Rapid Radiology-Pathology Correlation using the TIES NLP System with Manual Abstraction Tools

Matthew G Hanna, MD; Maria Habib, MD; Girish Chavan, MS; Eugene Tseytlin, MS; Liron Pantanowitz, MD; Rebecca S. Jacobson, MD, MS

Background

Diagnostic imaging has provided clinicians with increasingly sophisticated noninvasive representations of disease. Correlation of radiological interpretations to histopathologic diagnoses has been an important element for quality improvement and education programs [1]. In the field of breast imaging, correlation to pathology findings is mandated by law [2]. Traditional manual methods for abstracting this information can be time consuming and expensive. While current electronic medical record systems increase the availability of coded data, the radiology and pathology data needed for correlation is typically stored as free-text. Consequently, many health systems and practices must manually correlate mammography to subsequent pathologic findings, which is time consuming, expensive, and typically produces a prolonged delay in assessment. We describe our experience in developing a method to rapidly correlate radiology-pathology data from Breast Imaging-Reporting and Data System (BIRADS) categories and pathology diagnoses. As more robust clinical information processing tools become available, these correlations can also become more automated, ultimately benefitting both disease surveillance and population health strategies.

Case Presentation

Radiologists interpret large volumes of imaging studies with many diagnoses that cannot be definitively determined based on imaging findings alone. In order to provide optimal patient management, expedient and accurate communication of diagnoses between radiology and pathology findings must be established. At our institution, current workflow for radiology-pathology correlation of mammographic findings include multiple staff members accessing various clinical information systems to manually input correlative diagnosis information into Imagecast™ (GE Centricity). Radiology-pathology correlations are then conveyed as separate reports to each referring radiologist.

To pilot test an alternative quality assurance process, we used an existing natural language processing application, the Text Information Extraction System (TIES), and database comprised of 28.9 million radiology and pathology reports from our own institution [3-4]. TIES uses a variety of natural language processing methods to segment and annotate text data with coded concepts. We searched for radiology reports with a BIRADS concept longitudinally linked to pathology reports with breast biopsies within one year [Figure 1]. Using the TIES customizable manual annotation tool, we generated a data collection form in the TIES application (v5.6.2) directly suited for this correlation task. This process saves manually abstracted data associated with case sets and exported documents. Users can generate customizable data elements, which combine data definitions with a data field label for display on the form to the user. Form fields can be designated as text, numerical, Boolean, and categorical. After developing the BIRADS Radiology-Pathology Correlation QA form, we collected relevant data for each case and then used the results compare BIRADS categories with matched pathology reports from the aforementioned query [Figure 2].
Figure 1

Patient Event 1

All Reports

Report Type
PITT Radiology

Search Term
In section: Impression
Breast Imaging Reporting and Data System(C1511314)

Filtered Reports

Patient Event 2

All Reports

Report Type
PITT Pathology

Search Term
In section: Final Diagnosis
Biopsy of Breast(C0405352)

Filtered Reports

Within 1 Year(s)
Figure 2

EXAMINATION PERFORMED:

MR BREAST WITH/ WITHOUT CONTRAST BL. TEMP. "DATE DEC 11 2011" 0715-0845

CLINICAL HISTORY:

**NOTE**: One lesion in the infra mammary area and a palpable mass in the left breast and axillary lymph node with no palpable node in the right axilla

COMPARISON:

MR/Mammograms and ultrasonic examination "DATE DEC 11 2011"

TECHNIQUE:

Three planes (sagittal, axial and coronal) with and post dynamic contrast. MRI and ultrasound (US) images were obtained after administration of gadolinium diethylenetriamine penta-acetic acid (Gd-DTPA) and ultrasound images were obtained. Postprocessing images were obtained with CAD software

FINDINGS:

In 2 axillary left breast, there is a large heterogeneous enhancing mass measuring 3.5x3.5x2.0 cm suspicious for malignancy. There is a left axillary lymph node with suspicious for malignancy. There is no involvement of the left breast. There are no suspicious or abnormal areas of enhancement within the right breast.

IMPRESSION:

LARGEST HETEROGENEOUS ENHANCING UBIQUITATED MASS INFERIOR PSLEFT BREAST SUSPECTED FOR MALIGNANCY. MRI AND ULTRASOUND- GUIDED BIOPSY ON THE SAME DAY. SWB IN 3D AND PERFORMED.

2. LEFT AXILLARY NODE WITH HETEROGENEOUS ENHANCED CONTENT. THIS IS ALSO SUBSEQUENTLY FOUND OF SOME ULTRASOUND BIOPSY ON THE SAME DAY.

NO MR FINDINGS SUSPECTIVE FOR MALIGNANCY WITHIN THE RIGHT BREAST.

ASSESSMENT AND RECOMMENDATIONS

MRI BI-AXIAL category 5. HIGHLY SUSPECTIVE FOR MALIGANCY.

THIS MASS ULTRASOUND-ULTRASOUND GUIDED BIOPSY OF THE SAME DAY. ADDITIONALLY ULTRASOUND GUIDED BIPSY OF THE LEFT AXILLARY NODE WAS ALSO PERFORMED.

