Multiple Instance Learning for Histopathological Breast Cancer Images

P J Sudharshana, Caroline Petitjean, Fabio Spanhol, Luís Oliveira, Laurent Heutte, Paul Honeine

Pattern Recognition Letters
journal homepage: www.elsevier.com

Multiple Instance Learning for Histopathological Breast Cancer Images

P J Sudharshan[a,b], Caroline Petitjean[b,**], Fabio Spanhol[c], Luis Oliveira[c], Laurent Heutte[b], Paul Honeine[b]

[a]*Indian Institute of Information Technology - D&M, Jabalpur, India*
[b]*Normandie Université, Université de Rouen, LITIS, Rouen, France*
[c]*Federal University of Paraná, Department of Informatics (DInf), Curitiba, PR - Brazil*

ABSTRACT

Weakly supervised learning arises in many situations where the process of labeling the data is expensive. Multiple instance learning (MIL) provides an elegant framework to deal with this issue by organizing instances into bags, without the need to label all the instances. In this paper, we investigate the relevance of MIL for a computer-aided diagnosis system based on the analysis of histopathological breast cancer images. The experiments are conducted on the BreaKHis public dataset of about 8,000 microscopic biopsy images of benign and malignant breast tumors. By providing an extensive comparative analysis of MIL methods, it is shown that a recently proposed, non-parametric approach exhibits particularly interesting results. The comparison between MIL and single instance (conventional) classification reveals the relevance of the MIL paradigm.

1. Introduction

Supervised learning is a subfield of machine learning where a predictive function is inferred from a set of labeled training examples, in order to map each input instance to its output label. In a conventional setting, the training dataset consists of instances equipped with their corresponding labels. While instances are relatively easy to obtain, the expensive data-labeling process with human-based ground-truth descriptions remains the major bottleneck to have large-scale datasets. This issue gave rise to a novel paradigm in machine learning, with the so-called weakly supervised learning, namely when having a partially-labeled training dataset (Zhou, 2017).

Multiple Instance Learning (MIL) provides an elegant framework to deal with weakly supervised learning. In comparison with strong (*i.e.*, fully-labeled) supervised learning where every training instance assigned with a discrete or real-valued label, the rationale of MIL paradigm is that instances are naturally grouped in labeled bags, without the need that all the instances of each bag have individual labels. In the binary classification case, a bag is labeled positive if it has at least one positive instance; on the other hand, a bag is labeled negative if all its instances are negative (Foulds and Frank, 2010; Dietterich et al., 1997). With such training data grouped in labeled bags, MIL algorithms seek to classify either unseen bags (*i.e.,* bag-level classification) or unseen instances (*i.e.,* instance-level classification).

While the multiple instance paradigm arose in many domains prior to the 1990's, MIL was first described explicitly and studied by Dietterich *et al.* in 1997 (Dietterich et al., 1997). The original motivation in MIL is drug activity prediction, where experts provide activity labels to bags of molecules, labeling each individual molecule being costly and hard to set up. It turns out that the MIL is central in many relevant applications in various domains, such as in bioinformatics, text processing, computer vision and image processing, to name a few. Indeed, in many applications, ground-truth labeling is expensive in general and instances can be often grouped in bags, each bag having a set of partially-labeled instances. Of particular interest is image-based pathology classification for medical decision making, since it is relatively easy and part of the clinical protocol to take many images of some organs or tissues (physiology) under study; on the other hand, labeling each image is a time-consuming process dominated by human effort. MIL has indeed many applications in medical imaging, as shown in a recent review (Quellec et al., 2017).

In this paper, we focus on the classification of histopathological breast cancer images. Histopathological images are microscopic images of the tissue for disease examination.

**Corresponding author:
e-mail: caroline.petitjean@univ-rouen.fr (Caroline Petitjean)

Histopathological images prevail as the gold standard for cancer diagnosis, as well as many other diseases (Rubin et al., 2008). Preliminary work have shown the interest of MIL histopathological image classification applications on small datasets, see for example (Xu et al., 2014) for colon cancer and cytology. This work focuses on breast tumors, one of the most common types of cancer. In particular, we consider the recently established and public BreaKHis database (Spanhol et al., 2016b), which contains about 8,000 microscopic biopsy images of benign and malignant breast tumors, originating from 82 patients. While MIL is especially suitable for this application, no study has yet leveraged multiple instance learning for large datasets, with a comparative analysis of the state-of-the-art, as investigated in this paper.

The relevance of MIL for this type of application and dataset is naturally described in two different ways.

The first possibility is to divide each image into subimages or patches and to consider the image as a bag, while patches are the instances. In the field of natural scene images, this is related to region-based image categorization, where each instance encodes color, textural or spatial features related to that specific region (Herrera et al., 2016). In our binary setting, the image would be labeled "positive" (pathological) if it has at least one malignant patch; conversely, an image would be labeled benign if it does not have any portion labeled malignant. This multiple instance formalism is natural, since only a subset of the patches are labeled by experts, making it possible that entire images might be healthy whereas the patient is diagnosed with a tumor. This is not the case in the conventional strategy used so far, in a single instance classification setting with instances inheriting the label of their image.

