Wireless Capsule Endoscopy Image Classification

Samy Sekkai
Institut Galilée
Université Sorbonne Paris Nord
samy.sekkai@edu.univ-paris13.fr

***Abstract*—The use of computer-based solution for Wireless Endoscopic images classification is tool that can help medical practitioner reduce diagnosis time and, they can also be used as validation helper for established diagnosis. In this work, we trained three deep learning models for the detection of abnormal images collected from abdominal tract. Xception architecture showed better performance than VGG16 and VGG19. The model was used in a windows application for classification. The application allows the visualization and process of images in an intelligent way.**

Keywords—*kvasir, deep learning,, Gastrointestinal Tract.,*

I. Introduction

Capsule endoscopy also known as wireless or video capsule endoscopy(WCE) is a diagnostic procedure that involves swallowing a small capsule resistant to stomach enzymes to visualize the entire Gastrointestinal tract [1]. The capsule is about the size of vitamin tablet and contains a battery, a transmitter and LED light source and colour video camera. Sensors placed on the patient abdomens receive signals transmitted by the capsule. A wireless recorder places the waist receives and record the data sent by the sensors. The data can be visualised as images or videos on the monitor of the physicians. It allows inspection of the GI tract without discomfort to the patient or need for sedation, thus preventing the risks of conventional endoscopy [2]. Capsule endoscopy helps visualise section of the small intestine that cannot easily be seen or reaches with traditional endoscopic procedures. Capsule endoscopy is recommended for the diagnosis of Gi tract conditions such as Crohn's disease, Ulcerative colitis, Tumors, Polyps, Ulcers, Unexplained GI bleeding, etc. [1]. Some of the benefits of capsule endoscopy over standard endoscopy procedures include:

- No need for anaesthesia or sedation
- Non-invasive and painless
- No need for a hospital stay
- Helps with early and accurate diagnosis of GI problems

II. Related work

WCE procedure produces a huge number of images (50000 ~60000) [3][4]. Visualizing this amount of data by the physicians to generate a diagnosis is a tedious and time-consuming process; add to this the possibility of missing an information during the screening process leading to wrong diagnosis.

To support the physician in this task, several computers aided diagnosis works have been proposed. The most common are, anomaly detection, anomaly segmentation., and anomaly image classification [4].

- Anomaly detection: is about detecting the presence of anomaly as well as region and class,
- Anomaly segmentation: divides the image into multiple partitions by grouping similar pixels.
- Image Classification: classify whether the image is normal or not based on the learned features.

In this project we focus on the Image classification approach. Machine learning and Deep learning models are used for image classification. Recent methods are based on deep learning architecture given their ability to capture and extract high level semantic information. In a survey study made in [4] on medical image analysis, studies on classification were organized by objective, feature extraction approach, network architecture and dataset used. Most of the listed studies were based on CNN in terms of architecture and features. They used custom datasets which makes it difficult to reproduce the performance of their models. Plus, studies focus mostly on detecting a specific abnormality such as bleeding or polyps. A recent study [5] developed a deep learning model for anomaly classification Gastrointestinal Tract. the authors trained state of the art architectures with K-vasir dataset. Performances of the models are shown in table1.

Table 1:performance metrics of trained models [5]

Model name	Test accuracy
VGG16	98.325%
ResNet	92.313%
MobileNet	97.63%
InceptionV3	90.0%
Xception	98.275%r

Table1 shows that Xception and VGG models have higher accuracy than the other architectures. Thus, Xception and VGG16 architecture are selected for our project. We add to them VGG19 architecture since its architecture is derived from VGG16 and more recent.

III. DATASET

The initial data is downloaded from Kvasir source. Three dataset were collected hyper-kvasir [9], kvasirV2 (K2)[10],and kvasir-Capsule(KC)[11].The merged dataset contained 28 classes. According to the description of Kvasir dataset [10] images are mainly abdominal landmarks, or pathologies.

Anatomical Landmarks: recognizable feature within the GI tract that is easily visible through the endoscope. These features are vital for navigating and as a reference point to describe the location of a given finding. The landmarks are “Cecum”, “Z_line”, “and Pylorus”.

Pathological finding: is “abnormal feature within the gastrointestinal tract. Endoscopically, it is visible as a damage or change in the normal mucosa.” The main pathologies described are “Esophagitis”, “Polyps”, and “Ulcerative Colitis”.

