

# Automatic Diagnosis of Ovarian Carcinomas via Sparse Multiresolution Tissue Representation

Aïcha BenTaieb<sup>\*</sup>, Hector Li-Chang<sup>†</sup>, David Huntsman<sup>†</sup>, Ghassan Hamarneh<sup>\*</sup>

<sup>\*</sup> Medical Image Analysis Lab, Simon Fraser University, Burnaby, BC, Canada

<sup>†</sup> Departments of Pathology and Laboratory Medicine and Obstetrics and Gynaecology, University of British Columbia, Vancouver, Canada

**Abstract.** It has now been convincingly demonstrated that ovarian carcinoma subtypes are not a single disease but comprise a heterogeneous group of neoplasms. Whole slide images of tissue sections are used clinically for diagnosing biologically distinct subtypes, as opposed to different grades of the same disease. This new grading scheme for ovarian carcinomas results in a low to moderate interobserver agreement among pathologists. In practice, the majority of cases are diagnosed at advanced stages and the overall prognosis is typically poor. In this work, we propose an automatic system for the diagnosis of ovarian carcinoma subtypes from large-scale histopathology images. Our novel approach uses an unsupervised feature learning framework composed of a sparse tissue representation and a discriminative feature encoding scheme. We validate our model on a challenging clinical dataset of 80 patients and demonstrate its ability to diagnose whole slide images with an average accuracy of 91% using a linear support vector machine classifier.

**Keywords:** Histopathology, ovarian carcinomas, machine learning.

## 1 Introduction

Recent advances in epithelial ovarian cancer diagnosis have shown that morphologic subtypes of ovarian carcinomas (OC) are associated with distinct pathologic and molecular characteristics. This resulted in the introduction of a new grading system for OC diagnosis [12, 15]. The World Health Organization recommends characterizing OC as five distinct tumour types: High Grade Serous Carcinoma, Low Grade Serous Carcinoma, Endometrioid Carcinoma, Mucinous Carcinoma and Clear Cell Carcinoma (HGSC, LGSC, EN, MC and CC). At present, targeted therapies are being introduced for each subtype and successful treatments are highly correlated with the accurate classification of these subtypes [13].

Pathologists diagnose OC from tumour biopsies (Fig. 1). Samples of the tumorous tissue are collected and examined using Hematoxylin & Eosin (H&E) stained tissue sections at different microscope magnifications [10]. However, diagnosis from histopathology images is impaired by technical factors (e.g. lighting, staining variability, operator acquisition procedure) and by pathologists' experience. Non-expert pathologists often end up performing additional costly

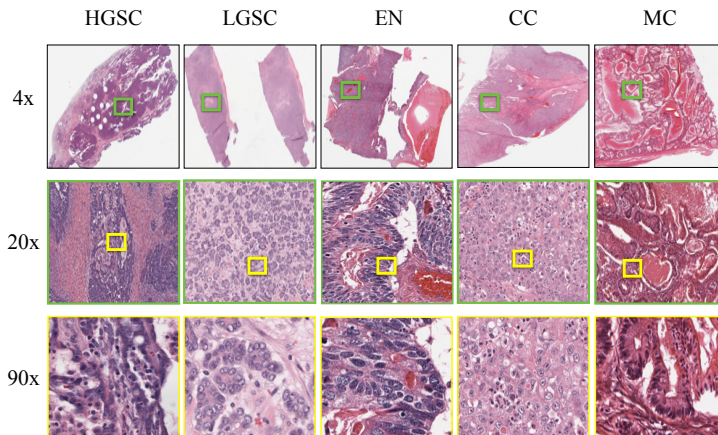


Fig. 1: Ovarian carcinoma subtypes at different microscope resolutions. HGSC: High Grade Serous Carcinoma, LGSC: Low Grade Serous Carcinoma, EN: Endometrioid carcinoma, CC: Clear cell Carcinoma and MC: Mucinous Carcinoma. Each carcinoma subtype is shown at 4x, 20x and 90x microscope zoom.

tests (e.g. immunohistochemistry) or asking for second opinions. The variability among tissues coupled with the limited knowledge of ovarian carcinomas subtypes translates into a moderate agreement among pathologists [11] and a high mortality rate among patients [1].