Second, the patient is considered as a bag, with the instances being its associated hundred of images or pieces of images, called patches. This makes full sense as: the diagnostic (*i.e.,* the label) is established only at the patient level. Furthermore, a patient diagnosed with a malignant tumor can still have some of its images described as tumor-free, *i.e.,* healthy, as just said; and a healthy patient has inevitably all of his images healthy. This hypothesis matches the MIL assumption. In natural scene image classification, this approach is related to facial recognition for example, several images of the same person taken from different angles (Herrera et al., 2016). Note that only the MIL paradigm can apprehend this type of situations.

We propose to tackle the problems of histopathological image classification and patient diagnosis through the benchmark of several MIL methods, as a first contribution. We consider the state-of-the-art of MIL methods. In particular we investigate the seminal Axis-Parallel Rectangle algorithm (APR) (Dietterich et al., 1997), and algorithms based on diversity density (DD) (Maron and Lozano-Pérez, 1998; Zhang and Goldman, 2001), k-NN (Citation-kNN) (Wang and Zucker, 2000) and Support Vector Machines (SVM) (Andrews et al., 2002), as well as a recently-proposed non-parametric algorithm (Venkatesan et al., 2015) and a deep learning approach revisiting Convolutional Neural Networks (CNN) for MIL (MILCNN) (Sun et al., 2016). As a second contribution we will study how MIL results compare to a single instance classification results, which is the only framework implemented on this data until now. Of course in this case we suppose that instances inherit labels from the bags. We will examine if it is preferable to cast this problem into a single instance one, or if MIL does indeed bring an added value, both at the image and patient levels (Alpaydin et al., 2015).

The remainder of this paper is organized as follows. Section 2 presents the MIL and provides a survey of MIL methods. Section 3 describes the BreaKHis dataset and the conducted experiments with the obtained results. Section 4 concludes the paper.

2. MIL methods: a brief overview

Under the standard MIL assumption, positive bags contain at least one positive instance, while negative bags contain only negative instances. We denote by L_B the label of a bag B, defined as a set of instances, each one described by its feature vector: $B = \{b_1, b_2, ..., b_N\}$. We denote by l_k the label of each instance b_k. We can now define the label of a bag, following the standard MIL assumption:

$$L_B = \begin{cases} +1 & \text{if } \exists l_k : l_k = +1; \\ -1 & \text{if } \forall l_k : l_k = -1. \end{cases} \quad (1)$$

There are other – more relaxed – assumptions, such as a bag is labeled positive when it contains a sufficient number of positive instances; since they are out of the scope of this paper, we refer the reader to (Foulds and Frank, 2010) for further reading.

MIL methods are usually divided into two groups, depending on how they exploit the information in the data (Amores, 2013). The first group consists of methods that consider the discriminant information at the instance level. Learning algorithms focus not at the larger scale of a bag, but at the local scale of instances. An advantage of these methods is that they can perform instance classification task when needed. However, they require that instances have a precise label, a requirement not all MIL problems meet. The instance level methods include APR, DD, SVM based approaches. The second group consists of the methods that consider the discriminative information to be in the bag level. These methods usually are more accurate, since they can model distribution and relations between classes (Carbonneau et al., 2016). An example of such methods is Citation-kNN (Wang and Zucker, 2000). For a review on MIL methods, we refer the reader to (Herrera et al., 2016; Amores, 2013; Carbonneau et al., 2016).

In remainder of this section, we briefly describe the well-established MIL methods that have been implemented and applied to the BreaKHis dataset.

2.1. Axis-parallel hyper rectangle (APR)

The MIL paradigm was first introduced in the seminal work of Dietterich et al. (1997), motivated mainly by an application in biochemistry. The goal is to predict whether a molecule will be binding to a given receptor or not. Each molecule, which can be considered as a bag, can take many different spatial conformations, namely the instances. The methodology to solve the MIL problem is to design an hyper rectangle (called

axis-parallel hyper rectangle (APR)) in the feature space aimed at containing at least one positive instance from each positive bag while excluding all the instances from negative bags. A molecule is classified as positive (resp. negative) if one (resp. none) of its instances belongs inside the APR.

2.2. Diverse Density (DD) and its variants

Diverse diversity (Maron and Lozano-Pérez, 1998) is closely related to the idea of the APR. The DD defines a function over the feature space, such that it is high at points that are both close to instances from positive bags, and far away from instances which are in negative bags. The DD algorithm attempts to find the local maxima of this function (called the positive instance targets or prototypes) by maximizing diverse density (i.e. conditional likelihood) over the instance space, using gradient ascent with multiple starting points. The DD approach has given rise to many variants, the most known is the Expectation-Maximization method (EM-DD) (Zhang and Goldman, 2001). In this variant, the DD measure is maximized iteratively with the EM algorithm.

