GastroNet: Abnormalities Recognition in Gastrointestinal Tract through Endoscopic Imagery using Deep Learning Techniques

Samira Lafraxo
LabSIV, Department of Computer Science
Faculty of Science, Ibn Zohr University
BP 8106, 80000 Agadir, Morocco
samira.lafraxo@gmail.com

Mohamed El Ansari
LabSIV, Department of Computer Science
Faculty of Science, Ibn Zohr University
BP 8106, 80000 Agadir, Morocco
melansari@gmail.com

***Abstract*—The human gastrointestinal (GI) tract may be infected by various diseases. If not detected at early stages, these abnormalities have the possibility to progress into gastric cancer, which is a common type of malignancies with yearly global cases exceeding one million. Endoscopy is a routinely used strategy for the examination of gastrointestinal tract diseases. During the examination, and due to many reasons like irregular morphologies, a huge number of frames, and exhaustion, gastrologists can miss some abnormalities. Thus, the automated classification of anomalies in endoscopic images is becoming necessary to assist medical diagnosis and reduce the cost and time of the medical process. Recent advances and high performance of deep learning techniques make it the best choice to adopt as a computer-aided-diagnosis strategy. In this paper, a novel deep learning model based deep convolutional neural network is proposed. Our model aims to automatically detect diseases from endoscopic images. The newly designed architecture is validated on the publicly available dataset KVASIR, which contains 8000 images. The results of our CNN approach compared to other well known pre-trained models showed important improvement and achieved 96.89% in terms of accuracy. The experiments demonstrated that the system can perform a high detection level without any human intervention.**

***Index Terms*—Gastrointestinal, Endoscopy, Deep learning, Convolutional neural network, Transfer learning, Automated detection**

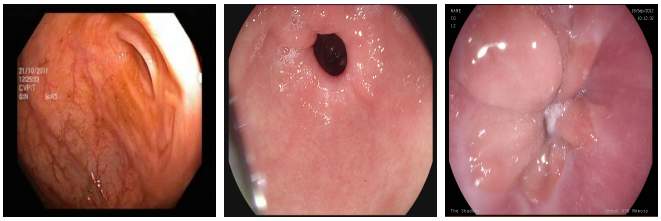
I. Introduction

The gastrointestinal tract is one of the most important parts of the body, this vital organ can be affected by different kinds of diseases. If not early and well treated, those abnormalities may be developed into cancerous cells. According to the World Health Organization (WHO), colorectal cancer is the third most common cause of death for both men and women. It represents 10% of announced cases of cancer [1]. Every year, about two 2.8 million GI cancers (colorectal, esophageal, stomach) are reported globally, with an important mortal rate of about 65% [2]. Because of imaging recent procedures, it becomes actually possible to visualize those parts of the human body which were inaccessible earlier. Endoscopy is one of these procedures, it is often used to examine the GI tract embedding a tube with a camera within it [3]. Therefore, endoscopic assessment of disease classification extremely depends on the gastroenterologists' decision, which means that findings may differ from an expert to another [4]. Manuel analysis of endoscopic data demands usually much time, requires a high level of concentration and also, can be sometimes wrong, depending on the expertise of the clinicians. Hence, automated recognition may be efficient in optimizing this process concerning time, cost, and quality of the classification [4]. In the field of artificial intelligence, deep learning-based methods have become recently the most popular technique to solve medical imaging problems. Assisted by the fast improvement of computational power GPU, deep learning has demonstrated high performance over classical machine learning methods [5]. The goal behind this study is to classify the different abnormalities in the gastrointestinal tract using endoscopic data. KVASIR dataset is used to train the deep learning based on convolutional neural network models. This dataset contains 8 classes (3 normal and 5 diseases) including Cecum, Pylorus, and Z-line (normal); Polyps, Ulcerative colitis, Esophagitis, Dyed-resection-margins, Dyed-lifted-polyps (diseases). The 8 classes are illustrated in Fig. 1, Fig. 2 and Fig. 3. The proposed system is compared with well known pre-trained architectures, then compared also to existing approaches. The structure of the paper is organized as follows. In Section II, we review existing methods for abnormalities classification. Section III is designed for the methods explanation. The experimental results are represented in Section IV. We finally conclude the paper in Section V.

