Deep Learning Algorithms for Tissue Identification in Hysteroscopies

Ana Martins[1]
up201405006@fc.up.pt
Francesco Renna[1]
frarenna@dcc.fc.up.pt
Mihaela Gotseva[2]
mihaela.gotseva@gmail.com
Hélder Ferreira[2]
helferreira78@gmail.com
Miguel Coimbra[3]
mcoimbra@dcc.fc.up.pt

[1] Instituto de Telecomumicações
Faculdade de Ciências da Universidade do Porto
Porto, PT
[2] Centro Hospitalar Universitário do Porto
Hospital de Santo António
Porto, PT
[3] INESC TEC
Faculdade de Ciências da Universidade do Porto
Porto, PT

Abstract

In this work, we present a comparison of different deep learning methods to classify images of uterine tissue collected from hysteroscopy exams. The considered solutions are based on the use of different convolutional neural networks and transfer learning strategies and they are applied to two distinct classification problems: i) fully automatic classification of hysteroscopy images and ii) semi-automatic classification of pre-selected portions of hysteroscopy images.

The obtained results testify the potential limitations of deep learning approaches in the presence of very limited training data in the detection of uterine polyps from normal endometrial tissue, where a maximum accuracy of 74% has been achieved. On the other hand, when applied to a semi-automatic task where significant portions of the images are pre-selected, the considered deep learning solutions achieve accuracy values above 92%, also in the presence of a reduced amount of training data.

1 Introduction

Hysteroscopy is a routine gynaecological procedure, which involves insertion of a small camera transvaginally into the uterine cavity in order to identify abnormalities, and in many cases treat them at the same time. As with any surgical procedure, there is a risk of complications, which is overall very low, however in some cases they can have serious long-term consequences. One of the most relevant complications is uterine perforation (UP). The reported incidence of UPs varies from country to country and is reported between 0.12 to 3% in Germany [2], Holland [3], and France [1]. The reason UPs are a concern is that in rare cases they can lead to major haemorrhage, which can require a life-saving hysterectomy. In other cases, UPs can be associated with injury to the bowel, bladder and ureters which often require additional surgical procedures and long-term treatment. In the context of pregnancy, UPs can lead to uterine dehiscence during pregnancy or delivery, which can be life-threatening for the mother and child. Another very rare long-term complication is the formation of fistulas between the abdomen and the uterus.

These rare, but potentially severe complications underline the need for creating computer assisted decision (CAD) systems for hysteroscopy, able to actively recognize the different kinds of tissues explored during the exam in order to further increase the safety of the procedure. A first step in this direction is represented by the development of a classification algorithm able to differentiate different types of uterine tissues from images collected during hysteroscopy.

Although, to the authors knowledge, there are no works in the literature that specifically addressed the problem of classifying images collected during hysteroscopy exams, deep convolutional neural networks (CNNs) are currently regarded as the state-of-the-art for several related biomedical image classification applications. For example, a study done for the classification of endoscopy images of small intestine tissue based on CNNs achieved higher classification sensitivity and shorter reading times than a conventional analysis done by gastroenterologists [5]. Similarly, deep neural networks have been shown to outperform doctors in the accurate differentiation of tiny colorectal polyps [4].

In this paper, we consider two classification tasks on hysteroscopy images that aim to discriminate between normal endometrial tissue and endometrial polyps (Figures 1 and 2). The first task consists in a fully automatic classification of hysteroscopy images, whereas the second task depicts a semi-automatic scenario where pre-selected cropped images are classified via deep CNNs.

2 Methodology

2.1 Materials

A total of 270 images of size 720×576 were collected from 25 patients during hysteroscopy exams perfomed in an outpatient clinic (OC) scenario. In addition, further 230 images were extracted from 11 videos of resolution 1440×1080 recorded during hysteroscopy exams performed under general anaesthetic (GA) in the operating room.

The images in the obtained dataset were divided into two classes by an experienced gynaecologist: normal endometrial tissue (Figure 1) and endometrial polyps (Figure 2). The first class contained 140 images of 13 patients from OC hysteroscopies plus 110 images extracted from hysteroscopy videos of 8 patients. Moreover, 130 images of 12 OC patients plus 120 video frames from 7 GA hysteroscopy patients were included in the second class (Table 1).