2.3. Citation-kNN

The Citation-kNN, an adaptation of k-nearest neighbors (k-NN) algorithm, is the first non-parametric approach (Wang and Zucker, 2000). The principle is to first apply the k-NN algorithm to bags, where the distance between bags is measured with the minimum Hausdorff distance. The latter is defined as the shortest distance between any two instances from each bag, namely

$$\mathrm{Dist}(A, B) = \min_{a_i \in A} \min_{b_j \in B} \|a_i - b_j\|$$

for any two bags A and B, where a_i and b_j are instances from each bag. This distance is used by a k-NN to classify a new bag, in the same sense as the regular k-NN approach. The citation-kNN method adds a final step that makes the process more robust: in addition to the nearest bags, the bags that count as their neighbors (called *citers*) are also considered.

2.4. mi-SVM and MI-SVM

Two alternative generalizations of the maximum margin idea used in SVM classification have been proposed in (Andrews et al., 2002). On one hand, the mi-SVM is based on the instance-level paradigm. Since the instance labels are not known, they are treated as hidden variables subject to constraints defined by their bag labels. The mi-SVM method attempts to recover the instance labels and, at the same time, to find the optimal discriminant function. On the other hand, the bag-level paradigm is adopted by the so-called MI-SVM. Its goal is to maximize the bag margin, defined between the positive instances of the positive bags, and the negative instances of the negative bags. In this setting, the bag is not represented by all its instance, but only by the "extreme" ones, in the same sense as *support vectors* in conventional SVM. Moreover, mi-SVM and MI-SVM inherit also the kernel trick, thus allowing to use linear, polynomial and RBF kernels.

2.5. Non-parametric MIL

This recent technique is designed as a modified version of the k-NN classifier (Venkatesan et al., 2015). The non-parametric MIL approach employs a new formulation based on distances to k-nearest neighbors. The idea is to parse the MIL feature space with a Parzen window technique, using different sized regions. Conversely to the majority vote used in k-NN, the vote contributions are the kernelized distances in the feature space. Non-parametric MIL has shown enhanced robustness to labeling noise on various datasets.

2.6. MILCNN

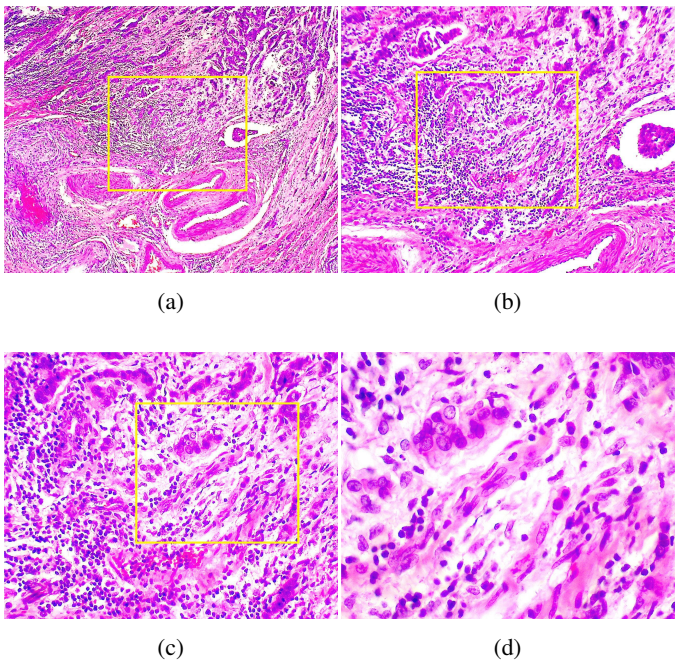
Deep learning networks have been overwhelming machine learning, pattern recognition and computer vision fields for a few years. MIL is no exception to this rule (Hoffman et al., 2016; Pathak et al., 2014; Kraus et al., 2016; Sun et al., 2016; Zhou et al., 2017; Wang et al., 2018). In (Sun et al., 2016), a Multiple Instance Learning Convolutional Neural Networks (MILCNN) is proposed. This framework was initially proposed for the data augmentation problem: in object detection, labels are not always preserved when the images are split for data augmentation. The proposed method considers data augmentation generated images as a bag, by combining a convolutional neural network (CNN) with a specific MIL loss function derived with respect to the bag.

3. Experiments and results

3.1. Description of the BreaKHis dataset

BreaKHis is a publicly available dataset of microscopic biopsy images of benign and malignant breast tumors (Spanhol et al., 2016b). The images were collected through a clinical study in 2014, to which all patients referred to the P&D Laboratory, Brazil, with a clinical indication of breast cancer were invited to participate. The institutional review board approved the study and all patients provided their written informed consent. All the data were anonymized. Samples were generated from the breast tissue biopsy slides, stained with hematoxylin and eosin (HE). The samples were collected by surgical open biopsy (SOB), prepared for histological study and labeled by pathologists of the P&D Lab. The diagnosis of each case were produced by experienced pathologists and confirmed by complementary exams such as immunohistochemistry analysis.