A system capable of automatically classifying OC subtypes from whole slide histopathology images would be valuable for several reasons. First, an automatic system provides computational abilities enabling rapid screening and learning from large scale multiresolution images (i.e. too large/detailed for a human observer to examine thoroughly). Second, such an automated system may act as a second reader while mimicking expert pathologists. Finally, it can benefit the diagnostic procedure by minimizing the inter-observer variability among pathologists while adding robustness to the diagnosis.

There is a vast literature on classification and grading of cancer from histopathology images [7, 8]. One widely used approach is to design a set of features, usually based on texture (e.g. SIFT, Gabor filters) or segmentation [14, 4]. More recent studies have applied unsupervised feature learning techniques to classify cancerous from non-cancerous regions of tissues [9, 3]. For cancer typing from a tissue section (i.e. classification of subtypes of ovarian cancer), the task has been shown to be more challenging [16]. Xu et al. [16] proposed to classify subtypes of colon cancer using multiple instance learning. This weakly-supervised framework uses hand-designed features and annotated data to first classify a tissue section as cancerous (or not). Then, a patch-based clustering is applied on cancerous tissue slides to identify different subtypes of colon cancer at different spatial locations. In OC diagnosis, this clustering framework is not fully suitable. In fact, different regions of OC tissues do not correspond to a different type but to different

grades of the same subtype, which increases the intra-class variability. Thus, OC diagnosis is a multiclass classification rather than a patch-based clustering problem.

To the best of our knowledge, no existing work has addressed the automatic typing of OC subtypes. The question we raise in this paper is whether it is possible to automate the analysis of OC subtypes despite the limited existing pathogenetic understanding of the disease, the high variability among patients and within tissues (in terms of staining and grades) and the low agreement between pathologists. To this end we make the following contributions: (i) we design an unsupervised feature learning technique based on a hybrid model combining a sparse multiresolution representation of tissue sections with a discriminative feature encoding scheme; (ii) we demonstrate that our technique achieves a better performance on OC than state-of-the-art unsupervised feature learning methods proposed in histopathology image classification; (iii) we show that our unified framework captures more complex and discriminative patterns of textures and shapes that are more suitable for a multiclass typing of tissues. Ultimately, we validate the proposed approach on real clinical data and show that our pipeline provides marked improvements over existing techniques.

## 2 Approach

At a high level, our approach proceeds as follows. We represent a tissue slide  $X$  with a set of unlabeled multiresolution patches  $[x_1, \dots, x_P]$  extracted at different spatial locations. For each patch  $x_i$  we learn a new sparse image representation using a multi-layer deconvolution network (DN) [17]. This representation is then used to encode a high-dimensional set of discriminative features  $\psi(x_i)$  via Fisher Vector Encoding (FVE) [5]. Finally, using a linear SVM, we predict a carcinoma subtype  $y_i \in \mathcal{Y}$  where  $\mathcal{Y} = \{HGSC, CC, EN, MC, LGSC\}$  for each multi-resolution patch of a tissue section. To infer a final carcinoma for the whole slide, we aggregate the classifiers' prediction probabilities,  $P(y_i|\psi(x_i))$ , for all patches. Next, we provide the details of these methodological components.

**Feature learning:** To begin, our goal is to find a new image representation more suitable for classification. This representation can then be used to extract robust local image descriptors. This is specifically challenging for OC as the appearance of the tissue widely differs at different locations and among patients. To overcome these challenges, we adopt an unsupervised feature learning approach which has shown to produce more robust features in presence of wide technical and biological variations [4]. Using a DN, we learn a set of filters that allow us to reconstruct the original image from convolutions with feature maps (Fig. 2-a). A feature map can be considered as an activation map where the values are filters' responses. These filters and feature maps are estimated using a unified optimization technique based on the convolutional decomposition of an image under a sparsity constraint [17]. Feature maps are inferred in a hierarchical fashion (Fig. 2-b) by stacking layers of sparse convolutions to form a multi-layer DN.

