Original Article

Clinically-inspired automatic classification of ovarian carcinoma subtypes

Aïcha BenTaieb[1], Masoud S Nosrati[1], Hector Li-Chang[2], David Huntsman[2], Ghassan Hamarneh[1]

[1]Department of Computing Sciences, Medical Image Analysis Lab, Simon Fraser University, Burnaby, [2]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada

E-mail: *Miss. Aïcha BenTaieb - abentaie@sfu.ca
*Corresponding Author

Abstract

Context: It has been shown that ovarian carcinoma subtypes are distinct pathologic entities with differing prognostic and therapeutic implications. Histotyping by pathologists has good reproducibility, but occasional cases are challenging and require immunohistochemistry and subspecialty consultation. Motivated by the need for more accurate and reproducible diagnoses and to facilitate pathologists' workflow, we propose an automatic framework for ovarian carcinoma classification. **Materials and Methods:** Our method is inspired by pathologists' workflow. We analyse imaged tissues at two magnification levels and extract clinically-inspired color, texture, and segmentation-based shape descriptors using image-processing methods. We propose a carefully designed machine learning technique composed of four modules: A dissimilarity matrix, dimensionality reduction, feature selection and a support vector machine classifier to separate the five ovarian carcinoma subtypes using the extracted features. **Results:** This paper presents the details of our implementation and its validation on a clinically derived dataset of eighty high-resolution histopathology images. The proposed system achieved a multiclass classification accuracy of 95.0% when classifying unseen tissues. Assessment of the classifier's confusion (confusion matrix) between the five different ovarian carcinoma subtypes agrees with clinician's confusion and reflects the difficulty in diagnosing endometrioid and serous carcinomas. **Conclusions:** Our results from this first study highlight the difficulty of ovarian carcinoma diagnosis which originate from the intrinsic class-imbalance observed among subtypes and suggest that the automatic analysis of ovarian carcinoma subtypes could be valuable to clinician's diagnostic procedure by providing a second opinion.

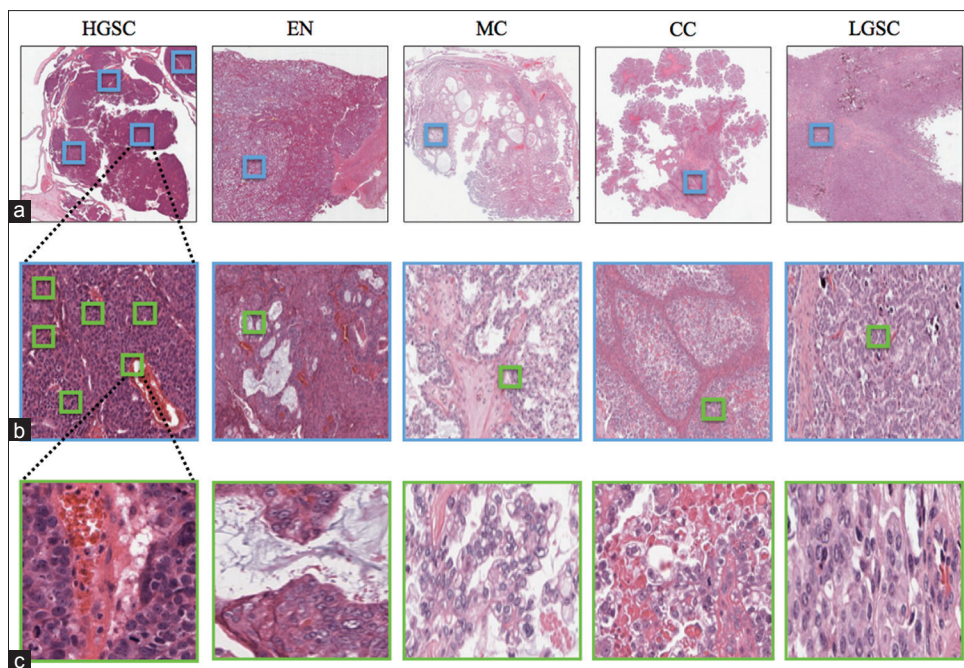**Key words:** Computer-aided diagnosis, machine learning, ovarian carcinoma

INTRODUCTION

It is now accepted that ovarian carcinomas are not a single disease, but consist of a heterogeneous group of several distinct histotypes.[1] The World Health Organization (WHO) recommends dividing ovarian carcinomas into five main epithelial types [Figure 1a]: High-grade serous carcinoma (HGSC),