978-1-7281-9015-0/20/$31.00 © 2020 IEEE

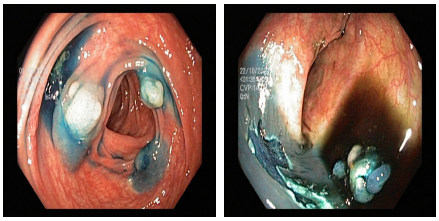
II. Related work

Over the last few decades, very promising results were obtained using computed-aided-diagnosis algorithms in the medical imaging area [6]–[9]. A study of the literature showed that various automatic approaches based on both handcrafted methods and deep learning techniques have been widely employed to detect and classify gastrointestinal tract abnormalities [10]–[14]. Naqvi *et al*. [10] implemented a system to recognize GI diseases, they used KVASIR dataset to evaluate



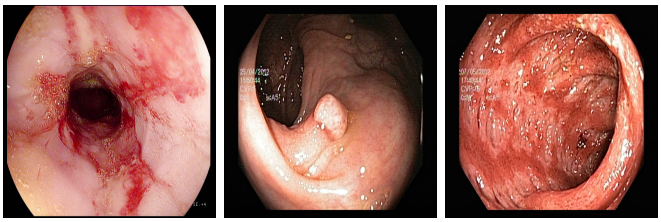
(a) (b) (c)

Fig. 1. Anatomical landmarks. (a) Cecum (b) Pylorus (c) Z-line



(a) (b)

Fig. 3. Polyp removal. (a) Dyed-lefted-polyps (b) Dyed resection margins



(a) (b) (c)

Fig. 2. Pathological Findings. (a) Esophagitis (b) Polyps (c) Ulcerative colitis

their work. The authors used 6 visual features generated the texture of the image using Haralick features and Local Binary Patterns (LBP). After applying feature selection, they train the model using kernel discriminant analysis and Logistic Regression. They achieved an F1-score of 0.75. [11] used Bidirectional Marginal Fisher Analysis (BMFA) to extract image features and then fed them to Support Vector Machine (SVM) for the classification task. Transfer learning with data augmentation was used by [12] on KVASIR dataset. Inception V3 was the pre-trained network utilized to fine-tuned the dataset. The model achieved an accuracy of 91.5%. Zhang *et al*. [13] proposed a CNN based approach for gastric pre-cancerous abnormalities classification of ulcer, erosion, and polyp. They used iterative reinforcement learning algorithm and SqueezeNet for the reduction of the model computational time and size. The overall accuracy obtained was 88.90%. In [14], Inception V3 and VVGNet pre-trained models on ImageNet dataset were used in the features extraction stage, and SVM was adopted for classification step, the combination of the extracted features yields very good results. Pogorelov *et al*. [15] worked on 17 different approaches but obtained the best accuracy employing a combination of the pre-trained ResNet50 and Logistic Model Tree (LMT) classifier. The main objective of this study is at the same time, to maximize the performance of our CNN model and minimize the computational time and resources for the task of classifying 8 classes as either disease state, medical procedures, or anatomical landmarks [15].

III. Methodology

In this section, we present our proposed gastrointestinal abnormalities detection system. The entire suggested scheme is depicted in Fig. 4. The first step consists of data pre-processing (resizing, normalization, histogram equalization) and data augmentation(mirroring, flipping, random rotation). Afterward, the processed images are fed to the deep learning architecture for the two main tasks of features extraction and classification. In this step two approaches are adopted, we implemented our own CNN architecture and optimize it by changing the hyper parameters at first. Then, we fine tuned the model based on most popular transfer learning architecture (VGG16, VGG19, MobileNet and Inception V3).

A. Proposed baseline CNN architecture

Convolutional neural network CNN [16] is a network that can be trained end-to-end without the need of any handcrafted methods, it allows the automatic extraction and selection of features and finally the classification or prediction. Generally, a CNN model is composed of five principal layers : input layer, convolutional layers, pooling layers, fully-connected layers and finally output layer. As showed in Table I, the adopted baseline CNN used in our experiments has the following architecture :

- Input layer : consists of 256×256 gastrointestinal images.
- Convolutional layers : the convolution operation is a multiplication of the input with the calculated weights (filters). In our case, we have used 4 convolutional layers with filter size of of 7×7 and zero padding.
- Max-pooling layers : max-pooling technique is used to down sample features maps resulting from convolutional layers, this is done by keeping the maximum value in each patch for every feature map. The max-pooling kernels are set to of 2×2 with a stride of 2.
- ReLu layers : we have used a ReLu layer after each convolutional layer to eliminate the negative values.
- Fully-connected layer : this layer gives the final prediction of each image, we choose sigmoid and softmax as fully connected layers.