As our objective is detection of abnormal images in the GI tract, we select the three pathological classes, and the three anatomical landmarks. Thus, our final dataset contains six classes (three normal, and three pathological). To get a balanced dataset we selected the lowest class occurrence as reference which is Esophagitis (1663 image), so each class will have 1600 sample in the dataset as shown in table 2.

For training the models, we follow the rule of 80:10:10; the dataset is split into train, validation, and test.

Table 2: dataset size

Data	Sample per class	Total
Train	1200	7200
Validation	200	1200
Test	200	1200

IV. ARCHITECTURES OF THE MODELS

For training the three architectures we added to the base architecture layers to adapt the output to the needs of our project.

A. VGG16

An algorithm for object classification and detection. One of the popular algorithms for image classification, it is easy to use with transfer learning [6], its architecture is shown in figure 1

- The number 16 in VGG16 refers to the number of layers with weights (i.e. learnable parameters layers). In VGG16 there are thirteen convolutional layers, five Max Pooling layers, and three Dense layers which sum up to 21 layers, but it has only sixteen weights.
- VGG16 input is a image tensor a size of (224, 244) with 3 RGB channels
- VGG16 has convolution layers of 3x3 filter with stride 1 and always uses the same padding and maxpool layer of 2x2 filter of stride 2; instead of having many hyper-parameters.
- The convolution and max pool layers are consistently arranged throughout the architecture.

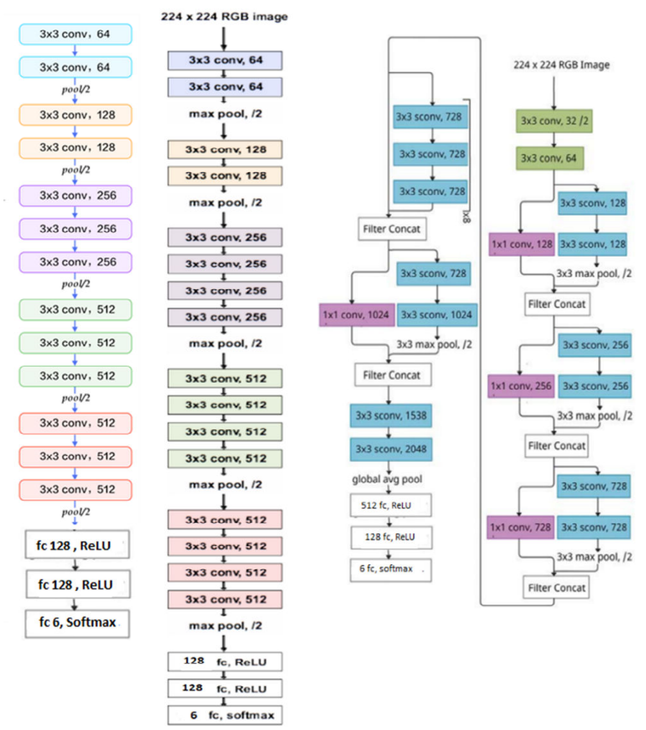

Figure 1: Network architectures with extra layers (from left to right: VGG16, VGG19, ,Xception)

- Conv-1 Layer has 64 number of filters, Conv-2 has 128 filters, Conv-3 has 256 filters, Conv 4 and Conv 5 has 512 filters.
- Three Fully Connected (FC) layers follow a stack of convolutional layers: the first two have 4096 neurons each, the third performs 1000-way ILSVRC classification and thus contains 1000 neurons (one for each class). The final layer is the soft-max layer.

For this project, we added to the base model three layers:

- Fully connected (128) with ReLu activation function,
- Fully connected (128) with ReLu activation function
- Fully connected (6) with softmax activation function.

B. VGG19

This architecture is similar to the one of VGG16; the difference resides on the number of layers in convolutional blocs. While VGG16 has 3 layers per bloc (Bloc 3, 4 and 5), VGG19 comes with 4 layers resulting in an overall of 19 learnable layers (with weights).[7]. Its architecture is shown in figure 1.

For this project, we added to the base model three layers:

- Fully connected (128) with ReLu activation function,
- Fully connected (128) with ReLu activation function
- Fully connected (6) with softmax activation function.

C. Xception

The model is inspired from CNN inception model, and it consists of [8] :

- 36 convolution layers that enable feature extraction.
- The layers are structured in 14 modules (one module is repeated eight times),
- Except the first and last one, all modules have linear direct connections to one another.
- The Depth Wise Separable Convolution (DWSC) is a spatial convolution that is executed independently of one another in parallel via each input channel and is followed by a point-bypoint 1×1 convolution that projects the output of the channel onto a new channel.

