodology and Data

In order to improve the diagnosis of inflammatory bowel disease, a methodology based on five phases is proposed. In the first phase, the authors collected images of the inflammatory bowel disease (i.e. Ulcerative colitis and Crohn's disease). In the second phase, data augmentation was applied to the training dataset. In the third phase, the convolutional networks were implemented and configured. In the fourth phase, the performance of the CNNs was evaluated using eight different metrics. Finally, the interpretability of the models used was implemented. This methodology is illustrated in Figure 1 and is further explained next.

Dataset: For the development of this study, two datasets of images referring to Crohn's disease and Ulcerative colitis were used, with the goal of creating a single dataset involving images of both diseases. LIMUC [10] dataset was used to get 1360 images of the Ulcerative colitis disease, while the CrohnIPI [3] dataset was selected to collect 1360 images of Crohn's disease. This last dataset has already been used in other works to develop tools for diagnosing the disease [6, 13, 14].

Data Augmentation: Before classifying the images, data augmentation was performed on the train set with horizontal and vertical Random Flip, Random Contrast with a factor of 0.15, Random Rotation with a factor of 0.2 and a Random Zoom with a portion of -0.2 for height and -0.3 for width [9].

Deep Neural Networks: The following architectures were used to classify images of inflammatory bowel disease: ResNet50, VGG16, and InceptionV3. These architectures have been pre-trained with the ImageNet dataset. Besides these architectures, a hybrid model was built where a CNN is combined with an LSTM [7]. The architectures used to classify the images were trained for 200 epochs, using Sparse Categorical Cross-entropy as the loss function, a batch size of 32, and Adam with a learning rate of 1E-05 was set as the optimizer [5]. During training, an Early Stopping callback was used with patience 5 to monitor the validation accuracy. The images were initially resized to 224 × 224 pixels.

Evaluation: In all experiments, the dataset was split into 70% for training, 20% for validation, and 10% for testing [2]. Classification metrics were selected to evaluate the network's on the test dataset. These include Accuracy, Loss, Precision, Recall, F1-Score, Area Under Curve (AUC), Mathew's Correlation Coefficient (MCC) and Inference time [1, 12].

Interpretability: In order to interpret the output of the neural networks and gain insights into the key features considered important for predicting outcomes, several algorithms were chosen. These include Grad-CAM, LIME, SHAP values, RISE, and Occlusion sensitivity. These algorithms play a vital role in analyzing and determining which specific parts of the images contribute to the predicted output.

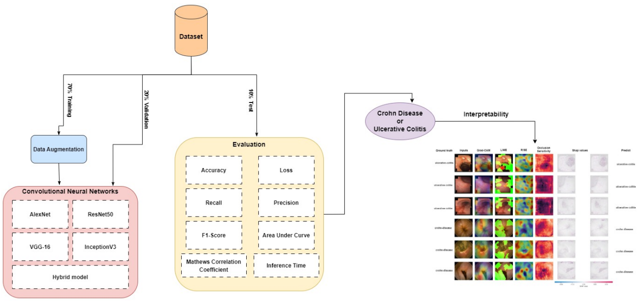

Figure 1: Experimental setup.

3 Results

In conducting this study, Tensorflow, version 2.8.0, Tfimm, version 0.6.13, and SHAP, version 0.40.0, were used. The programming environment for importing the libraries was Google Colab with the NVIDIA A100 GPU. Table 1 shows that the hybrid model was the architecture with the worst results among all the networks selected for this study. Even so, based on the AUC value, it shows a good ability to distinguish classes. It is also concluded that due to the complexity of the networks used in this study, the inference time is greater than 2 seconds.

After analysing the result of the network's interpretability, in Figures 2-5, it can be observed that they tend to interpret some characteristics of the images as decisive elements to predict the disease. The LIME algorithm and the Occlusion sensitivity, for the networks InceptionV3 and VGG16 have the ability to recognise, in ulcerative colitis images, zones

Table 1: Results obtained through the experience.

	ResNet50	VGG16	InceptionV3	Hybrid model
Acc	**1.0000**	**1.0000**	**1.0000**	0.9926
Loss	0.0044	**0.0000**	0.0067	0.3328
Recall	**1.0000**	**1.0000**	**1.0000**	**1.0000**
Precision	**1.0000**	**1.0000**	**1.0000**	0.9852
F1-Score	**1.0000**	**1.0000**	**1.0000**	0.9925
AUC	**1.0000**	**1.0000**	**1.0000**	0.9927
MCC	**1.0000**	**1.0000**	**1.0000**	0.9853
Inference time	2.44s	3.43s	2.94s	**2.23s**

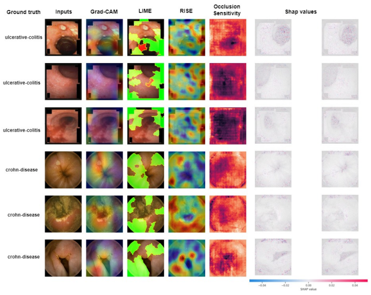

Figure 2: Interpretability and predicted output by the ResNet50 network.

of bleeding, as well as ulcers. In Crohn's disease, it can recognise the lesions caused by the disease. SHAP values and the RISE model on the same networks have also shown some evidence of this effect.