Figure 1: Patch extraction from tissue sections. (a) Tissue samples of the five main recognized ovarian carcinoma types. HGSC: High-grade serous carcinoma; EN: Endometrioid; MC: Mucinous; CC: Clear cell and LGSC: Low-grade serous carcinoma. (b) Low-resolution (×20) patch extraction. Twenty nonoverlapping patches were extracted automatically from each whole tissue slide. (c) High-resolution (×40) patch extraction. One hundred patches were extracted from each low-resolution patch

endometrioid (EN), clear cell (CC), mucinous (MC), and low-grade serous carcinomas (LGSC). These tumors not only differ at the molecular level but in many other aspects such as the response to treatment and aggressiveness.[2] Until very recently, all ovarian carcinomas were treated homogeneously, with surgery and/or common chemotherapy regimens depending on the disease stage, with disappointing results in many cases. It is estimated that 75% of patients with advanced disease stage experience recurrence after surgery and chemotherapy and ultimately die of the disease.[1] Thus, to improve outcomes for individual patients and have a better understanding of ovarian carcinomas, it is important to differentiate between these tumor types as accurately as possible.

While many clinical and biologic issues regarding ovarian carcinomas remain poorly understood, reproducible histopathological diagnosis of cancer is an important condition for successful treatment and prognosis.[3] Occasional cases present significant diagnostic challenges, as there remain ambiguities on how to define each subtype and how to characterize them efficiently from tissue sections. This results in imperfect inter-observer agreement[3] and reproducibility due to the subjective nature of the diagnostic procedure.

Automatic histopathology image analysis aims at tackling challenges observed in cancer diagnosis and covers different topics such as stain normalization, quantitative image analysis for cancer grading, automatic detection of

tissue components (e.g., nuclei and cytoplasm), image retrieval, etc., Several excellent reviews, such as Gurcan *et al.*[4] and Veta *et al.*,[5] have summarized existing methods proposed in this field.

One of the most common applications for automatic histopathology image analysis is the detection and grading of cancer. Doyle *et al.*[6] used textural and nuclear features for analyzing breast cancer histopathology images. They showed the importance of texture in classifying low and high-grades of breast cancer. Similarly, Al-Kadi[7] showed the importance of the combination of statistical and model-based textural features for meningioma tissues classification. Other works focused on segmentation-based features to describe the cytology and morphology of tissue components. Monaco *et al.*[8] proposed to extract statistics on glandular shapes from segmented images of prostate histology sections. They showed the effectiveness of these features in classifying benign versus malignant tumors. More recently, some works focused on learning features from pixel intensities to extract specific visual patterns. Cruz-Roa *et al.*[9] used a bag of words framework to detect biological structures from basal cell carcinoma. The goal was to detect this skin carcinoma type from tissue sections.

In this work, we demonstrate the usefulness of automatic image analysis and machine learning for ovarian carcinoma subtyping, by employing carefully chosen image-processing techniques to extract clinically-inspired discriminative features. One of the most common application for automatic histopathology image analysis

is the detection and grading of cancer.[6-9] Despite their effectiveness in detecting cancerous from noncancerous tissues, none of the existing works addressed the automatic identification of ovarian carcinoma subtypes.
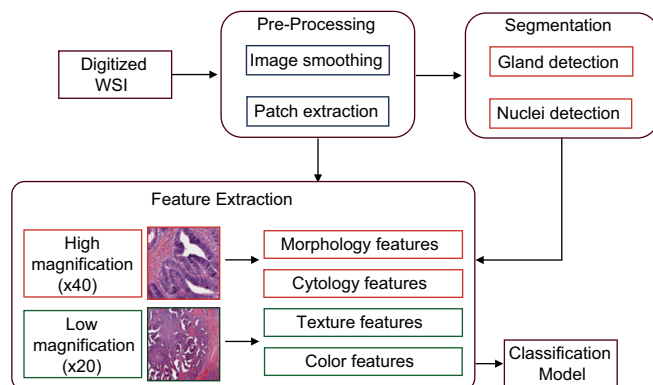
We describe our proposed automatic ovarian carcinoma classifier as a "translation" of pathologists' diagnostic procedure into a computer vision system that selects discriminative image features to perform an automatic diagnosis. An overview of the proposed model is shown in Figure 2. The framework includes four modules: Image preprocessing, image segmentation, feature extraction, and machine learning-based classification.

MATERIALS AND METHODS

A total of 80 representative slides from resection samples were used. The dataset, which is composed of 29 HGSC, 21 CC, 11 EN, 10 MC, and 9 LGSC images, was obtained (the dataset is available for review at http://www.gpecimage.ubc.ca/aperio/images/transcanadian/index.html) from a previously published trans-Canadian study on ovarian cancer classification.[3] Each of the eighty H and E slides was provided with a review diagnosis and labeled by expert pathologists.[3] The diagnoses from that study were derived according to 2003 WHO criteria with the following exceptions: Nuclear atypia and mitotic count were used to further classify serous carcinoma into high-grade and low-grade. MC cell type was characterized based on the presence of intracytoplasmic mucin in cells. The presence of glandular differentiation was accepted as part of high-grades serous carcinomas and not sufficient to the diagnosis of EN tumors which were characterized based on squamous differentiation.