More concretely, a single layer DN decompose an RGB patch  $x_i$  into a linear sum of  $K_1$  latent feature maps  $z_{k,1}^i$  convolved with filters  $f_{k,1}$  where  $k = [1, \dots, K_1]$ . This reconstruction relies on optimizing the following energy with respect to  $z$  and  $f$ :

$$E(x_i) = \lambda \left\| \sum_{k=1}^{K_1} z_{k,1}^i \otimes f_{k,1} - x_i \right\|_2^2 - \|z_{k,1}^i\|_1, \quad (1)$$

where the first term is the reconstruction error and the second term encourages sparsity in the latent feature maps.  $\lambda$  controls the balance between the contributions of the reconstruction and sparsity. The sign  $\otimes$  corresponds to the convolution operation. For a  $N \times M$  image and filters  $f_k$  of size  $H \times H$ , the resulting feature maps are of size  $(N + H - 1) \times (M + H - 1)$ . To better capture multiresolution patterns, we use a set of filters with adaptive sizes.

In practice, first layer filters of a DN  $f_{k,1}$  (learnt from minimizing  $E$  over all training set) are Gabor-like filters and represent low-level visual information from the image (Fig. 2-a). To capture more complex patterns, we learn a hierarchy of filters by stacking multiple layers of deconvolutions. The hierarchy is formed by treating the feature maps of layer  $l - 1$  as input to layer  $l$ . Each of these layers attempts to directly minimize the reconstruction error of the input patch while inferring sparse feature maps. During learning, we use the entire set of patches to seek for latent feature maps for each image and learnt filters. At layer  $l$ , we minimize the following energy function:

$$E_l(x) = \lambda \sum_{i=1}^P \sum_{j=1}^{K_{l-1}} \left\| \sum_{k=1}^{K_l} (z_{k,l}^i \otimes f_{k,l}) - z_{j,l-1}^i \right\|_2^2 + \sum_{i=1}^P \sum_{k=1}^{K_l} \|z_{k,l}^i\|_1 \quad (2)$$

where  $P$  is the total number of images,  $K_{l-1}$  and  $K_l$  are the total number of feature maps at layer  $l - 1$  and  $l$  respectively,  $z_{k,l}$  are the inferred feature maps at layer  $l$ , and  $f_{k,l}$  are the learnt filters. Feature maps  $z_{j,l-1}^i$  are inferred at layer  $l - 1$ . A max-pooling operation is applied on feature maps between layers. The energy function  $E_l(x; z, f)$  is biconvex with respect to  $z$  and  $f$  thus the optimization is solved using an iterative procedure [17]. At inference, we set the optimal  $z$  given a set of filters  $f$  and input images  $x$ . This corresponds to solving a convex energy function with a sparsity constraint. This step is optimized via stochastic gradient descent which showed to be efficient for large-scale problems [17].

Each patch is now represented by a hierarchical set of feature maps. We densely construct local image descriptors by splitting each feature map into overlapping quadrants of fixed size and pooling over the absolute value of activations in each quadrant. This pooling procedure adds translation invariance to the local descriptors [17]. We obtain a final set of local descriptors  $\phi(x_i)$  by concatenation of the pooled activations from each feature map (Fig. 2-c).

