Medical Image Analysis 39 (2017) 194–205

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

A structured latent model for ovarian carcinoma subtyping from histopathology slides

Aïcha BenTaieb[a,*], Hector Li-Chang[b], David Huntsman[b], Ghassan Hamarneh[a]

[a] Department of Computing Science, Medical Image Analysis Lab, Simon Fraser University, Burnaby, Canada
[b] Departments of Pathology and Laboratory Medicine and Obstetrics and Gynaecology, University of British Columbia, Vancouver, Canada

ARTICLE INFO

Article history:
Received 22 July 2016
Revised 15 April 2017
Accepted 27 April 2017
Available online 9 May 2017

Keywords:
Ovarian carcinoma
Subtyping
Digital pathology
Machine learning
Support vector machines
Latent representation

ABSTRACT

Accurate subtyping of ovarian carcinomas is an increasingly critical and often challenging diagnostic process. This work focuses on the development of an automatic classification model for ovarian carcinoma subtyping. Specifically, we present a novel clinically inspired contextual model for histopathology image subtyping of ovarian carcinomas. A whole slide image is modelled using a collection of tissue patches extracted at multiple magnifications. An efficient and effective feature learning strategy is used for feature representation of a tissue patch. The locations of salient, discriminative tissue regions are treated as latent variables allowing the model to explicitly ignore portions of the large tissue section that are unimportant for classification. These latent variables are considered in a structured formulation to model the contextual information represented from the multi-magnification analysis of tissues. A novel, structured latent support vector machine formulation is defined and used to combine information from multiple magnifications while simultaneously operating within the latent variable framework. The structural and contextual nature of our method addresses the challenges of intra-class variation and pathologists' workload, which are prevalent in histopathology image classification. Extensive experiments on a dataset of 133 patients demonstrate the efficacy and accuracy of the proposed method against state-of-the-art approaches for histopathology image classification. We achieve an average multi-class classification accuracy of 90%, outperforming existing works while obtaining substantial agreement with six clinicians tested on the same dataset.

1. Introduction

According to the World Health Organization, ovarian cancer is the fifth most common cancer type worldwide and its outcomes are the poorest among women (Prat, 2012). Clinical differences between histologic subtypes of ovarian cancer have long been recognized, but it is only recently that pathologists have been able to define carcinomas in a way that correlates well with clinical and molecular differences (Prat, 2012; Racoceanu and Capron, 2016). Currently, five main histologic types of ovarian carcinomas (cancers derived from epithelial cells) have been identified (Fig. 1): high-grade serous (HGSC), endometrioid (EN), clear cell (CC), mucinous (MC) and low-grade serous (LGSC). It is now recognized that these ovarian carcinoma subtypes can not (and should not) be treated equivalently and necessitate accurate classification for a personalized treatment.

Despite recent advances in the understanding of these histotypes, patients suffering from ovarian carcinomas still have poor prognostic rates. The success of cell-type-specific chemotherapy regimens and personalized treatments is contingent on a reliable and accurate subtyping or characterization of these cell-types from tissue sections.

Presently, clinical diagnosis of ovarian cancer involves the subtyping of ovarian carcinomas and is derived from the microscopic analysis of tissue sections, either from biopsies or resection specimens, that are mounted on glass slides, stained with hematoxylin and eosin (H&E) (Lalwani et al., 2011), and examined using light microscopy. Digitized tumor biopsies or whole slide images (WSI) are used by a small number of research centres and clinical laboratories, but their use (so-called "digital pathology") is expected to increase over time. Staining is used to highlight nuclei and the cellular content known as cytoplasm (the cells main biological components, which are naturally transparent) with various shades of blue and red.

During diagnosis, pathologists scan the tissue biopsy under a microscope seeking relevant abnormalities to diagnose each carci-

* Corresponding author.
E-mail address: abentaie@sfu.ca (A. BenTaieb).

