Clinically-Inspired Automatic Classification of Ovarian Carcinoma Subtypes

****

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. **Methods:** Our method is inspired by pathologists' workflow. We analyze 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 40 high-resolution histopathology images. The proposed system achieved a multiclass classification accuracy of 92.5% 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 serous carcinomas (High grade and low grade serous subtypes). **Conclusion:** Our results suggest that the automatic analysis of ovarian carcinoma subtypes could be valuable to clinician's diagnostic procedure by providing a second opinion and could guide them in determining more subtype-specific features.

Keywords: Ovarian carcinoma, computer-aided diagnosis, machine learning.

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 1-a): high-grade serous (HGSC), 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, in order 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 interobserver 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 in order 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 vs. malignant tumours. More recently, some works focused on learning features from pixel intensities to extract specific visual patterns. Caceido 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;7;8;9]. Despite their effectiveness in detecting cancerous from non-cancerous 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 pre-processing, image segmentation, feature extraction, and machine learning-based classification.

Material and Methods

Study Cases

A total of 40 representative slides from resection samples were used. The dataset, which is composed of 15 HGSC, 10 CC, 5 EN, 5 MC, and 5 LGSC images, was obtained [1] from a previously published trans-Canadian study on ovarian cancer classification[3]. Each of the 40 H&E slides was

---

[1]The dataset is available for review at http://www.gpec.ubc.ca/index.php/content=papers/transcanadian.php

4

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. Mucinous 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 endometrioid tumors which were characterized based on squamous differentiation.

Image Pre-processing and Segmentation

Each image is a single core of ovarian biopsy tissue. Tissue slides were digitized at multiple microscope magnifications with 100 to 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 analysed from 120 patches at two different microscope resolutions and different spatial locations. Patch extraction proceeds as follows (Figure 1-b). We randomly extracted a set of 20 rectangular non-overlapping 500×500 pixel patches (i.e. 5 million unique pixels) at 4x zoom and 100 patches of similar size at 20x 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 3-a.

Image Feature Extraction

A core component of the automatic ovarian carcinoma classifier is the feature extraction. We automatically extracted specific features at two different magnifications. Low magnification features (4x) were designed to describe the architectural organization of the tissue via quantifica-

tion of color, texture and shape characteristics, while higher magnification features (20x) quantify the cytology and morphology of nuclei and cytoplasmic structures. Table 1 summarizes the set of features used in the automatic ovarian carcinoma classifier.

1) Low magnification features (4x):

Low magnification features comprise color and texture features calculated from the 4x 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 pre-processing 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 5 clusters. Then, we defined the distance measure ($d_{i,j}$ (1)) between distributions of each image in the dataset and the centroids of each cluster.

$$d_{ij} = 1 - exp(-\alpha ||b_i - c_j||), \tag{1}$$

where b_i is the color histogram of image i, c_j is the histogram corresponding to the centroid of the jth cluster, and $||.||$ is the L_2 distance between two color histograms. Here, α 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.

2) High magnification features (20x):

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 (NCR) 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 3-b).

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 4-a). 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 4-b). The convex hull[17] of the detected gland was used to extract statistical measures describing the gland's area, circularity ($\frac{4\pi * \text{lumen area}}{\text{lumen perimeter}}$) and thickness between the lumen and its surrounding nuclei. We also quantified the nuclei abundance around each gland by counting the number of detected nuclei.

Furthemore, we quantified the hierarchical vs non-hierarchical organization of neighboring glands within a patch (Figure 4-b). For this purpose, we constructed a network whose nodes correspond to centroids of lumen regions and whose edges connect neighboring centroids (Figure 4-b). 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).

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

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

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 in each image.

Image Classification

The final feature vector for each image, computed based on the aforementioned descriptions, was of 179 dimensions (forming the feature space, see 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 SVM classifier. Each of these modules favors better discrimination between each carcinoma subtype.

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:

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

where I^m, I^n are two subjects represented by patches from their WSI. x_i^m, x_j^n are the ith and j^{th} patches described by their feature vector, for subjects m and n respectively. The dissimilarity matrix is of size $N \times N$ 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 subtype/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 support vector machine (SVM)[19]. More concretely, the classifier learns a set of hyper-parameters 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.

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 40 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 35 remaining patients. Also, to test the sensitivity of the method to different tissue patches, we repeated these tests five 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 92.5%. 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 61.2% to 92.5%).

Class Confusion

We show the (asymmetric) confusion matrix (Figure 5-b) to better visualize which carcinoma classes (subtypes) are being confused with each other by our system. The automatic classifier's confusion agrees with clinicians' confusion[1] for complex serous carcinoma cases.

In our dataset, HGSC is the most frequent type (37.5% of samples) as observed in real-life clinical settings. This class imbalance partially affects the classification accuracy in favor of this specific class, explaining its confusion with less represented classes. Figure 5-a shows samples of the misclassified subjects from the multi-class experiment. We can observe the highly diverse phenotypic variability in each class. As a matter of fact, the dataset presents high variability of staining, tissue morphology and screening conditions that adds 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.) in order 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 show how well the model can distinguish serous carcinomas (HGSC, LGSC), which are clinically the most frequent yet misdiagnosed carcinomas, from all other subtypes.