B. Transfer Learning

Transfer learning [17] refers to training a network on large well known dataset and then fine tuned it towards the small given new dataset by allowing last layers adaptation. In this work, we adopted the technique of transfer learning and used a number of popular architecture : MobileNet, VGGNets and Inception V3. And fine tuned them using endoscopic images.

1) MobileNet V2: MobileNet V2 [18] is a based CNN model which has an image input size of of 224×224 and made 54 layers. It is based on depthwise separacle convolutions which consists of applying two 1D convolutions with two

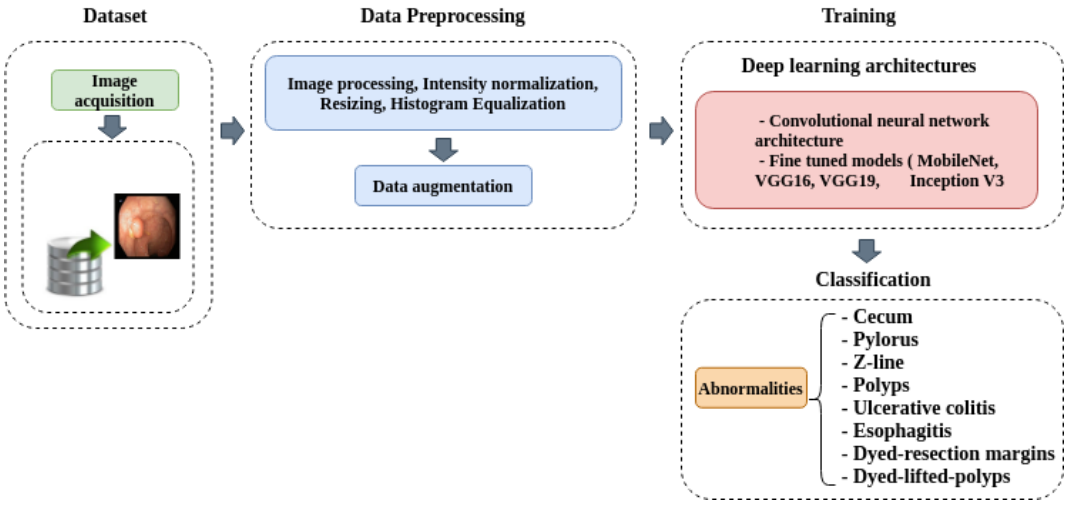


Fig. 4. Block diagram of the proposed method.

TABLE I
CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 63, 63, 32)	4736
activation_1 (Activation)	(None, 63, 63, 32)	0
conv2d_2 (Conv2D)	(None, 57, 57, 32)	50208
activation_2 (Activation)	(None, 57, 57, 32)	0
conv2d_3 (Conv2D)	(None, 51, 51, 32)	50208
activation_3 (Activation)	(None, 51, 51, 32)	0
max_pooling2d_1 (MaxPooling2)	(None, 25, 25, 32)	0
dropout_1 (Dropout)	(None, 25, 25, 32)	0
flatten_1 (Flatten)	(None, 20000)	0
dense_1 (Dense)	(None, 128)	2560128
activation_4 (Activation)	(None, 128)	0
dropout_2 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 8)	1032
activation_5 (Activation)	(None, 8)	0

kernels. This requires less memoru and parameters and gives an efficient small model.

2) VGGNets: VGG (Visual Geometry Group) is a convolutional neural network model proposed by simonyan and Zisserman In [19]. The model achieved 92.7% of accuracy on ImageNet, a dataset of 14 million images associated to 1000 classes. If differs from AlexNet by changing large kernel-sized filters with of 3×3 kernel sized. The most poplar VGGNet networks are VGG16 and VGG19 which contains 16 and 19 layers layers respectively.

3) Inception V3: Inception V3 [20] is an extended version of the familiar GoogleNet which had obtained good clasification results in various biomedical applications. It proposed a model that concatenates multiples different sized convolutional filter into one new filter. It has block of parallel convolutional layers with different size of filters.

