Effect of feature information-aided review on pathology trainee performance for ovarian cancer subtyping: an observer study

Marios A Gavrielides, Ph.D.
Meghan Miller, B.S.
Jeffrey Seidman, M.D.

Ian Hagemann, M.D.
Heba Abdelal, MD.
Zahra Alipour, M.D.
Jie-Fu Chen, M.D.
Behzad Salari, MD.
Lulu Sun, M.D., Ph.D.
Huifang Zhou, M.D., Ph.D.

Center for Devices and Radiological Health (CDRH) / U.S. Food and Drug Administration (FDA)/Silver Spring, MD

Department of Pathology and Immunology/Washington University School of Medicine/St. Louis, MO
Overview

• Feature information-aided review:
  • a type of Clinical Decision Support System (CDSS) where pathologists are provided with information on histological content
Overview

• Background on ovarian cancer subtypes
• Observer studies on pathologist concordance for ovarian cancer subtyping
• A pilot study with pathology trainees utilizing a CDSS for this task
Background - Ovarian Cancer types

- Types of ovarian cancer:
  - Epithelial ovarian cancer (90% of all OC, 75% all OV tumors)
  - Germ cell ovarian cancer
  - Stromal cell ovarian cancer
  - Small cell carcinoma
Background - Ovarian Cancer subtypes

- Epithelial ovarian carcinomas are further classified into histologic subtypes:

<table>
<thead>
<tr>
<th>Subtype</th>
<th>Prevalence*</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-grade serous (HGSC)</td>
<td>~66%</td>
</tr>
<tr>
<td>Low-grade serous (LGSC)</td>
<td>~5%</td>
</tr>
<tr>
<td>Endometrioid (EM)</td>
<td>~7%</td>
</tr>
<tr>
<td>Clear-cell (CC)</td>
<td>~8%</td>
</tr>
<tr>
<td>Mucinous (MUC)</td>
<td>~3%</td>
</tr>
<tr>
<td>Mixed (2+ subtype components &gt;10%)</td>
<td>~2%</td>
</tr>
<tr>
<td>Other</td>
<td>&lt;7%</td>
</tr>
<tr>
<td>Carcinosarcomas (CS) (~6%), Brenner tumor, Seromucinous, Undifferentiated)</td>
<td></td>
</tr>
</tbody>
</table>

Histologic subtypes of ovarian cancer

• Subtypes are now considered as essentially distinct diseases:
  • different origins, genetic alterations, clinicopathological features
  • different prognosis and response to treatment

Ovarian Carcinoma Subtypes Are Different Diseases: Implications for Biomarker Studies

Martin Köbel¹,², Steve E. Kalloger¹, Niki Boyd¹, Steven McKinney¹, Erika Mehl¹, Chana Palmer³, Samuel Leung¹,
Potential for subtype-specific treatment

- Historically, subtype has **not** been a factor in treatment selection
- Ongoing trials focus on subtype-specific treatment
  - Goal: improve the survival rate of ovarian cancer patients which has remained stagnant over the last 20 years
- Important to assign ovarian cancer subtypes accurately and reproducibly
  - Subtypes often have overlapping histologic features and some subtypes are rare
    - performance for this task depends on the pathologist’s expertise
  - Previous/ongoing work: examined pathologist concordance and reasons for discrepancies
Observer studies on pathologist concordance for OC subtyping

A. Expert review of subtype on microscope (JHU, FDA)
   • Reader study #1: independent case-based subtype diagnosis (114 cases)
   • Reader study #2: independent section-based subtype diagnosis (225 slides)
   • Reader study #3: consensus review of discrepant cases (21 cases)
     • Major source of discrepancy: identification and relative importance of histologic features (manuscript in preparation)

B. Pathologist review of patterns on microscope and WSI (UMD, GW, FDA)
   • Reader study #4: independent section-based review on microscope (225 cases) in multiple sites
   • Reader study #5: independent section-based review on WSI
     • Indicate presence/absence of 8 selected histologic features
     • Provide subtype
Histologic features for differential OC subtype classification

• Set of 8 features selected with Dr. Seidman and extensive literature search

1. High grade nuclear atypia:
2. High mitotic count:
3. Intracytoplasmic mucin
4. Hyalinized stroma
5. Clear cell architectural patterns: i.e. presence of hobnail cells, tubulocystic, or tubulopapillary patterns
6. Sarcomatous component
7. Squamous differentiation
8. Features suggestive of Endometriosis

• Inexperienced pathologists often have difficulty identifying or characterizing these features (observer study to be completed in May)
CDSS for ovarian cancer subtyping

• Overall project goals:
  • Develop methods for characterizing histology content in an accurate and reproducible manner:
  • Assess the robustness of such methods so that validation results can be generalized in clinical setting

• Pilot study: a CDSS to assist pathology trainees for ovarian cancer subtyping
Clinical Decision Support Systems (CDSS) in digital pathology

- Many tools have been developed for digital pathology, very few have been evaluated with pathologist in the loop
  - Gavrielides et al (Archives PathLabMed 2009): HER2 assessment with computer-extracted quantitative values of membrane staining
  - Vandenberghe et al. (Sci reports, 2017): Discordances between CAD and unaided led to re-classification of HER2 status in 8/12
  - Steiner et al (AmJSurgPathol): CAD for detection of metastases
    - Benefit for micrometastases detection (N=19), 83% to 91%

- No studies focused on evaluating effect of clinical decision support tools on primary diagnosis
Feature information-aided review as a CDSS for OC subtyping

Unaided review
Feature information-aided review as a CDSS for OC subtyping

Aided review
CDSS for ovarian cancer subtyping

• CDSS-aided review:
  • Textual indication on presence of any of 8 histologic patterns important for differential subtype classification
  • Histologic features identified in separate study by expert in Gynecologic Pathology
  • Intended use: alert observer on presence of histologic features that may have been overlooked during WSI review
CDSS for ovarian cancer subtyping