Image Preprocessing and Segmentation

Each image is a single core of ovarian biopsy tissue. Tissue slides were digitized at multiple microscope magnifications with 100–900 millions of pixels per image and an average of 650 ± 50 million pixels. To process and reduce the large amount of information embedded in these multi-resolution digitized histopathology slides,

every image was automatically analyzed from 120 patches at two different microscope resolutions and different spatial locations. Patch extraction proceeds as follows [Figure 1b and c]. We randomly extracted a set of twenty rectangular nonoverlapping 500 × 500 pixel patches (i.e., 5 million unique pixels) at ×20 zoom and 100 patches of similar size at ×40 zoom (25 million pixels). Regions where the majority of pixels lie outside tissue (i.e., background pixels) were not selected. Background pixels, which appear white, were detected using a threshold on pixels' intensities. This procedure was automatically performed and did not require any user interaction.

Images were normalized with respect to staining variations using a state-of-the-art staining normalization technique.[10] Image segmentation using the graph cuts method[11] was then performed to detect epithelial nuclei and cellular structures. A sample segmentation result is shown in Figure 3a. While there exist more advanced techniques for nuclei segmentation, we chose to mimic pathologist's rough estimation of nuclei density which generally relies on a visual assessment of the amount of stain observed in the tissue slide.

Image Feature Extraction

A core component of the automatic ovarian carcinoma classifier is the feature extraction. Based on discussions with pathologists describing their workflow and diagnostic procedure, we resorted to using the following
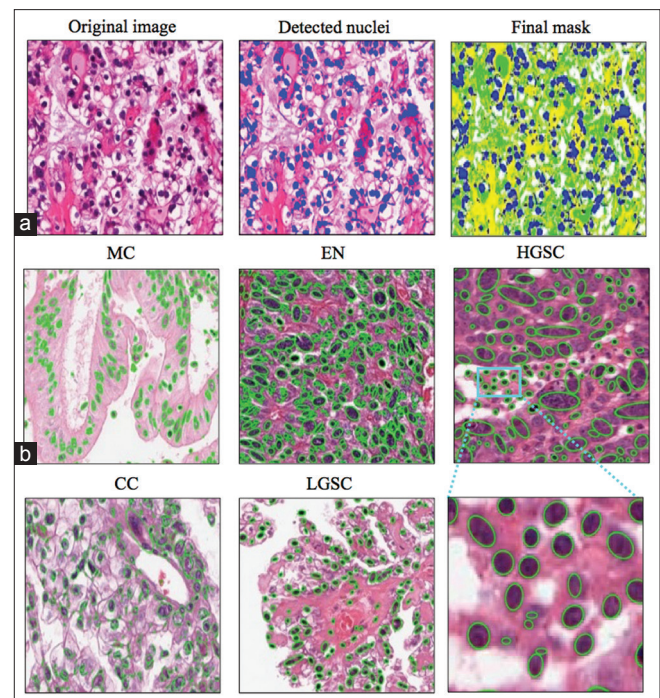
Figure 3: Segmentation results. (a) Segmentation output. The image is partitioned into clusters of similar color to detect the principal tissue components. We create a mask of each tissue component: Nuclei (blue), cytoplasm (green), stroma (yellow) and background (white). (b) Nuclei shape analysis using ellipse (green) fitting. The ellipse shape approximates each nucleus radius, elongation and area

Figure 2: Overview of the proposed automatic ovarian carcinoma classification pipeline

features extracted at two different magnifications. Low magnification features (×20) were designed to describe the architectural organization of the tissue via quantification of color, texture and shape characteristics while higher magnification features (×40) quantify the cytology and morphology of nuclei and cytoplasmic structures. Table 1 summarizes the set of features used in the automatic ovarian carcinoma classifier.

Low magnification features (×20)
Low magnification features comprise color and texture features calculated from the ×20 patches. Color appearance reflects the nucleic acids and proteins prevalence in each image and are computed as the mean, the standard deviation, the 5th and 95th percentile of each color channel (red, green and blue) as well as the ratio of red over blue channels.

Color distribution ranges are often similar for some subtypes such as EN and HGSC, which are more prone to misclassification. To capture color differences more efficiently, we introduce our color dissimilarity feature. Color dissimilarity is quantified in an unsupervised manner. After preprocessing the images, we cluster the dataset based on the color histogram of images using k-means. The number of clusters is defined via cross validation. In our experiments, we set it to five clusters. Then, we defined the distance measure (d_{ij} (1)) between distributions of each image in the dataset and the centroids of each cluster.

$$d_{ij} = 1 - e^{-(\alpha \| b_i - c_i \|)} \tag{1}$$

Where b_i is the color histogram of image i, c_j is the histogram corresponding to the centroid of the jth cluster, and $\| \bullet \|$ is the L2 distance between two color histograms. Here, is a coefficient for normalizing the sensitivity of our dissimilarity measure that we set to 0.5 empirically.