IV. RESULTS AND DISCUSSION

In this section, experiments are carried out to evaluate the performance of our models using both CNN architecture and transfer learning networks. Details about the implementation of the method are also described. Finally, the performance of the approach is compared with state-of-the-art results.

A. Dataset

For our experiments, we utilized the recently published, publicly available KVASIR dataset [15]. This dataset has been created for the purpose of improving systems working on the automated classification, detection and localization of endoscopic abnormalities captured on the GI tract KVASIR dataset. It has 8 classes showing several diseases and normal anatomical landmarks. The whole dataset has 8000 images where 1000 images belong to each class. Images of the dataset have resolution from of 720×576 up to of 1920×1072.

B. Experimental setup

Our method has been implemented using python and keras library with tensorflow backend on an Intel (R) core (TM) i7 Gen GHz processor. While preprocessing operations were conducted using Graphical Processing Unit (GPU) NVIDIA GEFORCE GTX 1050 Ti and RAM with 8 GB and 4 GB respectively. Training steps were attended using NVIDIA GPUs available at Google Colab : TITAN RTX Tesla with 25GB memory. To train our model, we set the number of epochs, batch size and learning rate to 300, 1000, 0.001 respectively. We used categorical cross-entropy as loss function, and Adam as an optimizer for the cross-entropy function.

C. Evaluation Metrics

The four key metrics used to evaluate our proposed approach on the KVASIR endoscopic dataset are mentioned bellow.

Accuracy : the ratio of sample that were correctly classified overall prediction.

$$Accuracy = \frac{TruePositive + TrueNegative}{TP + FP + FN + TN} \quad (1)$$

Precision : The number of positive observations, that are correctly identified to the total predicted positive samples.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (2)$$

Recall : the ration of data points correctly predicted as positive to the total true positives in the dataset.

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (3)$$

F1-measure : this metric represents the harmonic mean of recall and precision.

$$F1 - score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (4)$$

D. Experimental results

First, we have performed an evaluation of the selected approach using the augmented dataset randomly splitting into 80% training set and 20% test set. An overview of the results obtained can be found in Table II where we present the measured performance metrics. The comparison between different results shows the superiority of our proposed CNN network in terms of accuracy, precision , recall and F1-score. From this table, we can also observe that our model is more efficient when talking about computational time and number of trainable parameters. Training and validation curves for accuracy and loss values are depicted in Fig. 5.

TABLE II
PERFORMANCE OF DIFFERENT STUDIED ARCHITECTURES IN (%).

Method	ACC	PREC	REC	F1-measure	Parameters
MobileNet	70.09	80.89	80.89	80.89	5.856.968
VGG 16	75.30	56.81	75.37	64.78	16.079.746
VGG 19	78.06	15.05	12.25	13.50	16.081.288
Inception V3	92.89	77.92	68.46	72.88	23.050.344
Our method	96.89	87.79	87.46	87.62	2.666.312

Furthermore, we also present the confusion matrix of the proposed CNN in Fig. 6. The obtained confusion matrix shows that the model can distinguish the dyed lifted polyps and the dyed resection margins very well. Z-line and Pylorus images are also well detected. However, The model is presenting some confusions when identifying Esophgitis images and misclassified them as Z-line, it is been also quite confused between Polyrus and Cecum. Those cases of misclassification may be due to the similarity between the features presented in some classes.

As it can be clearly remarkable through Table III, the results achieved in our study are competitive with the results presented in the previous works with an overall accuracy of 96.89%. Nevertheless, these results can not be comparable because the training and test sets are not the same. Future work suggests to integrate different networks in one ensemble model and improve the system to achieve better results.

TABLE III
OUR METRICS COMPARED TO THE ONES REPORTED IN POGORELOV WORK.