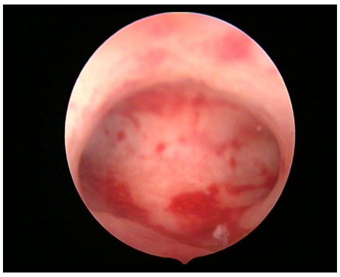

Figure 1: Example of image of normal endometrium tissue.

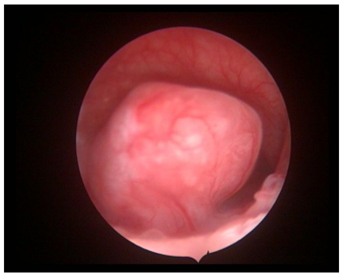

Figure 2: Example of image of endometrial polyp .

The dataset for the semi-automatic classification task was generated from the previous dataset (500 images from 40 patients) by cropping four different significant portions from each image.

Tissues	N° of images	Cropped images	N° of patients
Normal endometrial	140+110	1000	13+8
Endometrial polyp	130+120	1000	12+7

Table 1: Division of images into normal endometrial tissue and endometrial polyp classes

2.2 CNN architectures and training

In this work, two different convolutional neural network architectures were considered: VGG-16 and ResNet-50, as they demonstrate excellent performance in a variety of related biomedical image classification tasks. In addition, the sets of weights of these architectures trained over the ImageNet dataset are publicly available.

VGG-16 and ResNet-50, as they demonstrate excellent performance in a variety of related biomedical image classification tasks.

Thus, for each network architecture, VGG-16 and Resnet-50, three different transfer learning schemes were considered: i) combination of feature extraction and fully connected layers (FE+FC), ii) combination of feature extraction and suppor vector machines (FE+SVM), and iii) fine tuning of convolutional and fully connected layers (FT+FC) , where in all three configurations, the networks were pre-trained over the ImageNet dataset.

Feature extraction consists of freezing the convolutional base of the pre-trained model to prevent the weights of these layers from being updated during training. On the other hand, the fully connected layers of the network are trained from scratch with the data of the considered task, to allow adaptation of the classification to the data set and the analyzed classes. When combining feature extraction with a support vector machines (SVM), the features obtained from the convolutional layers of the pre-trained networks are used as input of an SVM which is trained over the data of the considered task.

In the case of fine adjustment, only the four layers of the convolutional base are frozen for both VGG-16 and ResNet-50. The remaining convolutional layers and fully connected layers are fine tuned using the data of the considered task in order to extract features more related to the particular classification task and to allow better adaptation of the classifier. Note that re-training some of the convolutional layers with the dataset of the specific task considered allows a greater adaptation of the network for the classification objective, but reduces the robustness against overfitting, given the greater number of parameters trained with the small size dataset.

In order to better cope with the reduced size of the available training dataset, data augmentation is applied to all the training configurations. In particular, for each of the training images, 5 different transformations were considered including rotations, mirroring, zooming, and brightness level adjustment.

All networks were trained for 50 epochs, using the Adam optimizer with learning rate 0.0001 and mini-batch size of 32. Additionally, a dropout of 0.4 was used in two layers for each network, between the fully connected layers.

3 Results

In this section, we report the classification results obtained with the different CNN-based setups described in Section 2.2 for the fully-automatic and semi-automatic classification of endometrial images. The classification performance is evaluated using the following metrics: accuracy, precision, recall, and F1-score.

For both classifications task, the images in the dataset were randomly divided into 80% training images and 20% test images, guaranteeing that images from patients in the test set could not be included in the training set. The classification results for this task are reported in Table 2.

	VGG-16			ResNet-50		
Valores	**FE+FC**	**FT+FC**	**FE+SVM**	**FE+FC**	**FT+FC**	**FE+SVM**
Accuracy	0.67	0.54	0.64	0.70	0.74	0.70
Precision	0.69	0.52	0.60	0.70	0.67	0.64
Recall	0.62	0.90	0.86	0.70	0.94	0.92
F1-score	0.65	0.66	0.70	0.70	0.78	0.75

Table 2: Comparison of different transfer learning techniques applied to the VGG-16 and ResNet-50 architectures for the fully automatic classification.