**Fisher Vector Encoding:** Encoding features from the local descriptors  $\phi(x_i)$  is a critical step for designing discriminative features [6]. We use FVE to define a mapping or “encoding” of the descriptors  $\phi$  into a higher dimensional

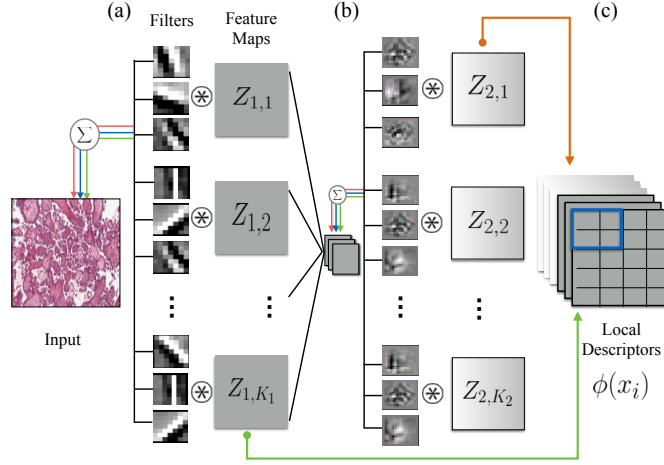


Fig. 2: Representation of a 2-layer DN. Feature maps  $Z_{j,k}$ , convolved with learnt low and mid-level filters, form a sparse representation of an image.  $K_1$  and  $K_2$  refer to the number of feature maps in each layer. Aggregation of the feature maps is used to construct dense local descriptors  $\phi$ . In our implementation, we use a 3-layer DN.

feature space. Here, our assumption is that higher dimensional features facilitate class-separation and enable us to use simple linear classifiers. To do this, we first learn a codebook of Gaussian Mixture Models (GMM) from the total set of local descriptors. Each Gaussian models the probability  $P(\phi|\theta)$ , where  $\theta = (\mu_g, \sigma_g, \pi_g : g = 1, \dots, G)$  are the parameters representing the mean, covariance and prior probabilities of the Gaussian distribution. The GMM can be thought of as a soft dictionary of words in a bag-of-words (BoW) scheme. To encode features from dense local descriptors, we used FVE to compute the average first and second order differences between descriptors and GMM centers. For each GMM  $g$  and descriptor  $d \in \phi$ , we compute the following vectors:

$$\begin{aligned}
 \mathcal{G}_{\mu,g} &= \frac{1}{\mathcal{D}\sqrt{\pi_g}} \sum_{k=1}^{\mathcal{D}} \gamma_d(g) \frac{(d - \mu_g)}{\sigma_g}, \\
 \mathcal{G}_{\sigma,g} &= \frac{1}{\mathcal{D}\sqrt{2\pi_g}} \sum_{k=1}^{\mathcal{D}} \gamma_d(g) \left[ \frac{(d - \mu_g)^2}{\sigma_g^2} - 1 \right], \\
 \gamma_d(g) &= \frac{\pi_g \mathcal{N}(d; \mu_g, \sigma_g)}{\sum_{i=1}^G \pi_i \mathcal{N}(d; \mu_i, \sigma_i)},
 \end{aligned} \tag{3}$$

$\mathcal{D}$  is the total number of local descriptors per patch and  $d$  represents one local descriptor for patch  $x_i$ .  $\gamma_d(g)$  is a weight for the  $g^{th}$  Gaussian distribution. We note  $\psi(x_i) = [\mathcal{G}_{\mu,1}, \dots, \mathcal{G}_{\mu,G}, \mathcal{G}_{\sigma,1}, \dots, \mathcal{G}_{\sigma,G}]$  the final feature vector for a given patch in each training image. The dimensionality of the feature vector is  $2 \times \mathcal{D} \times G$

where  $G$  is the codebook size (number of Gaussians).

**Classification:** Given the final feature representation for each patch in all training images, we train a linear multiclass SVM classifier. When used with high-dimensional feature vectors, linear SVM is more suitable than other non-linear classifiers (e.g. Random Forests or Kernel SVM) as it has faster training time and is less likely to overfit to the data. To predict a carcinoma subtype for a novel tissue slide  $X$ , we extract the local descriptors  $\phi$  for all patches  $x_i$  of this new tissue. Then, we encode features  $\psi$  via FVE and classify these features using the trained SVM. The classifier’s output probabilities  $P(y_i|\psi(x_i))$  are aggregated using geometric mean to infer a final tissue label  $P(y_i|X)$ .

