Histopathological classification of precursor lesions of esophageal adenocarcinoma: A Deep Multiple Instance Learning Approach

Jakub M. Tomczak **Maximilian Ilse** **Max Welling**
University of Amsterdam

Marnix Jansen
University College Hospital, London

Helen G. Coleman
Queen's University Belfast

Marit Lucas **Kikki de Laat** **Martijn de Bruin** **Henk Marquering**
Myrtle J. van der Wel **Onno J. de Boer** **C. Dilara Savci Heijink** **Sybren L. Meijer**
the Academic Medical Center, Amsterdam

Abstract

In this paper, we hypothesize that morphological properties of nuclei are crucial for classifying dysplastic changes. Therefore, we propose to represent a whole histopathology slide as a collection of smaller images containing patches of nuclei and adjacent tissue. For this purpose, we use a deep multiple instance learning approach. Within this framework we first embed patches in a low-dimensional space using convolutional and fully-connected layers. Next, we combine the low-dimensional embeddings using a multiple instance learning pooling operator and eventually we use fully-connected layers to provide a classification. We evaluate our approach on esophagus cancer histopathology dataset.

1 Introduction

Over the last decades, a steep increase in the incidence of esophageal adenocarcinoma (EA) is observed. EA has a dismal prognosis. Barrett's esophagus is the only known precursor lesion and progresses through a metaplasia-dysplasia-carcinoma sequence. A histopathological diagnosis of low grade dysplasia in surveillance biopsies is one of the strongest independent risk factors for progression. Reported progression rates to high grade dysplasia and adenocarcinoma, however, vary greatly from 1- 40% [1]. This diagnostic variation results in over-interpretation of clinical non-relevant lesions and overtreatment at one hand and under-interpretation of progressive lesions and diagnostic delay at the other [1]. International guidelines now mandate a second opinion whenever the diagnosis of dysplasia is considered.

Histopathology is a part of a complex clinical medicine in whici patient tailored diagnostic work-up requires doctors to handle large volumes of data from imaging but also macro level physiology, laboratory test, and, increasingly,"-omic" data. As a result, this complexity leads to unwanted high diagnostic variability in patient care. There is now an urgent clinical need for next generation diagnostic methods for optimal risk stratification.

Deep learning is currently the state-of-the-art machine learning technique, which was proven to be successfully applied in numerous field of medical imaging [3]. We propose to utilize deep learning to classify H&E slides. However, processing whole slides could be computationally demanding, hence, we focus on nuclei and surrounding tissue. Our hypothesis is that **morphological properties**

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

of nuclei are crucial for classifying dysplastic changes. As a result, we process collections of image patches containing nuclei using the idea of multiple instance learning [4] with deep learning for feature extraction and classification.

2 Methodology

Problem formulation In the classical (binary) supervised learning problem one aims at finding a model that predicts a value of a target variable, $y \in \{0, 1\}$, for a given instance, $\mathbf{x} \in \mathbb{R}^D$. In the considered case, however, instead of a single instance there is a bag of i.i.d. instances, $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$. We assume that K could vary for different bags. There is also a single binary label Y associated with the bag. Individual labels of instances, *i.e.*, $y_1, \ldots, y_K$ and $y_k \in \{0, 1\}$, for $k = 1, \ldots, K$, remain unknown during training. Further, the bag label is determined by the individual labels that could be formulated in the following manner:

$$Y = \max_k \{y_k\}. \tag{1}$$

The considered problem is known as the **multiple instance learning** (MIL). Since the individual labels are unknown, there are two main MIL approaches, namely, *the instance level approach* that tries to predict individual labels and next combine them by a MIL pooling operator, and *the embedding-level approach* that embeds instances to a low-dimensional representation and further combine them to a fixed size vector that is fed to a final classifier. The individual labels are latent, therefore, it is advocated to use the latter approach in order to avoid a possible bias of determining y's.

In this paper, we propose to use the embedding-level approach with four different MIL pooling operators, namely, the MAX-operator:

$$\forall_{m=1,\ldots,M} : z_m = \max_{k=1,\ldots,K} \{\mathbf{h}_{km}\}, \tag{2}$$

the MEAN-operator:

$$\mathbf{z} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{h}_k, \tag{3}$$

and the weighted mean operator:

$$\mathbf{z} = \sum_{k=1}^{K} a_k \mathbf{h}_k, \tag{4}$$

with weights determined by the ATTENTION mechanism [2]:

$$a_k = \mathrm{softmax}\big(\mathbf{w}^\top \tanh(\mathbf{V}\mathbf{h}_k^\top)\big), \tag{5}$$

where $\mathbf{w} \in \mathbb{R}^L$ and $\mathbf{V} \in \mathbb{R}^{L \times M}$ are parameters and $\mathrm{softmax}$ is the softmax non-linearity, or weights determined by the GATED ATTENTION mechanism [2]:

$$a_k = \mathrm{softmax}\Big(\mathbf{w}^\top \big(\tanh(\mathbf{V}\mathbf{h}_k^\top) \odot \mathrm{sigm}(\mathbf{U}\mathbf{h}_k^\top)\big)\Big), \tag{6}$$

where $\mathbf{U} \in \mathbb{R}^{L \times M}$ are parameters, $\odot$ is an element-wise multiplication and $\mathrm{sigm}(\cdot)$ is the sigmoid.