Images were acquired in RGB color space, with a resolution of 752×582 using magnifying factors of 40×, 100×, 200× and 400×. Fig. 1 shows these 4 magnifying factors on a single image. This image is acquired from a single slide of breast tissue containing a malignant tumor (breast cancer). The highlighted rectangle (manually added for illustrative purposes only) is the area of interest selected by pathologist to be detailed in the next higher magnification. To date, the database is composed of 7,909 images divided into benign and malignant tumors. Table 1 summarizes the image distribution. For more information about the dataset, we refer to (Spanhol et al., 2016b).



(a) (b)

(c) (d)

Fig. 1: A slide of breast malignant tumor seen in different magnification factors of the same image: (a) 40×, (b)100×, (c) 200×, and (d) 400×

Table 1: Image distribution by magnification factor and class

Magnification	Benign	Malignant	Total
40×	625	1,370	1,995
100×	644	1,437	2,081
200×	623	1,390	2,013
400×	588	1,232	1,820
Total	2,480	5,429	7,909
# Patients	24	58	82

3.2. Experimental protocol

Following the standard labeling convention in use in medical studies, the label "positive" (resp. "negative") refers to malignant (resp. benign) images. The BreaKHis dataset has been randomly divided into a training set (70%) and a testing set (30%), in which patients used to build the training set are not used for the testing set. We used this division and the predefined 5-fold cross-validation consistently with the protocol described in (Spanhol et al., 2016a). We computed the average rate over runs. Note that the folds are publicly available and allow for a fair comparison of methods. To handle the image high resolution (752×582) and to augment data for training, images were divided into 64×64 patches. Thousand patches were randomly extracted from each input image for training. Each patch is described with a 162-long feature vector of Parameter-Free Threshold Adjacency Statistics (PFTAS) features (Hamilton et al., 2007; Coelho et al., 2010). There features have shown particularly relevant for this dataset, when assessed against many other such as local binary patterns (LBP), completed LBP, local phase quantization, gray-level co-occurrence matrices, as well as computer vision features such as ORB (oriented FAST and rotated BRIEF) (Spanhol et al., 2016b).

Twelve MIL methods were evaluated on the BreaKHis dataset, as described in the Section 2: APR, DD and EM-DD, citation-kNN, mi-SVM and MI-SVM, both with linear, polynomial and RBF kernels, non-parametric MIL, and MILCNN. For all methods except the non-parametric and the MILCNN, we used the implementation of the J. Yang's MIL Library[1] with MATLAB 2017a. The non-parametric MIL algorithm was obtained from the author's website[2]. For the implementation of MILCNN in Python, Keras and Theano were used (Chollet, 2015). The hyper-parameters for each method were optimized using grid search for the BreaKHis dataset as shown in Appendix A.

In the following, we first show the benchmark of MIL methods, and then assess the best MIL method against single instance classification frameworks.

3.3. Results

MIL benchmark on BreaKHis dataset

We provide results for two differents settings, as aforementioned. In the first setting, each patient is considered as a bag, which is labeled with its diagnosis. This is possible with our dataset, since several hundreds of images are available for each patient, as shown in Table 1. In the second setting, we consider each image as a bag; in this case, the instances are the patches.

As expected (see Table 2 and Fig. 2 and 3), DD-based approaches and APR yield the poorest results which leads us to think that positive instances are not clustered in a single area of the feature space. For SVM-based approaches, MI-SVM leads to enhanced results, which shows that a bag level paradigm is better suited to the data. At last, best classification rates are reported with the non-parametric MIL approach.

MIL vs single instance learning

For these experiments, we collected results obtained from single instance classification setting, using state-of-the-art classifiers such as 1-NN, quadratic discriminant analysis (QDA), random forest, and SVM. Hyperparameters of these classifiers were tuned using grid search and only the best results were retained. These classifiers take as input the PFTAS feature vector describing each image. For the CNN approach, we used AlexNet (Krizhevsky et al., 2012). Decisions are taken on each patch and are fused together using the Max Fusion Rule.

Unsurprisingly, the CNN performs better than other machine learning models trained with hand-crafted textual descriptors (in accordance with (Han et al., 2017); however, their results are not comparable since they do not use the same folds), see Fig. 4 and Fig. 5. We observe that the non-parametric MIL brings interesting improvements for all magnification factors (except the 400×) at patient level. This suggests that instances, namely patches, provide only partial, complementary information for the image or the patient level (Alpaydin et al., 2015), and that a bag-based analysis is fully valuable for the analysis of histopathology images.

[1] CMU MIL toolbox: `http://www.cs.cmu.edu/~juny/MILL/`

[2] `https://github.com/ragavvenkatesan/np-mil`

Table 2: Accuracy rate at respective levels. Best results columnwise are in bold.