### 3 Experiments and Discussion

We evaluated our approach on a dataset composed of 80 patients (29 HGSC, 21 CC, 11 EN, 10 MC and 9 LGSC). Each tissue was labeled by two expert pathologists [10]. We extracted 50 patches at 20x and 90x resolution on every whole slide image. Using these colour patches, we trained a 3-layer DN with filter sizes  $7 \times 7$  for 20x patches and  $10 \times 10$  for 90x patches, at each layer. In all our experiments, the SVM classifier was trained on 40 patients (selected randomly) and tested on the rest. We compared the different components of our method to existing approaches used in histopathology image classification. All parameters inherent to each technique were determined via cross-validation on the training set. Table 1 reports our results for the following experiments.

**Sparse tissue representation:** We tested the discriminative ability of our features. A sparse representation was obtained from the feature maps of a multi-layer DN from which we densely constructed local descriptors. These descriptors were then vector-quantized using a BoW representation [2] and used to classify each patch. We used the traditional SIFT local descriptors and Sparse Coding (SC) [3] as comparison. We extracted  $16 \times 16$  multiresolution samples from each image on a grid with step size 8 pixels to generate dense SIFT descriptors and SC dictionary. The same BoW quantization was applied on top of both descriptors. We observe in Table 1 that our representation allows for better discrimination between OC subtypes even when used with a simple quantization technique (BoW). Additionally, we see that higher layers of the DN induce a discrimination gain. This is linked to the convolutional property of the DN, which enables the model to learn more complex representations at higher layers (Fig. 3).

**Feature Encoding:** We also tested the performance of FVE compared to the traditional BoW with Spatial Pyramid Matching (SPM) pooling technique [4]. Table 1 shows the significant gain achieved using FVE instead of BoW encodings. Using a relatively small number of GMMs, FVE gives an accuracy of 92.1%. In contrast, similar accuracy could not be achieved using BoW even with a larger number of visual words (e.g. 1024).

**Classification:** We report the final multiclass accuracy after using the geometric mean to aggregate the classifiers probabilities and predict a tissue label

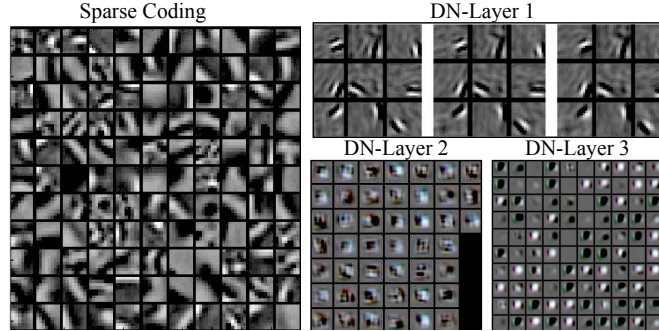


Fig. 3: Filters learnt using SC and multi-layer DN. We note how filters obtained with DN are more diverse at higher layers and span different orientations more uniformly.

from multiresolution patches. We note how this fairly simple aggregation technique results in performance gain in all experiments. We also report the mean Area Under ROC Curve (AUC) after one-against-all classification to better appreciate the differences between methods.

Features	Encoding	Pooling	Dictionary Size	Accuracy (Patches)	Accuracy (Tissues)	Mean AUC (Tissues)
SIFT	BoW	–	1024	32.5%±0.8	37.5%±0.6	–
SIFT	BoW	SPM	512	57.5%±0.5	61.5%±0.8	–
<b>SIFT</b>	<b>FVE</b>	<b>SPM</b>	<b>64</b>	66.7%±0.5	<b>68.5%±0.5</b>	0.67
SC	BoW	–	1024	57.5%±0.4	59.5%±0.5	–
<b>SC</b>	<b>BoW</b>	<b>SPM</b>	<b>1024</b>	68.0±2.0	<b>71.5%±1.5</b>	0.71
DN-1	BoW	–	256	35%±0.1	37.5%±0.8	–
DN-2	BoW	–	256	42.8±0.2	45.5%±0.8	–
DN-3	BoW	–	1024	48.0±0.8	51.5%±0.9	–
DN(1-2)	FVE	SPM	64	67.9±1.5	69.5%±1.2	–
DN(2-3)	FVE	SPM	64	83.7±1.2	85.0%±0.9	–
<b>DN(1-3)</b>	<b>FVE</b>	<b>SPM</b>	<b>128</b>	89.4%±1.7	<b>91.0%±1.0</b>	<b>0.86</b>