For this project, we added to the base model three layers:

- Fully connected (128) with ReLu activation function,
- Fully connected (128) with ReLu activation function
- Fully connected (6) with softmax activation function

V. RESULTS

All the results were obtained from training the model in whith python in jupyterlab environment.

The classes were codified as follow:

Table 3: Classes codification

Class	ID
Cecum	0
Esophagitis	1
Polyp	2
Pylorus	3
Ulcerative_colitis	4
Z_Line	5

A. VGG16

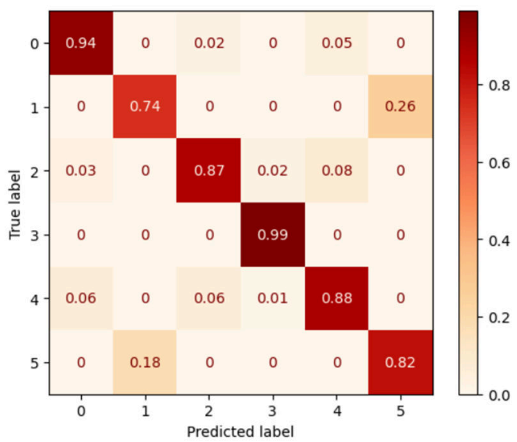

Figure 2: VGG16 normalised confusion matrix for test set

Table 4: VGG16 performance

	Accuracy	Precision	Recall
Train	0.951	0.949	0.952
Validation	0.853	0.852	0.856
Test	0.853	0.856	0.85

B. VGG19

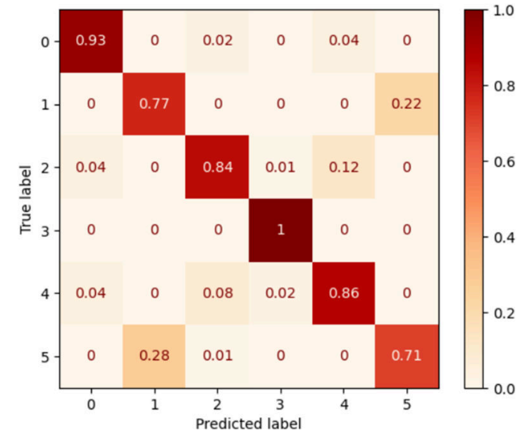

Figure 3:VGG19 normalised confusion matrix for test set

Table 5:VGG19 performance

	Accuracy	Precision	Recall
Train	0.998	0.998	0.998
Validation	0.857	0.856	0.858
Test	0.852	0.855	0.85

C. Xception

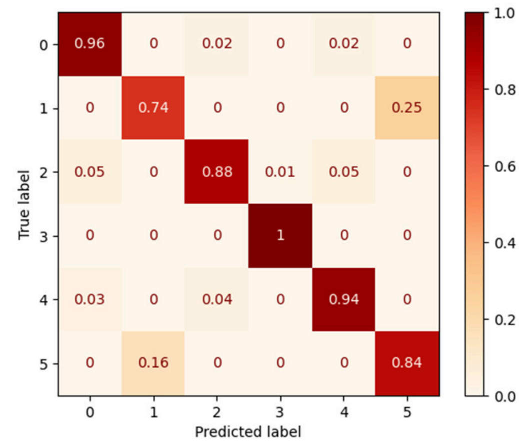

Figure 4: Xception normalised confusion matrix for test set

	Accuracy	Precision	Recall
Train	0.997	0.997	0.997
Validation	0.882	0.881	0.883
Test	0.882	0.881	0.883

The comparison between results of the models shows that Xception model outperforms VGG16 and VGG19 in validation and tests for all metrics.

To analyse, more, the performance of Xception model, we calculate the precision and Recall of the model for each class, results are shown in table 6.

For normal classes, ***Cecum*** and ***Pylorus*** show the highest values for Recall and Precision, meaning the model can classify almost all the sample images of the two classes correctly (Recall). At the same time, it has low confusion with other classes, 92% of images classified as ***Cecum*** were ***Cecum***, and 98% of the images classified as ***Pylorus*** were ***Pylorus*** images.