	Patient as bag				Image as bag			
	40×	100×	200×	400×	40×	100×	200×	400×
Iterated-discrim APR (Dietterich et al., 1997)	73.8 ± 3.8	66.5 ± 4.1	84.2 ± 4.9	68.0 ± 5.6	70.4 ± 2.4	65.1 ± 5.0	81.3 ± 5.5	67.3 ± 4.9
DD (Maron and Lozano-Pérez, 1998)	70.5 ± 6.1	64.5 ± 4.3	68.3 ± 3.6	71.2 ± 3.3	71.2 ± 5.9	66.1 ± 5.4	66.7 ± 2.9	70.8 ± 3.8
EM-DD (Zhang and Goldman, 2001)	78.3 ± 5.6	80.6 ± 5.2	77.1 ± 6.3	78.7 ± 5.7	73.1 ± 5.4	76.4 ± 4.8	78.2 ± 5.2	76.2 ± 5.6
Citation-kNN (Wang and Zucker, 2000)	73.7 ± 4.6	72.8 ± 5.4	75.7 ± 3.1	77.2 ± 3.6	73.1 ± 4.3	73.0 ± 5.7	71.3 ± 3.5	78.7 ± 3.1
mi-SVM Linear (Andrews et al., 2002)	79.5 ± 4.3	83.4 ± 4.6	83.6 ± 4.7	81.0 ± 5.2	72.6 ± 4.4	80.6 ± 3.7	80.1 ± 4.9	78.2 ± 5.3
mi-SVM poly (Andrews et al., 2002)	75.2 ± 6.1	79.8 ± 4.8	76.5 ± 3.9	68.5 ± 5.1	75.6 ± 5.7	78.7 ± 4.0	75.2 ± 5.6	69.2 ± 4.8
mi-SVM RBF (Andrews et al., 2002)	77.8 ± 1.6	75.4 ± 1.5	73.8 ± 2.3	72.9 ± 3.4	77.9 ± 2.2	77.3 ± 2.1	74.6 ± 2.9	71.4 ± 3.9
MI-SVM Linear (Andrews et al., 2002)	85.6 ± 5.6	82.1 ± 5.9	84.6 ± 4.8	80.9 ± 4.9	79.5 ± 4.1	78.2 ± 4.4	80.8 ± 4.7	78.9 ± 5.1
MI-SVM poly (Andrews et al., 2002)	84.8 ± 2.7	82.5 ± 4.6	83.9 ± 4.2	81.3 ± 4.2	86.2 ± 2.8	82.8 ± 4.8	**81.7 ± 4.4**	**82.7 ± 3.8**
MI-SVM RBF (Andrews et al., 2002)	79.0 ± 2.1	71.9 ± 2.9	76.2 ± 1.9	73.0 ± 3.5	78.3 ± 3.2	72.2 ± 3.0	76.8 ± 1.6	71.9 ± 2.4
Non-parametric (Venkatesan et al., 2015)	**92.1 ± 5.9**	**89.1 ± 5.2**	**87.2 ± 4.3**	82.7 ± 3.0	**87.8 ± 5.6**	**85.6 ± 4.3**	80.8 ± 2.8	**82.9 ± 4.1**
MILCNN (Sun et al., 2016)	86.9 ± 5.4	85.7 ± 4.8	85.9 ± 3.9	**83.4 ± 5.3**	86.1 ± 4.2	83.8 ± 3.1	80.2 ± 2.6	80.6 ± 4.6

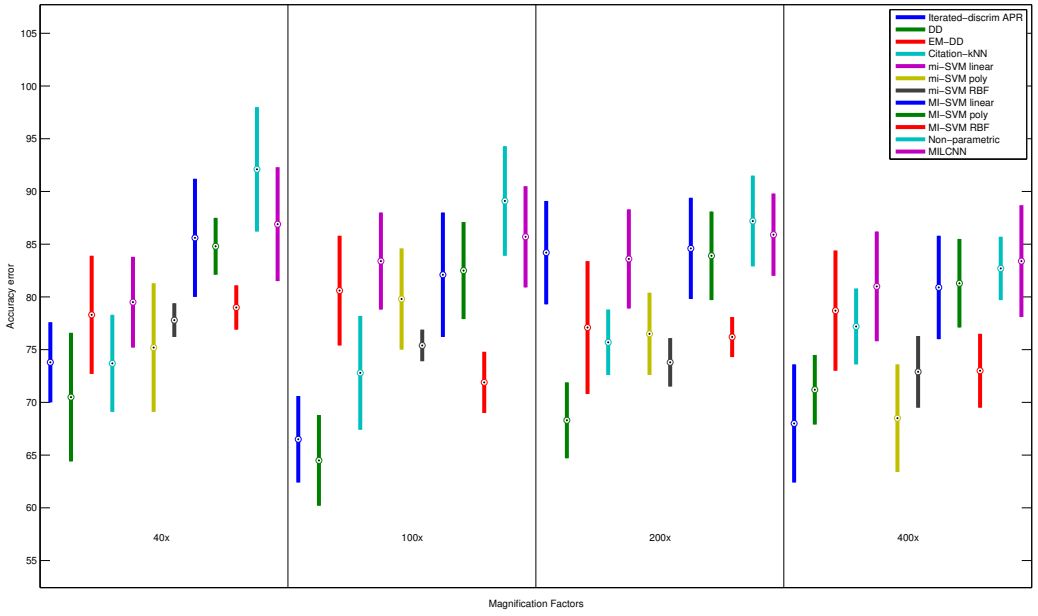


Fig. 2: Accuracy results of MIL benchmark with patient as bag (left part of Table 2)

4. Conclusions and future work

Multiple instance learning provides a classification framework that is particularly adapted to computer-aided diagnosis based on histopathological image analysis. In the case of the BreaKHis dataset, several hundreds of images are available per patient. The patient can thus be considered as a bag, which is labeled with its diagnosis.