OVERALL ASSESSMENT BI-AXIAL category 5.

Dated by: [Signature] on [DATE] at [TIME] AM

[MARS EXAM, TYPE]

MR BREAST WITH/ WITHOUT CONTRAST BL.

[MARS REPORT_SUBTYPE]

PROCEDURE

[MARS DX]

NEXT VISIT: [RECOMMENDATION] "PLEASE HET 3 MONTH"
<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient ID (For Patient)</td>
<td></td>
</tr>
<tr>
<td>Laterality (Finding 1) (For Patient)</td>
<td>Left</td>
</tr>
<tr>
<td>Location, O'clock, DFN (Finding 1) (For Patient)</td>
<td>2</td>
</tr>
<tr>
<td>BIRADS Category (Finding 1) (For Patient)</td>
<td>5</td>
</tr>
<tr>
<td>Pathology Risk (Finding 1) (For Patient)</td>
<td>Malignant</td>
</tr>
<tr>
<td>Pathology Dx (Finding 1) (For Patient)</td>
<td>IDC</td>
</tr>
<tr>
<td>Rad-Path Correlation (Finding 1) (For Patient)</td>
<td>True</td>
</tr>
<tr>
<td>Laterality (Finding 2) (For Patient)</td>
<td></td>
</tr>
<tr>
<td>Location, O'clock, DFN (Finding 2) (For Patient)</td>
<td></td>
</tr>
<tr>
<td>BIRADS Category (Finding 2) (For Patient)</td>
<td></td>
</tr>
</tbody>
</table>
Outcome

Using the TIES application we generated a query to collect the relevant cohort. We retrieved a total of 2953 reports from 938 patients. Figure 3 shows the TIES display of demographic information [Figure 3]. Total query time was 1 minute and 28 seconds. To compare data extraction times across multiple modalities, a novice user of the TIES application and clinical information systems correlated matched radiology and pathology patient reports from 145 patients. The form was developed for input up to two biopsied mammographic findings.

Figure 3

Demographic Distribution

Data elements included in the form were as follows:
- Laterality (i.e. right, left)
- Position (i.e. location, o’clock, distance from nipple)
- BIRADS category (i.e. 0, 1, 2, 3, 4, 4A, 4B, 4C, 5, 6)
- Pathology Diagnosis (i.e. invasive ductal carcinoma, duct carcinoma in-situ, etc)
- Pathology Risk Classification (i.e. benign, high risk, malignant)
- Correlation (true, false)

To compare the QA form developed in TIES with available manual extraction tools, we evaluated the current workflow of searching the clinical information systems. The time measured excluded patient search query times for both comparative modalities, however included searching within each patient’s electronic medical record for corresponding imaging and pathology reports. The data elements above were used as benchmark criteria for data extraction for all evaluations. Consequently, we also performed the same task using the traditional method of searching the RIS and LIS, and then combining the data manually using Microsoft Excel. Calculating current workflow, the average time of data extraction from the hospital EMR and LIS averaged 187 seconds for each patient.

Using TIES, we evaluated two methods of data extraction to identify meaningful utilization of the custom form creation tool. We used TIES to aggregate matched patient reports (patient radiology and pathology reports are grouped and contiguously displayed) and replicated an Excel spreadsheet for data input similar to the non-TIES data extraction; this method averaged 85 seconds per patient. Using the TIES manual annotation tool with a custom developed BIRADS radiology-pathology QA form, data elements were extracted on average in 62 seconds per patient, and were able to subsequently be exported to Microsoft Excel as an Excel Workbook file.
Additionally, using the data elements and exported data from the TIES QA form fields comprising of patient clinical information, we were able to facilitate rapid collection of a large data set from radiology-pathology correlative information. Of the 145 patients, the radiology and pathology correlations are listed in Figure 4. The specific BIRADS correlation information can be directly viewed from each definitive QA form; however the collective results [Figure 4] can be readily viewed for population studies.

**Figure 4**

<table>
<thead>
<tr>
<th>Radiology</th>
<th>Pathology</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIRADS</td>
<td>n=145</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
</tr>
</tbody>
</table>

**Discussion**

Natural language processing engines have undergone significant development in recent decades [5]. TIES is an open-source NLP data analytical tool for cohort assessment. Aggregation of patient cohorts in case sets enables investigators to develop custom forms to extract discrete data from patient reports via manual annotation. Creating customizable forms for analysis of structured data of patient clinical information can improve data collection turnaround time. The quality assurance processes may benefit from having readily matched patient reports for correlation of pathology and radiology information using systems like TIES.

**Conclusion**

Radiology-pathology correlation is crucial to the quality analytics of health care. Conventional clinical information systems are inundated with text based input. NLP solutions exist and can be readily used for mandated quality assurance. TIES customizable forms for BIRADS radiology-pathology correlation proves to be a rapid and robust analytic tool.
References


Keywords

pathology correlation, quality, natural language processing