Resources Required for Semi-Automatic Volumetric Measurements in Metastatic Chordoma: Is Potentially Improved Tumor Burden Assessment Worth the Time Burden?

Suyash Mohan, MD, PDCC, University of Pennsylvania; Emmanuel Botzolakis, MD, PhD; Ilya M. Nasrallah, MD, PhD; Edward Herskovits, MD, PhD; Harold Kundel, MD; Nick R. Bryan, MD, PhD (Presenter)

Hypothesis

We hypothesize that the diagnostic accuracy of the differential diagnosis (DDx) generated by our expert system, using key features (KFs) extracted by radiologists, would perform at the level of an intermediate level (PGY-2 and PGY-3) radiology resident.

Introduction

Despite an explosive growth in digital imaging and computational techniques throughout medicine, applications that assist in formulating radiological differential-diagnosis (DDx) are limited.

We have developed a Bayesian Expert System (BES), RadDx, which takes as input key features (KFs) determined by radiologists from adult brain MRI examinations, and outputs a rank-ordered DDx list.

We hypothesize (as stated above) that the diagnostic accuracy of the DDx generated by the model, using KFs extracted by radiologists, would perform at the level of an intermediate level (PGY-2 and PGY-3) radiology resident.

Methods

13 neuroradiology attendings and 11 neuroradiology fellows selected a set of 450 MRI examinations from the radiology archives of our tertiary-care teaching hospital, representing the spectrum of adult neuroradiology diagnoses (80) including ‘normal’. The consensus diagnoses for these cases, based upon the clinical, radiologic and pathologic records, where available, were taken as the ground truth (GT). A total of 33 diagnosis-blind readers, in 3 level-of-training groups, 13 neuroradiology attendings, 11 neuroradiology fellows and 9 radiology residents, viewed the cases with no clinical information except age on a PACS workstation. They reported predefined signal and spatial KFs, as well as their 3 highest probability diagnoses for each of these cases on a specially designed web-based interface. Six reads were obtained for each case, two each from each group. Each reader’s KFs were entered into the BES, which then generated a probability rank-ordered DDx (BES-DDx). We compared Radiologist Group-DDx and BES/Radiologist Group-DDx to the GT diagnoses using Obuchowski/Rockette (OR) Dorfman/Berbaum/Metz (DBM) MultiReaderMultiCase (MRMC) receiver operating characteristic (ROC) analysis.

Results

1,803 cases have been reviewed. Unbiased AUCs for Radiologists’ Normal vs Abnormal on First Choice; Correct Diagnosis in Top 3 Choices; Correct Diagnosis on First Choice; and the more general Choice Using Rating Scale results are shown in Figure 1 and Table 1. As expected, greater diagnostic accuracy was
associated with higher levels of training. These results show that all levels-of-training groups discriminate Normal from Abnormal fairly well, but the ability to make the correct diagnosis is very training level dependent. This probably relates to the large number of disease choices and relative rarity of some of the diseases. The AUCs corrected for bias, using OR-DBM-MRMC show that the performance of the BES was also training level dependent, but always below the level of the corresponding radiologist. However, the Choice Using Rating Scale AUC of the attending BES DDx (0.86) was significantly (p<.05) greater than that of the self-generated resident DDx (0.74) and equal to that of the fellow’s (0.86). In summary, in terms of diagnostic accuracy, these results show that the attending BES DDx approximates that of a self-generated DDx of a neuroradiology fellow and is better than that of a radiology resident.

![Figure 1: OR-DBM-MRMC ROCs using PROPROC for curve fitting. Reader/ BES Diagnosis vs. Ground Truth](image-url)
Table 1: Sensitivity, Specificity and AUCs of Reader/ BES Diagnosis vs. Ground Truth

<table>
<thead>
<tr>
<th>Attending</th>
<th>Nor vs Abn 1st Choice</th>
<th>Any Choice Correct</th>
<th>First Choice Correct</th>
<th>Rating Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sens</td>
<td>Spec</td>
<td>AUC</td>
<td>Sens</td>
</tr>
<tr>
<td>Rad DDx</td>
<td>0.93</td>
<td>0.83</td>
<td>0.76</td>
<td>0.59</td>
</tr>
<tr>
<td>BES DDx</td>
<td>0.90</td>
<td>0.80</td>
<td>0.71</td>
<td>0.39</td>
</tr>
<tr>
<td>Fellow</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sens</td>
<td>Spec</td>
<td>AUC</td>
<td>Sens</td>
</tr>
<tr>
<td>Rad DDx</td>
<td>0.94</td>
<td>0.63</td>
<td>0.63</td>
<td>0.57</td>
</tr>
<tr>
<td>BES DDx</td>
<td>0.90</td>
<td>0.66</td>
<td>0.63</td>
<td>0.36</td>
</tr>
<tr>
<td>Resident</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sens</td>
<td>Spec</td>
<td>AUC</td>
<td>Sens</td>
</tr>
<tr>
<td>Rad DDx</td>
<td>0.90</td>
<td>0.58</td>
<td>0.59</td>
<td>0.39</td>
</tr>
<tr>
<td>BES DDx</td>
<td>0.87</td>
<td>0.62</td>
<td>0.60</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Discussion

RadDx, a novel radiology-specific expert system, has the potential for reducing the variability of radiology reports, improving report quality, and increasing the efficiency of practicing radiologists.

Conclusion

A neuroradiological diagnostic decision-support application was developed and evaluated using a set of 450 diagnosis-proven clinical neuroradiology MRI examinations. As expected, a radiologist’s diagnostic accuracy increased as a function of level of training. BES performance was below that of the radiologists, but using KFs extracted by Attendings, the BES performed at a resident level. Based on previous work, optimization of the BES is expected to improve its performance, with potential to reduce the variability of radiology reports, improve report quality, reduce missed diagnoses and increase the efficiency of practicing radiologists.

References


Keywords

Clinical Diagnostic Decision Support, Radiology, Bayesian Expert System, Neuroradiology