Our MIL benchmark shows that the recently proposed non-parametric MIL is particularly efficient for the tasks of patient and image classification. Patient classification rates can reach up to 92.1% for the 40× magnification factor, a level never reached by conventional classification frameworks, which enhances the fact that instances are complementary and can be fruitfully considered in a MIL framework. MIL can thus leverage digital histopathological image classification and analysis to improve computer-aided diagnosis.

As future work, we are currently engaged in experimenting other deep learning frameworks. With the acceleration of proposals in this area, no doubt that a more efficient networks will be proposed in the near future. We also want to investigate MIL for histopathological image segmentation. MIL can indeed be an adequate framework to find location of malignant region position in histopathological images (Pathak et al., 2014; Xu et al., 2014; Kraus et al., 2016). Since manual labeling is too long, MIL can help in pixel labeling and clustering and can serve as a feedback to the pathologist. The image is considered as a bag and the pixels as instances.

Appendix A. Method hyper-parameterization

For non-parametric MIL (Venkatesan et al., 2015):

- Averaged accuracy over 100 runs

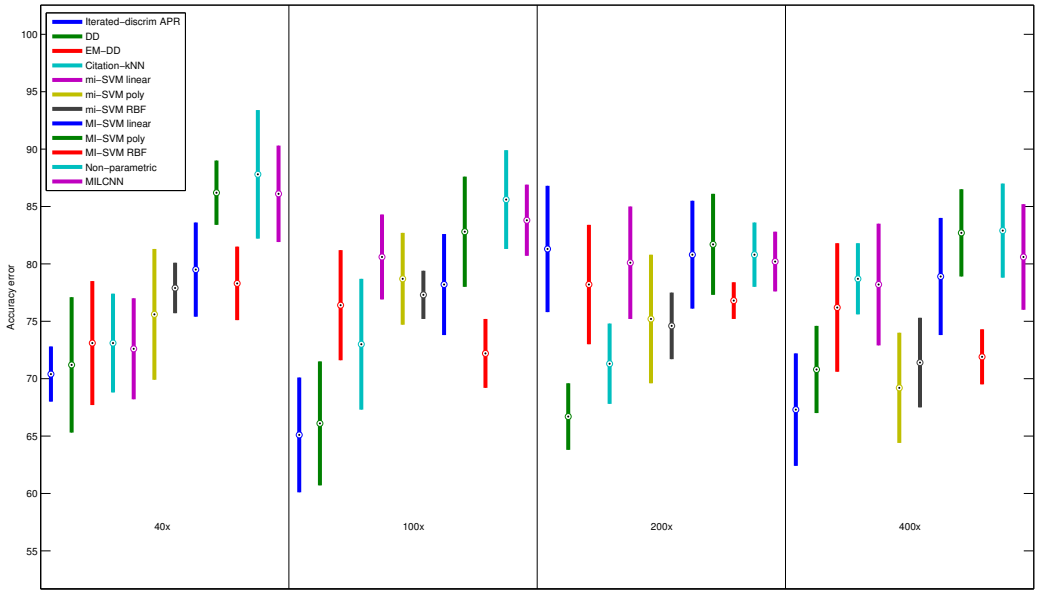


Fig. 3: Accuracy results of MIL benchmark with image as bag (right part of Table 2)

Table 3: Comparison of MIL (non-parametric) vs single instance classification (SIL). Best results columnwise are in bold.

		Patient as bag (MIL) or level (SIL)				Image as bag (MIL) or level (SIL)			
		40×	100×	200×	400×	40×	100×	200×	400×
MIL	Non-parametric	**92.1 ± 5.9**	**89.1 ± 5.2**	**87.2 ± 4.3**	82.7 ± 3.0	**87.8 ± 5.6**	**85.6 ± 4.3**	80.8 ± 2.8	**82.9 ± 4.1**
SIL	CNN	90.0 ± 6.7	88.4 ± 4.8	84.6 ± 4.2	**86.1 ± 6.2**	85.6 ± 4.8	83.5 ± 3.9	**83.1 ± 1.9**	80.8 ± 3.0
	1-NN	80.9 ± 2.0	80.7 ± 2.4	81.5 ± 2.7	79.4 ± 3.9	79.1 ± 2.1	77.8 ± 3.0	79.6 ± 1.9	77.6 ± 4.0
	QDA	83.8 ± 4.1	82.1 ± 4.9	84.2 ± 4.1	82.0 ± 5.9	82.8 ± 3.6	80.7 ± 4.9	**83.3 ± 3.0**	80.5 ± 5.6
	RF	81.8 ± 2.0	81.3 ± 2.8	83.5 ± 2.3	81.0 ± 3.8	80.2 ± 1.9	80.4 ± 3.8	82.4 ± 2.3	80.0 ± 4.5
	SVM	81.6 ± 3.0	79.9 ± 5.4	85.1 ± 3.1	82.3 ± 3.8	79.9 ± 3.7	77.1 ± 5.5	84.2 ± 1.6	81.2 ± 3.6