Method	ACC	PREC	REC	F1-measure
6 Layer CNN	0.914	0.661	0.640	0.651
3 Layer CNN	0.959	0.589	0.408	0.453
Inception V3 TFL	0.924	0.698	0.689	0.693
2 GF Random Forrest	0.928	0.713	0.715	0.740
2 GF Logistic Model Tree	0.926	0.706	0.707	0.705
6 GF Random Forrest	0.933	0.732	0.732	0.727
6 GF Logistic Model Tree	0.937	0.748	0.748	0.747
Baseline (JCD RF)	0.927	0.708	0.710	0.706
Our proposed method	0.968	0.874	0.877	0.876

V. CONCLUSION

In this paper, we present five different approaches for the multi-class classification of the gastrointestinal tract abnormalities. The proposed methods are based on CNN architecture and pre-trained models with transfer learning mechanism. Our CNN network showed better performance compared to pre-trained architectures fine tuned to the KVASIR dataset. Compared to other state-of-the-art methods, our approach demonstrated satisfactory results, achieving 96.89% in terms of accuracy. As a future work, we plan to work with Wireless Capsule Endoscopy (WCE) to detect diseases in the GI tract.

REFERENCES

[1] M. Ervik, F. Lam, J. Ferlay, L. Mery, I. Soerjomataram, F. Bray *et al.*, "Cancer today. lyon, france: International agency for research on cancer," *Cancer Today*, vol. 3, pp. 235–248, 2016.

[2] T. De Lange, P. Halvorsen, and M. Riegler, "Methodology to develop machine learning algorithms to improve performance in gastrointestinal endoscopy," *World journal of gastroenterology*, vol. 24, no. 45, p. 5057, 2018.

[3] S. Nadeem, M. A. Tahir, S. S. A. Naqvi, and M. Zaid, "Ensemble of texture and deep learning features for finding abnormalities in the gastro-intestinal tract," in *International Conference on Computational Collective Intelligence*. Springer, 2018, pp. 469–478.

[4] C. Gamage, I. Wijesinghe, C. Chitraranjan, and I. Perera, "Gi-net: anomalies classification in gastrointestinal tract through endoscopic imagery with deep learning," in *2019 Moratuwa Engineering Research Conference (MERCon)*. IEEE, 2019, pp. 66–71.

[5] M. Kirkerød, R. J. Borgli, V. Thambawita, S. Hicks, M. A. Riegler, and P. Halvorsen, "Unsupervised preprocessing to improve generalisation for medical image classification," in *2019 13th International Symposium on Medical Information and Communication Technology (ISMICT)*. IEEE, 2019, pp. 1–6.

[6] S. Charfi and M. El Ansari, "Computer-aided diagnosis system for colon abnormalities detection in wireless capsule endoscopy images," *Multimedia Tools and Applications*, vol. 77, no. 3, pp. 4047–4064, 2018.

[7] S. Charfi, M. El Ansari, and I. Balasingham, "Computer-aided diagnosis system for ulcer detection in wireless capsule endoscopy images," *IET Image Processing*, vol. 13, no. 6, pp. 1023–1030, 2019.

[8] M. Souaidi and M. El Ansari, "Multi-scale analysis of ulcer disease detection from wce images," *IET Image Processing*, vol. 13, no. 12, pp. 2233–2244, 2019.

[9] M. Souaidi, A. A. Abdelouahed, and M. El Ansari, "Multi-scale completed local binary patterns for ulcer detection in wireless capsule endoscopy images," *Multimedia Tools and Applications*, vol. 78, no. 10, pp. 13 091–13 108, 2019.

[10] S. S. A. Naqvi, S. Nadeem, M. Zaid, and M. A. Tahir, "Ensemble of texture features for finding abnormalities in the gastro-intestinal tract." in *MediaEval*, 2017.

[11] Y. Liu, Z. Gu, and W. K. Cheung, "Hkbu at mediaeval 2017 medico: Medical multimedia task," 2017.