Table 1: Performance of different methods on ovarian carcinomas dataset. We report the average accuracy of prediction on patches and tissue sections. Experiments were repeated 3 times on shuffled training and test sets of 40 patients each. The best dictionary size is shown for each experiment. Mean AUC is shown using one-against-all.

## 4 Conclusion

This paper is the first work to automate ovarian carcinomas subtypes classification from histopathology images. We proposed a method that learns robust features by capturing complex tissue patterns. The unsupervised nature of our feature learning framework enabled us to discover discriminative patterns from limited data. Our approach proved more suitable for the classification of OC and

outperformed existing methods. Future work will focus on combining this hybrid feature representation with latent classification models to discover discriminative regions on tissues. We are also interested in experimenting with larger datasets, as new problems arise when more data becomes available, which together motivate new solutions to be developed.

**Acknowledgment:** We would like to thank NSERC for their financial support and our collaborators at the British Columbia Cancer Agency for providing insight and expertise that greatly assisted the research.

## References

1. JP Baak et al. Interpathologist and intrapathologist disagreement in ovarian tumor grading and typing. *Analytical and Quantitative Cytology and Histology*, 1986.
2. JC Caicedo et al. Histopathology image classification using bag of features and kernel functions. In *AI in Medicine*, pages 126–135. 2009.
3. H Chang et al. Characterization of tissue histopathology via predictive sparse decomposition and spatial pyramid matching. In *MICCAI*, pages 91–98, 2013.
4. H Chang et al. Classification of tumor histology via morphometric context. In *CVPR*, pages 2203–2210, 2013.
5. Chatfield et al. The devil is in the details: an evaluation of recent feature encoding methods. 2011.
6. A Coates and AY Ng. The importance of encoding versus training with sparse coding and vector quantization. In *ICML*, pages 921–928, 2011.
7. C Demir and B Yener. Automated cancer diagnosis based on histopathological images: a systematic survey. *Rensselaer Polytechnic Institute, Tech. Rep*, 2005.
8. MN Gurcan et al. Histopathological image analysis: A review. *Reviews in Biomedical Engineering*, pages 147–171, 2009.
9. J Han et al. Comparison of sparse coding and kernel methods for histopathological classification of glioblastoma multiforme. In *ISBI*, pages 711–714, 2011.
10. M Köbel et al. Diagnosis of ovarian carcinoma cell type is highly reproducible: a transcanadian study. *The American Journal of Surgical Pathology*, 34(7):984–993, 2010.
11. A Malpica et al. Interobserver and intraobserver variability of a two-tier system for grading ovarian serous carcinoma. *The American Journal of Surgical Pathology*, pages 354–357, 2007.
12. WG McCluggage. Morphological subtypes of ovarian carcinoma: a review with emphasis on new developments and pathogenesis. *Pathology Journal of the RCPA*, pages 420–432, 2011.
13. WG McCluggage. Ten problematical issues identified by pathology review for multidisciplinary gynaecological oncology meetings. *Journal of Clinical Pathology*, pages 420–432, 2011.
14. S Naik et al. Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In *ISBI*, pages 284–287, 2008.
15. RA Soslow. Histologic subtypes of ovarian carcinoma: an overview. *International Journal of Gynecologic Pathology*, pages 161–174, 2008.
16. Y Xu et al. Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering. In *CVPR*, pages 964–971, 2012.
17. MD Zeiler et al. Deconvolutional networks. In *ICCV*, pages 2528–2535, 2010.