Evaluation Challenges for Bridging Semantic Gap: Shape Disagreements on Pulmonary Nodules in the Lung Image Database Consortium

William H. Horsthemke, DePaul University, USA
Daniela S. Raicu, DePaul University, USA
Jacob D. Furst, DePaul University, USA

ABSTRACT

Evaluating the success of prediction and retrieval systems depends upon a reliable reference standard, a ground truth. The ideal gold standard is expected to result from the marking, labeling, and rating of domain experts. However, experts disagree and lack of agreement challenges the development and evaluation of image-based feature prediction. This paper addresses the success and limitations in bridging the semantic gap between CT-based pulmonary nodule image features and radiologists' ratings of diagnostic characteristics. The prediction of diagnostic characteristics promises to automatically annotate images with medically meaningful descriptions usable for indexing and retrieving in content-based image retrieval (CBIR) and computer-aided diagnosis (CADx). Successful results in predicting texture characteristics will be contrasted with less successful results for boundary shapes. The two primary differences in agreement between radiologists will be discussed; the first concerns agreement about the existence of a nodule, while the second considers the variability in radiologists' ratings.

Keywords: decision support; evidence-based medicine image; image analysis; image processing; medical imaging systems; retrieval; semantic data model

INTRODUCTION

Increased utilization of diagnostic imaging imposes a growing demand for improving the efficiency of radiology departments. Picture Archiving and Communication Systems (PACS) technology helps manage the operational demands but new technologies are needed to address the diagnostic demand. Recent research (Sluimer, Schilham, Prokop, & Ginneken, 2006) and commercial success (Doi, 2005) with computer-aided detection (CAD) and diagnosis (CADx) have demonstrated the efficacy of using CAD(x) as a second reader to augment the diagnostic process. While increasingly accurate
in detection and diagnosis, CAD(x) rarely offers supporting guidance about the rationale for the diagnosis or supplies descriptive annotations about medically meaningful diagnostic characteristics (Doi, 2005). Two areas of research attempt to address this deficiency, one focuses on standardizing diagnostic terminology with notable success in the mammography community (American College, 2003) while the other, semantic mapping, focuses on automatically labeling images with usable, medically meaningful, diagnostic descriptors. Semantic mapping attempts to extract image features and build predictive models of diagnostic characteristics for labeling images.

Semantic mapping promises to add clinically relevant diagnostic evidence to support the medical decision maker using CAD(x)-based tools as well as content-based retrieval (CBIR) systems (Bui, Taira, Dionisio, Aberle, El-Saden, Kangarloo, 2002). Case-based reasoning is a major aim of CBIR and approaches include the display of relevant images based entirely upon the image contents (El-Naga, Yongyi, Galatsanos, Nishikana, & Wernick, 2004). Extending this approach to automatically extract clinically relevant diagnostic characteristics will allow for retrieval of past cases with similar characteristics (Bui, et al., 2002).

The semantic mapping approach discussed in this paper follows the methodology introduced by Raicu, Varuthbangkul, Cisneros, Furst, Channin, & Armato III (2007) of combining measurements of low-level image features and radiologists’ outlines of pulmonary nodules to predict each radiologists’ ratings for a set of diagnostic characteristics such as texture, margin, subtlety, spiculation, etc. These characteristics reflect image-based radiologist-interpreted image features useful for diagnostic decision making. Using radiologist-defined ratings for diagnostic characteristics as the target for classification and prediction models, this approach aims to bridge the semantic gap between low-level image features and medically meaningful descriptions of images. The predicted ratings for diagnostic characteristics promise to enhance medical decision making both with the CADx framework as well as content-based image retrieval. For CADx, the ratings of diagnostic characteristics offer evidence for radiologists in decision making. For CBIR, the ratings will enhance indexing and retrieval relevance by offering queries based upon diagnostic criteria and providing additional diagnostic evidence for retrieved images.

In this paper, we extend previous work on semantic mapping of image features to predict radiologists’ ratings of diagnostic characteristics of pulmonary nodules by focusing on building boundary-based shape descriptor models directly from radiologist-drawn outlines on the assumption that their outlines best represent their perception of the boundary of the nodules. Through example cases, we verify the approach and show that the shape metric varies accordingly. After applying the shape feature method, two predictive methods are applied but fail to predict the radiologist ratings for three (3) shape characteristics. In discussing the results we show that the variability in their ratings for shape characteristics is much higher than for other diagnostics characteristics such as texture which have been successfully mapped by Raicu, et al., (2007). We conclude with strategies for combining the radiologists’ outlines and ratings in an attempt to produce a usable shape descriptor model.


RELATED WORK

Evaluation remains a significant challenge to the development of semantic mapping in medical imaging. The challenges reside in the variability in diagnostic opinions among radiologists (Armato, McLennan, McNitt-Gray, Meyer, Yankelevitz, Aberle, et al., 2004) and