Conclusion

In this study, we examined the performance of an automatic classification system in predicting ovarian carcinoma subtypes on a clinically-derived dataset of 40 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 92.5% on a multi-class classification of ovarian epithelial subtypes. The proposed pipeline is fully automatic with a quasi instantaneous test phase ($\leq$ 1 second on an Intel CPU, E8400 @ 3.00GHz machine). The results shown suggest that such an automatic system has the potential to provide faster, cheaper, more robust, available and accurate systems to facilitate diagnosis or act as an additional decision support tool for pathologists. On one hand, reported results highlight the difficulty of this task given the large number of features necessary to efficiently discriminate between carcinomas, and explain the imperfect inter-observer agreement between pathologists. On the other hand, this work highlights the advantages of automated ovarian cancer subclassification to overcome this clinical challenge. We believe that the technology developed in this study may also help in identifying intrinsic patterns in the tissue based on their visual appearance. This contribution could be helpful to pathologists to define more subtype-specific features for cell-type characterization and gain more insight and understanding of the disease.

It should be pointed out that the automatic estimation has some drawbacks. As highlighted in the results, automatic systems' performance can be negatively affected by heterogeneous processing and staining, severe or atypical cases and digitization occlusions. However, robustness toward these situations is likely to be achieved by training the system on larger datasets or adding user supervision during the patch extraction, which is the focus of our future work.

References

1. J Prat. New insights into ovarian cancer pathology. *Annals of Oncology*, 23(suppl 10):111–117, 2012.

2. C Gilks and J Prat. Ovarian carcinoma pathology and genetics: recent advances. *Human pathology*, 40(9):1213–1223, 2009.

3. M Köbel, S Kalloger, P Baker, C Ewanowich, J Arseneau, V Zherebitskiy, S Abdulkarim, S Leung, M Duggan, D Fontaine, 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.

4. M Gurcan, L.E Boucheron, A Can, A Madabhushi, N.M Rajpoot, and B Yener. Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, 2:147–171, 2009.

5. M Veta, J.P Pluim, P.J Van Diest, and M.A Viergever. Breast cancer histopathology image analysis: a review. *IEEE transactions on bio-medical engineering*, 61(5):1400–1411, 2014.

6. S Doyle, S Agner, A Madabhushi, M Feldman, and J Tomaszewski. 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*, pages 496–499, 2008.

7. O.S Al-Kadi. Texture measures combination for improved meningioma classification of histopathological images. *Pattern recognition*, 43(6):2043–2053, 2010.

8. J Monaco, J Tomaszewski, M Feldman, I Hagemann, M Moradi, P Mousavi, A Boag, C Davidson, P Abolmaesumi, and A Madabhushi. High-throughput detection of prostate cancer in histological sections using probabilistic pairwise markov models. *Medical image analysis*, 14(4):617–629, 2010.

9. A Cruz-Roa, J.C Caicedo, and F.A González. Visual pattern mining in histology image collections using bag of features. *Artificial intelligence in medicine*, 52(2):91–106, 2011.

10. M Macenko, M Niethammer, J.S Marron, D Borland, John T Woosley, X Guan, C Schmitt, and NE Thomas. A method for normalizing histology slides for quantitative analysis. In *International Symposium on Biomedical Imaging*, volume 9, pages 1107–1110, 2009.

11. Y Boykov and G Funka-Lea. Graph cuts and efficient nd image segmentation. *International journal of computer vision*, 70(2):109–131, 2006.

12. S.O Shim and T.S Choi. Image indexing by modified color cooccurrence matrix. In *IEEE International Conference on Acoustics, Speech, and Signal Processing,*, volume 3, pages III–577, 2003.

13. R.M Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979.

14. D Clausi and M Jernigan. Designing gabor filters for optimal texture separability. *Pattern Recognition*, 33(11):1835–1849, 2000.

15. H Tamura, S Mori, and T Yamawaki. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics*, 8(6):460–473, 1978.

16. H Irshad. Automated mitosis detection in histopathology using morphological and multi-channel statistics features. *Journal of pathology informatics*, 4, 2013.

17. Rafael C Gonzalez. *Digital image processing*. Pearson Education India, 2009.

18. Q Gu, Z Li, and J Han. Linear discriminant dimensionality reduction. In *Machine Learning and Knowledge Discovery in Databases*, pages 549–564. 2011.

19. C Cortes and V Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

20. E Hüllermeier. Uncertainty in clustering and classification. In *Scalable Uncertainty Management*, pages 16–19. Springer, 2010.

Figure 1 (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 (20x) patch extraction. 20 non-overlapping patches were extracted automatically from each whole tissue slide. (c) High-resolution (40x) patch extraction. 100 patches were extracted from each low-resolution patch.

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

Figure 3 (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 4 (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).

Figure 5 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.