It can be observed that the classification performance is, in general, not very satisfactory, even if a slight advantage is obtained when using the ResNet-50 architecture. The poor performance registered is mainly caused by the lack of a larger training set, thus leading to significant overfitting, and by the presence of specific features in the images that can lead to misclassification. In particular, several errors are observed in the classification of images of normal endometrial tissue, since a significant portion of them contains the channels of the fallopian tubes (Figure 3), which are often confounded with the presence of polyps. On the other hand, the proposed algorithms often fail in detecting small polyps from images (Figure 4).

Table 3 contains the results obtained when applying the CNN-based methods described in Section 2.2 to the semi-automatic task of classifying pre-selected cropped images from the original dataset. In this case the proposed, architectures are able to achieve significantly better performance, thus guaranteeing reliable discrimination between endometrial polyps and normal tissue.

Figure 3: Example of image from normal endometrium with fallopian tube channel.

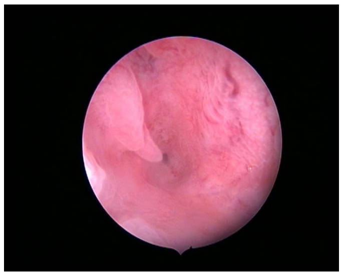

Figure 4: Example of image with the presence of a small polyp.

	VGG-16			ResNet-50		
Valores	**FE+FC**	**FT+FC**	**FE+SVM**	**FE+FC**	**FT+FC**	**FE+SVM**
Accuracy	0.96	0.93	0.92	0.95	0.95	0.96
Precision	0.97	0.89	0.89	0.92	0.92	0.94
Recall	0.94	0.97	0.96	0.99	0.98	0.99
F1-score	0.96	0.93	0.93	0.96	0.95	0.97

Table 3: Comparison of techniques in the transfer learning application to the VGG-16 and ResNet-50 architecture for the semi-automatic classification.

The data considered for this tasks are portions of the original images, which may facilitate the training of the network due to the 4x increase of the dataset size. This has allowed the network to extract the necessary characteristics in order to distinguish the classes. Moreover, the considered cropped images represent lower-dimensional data with reduced variability, thus simplifying the corresponding classification task.

4 Conclusion

The problem of classifying images obtained from an hysteroscopy exam using CNN-based classifier was considered. Different network architectures and transfer learning techniques were tested to discriminate normal endometrial tissue images from endometrial polyps.

When considering a fully automatic classification of hysteroscopy images, the use of fine tuning on a ResNet-50 architecture pre-trained over the ImageNet dataset is shown to provide interesting classification results even in the presence of a strictly reduced training dataset.

On the other hand, classification of pre-selected portions cropped from the original images is shown to be reliably performed even with such a small training dataset, due to the reduced variability of the considered samples.

Acknowledgments

This work is funded by FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/50008/2020.

References

[1] Aubert Agostini et al. Risk of uterine perforation during hysteroscopic surgery. *The Journal of the American Association of Gynecologic Laparoscopists*, 9:264–7, 08 2002. doi: 10.1016/S1074-3804(05)60401-X.

[2] Burkhard Aydeniz et al. A multicenter survey of complications associated with 21 676 operative hysteroscopies. *European journal of obstetrics, gynecology, and reproductive biology*, 104:160–4, 10 2002. doi: 10.1016/S0301-2115(02)00106-9.

[3] Frank William Jansen et al. Complications of hysteroscopy: a prospective, multicenter study. *Obstetrics and gynecology*, 96 2:266–70, 2000.

[4] Peng-Jen Chen et al. Accurate classification of diminutive colorectal polyps using computer-aided analysis. *Gastroenterology*, 154, 10 2017. doi: 10.1053/j.gastro.2017.10.010.

[5] Zhen Ding et al. Gastroenterologist-level identification of small bowel diseases and normal variants by capsule endoscopy using a deep-learning model. *Gastroenterology*, 2019.