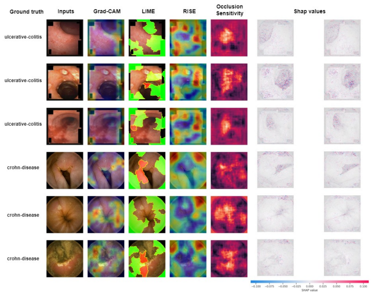

Figure 3: Interpretability and predicted output by the VGG16 network.

4 Conclusions

The findings of this study highlight the utility of deep learning architectures in assisting gastroenterologists with the early diagnosis of inflammatory bowel disease. The classification metrics demonstrate that the employed architectures achieved a good level of accuracy in disease classification. However, upon analyzing the interpretability of the architectures utilized in this study, it becomes evident that certain interpretability models partially or even completely disregard the mucosal part of the intestine.

Future research will prioritize medical validation using new images encompassing both types of disease. This validation will enable a comprehensive understanding of which architectures prove most interpretable in identifying the specific type of inflammatory bowel disease, based on the obtained results and the insights of experienced gastroenterologists. Also, they could develop studies where vision transformers are applied to the classification of images as an alternative to CNNs. These models demonstrate to be accurate and have a good performance in the classification of images [8].

References

[1] D Chicco and G Jurman. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21:6, 2020.

[2] M Chierici, N Puica, M Pozzi, A Capistrano, MD Donzella, A Colangelo, V Osmani, and G Jurman. Automatically detecting Crohn's disease and Ulcerative Colitis from endoscopic imaging. *BMC Medical Informatics and Decision Making*, 22(S6):300, 2022.

[3] CrohnIPI. CrohnIPI. `https://crohnipi.ls2n.fr/en/crohn-ipi-project/` (accessed Feb. 21, 2023).

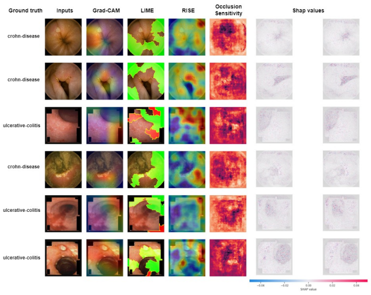

Figure 4: Interpretability and predicted output by the InceptionV3 network.

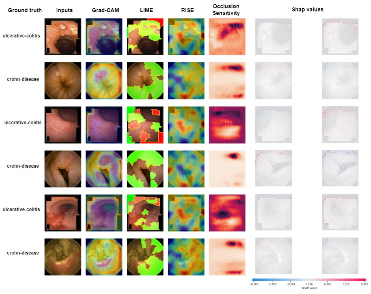

Figure 5: Interpretability and predicted output by the hybrid model.

[4] Doença inflamatória do intestino CUF. Doença inflamatória do intestino CUF. `https://www.cuf.pt/saude-a-z/doenca-inflamatoria-do-intestino` (accessed Feb. 21, 2023).

[5] MN Khan, MA Hasan, and S Anwar. Improving the Robustness of Object Detection Through a Multi-Camera–Based Fusion Algorithm Using Fuzzy Logic. *Frontiers in Artificial Intelligence*, 4: 638951, 2021.

[6] A Maissin, R Vallée, M Flamant, M Fondain-Bossiere, CL Berre, A Coutrot, N Normand, H Mouchère, S Coudol, C Trang, and A Bourreille. Multi-expert annotation of Crohn's disease images of the small bowel for automatic detection using a convolutional recurrent attention neural network. *Endoscopy Int Open*, 09:E1136–E1144, 2021.

[7] J Maurício and I Domingues. Deep Neural Networks to distinguish between Crohn's disease and Ulcerative colitis. In *11th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*, 2023.

[8] J Maurício, I Domingues, and J Bernardino. Comparing vision transformers and convolutional neural networks for image classification: A literature review. *Applied Sciences*, 13(9):5521, 2023.

[9] J. Maurício and I. Domingues. Knowledge distillation of vision transformers and convolutional networks to predict inflammatory bowel disease. In *26th Iberoamerican Congress on Pattern Recognition*, 2023 (submitted).

[10] G Polat, HT Kani, IE, YO Alahdab, A Temizel, and O Atug. Labeled Images for Ulcerative Colitis (LIMUC) Dataset, 2022.

[11] SS Seyedian, F Nokhostin, and MD Malamir. A review of the diagnosis, prevention, and treatment methods of inflammatory bowel disease. *J of Medicine and Life*, 12:113–122, 2019.

[12] M Turan and F Durmus. UC-NfNet: Deep learning-enabled assessment of ulcerative colitis from colonoscopy images. *Medical Image Analysis*, 82:102587, 2022.

[13] R Vallée, A Maissin, A Coutrot, H Mouchère, A Bourreille, and N Normand. CrohnIPI: An endoscopic image database for the evaluation of automatic Crohn's disease lesions recognition algorithms. In *Medical Imaging: Biomedical Applications in Molecular, Structural, and Functional Imaging*, page 61. SPIE, 2020. ISBN 9781510634015.

[14] R Vallée, A Coutrot, N Normand, and H Mouchère. Influence of expertise on human and machine visual attention in a medical image classification task. In *European Conference on Visual Perception*, 2021.