Texture features, on the other hand, capture the organization of cells and structural patterns observed in the tissue, e.g., organized, disorganized, homogeneous, grainy, striped, or checkered. Using image-processing
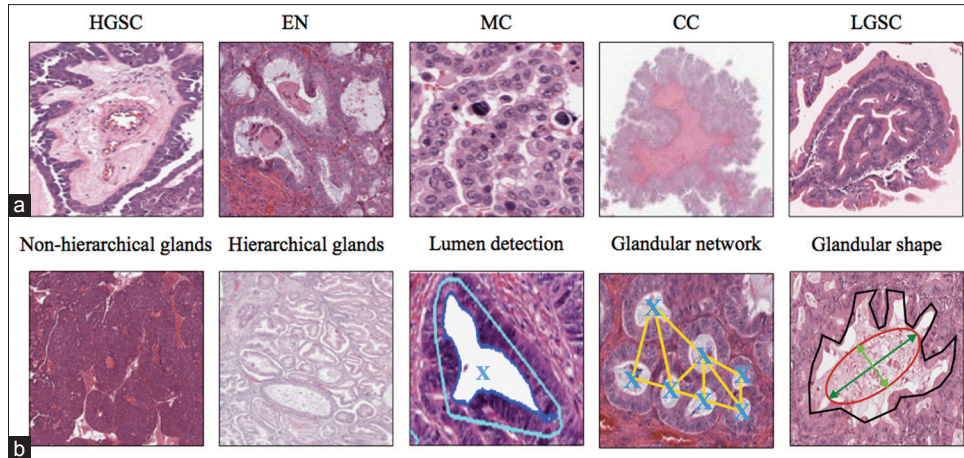
techniques, we computed two types of texture features relevant to histopathology images: A multispectral color-texture and a Gabor filter-based feature.

The multispectral color-texture approach has been shown to increase the performance of classification applications.[12] This approach computes several co-occurrence matrices which capture the number of color pairs between adjacent pixels. We computed a co-occurrence matrix on each isolated labeled region of interest from the segmented images: Nuclei, cytoplasm and stroma and their respective pairs. This resulted in six matrices: Nuclei, cytoplasm, stroma and pairwise combinations of these three structures. These matrices were then used to compute texture statistics corresponding to the four first Haralick features[13] which describe the average spatial relationship observed between tissue structures (nuclei, cytoplasm, stroma and their respective pairs). To complement our texture analysis, we used a filtering-based approach that showed excellent results for texture characterization.[4] Gabor filter banks[14] decompose an image based on its texture for classification purposes. Multi-channel Gabor filters (i.e., red, green and blue), mimics the human visual system. Here, the aim is to measure the response of each image to a particular filter described with a specific frequency and orientation. In our implementation, we used 38 filters. Each filter gave a different response when applied to the image. We used these responses to compute statistical texture measures corresponding to Tamura's texture features[15,16] that were designed to discriminate between different textures more accurately.

High magnification features (×40)
At higher magnification, we used the segmented images to compute a set of features describing the morphology and cytology of cells and nuclei in the tissue. The median nuclear density and nuclear to cytoplasm ratio was quantified on each segmented image by counting the number of automatically detected nuclei and cytoplasmic structures. We also characterized the nuclear shape by fitting an ellipse to each segmented nuclei and computing the ellipse major and minor axis lengths, eccentricity and area [Figure 3b].

To quantify the cytology, we studied the glandular organization and shape of cells. Cells in tumors organize in circular or elliptic configurations and cluster into groups of similar shape and size defining a gland [Figure 4a]. These glands are organized in a similar fashion with three main components: A central empty region (lumen), a cytoplasm and a nucleus. We detected glands by automatically looking for lumen regions surrounded by nuclei [Figure 4b]. The convex hull[17] of the detected gland was used to extract statistical measures describing the gland's area, circularity $\frac{4\pi \times \text{lumen area}}{\text{lumen perimeter}}$ and thickness between the lumen and its surrounding

Table 1: Features used for classification

Magnification	Feature type	Dimensionality	Biological appearance
Low (x20)	Color	14	Blue VS Red/Pink
	Texture	138	Organized VS Disorganized
High (x40)	Morphology	5	Nuclei Size, Appearance, Density
	Cytology	22	Glandular Organization, Hierarchical VS Non-hierarchical

Figure 4: Glands analysis. (a) Glandular patterns characteristic of each cell-type. (b) Automatic gland detection on images from tissue slides and features extracted from glands. The glandular network representing neighboring glands is formed by nodes (blue X) corresponding to detected glands and edges linking neighboring glands (yellow lines)

nuclei. We also quantified the nuclei abundance around each gland by counting the number of detected nuclei.