• Reader study: unaided and CDSS-aided subtype diagnosis
  • 90 whole slide images (WSI) from 75 OC patients
    • 56 cases with unanimous consensus truth, 34 with majority consensus
      • challenging set (mean inter-expert concordance, 3 experts: 74.1%)
    • 17 HGSC, 15 each (LGSC, MUC, EM, CC), 13 CS
  • 6 pathology residents (2nd and 3rd year) at Wash Univ Medical Center
    • Interested in this diverse dataset and to compare their scores to the experts’ scores following the study
  • Two sessions, 3-week minimum washout, randomized order shuffled between sessions
  • Same calibrated monitor, same interface (PathXL Tutor), same work station
  • Two-step training on OC subtyping using same interface
    • Interactive tutorial showing patterns and full WSI examples
    • Quiz on 20 WSI
### Percent correct agreement w/ expert panel

<table>
<thead>
<tr>
<th>Pathologist</th>
<th>Unaided % correct</th>
<th>Aided % correct</th>
<th>Difference (Aided-Unaided)</th>
<th>95% C.I.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>62.2%</td>
<td>74.4%</td>
<td>12.2%</td>
<td>[1.1 23.3]</td>
</tr>
<tr>
<td>2</td>
<td>55.6%</td>
<td>58.9%</td>
<td>3.3%</td>
<td>[-7.8 14.4]</td>
</tr>
<tr>
<td>3</td>
<td>57.8%</td>
<td>65.6%</td>
<td>7.8%</td>
<td>[-4.4 20.0]</td>
</tr>
<tr>
<td>4</td>
<td>70.0%</td>
<td>61.1%</td>
<td>-8.9%</td>
<td>[-22.2 4.4]</td>
</tr>
<tr>
<td>5</td>
<td>61.1%</td>
<td>78.9%</td>
<td>17.8%</td>
<td>[6.7 28.9]</td>
</tr>
<tr>
<td>6</td>
<td>70.0%</td>
<td>77.8%</td>
<td>7.8%</td>
<td>[-3.3 18.8]</td>
</tr>
<tr>
<td>Average</td>
<td>62.8%</td>
<td>69.4%</td>
<td>6.7%</td>
<td>[-2.8 16.1]</td>
</tr>
</tbody>
</table>

- 5/6 observers had increase in agreement with expert panel (for 2 observers increase was statistically significant)
- 1/6 observers had decrease in agreement (non-stat. significant)

*95% confidence intervals (95% C.I.) derived w/ bootstrap resampling
## Percent correct agreement w/ expert panel across subtypes

<table>
<thead>
<tr>
<th>Subtype</th>
<th>Unaided pairwise mean (std)</th>
<th>Aided pairwise mean (std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High grade Serous</td>
<td>51.0% (9.6%)</td>
<td>57.8% (21.5%)</td>
</tr>
<tr>
<td>Low grade serous</td>
<td>68.9% (9.1%)</td>
<td>71.1% (20.1%)</td>
</tr>
<tr>
<td>Endometrioid</td>
<td>67.8% (12.9%)</td>
<td>63.3% (13.8%)</td>
</tr>
<tr>
<td>Mucinous</td>
<td>48.9% (18.2)</td>
<td>54.4 (20.4%)</td>
</tr>
<tr>
<td>Clear Cell</td>
<td>68.9% (18.7%)</td>
<td>77.8% (16.1%)</td>
</tr>
<tr>
<td>Carcinosarcoma</td>
<td>74.4% (13.5%)</td>
<td>97.4% (4.0%)</td>
</tr>
</tbody>
</table>

- Largest benefit for subtypes with more definitive histologic features
  - detection typically is indicative of subtype
- Negative effect for EM: most cases had overlapping features (for 7/15 cases, histologic features provided included mucin)
  - detection is not enough, requires strategy for dealing with contradictory info
Inter-observer agreement

<table>
<thead>
<tr>
<th>Unaided % correct mean (std)</th>
<th>Aided % correct mean (std)</th>
<th>Difference in mean (Aided-Unaided)</th>
<th>95% C.I.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>53.5% (4.4%)</td>
<td>62.7% (10.7%)</td>
<td>9.2%</td>
<td>[-13% 38%]</td>
</tr>
</tbody>
</table>

- Potential for improving inter-observer agreement for trainees
- Benefit varied widely among observer pairs
  - inadequate training: observer training was limited to the clinical task not on the CDSS
  - lack of trust (indicated in post-study survey)
Discussion: lessons learned

• Trainees benefited from information on missed histologic features that were specific to a subtype
  • For overlapping features, localization and characterization of findings would be more informative
  • Findings can inform development of tools describing histology content

• Focusing training on the interaction with the CDSS could build trust in its capabilities
  • Improve consistency of CDSS benefit across users

• Information on histologic features was based on one expert’s observation
  • Provided information derived from a well-validated automated algorithm might increase trust of users
Summary

• We have conducted a pilot study on the use of a CDSS for ovarian cancer subtype diagnosis
  • 1\textsuperscript{st} CDSS for primary diagnosis to the best of our knowledge

• Considering this was a challenging set of cases and limitations of this particular implementation, results showed potential of CDSS to improve performance of inexperienced pathologists
  • close knowledge gap for complex tasks

• Future work:
  • development of AI/ML methods for supporting clinical decision making
  • assessing conditions for generalizability in the clinical setting
Acknowledgment

• Pathologists at FDA, JHU, GW, UMD, and Wash Univ for their precious time and expertise

• FDA Critical Path for initial funding for this project

• Comments? marios.gavrielides@fda.hhs.gov