Analysis of Influence of Additional Diagnostic Clues During Pathology Diagnosis
Mikhail Kovalenko, MS, Richard Hammer, MD, Dmitry Shin, PhD
Department of Pathology and Anatomical Sciences, University of Missouri-Columbia

Abstract

- We attempt to quantify the role of additional diagnostic clues in the process of pathology diagnosis in order to reduce the possibility of pitfalls.
- We use Apriori method of association rule mining with custom Perl extensions to collect sets of diagnostic clues that show an improvement in diagnostic quality.
- We look at the resulting information gain measure in an attempt to gauge the impact of additional factors in diagnostic process.

Introduction

- In cancers of hematopoietic origin, an accurate diagnosis depends on an in-depth analysis of clinical information as well as the correct evaluation of morphological patterns with a number of overlapping features. The amount of practitioner’s clinical experience also plays a role in the quality of diagnostic outcome.
- The presence of overlaps and personal opinion in diagnostic clue interpretation make the diagnostic process a subjective exercise and leads to diagnostic pitfalls and to difficulties in defining a unified set of best practices in diagnosis of hematopathological cancers.
- Our ongoing research is focused on revealing diagnosis-related details and heuristics that can be used to quantify and potentially improve the diagnostic process in pathology using whole-slide imaging (WSI) and analytical tools. In pursuit of this goal we have developed PathEdEx informatics framework and WSI platform that features realistic diagnostic workflow, WSI viewing capability, and gaze tracking capture and is used to record user activities related to diagnosing cancerous tissue slides, including visual and non-visual diagnostic clues, and to convert them into highly explanatory semantic format.
- Here we look at the influence of additional diagnostic factors that, when supplemented routinely used disease-specific clues, can decrease the chance of diagnostic pitfalls and improve diagnostic outcomes.

Methods

- We used PathEdEx to record clinical pathologists’ activities in order to capture visual diagnostic clues for subsequent informatics analysis.
- We extended association rule (AR) mining techniques to extract the interesting results, then measured information gain from the additional diagnostic clues.
- A transaction is defined as a complete diagnostic session for a given patient case.

Results and Discussion

- We induced association rules based on the collections of cell types covered by a user’s gaze track and the resulting diagnostic decision made by the pathologist in a simulated session.
- The clues were assembled into itemsets representing each instance of case consideration and processed using R’s arules package.
- The induced rules were further processed by a Perl script that retained only the association rules in the form of:

\[ \text{Item}_1, \text{Item}_2, \text{Item}_3, ..., \text{Item}_n \Rightarrow \text{Diagnosis} \]

with the diagnosis on the right-hand side. Another Perl script arranged all results into sorted lists, grouped by diagnosis, that started with rulesets containing one diagnostic clue and added rules with more clues until the confidence of the ruleset was maximized (Figure 1).
- Kullback–Leibler divergence (KLD) was calculated to judge information gain that is achieved when a more specific set of clues \( Y \) is used instead of a less specific one \( X \) according to:

\[ D_{KL}(Y||X) = \sum_{i=1}^{N} \ln \left( \frac{Y_i}{X_i} \right) Y_i \]

where \( D_{KL} > 0 \) to achieve gain.

- Two expert pathologists, four pathology residents and two PSFs diagnosed eight anonymized cancer cases. We only considered biological features that were noted in the tissue by the examining pathologist as relevant to the diagnosis. We did not attempt to collect any mental decisions or side commentary. Low-power microscopic observations also were not considered.

<table>
<thead>
<tr>
<th>LH</th>
<th>RHS</th>
<th>Support</th>
<th>Confidence</th>
<th>Lift</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;H&amp;E&gt;</td>
<td>Irregular lymphocyte</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;H&amp;E&gt;</td>
<td>Irregular lymphocyte, &lt;H&amp;E&gt; Small round lymphocyte</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;H&amp;E&gt;</td>
<td>Irregular lymphocyte, &lt;H&amp;E&gt; Small round lymphocyte, &lt;H&amp;E&gt; Clear cells</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;H&amp;E&gt;</td>
<td>Histiocyte</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;H&amp;E&gt;</td>
<td>Histiocyte, &lt;H&amp;E&gt; Small round lymphocyte</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;H&amp;E&gt;</td>
<td>Histiocyte, &lt;H&amp;E&gt; Small round lymphocyte, &lt;H&amp;E&gt; Reed-Sternberg Cell</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;H&amp;E&gt;</td>
<td>Large centroblast</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;H&amp;E&gt;</td>
<td>Large centroblast, &lt;H&amp;E&gt; Monocytoid B cell</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;H&amp;E&gt;</td>
<td>Large centroblast, &lt;H&amp;E&gt; Monocytoid B cell, &lt;H&amp;E&gt; Small centrocyte</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Three sets of results of additional diagnostic clue analysis (fa, fb, and fc top to bottom).

Conclusion

We conclude that the approach presented here can be used to uncover additional diagnostic clues with the goal to reduce diagnostic pitfalls, given that the appropriate rule-pruning procedures are performed.

Acknowledgements

This work was supported by NIH NLM T32 NRSA Institutional Research Training Grant 1T32LM012410-01.