Furthermore, we quantified the hierarchical versus nonhierarchical organization of neighboring glands within a patch [Figure 4b]. For this purpose, we constructed a network whose nodes correspond to centroids of lumen regions and whose edges connect neighboring centroids [Figure 4b]. In the constructed network, we grouped nodes into connected components representing neighboring glands. We then computed the shape similarity (e.g., gland size and eccentricity) across glands from the same connected component and neighboring connected components. We define three measurements for to describe the shape similarity between connected components:

- The average number of elements in each connected component, which reflects glands proximity in the tissue
- Shape similarity s_1 which captures the difference between a connected component and all components in terms of circularity and thickness defined to quantify each single gland (2)
- Size similarity s_2 which measures the average lumen area difference between a connected component and all components (3).

$$s_1(i) = \frac{1}{N}\sum_{j=1}^{N} LSV_i - LSV_j \qquad (2)$$

$$s_2(i) = \frac{1}{N}\sum_{j=1}^{N} \left\| A_i - A_j \right\|^2 \qquad (3)$$

Where LSV_i and LSV_j are the feature vectors quantifying shape measurement (glands circularity and thickness) on the ith and j^{th} connected components. A_i is the average area of lumens in the connected component i. N is the number of connected components per image.

Image Classification

The final feature vector for each image, computed based on the aforementioned descriptions, was of 179 dimensions (forming the feature space, [Table 1]). These features were used in a machine learning framework to train a classification model. This model was carefully designed to overcome the specific difficulties (high intra-class and low inter-class variance) observed in ovarian carcinoma classification. In fact, the classification model relies on four specific modules: Dissimilarity matrix, feature selection, dimensionality reduction and support vector machine (SVM) classifier. Each of these modules favors better discrimination between each carcinoma subtype. We opted for a linear classification method with a linear SVM classifier and dimensionality reduction technique as these techniques have generally fast computation times and most importantly, are less likely to overfit to our relatively small dataset.

We introduce a dissimilarity matrix that allows us to separate classes in feature space more effectively. This dissimilarity is defined as the distance between each pair of patients' feature vector. The dissimilarity

coefficient for each patient is computed based on the minimum sum distance as follows:

$$D(I^m, I^n) = \sum_i \min_j d(x_i^m, x_j^n) + \sum_j \min_i d(x_j^m, x_i^n) \qquad (4)$$

Where I^m, I^n are two subjects represented by patches from their WSI. are the ith and j^{th} patches described by their feature vector, for subjects m and n respectively. The dissimilarity matrix is of size N × N where N is the total number of subjects. This dissimilarity matrix represents the final feature representation of our training set.

The performance of the classifier intrinsically depends on the separability of each sub-type/class in this feature space. On top of the dissimilarity matrix, we apply linear

discriminant dimensionality reduction[18] that integrates Fisher score for feature selection and linear discriminant analysis for dimensionality reduction. Using a dissimilarity matrix allows for a better discrimination of subtypes by selecting a subset of dissimilar and relevant features. The feature selection and dimensionality reduction modules speed up the computation time.

The final module of the classification model is a linear classifier based on a trained multi-class SVM.[19] More concretely, the classifier learns a set of hyperparameters defining four distinct separating hyperplanes (i.e., discriminate between each pair of ovarian carcinoma subtype) in the high-dimensional feature space. The classifier was trained on features and class labels of training images that always excluded the novel image to be classified.

To test the classifier on an unseen image, all 179 clinically-inspired features were extracted from the new image and classified using the trained classification model which outputs a predicted carcinoma for the given novel image.

Implementation Details
Each whole slide image was preprocessed using the VIPS library (http://www.vips.ecs.soton.ac.uk/index.php?title=Libvips) to extract patches at different magnifications, using the built-in function dzsave to create image pyramids from a whole slide image. Image segmentation and feature extraction were implemented in Python 2.7 (https://www.python.org) using the OpenCV v3.0 library (http://opencv.org), using functions calcHist to extract our color features, cvGrayscaleMat updated to compute the multispectral color-texture gray-level co-occurrence matrix, getGaborKernel and filter 2D to compute the Gabor filter responses and the function CV connected components to quantify the glandular organization of tissues. All modules of our classification pipeline were developed using scikit-learn (http://scikit-learn.org) and the sklearn.svm. LinearSVC package (http://scikit-learn.org) to train our multiclass linear SVM. All code was tested on a CPU Core 2 Duo E8400 @ 3.00 GHz machine. To facilitate direct comparison, we release our model and data at the following URL: http://199.60.17.63.