http://dx.doi.org/10.1016/j.media.2017.04.008
1361-8415/

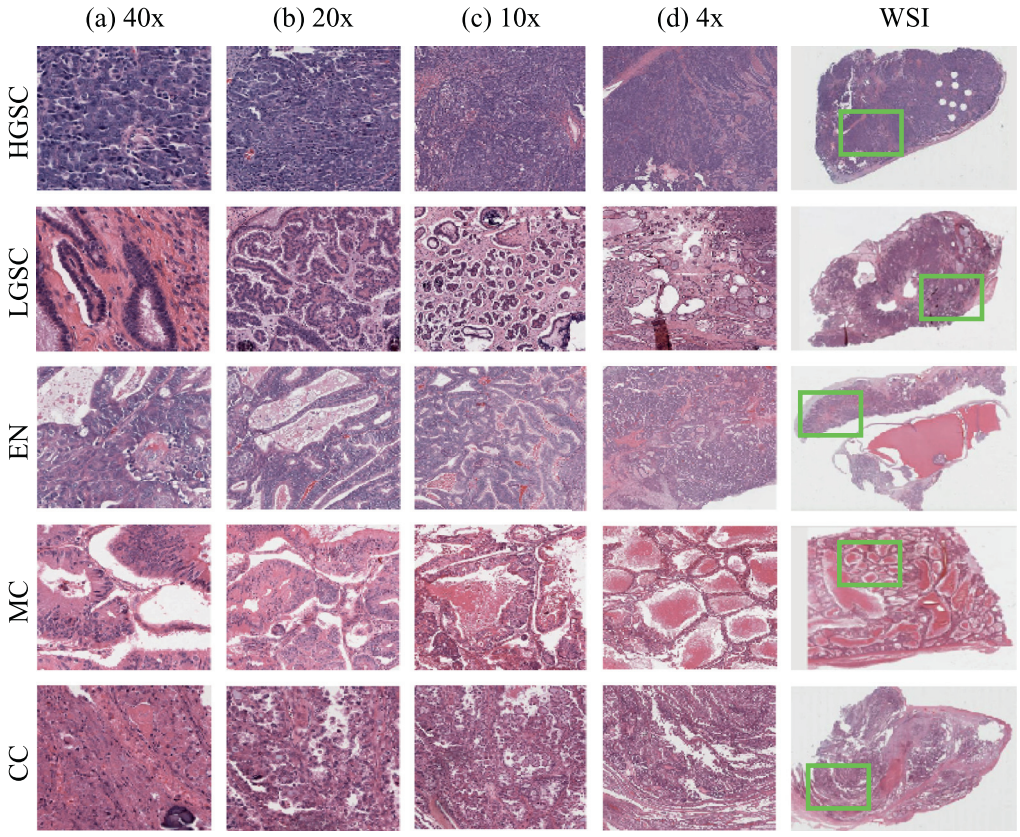

Fig. 1. Whole slide images of ovarian carcinoma subtypes. HGSC: High Grade Serous Carcinoma, LGSC: Low Grade Serous Carcinoma, EN: Endometrioid carcinoma, MC: Mucinous Carcinoma, CC: Clear cell Carcinoma. Columns from left to right correspond to the appearance of tissues at a selected region on the WSI (green box) for decreasing magnification levels. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

noma type. These abnormalities appear at multiple magnifications. At lower magnifications, tissues organization resulting from cell proliferation leads to specific architectural patterns that are recognized as malignant and suggest cancer subtypes. At high magnification, cellular appearances confirm the histologic subtype, and nuclear shape and size are often seen as indicators of the risk of cancer progression. Thus, a clinical diagnosis or subtyping of ovarian carcinomas is the outcome of a visual-cognitive combination of these magnification-specific cues extracted from tissues.

The nature of the diagnostic procedure implies an inherent element of interpretation and hence subjectivity, and major errors can occur in pathology that have the potential of being undetected without appropriate safeguards (such as automatic review of new malignant diagnoses). The rate of major errors has been estimated to be in the range of 1.5 to 5% (Frable, 2006). In a recent study (Gavrielides et al., 2015) involving 114 patients and three expert pathologists, it was found that pathologists disagree on ovarian cancer cell-type classification on average 13% of the time, with a maximum disagreement on MC (21.4%) and EN (10%) cases. In practice, pathologists often end up scanning large amount of tissues for a diagnosis. When multiple tissue sections are not available, molecular features are required (e.g. p53 staining). Finally, challenging cases often require additional resource-intensive tests (e.g. immunohistochemistry) or asking for expert pathologists' opinion before agreeing on a diagnosis. Consequently, there is currently a need for faster, more robust, and reproducible systems that would complement and assist pathologists and clinicians during the diagnosis of ovarian carcinomas (Hipp et al., 2011).

Nonetheless, designing such automatic systems is a challenging task, as ovarian carcinomas are diverse and exhibit large intra-class variation. Besides, WSIs represent a computational hurdle as they contain large amount of information but may be composed of only a small number of important regions, while the remaining parts are irrelevant for classification. In practice, pathologists can easily spot these irrelevant regions (e.g fibrous tissue or apoptotic cells common to many types of cancers) and discard them during their analysis. However, building a computational model that can correctly identify and categorize these regions of interest without the need for extra manual annotation is challenging. Arguably, such a model must reason about which *combination* of spatial regions, and at what magnification levels, diagnostically relevant evidences occur.

Ovarian carcinomas are the result of an abnormal growth of epithelial cells. Epithelial tissues are formed by an ensemble of similar cells whose core is a nuclei and a cytoplasm enclosed in a membrane. Fig. 1 presents examples of clinico-pathologic features observed on each ovarian carcinoma subtype at different levels of magnification. First, at 40× and 20×, an abnormally high proliferation of nuclei is observed (Fig. 1a and b). This cellular growth further causes characteristic glandular organization and a solid appearance to the tissue that can be visualized under 10× and 4× magnification (Fig. 1c and d). Analysis of tissues at an isolated magnification can rarely lead to an effective characterization of the tumour type. In the case of HGSC and LGSC (Fig. 1), the highest magnification is ambiguous as it shows similar nuclei grade and proliferation for both carcinoma subtypes. At the other end of the spectrum, the lowest magnification shows how these nuclei formed cells that organize into glands with specific patterns distinctive of each tumour subtype (e.g. micropapillary in the case of HGSC vs. [macro]papillary for LGSC). At 40×, EN, CC and MC show subtle variations with a few to many mitotic figures and a lack of nuclear atypia. However, at 20×, papillary patterns with little cell