- Range of k for grid search: 50 (1-50) using elbow method
- No. of Tsteps: 3000
- Distance Method: Euclidean

For APR (Dietterich et al., 1997):

- Kernel Width: 0.999
- Outside Probability: 0.023
- GridNum: 25000

For DD (Maron and Lozano-Pérez, 1998):

- Scaling: 1
- Aggregate: average
- Threshold: 0.5
- No. of runs: 100

For EM-DD (Zhang and Goldman, 2001):

- Scaling: 1
- Aggregate: average
- Threshold: 0.5
- No. of runs: 500
- Iteration Tolerance: 0.08

For Citation-kNN (Wang and Zucker, 2000):

- Bag Distance Type: minimum
- Instance Distance Type: Euclidean
- Reference nodes considered: 5
- CiterRank: 11

For mi-SVM (Andrews et al., 2002):

- Kernel: Linear, poly, RBF
- KernelParam - NA/degree/gamma: (NA), 4, 0.32
- CostFactor: 1/0.96/1
- NegativeWeight: 1/1/1
- Threshold: 0.5/0.55/0.5

For MI-SVM (Andrews et al., 2002):

- Kernel: Linear, poly, RBF
- KernelParam - NA/degree/gamma: (NA), 5, 0.17
- CostFactor: 1/1/1
- NegativeWeight: 1/1/1
- Threshold: 0.5/0.5/0.5

For MILCNN (Sun et al., 2016), the structure is the same as that of the MILCNN for CIFAR10/CIFAR100.

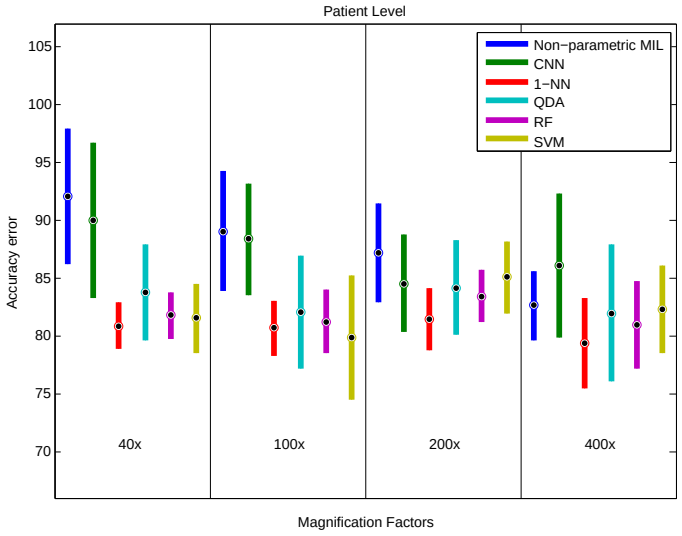


Fig. 4: Accuracy results: MIL vs SIL at patient level (left part from Table 3)

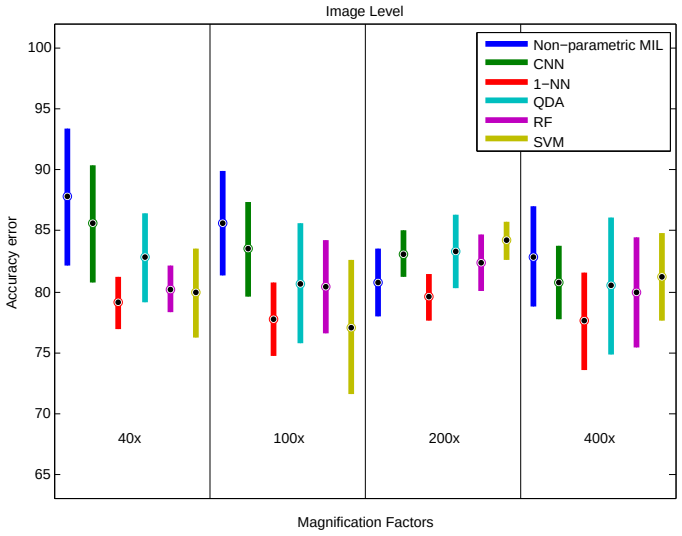


Fig. 5: Accuracy results: MIL vs SIL at image level (right part from Table 3)

Acknowledgment

The authors acknowledge the CRIANN (Centre des Ressources Informatiques et Applications Numérique de Normandie, France) for providing computational resources.

References

Alpaydin, E., Cheplygina, V., Loog, M., Tax, D.M.J., 2015. Single- vs. multiple-instance classification. Pattern Recognition 48, 2831–2838.