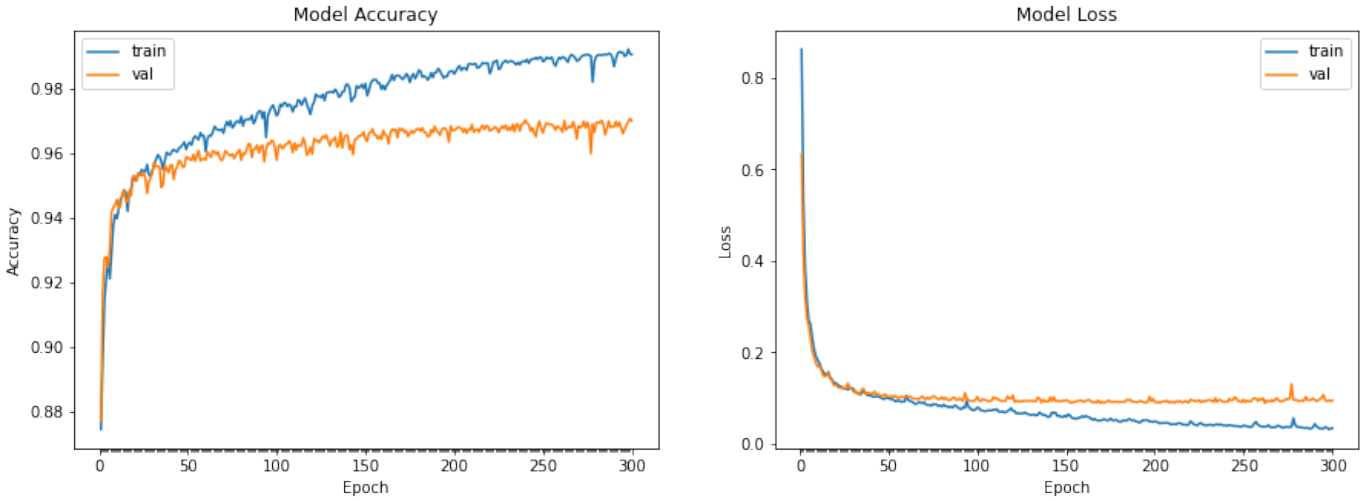


Fig. 5. Accuracy and loss model curves for GI tract diseases classification.

[12] A. Asperti and C. Mastronardo, "The effectiveness of data augmentation for detection of gastrointestinal diseases from endoscopical images," *arXiv preprint arXiv:1712.03689*, 2017.

[13] X. Zhang, W. Hu, F. Chen, J. Liu, Y. Yang, L. Wang, H. Duan, and J. Si, "Gastric precancerous diseases classification using cnn with a concise model," *PloS one*, vol. 12, no. 9, p. e0185508, 2017.

[14] T. Agrawal, R. Gupta, S. Sahu, and C. Y. Espy-Wilson, "Scl-umd at the medico task-mediaeval 2017: Transfer learning based classification of medical images." in *MediaEval*, 2017.

[15] K. Pogorelov, K. R. Randel, C. Griwodz, S. L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D.-T. Dang-Nguyen, M. Lux, P. T. Schmidt *et al.*, "Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection," in *Proceedings of the 8th ACM on Multimedia Systems Conference*, 2017, pp. 164–169.

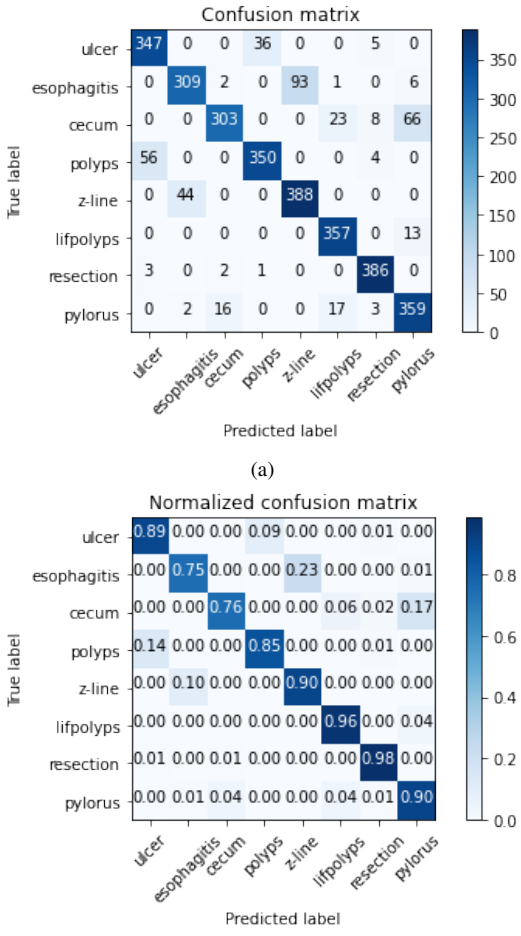
[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[17] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[20] S. Tsang, "Inception-v3–1st runner up (image classification) in ilsvrc 2015," 2015.



(a)

(b)

Fig. 6. Confusion matrix and normalized confusion matrix for GI tract diseases classification.