From the normalized confusion matrix, we have 2% of ***Cecum*** images classified as ***Polyp*** and 2% classified as ***Ulcerative_colitis***. In the case of ***Pylorus*** only one sample was misclassified as ***Ulcerative_colitis***.

The model has the highest confusion between *Z_line* and *Esophagitis*. 16% of *Z_line* images are misclassified as *Esophagitis*, and 25% of *Esophagitis* are misclassified as *Z_line*. the problem is the 25% , this mean that high number of Esophagitis anomaly are not Detected.

Table 6: Xception performance per class for test set

Class		Precision	Recall
Cecum	0	0.92	0.99
Esophagitis	1	0 .82	0.74
Polyp	2	0.95	0.88
Pylorus	3	0.99	0.99
Ulcerative_colitis	4	0.94	0.94
Z_Line	5	0.77	0.84

The trained Xception model is used in windows desktop application to classify WCE images. This application is used as decision aid tool that a medical practitioner can use to help him/her identify a pathology in the images. A screenshot of the application user interface is shown in figure 5.

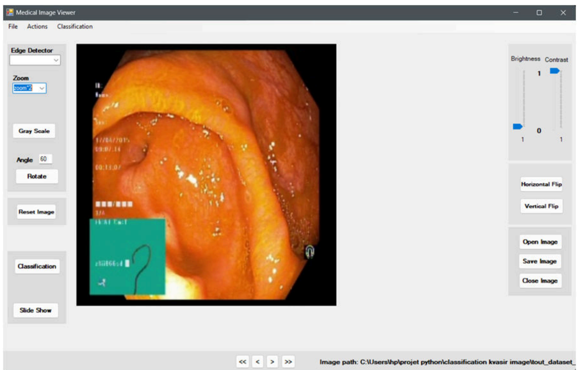

Figure 5: Image visualization application User interface

VI. CONCLUSION

In this work we trained three deep learning models (VGG16, VGG19, Xception) on kvasir Wireless Endoscopic images for the dectection of images with pathologies. Six classes were used 3 as abdomina landmarks and thre others as pathologies. Xception model showed the best performances with all the metrics. The Xception model was used as WCE image classifier in a windows application.

An important limitation in this work is that the classifier model; presents high confusion for some classes(Z_line and Ulcerative_colitis is the highest). We also used a small sized dataset which was not enough for the model to train better

VII. FUTURE WORK

To improve the performance of the model, we plan to apply data augmentation, to increase the size of data. Multiple augmentation techniques can be applied to generate synthetics images from the existing ones to increase the size of the dataset in order to improve the performance.

We added three layers at the ouput of the base model, we plan to test other possible fine tuninge to improve the performnce.

REFERENCES

[1] Silva, J., Histace, A., Romain, O., Dray, X., & Granado, B. (2014). Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. International journal of computer assisted radiology and surgery, 9, 283-293

[2] Bashar, M. K., Kitasaka, T., Suenaga, Y., Mekada, Y., & Mori, K. (2010). Automatic detection of informative frames from wireless capsule endoscopy images. Medical Image Analysis, 14(3), 449-470.

[3] Muruganantham, P., & Balakrishnan, S. M. (2021). A survey on deep learning models for wireless capsule endoscopy image analysis. *International Journal of Cognitive Computing in Engineering*, *2*, 83-92.

[4] Muhammad, K., Khan, S., Kumar, N., Del Ser, J., & Mirjalili, S. (2020). Vision-based personalized wireless capsule endoscopy for smart healthcare: taxonomy, literature review, opportunities and challenges. Future Generation Computer Systems, 113, 266-280.

[5] Dheir, I. M., & Abu-Naser, S. S. (2022). Classification of anomalies in gastrointestinal tract using deep learning.

[6] https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918#:~:text=VGG16%20is%20object%20detection%20and,to%20use%20with%20transfer%20learning.

[7] Leonardo, M. M., Carvalho, T. J., Rezende, E., Zucchi, R., & Faria, F. A. (2018, October). Deep feature-based classifiers for fruit fly identification (Diptera: Tephritidae). In 2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI) (pp. 41-47). IEEE.

[8] Westphal, E., & Seitz, H. (2021). A machine learning method for defect detection and visualization in selective laser sintering based on convolutional neural networks. Additive Manufacturing, 41, 101965..

[9] https://datasets.simula.no/hyper-kvasir/

[10] https://datasets.simula.no/kvasir/

[11] https://datasets.simula.no/kvasir-capsule/