RESULTS AND DISCUSSION

Multiclass Classification Accuracy
We carried out a leave-one-patient-out cross-validation to test the sensitivity of our method to the training data used. While iterating over all 80 patients slides, we randomly removed five patients (one from each class) from the dataset and used them as a test set. At each round, the training set corresponds to the 75 remaining patients. Also, to test the sensitivity of the method to

different tissue patches, we repeated these tests 5 times using five different sets of automatically randomly sampled patches from each tissue slide. Table 2 reports the mean and standard deviation values of the accuracy, sensitivity, specificity and precision of our classifier. At test stage, given a new tissue sample, the automatic ovarian carcinoma classifier was able to predict a carcinoma type with an average accuracy of 95.0%. As shown in Table 2, each of the classifier's modules played a critical role in the final classification accuracy. We observed a significant improvement in the classification accuracy as we added each of the modules: Dissimilarity matrix, feature selection and dimensionality reduction (from 72.5% to 95.0%).

Using cross-validation not only allow us to estimate how well our model will generalize to new test sets, it also shows the sensitivity of the model to different training sets. In fact, when using an iterated 3-fold cross-validation on our dataset of eighty patients, we report a variance of 2.0% across iterations. This reported variance implies that our model is relatively robust to training set changes. While these results are helpful to understand the characteristics of the model, they are still limited to the study cases used in this work and can not be interpreted as conclusive clinical validations.

To validate the statistical significance of our results, we compare our results with a pure chance predictor. A random predictor has an average classification accuracy of 20% on our 5-class dataset. We test whether our classifiers output predictions are statistically different from a random predictor using a paired Fisher exact test in which the null hypothesis assumes there is no nonrandom association between our classifier and a random predictor. We reject this null hypothesis with a $P = 5e^{24}$, confirming that our classifiers predictions are significantly different from the result of a pure chance predictor. We also obtain a correlation coefficient of 0.88 ($P = 0.0001$) between the classifiers predicted classes and the ground truth labels. Again, these results attest of the statistical significance of our method compared to random predictions.

Class Confusion
We show the (asymmetric) confusion matrix [Figure 5b] to better visualize which carcinoma classes (subtypes) are being confused with each other by our system. The automatic classifier's confusion agrees with clinicians'

Table 2: Multi-class classification performance

All features with	Accuracy (%)	F1 score	Recall	Precision
-	72.5±1.5	0.65	0.62	0.70
Dissimilarity matrix	78.5±1.5	0.70	0.68	0.72
Feature selection	88.5±1.5	0.86	0.85	0.87
Dimensionality reduction	95.0±1.5	0.93	0.90	0.96
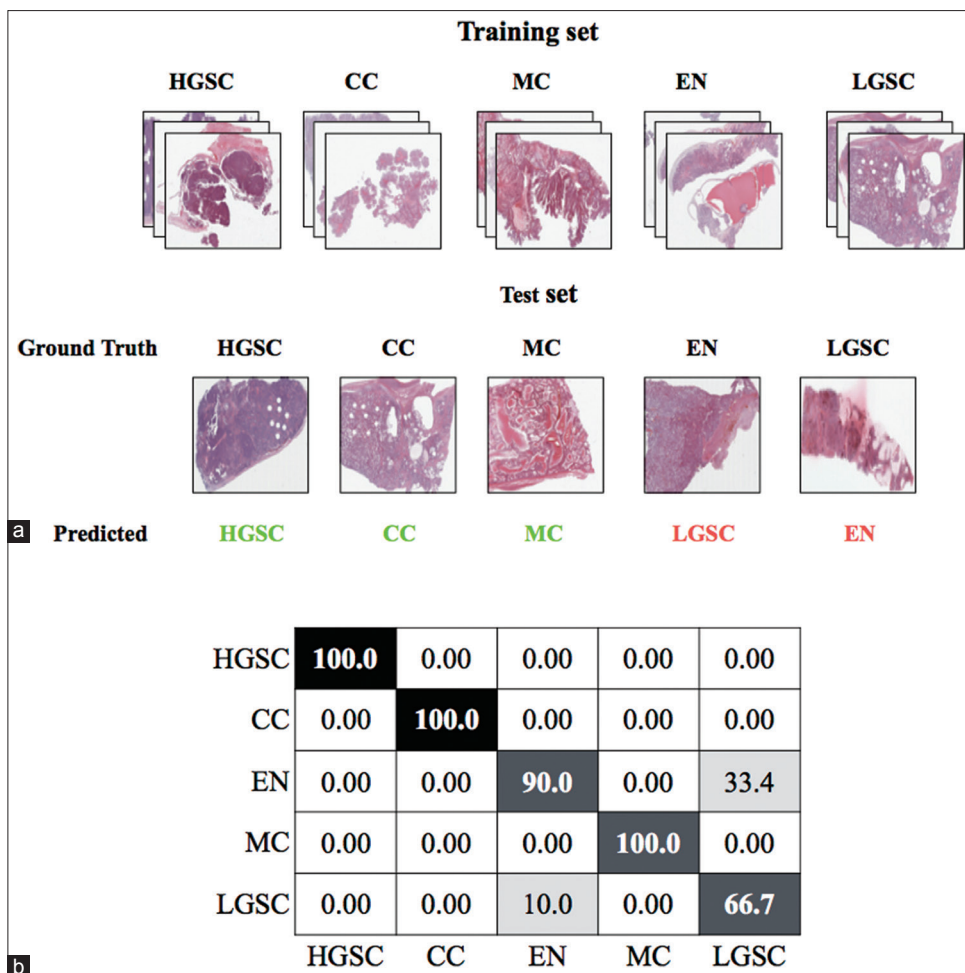
Figure 5: Class-confusion. Multi-class classification performance. (a) Misclassified tissue samples. The first row corresponds to sample images from the training set. The final row shows the predicted class labels. We observe a high range of color, tissue and staining variability in the misclassified samples. (b) Confusion matrix after leave-one-out cross validation. Each column of the matrix represents the predicted class, while each row represents the ground truth class