Amores, J., 2013. Multiple instance classification: Review, taxonomy and comparative study. Artificial Intelligence 201, 81 – 105.

Andrews, S., Tsochantaridis, I., Hofmann, T., 2002. Support vector machines for multiple-instance learning, in: Proceedings of the 15th International Conference on Neural Information Processing Systems, MIT Press, Cambridge, MA, USA. pp. 577–584.

Carbonneau, M.A., Cheplygina, V., Granger, E., Gagnon, G., 2016. Multiple instance learning: A survey of problem characteristics and applications. CoRR abs/1612.03365.

Chollet, F., 2015. Keras: Theano-based deep learning library. Code: https://github. com/fchollet. Documentation: http://keras.io .

Coelho, L.P., Ahmed, A., Arnold, A., Kangas, J., Sheikh, A.S., Xing, E.P., Cohen, W.W., Murphy, R.F., 2010. Structured Literature Image Finder: Extracting Information from Text and Images in Biomedical Literature. Springer. pp. 23–32.

Dietterich, T.G., Lathrop, R.H., Lozano-Prez, T., 1997. Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence 89, 31 – 71.

Foulds, J., Frank, E., 2010. A review of multi-instance learning assumptions. The Knowledge Engineering Review 25, 125.

Hamilton, N., S Pantelic, R., Hanson, K., Teasdale, R., 2007. Fast automated cell phenotype classification 8, 110.

Han, Z., Wei1, B., Zheng, Y., Yin, Y., Li, K., Li, S., 2017. Breast cancer multi-classification from histopathological images with structured deep learning model. Scientific Reports 7.

Herrera, F., Ventura, S., Bello, R., Cornelis, C., Zafra, A., Sánchez-Tarragó, D., Vluymans, S., 2016. Multiple instance learning, in: Multiple Instance Learning. Springer, pp. 17–33.

Hoffman, J., Wang, D., Yu, F., Darrell, T., 2016. Fcns in the wild: Pixel-level adversarial and constraint-based adaptation. arXiv preprint arXiv:1612.02649 .

Kraus, O.Z., Ba, J.L., Frey, B.J., 2016. Classifying and segmenting microscopy images with deep multiple instance learning. Bioinformatics 32, i52–i59.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, pp. 1106–1114.

Maron, O., Lozano-Pérez, T., 1998. A framework for multiple-instance learning, in: Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems 10, pp. 570–576.

Pathak, D., Shelhamer, E., Long, J., Darrell, T., 2014. Fully convolutional multi-class multiple instance learning. CoRR abs/1412.7144.

Quellec, G., Cazuguel, G., Cochener, B., Lamard, M., 2017. Multiple-instance learning for medical image and video analysis. IEEE reviews in biomedical engineering 10, 213–234.

Rubin, R., Strayer, D., Rubin, E., McDonald, J., 2008. Rubin's Pathology: Clinicopathologic Foundations of Medicine. Lippincott Williams & Wilkins.

Spanhol, F.A., Oliveira, L.S., Petitjean, C., Heutte, L., 2016a. Breast cancer histopathological image classification using convolutional neural networks, in: 2016 International Joint Conference on Neural Networks (IJCNN), pp. 2560–2567.

Spanhol, F.A., Oliveira, L.S., Petitjean, C., Heutte, L., 2016b. A dataset for breast cancer histopathological image classification. IEEE Transactions on Biomedical Engineering 63, 1455–1462.

Sun, M., Han, T.X., Liu, M.C., Khodayari-Rostamabad, A., 2016. Multiple instance learning convolutional neural networks for object recognition, in: International Conference on Pattern Recognition (ICPR), pp. 3270–3275.

Venkatesan, R., Chandakkar, P.S., Li, B., 2015. Simpler non-parametric methods provide as good or better results to multiple-instance learning, in: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 2605–2613. doi:10.1109/ICCV.2015.299.

Wang, J., Zucker, J.D., 2000. Solving the multiple-instance problem: A lazy learning approach, in: Proceedings of the Seventeenth International Conference on Machine Learning, pp. 1119–1126.

Wang, X., Yan, Y., Tang, P., Bai, X., Liu, W., 2018. Revisiting multiple instance neural networks. Pattern Recognition 74, 15 – 24.

Xu, Y., Zhu, J.Y., Eric, I., Chang, C., Lai, M., Tu, Z., 2014. Weakly supervised histopathology cancer image segmentation and classification. Medical image analysis 18, 591–604.

Zhang, Q., Goldman, S.A., 2001. Em-dd: An improved multiple-instance learning technique, in: In Advances in Neural Information Processing Systems, MIT Press. pp. 1073–1080.

Zhou, L., Zhao, Y., Yang, J., Yu, Q., Xu, X., 2017. Deep multiple instance learning for automatic detection of diabetic retinopathy in retinal images. IET Image Processing .

Zhou, Z.H., 2017. A brief introduction to weakly supervised learning. National Science Review 00, 1 – 10. doi:10.1093/nsr/nwx106.