Moreover, we parameterize the model using neural networks. First, each instance is encoded into a low-dimensional embedding using a series of convolutional layers with pooling layers (CNN) followed by fully connected layers (FCN). Next, the low-dimensional embeddings are transformed into a single fixed-size representation by a MIL pooling. Eventually, fully connected layers are utilized to provide final prediction. The whole architecture is schematically presented in Figure 1.

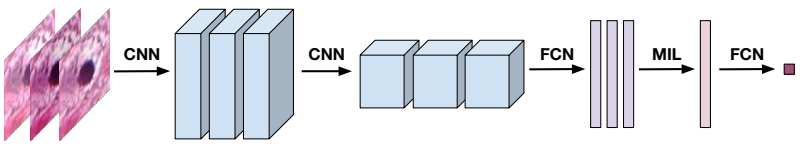

Figure 1: A schematic representation of the deep multiple instance learning approach.

3 Experiment

Data The dataset was collected at the Academic Medical Center in Amsterdam and it contains histopathology slides of 21 patients with annotations of healthy regions and dysplastic changes (low grade and high grade dysplasia). Biopsies and endoscopic mucosal resection specimens of the lower part of the esophagus were digitalized using the Philips Ultrafast scanner with a $40\times$ objective (.25μm/pixel). The digitalized slides were manually delineated by two expert observers and subsequently checked by an expert GI pathologist using an in-house developed delineation tool. When uncertainty arose as to whether a gland was benign of malignant, the immunohistochemically stained slides with p53 were consulted, (iv) and integration of the proposed approach for histopathology slides with other sources of patient data (*e.g.*, genomics). In the experiment, each slide was divided into smaller images (500×500 pixels). Next, each small image was represented as a collection of patches (27×27 pixels) and a patch contains a nucleus in the center and adjacent tissues.[1] Eventually, each image was labeled as $Y = 1$ if it contained dysplastic changes; otherwise it was labeled as $Y = 0$. We divided the dataset into fixed subsets: training set (9 patients), validation set (6 patients) and test set (6 patients).

Table 1: An average with a standard error of the test accuracy of our approach with different MIL pooling operators and the majority class. All experiments were repeated 5 times.

MIL pooling	Accuracy
MAJORITY CLASS	52.1%
MAX	$58.7\% \pm 1.8\%$
MEAN	$67.2\% \pm 2.5\%$
ATTENTION	$65.9\% \pm 3.5\%$
GATED ATTENTION	$\mathbf{70.1}\% \pm 1.7\%$

Discussion The results of the experiment are presented in Table 1. First, we notice that the choice of the MIL pooling operator makes a huge difference in terms of the classification accuracy. Second, MEAN and ATTENTION perform similarly (the difference is not statistically significant). However, the GATED ATTENTION outperforms all other operators. Third, the best accuracy is at the level of around 70% that indicates a potential usefulness in clinical practice.

4 Final remarks

In this paper, we present a first step towards a diagnostic histopathology tool for reliable and reproducible classification of direct precursor lesions of esophageal adenocarcinoma. Our preliminary results indicate usefulness of the proposed deep multiple instance learning approach. Nevertheless, more thorough evaluation is necessary. We aim at the following steps in our future work: (i) more thorough evaluation using cross-validation methodology, (ii) collecting more data, (iii) in the current approach we hypothesize that nuclei contain necessary information, however, morphological properties of glandular architecture or patterns of cells are of diagnostic importance as well, therefore, we will consider larger patchesto integrate these features.

Acknowledgments JMT was funded by the European Commission within the MSC-IF (Grant No. 702666). MI was funded by the NWO (Grant DLMedIa). ML was funded by ITEA3 (Grant No. ITEA151003).

References

[1] L.C. Duits et al. Patients with Barrett's esophagus and confirmed persistent low-grade dysplasia are at increased risk for progression to neoplasia. *Gastroenterology*, 152(5):993–1001, 2017.

[2] M. Ilse, J.M. Tomczak, and M. Welling. Attention-based deep multiple instance learning. *arXiv preprint arXiv:1802.04712*, 2018.

[3] G. Litjens et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

[4] G. Quellec, G. Cazuguel, B. Cochener, and M. Lamard. Multiple-instance learning for medical image and video analysis. *IEEE Reviews in Biomedical Engineering*, 2017.

[1] We used CellProfiler (`http://cellprofiler.org/`) to segment nuclei out.