confusion[1] for complex serous carcinoma cases. Moreover, to test our classifier's robustness against class-imbalance during training, we repeatedly performed a series of leave-one-out cross-validations. We used a test set of five patients (1 per class) and gradually increased the imbalance between the classes in the training data. Each experiment involved randomly selecting the training and test set from the total dataset. The distribution of classes in the training set was chosen to create imbalance ratios of 1:1 (balanced), 1:2, 1:3 and 1:7 between the most and least represented classes. We carried 1000 repetitions of each experiment. We obtain an average multiclass classification accuracy of 70.0%, 66.9%, 66.7% and 66.8% for training set imbalance ratios of 1:1, 1:2, 1:3 and 1:7, respectively.

While we expect the class-imbalance to affect the average multiclass accuracy (the classification drops by 3.1% when training on imbalanced datasets), this effect did not increase with the increase of imbalance between classes. In particular, the classification accuracy stabilizes despite the increase in class-imbalance (66% for an imbalance ratio of 1:2, 1:3 and 1:7). These results may be attributed to the SVM classifiers ability to be robust to highly imbalanced training datasets.

Figure 5a shows samples of the misclassified subjects from the multi-class experiment. We can observe the highly diverse phenotypic variability in each class, which is to be expected since the dataset presents high variability of staining, tissue morphology and screening conditions that add to the complexity of the task.

Uncertainty of Prediction

As a second experiment, we trained 20 (5 × 4) binary classifiers for each subtype pairs (HGSC vs. CC; HGSC vs. EN; HGSC vs. MC; etc.,) to quantify how well the proposed features discriminate between subtypes. In this case, the training set was composed of half of the subjects of both classes involved and the other remaining half comprised the test set. We present the accuracy for each binary classifier in

Table 3. In addition, we report a measure of prediction uncertainty[20] in Table 4.

These pairwise classification accuracy results (ranging from 100% for HGSC vs. CC to 55% for LGSC vs. MC and EN) highlight the complexity of ovarian carcinoma diagnosis. In fact, due to the high imbalance between HGSC and most of the other subtypes, binary classifiers tend to overfit to the most represented class. As observed in Table 3, HGSC cases, which are the most frequent among patients (comprising 36.25% of our 5-class data) yet generally clinically misdiagnosed, can be successfully identified from all other subtypes (average of 88% accuracy for HGSC vs. others and no confusion with other subtypes. However, MC and LGSC are more often misclassified by the SVM classifier (average of 66.5% accuracy for MC vs. others and 73.75% for LGSC vs. others). These results also show that the set of features used in this study seem to be relevant for HGSC, EN and CC carcinoma (which are usually characterized by the abundance of nuclei, cellular structures grouped to form island shapes) but might not be sufficient to discriminate MC and LGSC.

CONCLUSION

In this study, we examined the performance of an automatic classification system in predicting ovarian carcinoma subtypes on a clinically-derived dataset of eighty patients. In contrast to other studies, this work is the first attempt of automating ovarian carcinoma subtypes classification from histopathology images. The automatic classifier was designed based on the combination of expert knowledge on ovarian carcinomas and state-of-the-art computer vision techniques for histopathology image analysis, eliminating the subjectivity that usually affects the diagnostic procedure. Our automatic system achieved an average accuracy of 95.0% on a multi-class classification of ovarian epithelial subtypes. The proposed pipeline is fully automatic with a quasi instantaneous test phase (~1s on an Intel CPU, E8400 @ 3.00GHz machine).

Table 3: Binary classification accuracy results

Accuracy (%)	HGSC	CC	EN	MC	LGSC
LGSC	95.0	90.0	55.0	55.0	
MC	85.0	70.0	56.0		
EN	75.0	68.0			
CC	100.0				
HGSC					

Table 4: Uncertainty of preparation per carcinoma subtype

Classes	HGSC	CC	EN	MC	LGSC
Uncertainty (%)	12.0	2.0	18.0	13.0	28.0

The results reported in this study are promising but are so far only preclinical. Further investigations should be made on larger cohorts of patients and using independent test sets to make any conclusive comments about the suitability of such automatic systems in clinical practice.

It should be pointed out that the proposed automatic prediction approach has some drawbacks. As highlighted in the results, the automatic systems' performance can be negatively affected by heterogeneous processing and staining, severe or atypical cases and digitization occlusions. Moreover, as this work is a first attempt to evaluate the performance of an automatic system for ovarian carcinoma diagnosis, further investigations must be performed on larger cohorts with an independent test set to fully evaluate the potential of such systems in real practice.

Also, while exploration of more advanced automatic feature learning (e.g., auto-encoders) and machine learning models (e.g., deep learning) may improve the classification accuracy and is an important future work, it may also result in a less intuitive automatic pipeline (not biologically or clinically-inspired) and raises the question of whether such well-performing black-box techniques will be trusted and useful to pathologists and clinicians. Future work should explore leveraging histopathology image classification methods that were not designed for ovarian carcinoma.

Finally, although robustness toward the heterogeneous appearance of tissue slides is likely to be achieved by training the system on larger datasets or adding user supervision during the patch extraction, handling the class-imbalance intrinsic to ovarian carcinoma diagnosis might require the design of more elaborate class-specific features. This will be the focus of our future works.

Conflicts of Interest
There are no conflicts of interest.

REFERENCES

1. Prat J. New insights into ovarian cancer pathology. Ann Oncol 2012;23 Suppl 10:x111-7.
2. Gilks CB, Prat J. Ovarian carcinoma pathology and genetics: Recent advances. Hum Pathol 2009;40:1213-23.
3. Köbel M, Kalloger SE, Baker PM, Ewanowich CA, Arseneau J, Zherebitskiy V, *et al.* Diagnosis of ovarian carcinoma cell type is highly reproducible: A transcanadian study. Am J Surg Pathol 2010;34:984-93.
4. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: A review. IEEE Rev Biomed Eng 2009;2:147-71.
5. Veta M, Pluim JP, van Diest PJ, Viergever MA. Breast cancer histopathology image analysis: A review. IEEE Trans Biomed Eng 2014;61:1400-11.

6. Doyle S, Agner S, Madabhushi A, Feldman M, Tomaszewski J. Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE; 2008.

7. Al-Kadi OS. Texture measures combination for improved meningioma classification of histopathological images. Pattern Recognit 2010;43.6:2043-53.

8. Monaco JP, Tomaszewski JE, Feldman MD, Hagemann I, Moradi M, Mousavi P, *et al.* High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models. Med Image Anal 2010;14:617-29.

9. Cruz-Roa A, Caicedo JC, González FA. Visual pattern mining in histology image collections using bag of features. Artif Intell Med 2011;52:91-106.

10. Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, *et al.* A method for normalizing histology slides for quantitative analysis. In: International Symposium on Biomedical Imaging. Vol. 9. IEEE Press, Piscataway, NJ, USA; 2009. p. 1107-10.

11. Boykov Y, Funka-Lea G. Graph cuts and efficient ND image segmentation. Int J comput vis 2006;70:109-131.

12. Shim SO, Choi TS. Image indexing by modified color co-occurrence matrix. In: IEEE International Conference on Acoustics, Speech, and Signal Processing; 2003.

13. Haralick, Robert M. Statistical and structural approaches to texture. Proceedings of the IEEE. 1979;67:786-804.

14. Clausi D, Jernigan M. Designing gabor filters for optimal texture separability. Pattern Recognit 2000;33;1835-49.

15. Tamura H, Mori S, Yamawaki T. Textural features corresponding to visual perception. IEEE Transactions on Systems, Man and Cybernetic 1978;8;460-73.

16. Irshad H. Automated mitosis detection in histopathology using morphological and multi-channel statistics features. J Pathol Inform 2013;4:10.

17. Gonzalez RC. Digital Image Processing. Pearson Eductation; 2009.

18. Gu Q, Li Z, Han J. Linear discriminant dimensionality reduction. In: Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg; 2011.

19. Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995;20:273-97.

20. Hüllermeier E. Uncertainty in clustering and classification. In: Scalable Uncertainty Management. Springer Berlin Heidelberg; 2010. p. 16-19.