

## Chapter 93

# Evaluation Challenges for Computer–Aided Diagnostic Characterization: Shape Disagreements in the Lung Image Database Consortium Pulmonary Nodule Dataset

**William H. Horsthemke**  
*DePaul University, USA*

**Daniela S. Raicu**  
*DePaul University, USA*

**Jacob D. Furst**  
*DePaul University, USA*

**Samuel G. Armato III**  
*University of Chicago, USA*

### ABSTRACT

*Evaluating the success of computer-aided decision support systems depends upon a reliable reference standard, a ground truth. The ideal gold standard is expected to result from the marking, labeling, and rating by domain experts of the image of interest. However experts often disagree, and this lack of agreement challenges the development and evaluation of image-based feature prediction of expert-defined “truth.” The following discussion addresses the success and limitation of developing computer-aided models to characterize suspicious pulmonary nodules based upon ratings provided by multiple expert radiologists. These prediction models attempt to bridge the semantic gap between images and medically-meaningful, descriptive opinions about visual characteristics of nodules. The resultant computer-aided diagnostic characterizations (CADc) are directly usable for indexing and retrieving in content-based medical image retrieval and supporting computer-aided diagnosis. The predictive performance of CADc models are directly related to the extent of agreement between radiologists; the models better predict radiologists’ opinions when radiologists agree more with each other about the characteristics of nodules.*

## INTRODUCTION

Computer-aided decision support in medical imaging has focused primarily on the challenging problems of detecting and diagnosing suspicious lesions such as pulmonary nodules, which are often missed or misinterpreted by radiologists. Although automated decision support methods such as detection (CADe) and diagnosis (CADx) offer valuable diagnostic information about the presence or absence of suspicious lesions and perhaps probabilities about the likelihood of malignancy, together, these CAD(x)—CADe or CADx—systems rarely describe the lesion or offer additional information to support the radiologist in making their decision (Doi, 2005). This black-box approach generates skepticism against CAD(x) systems according to researchers developing commercial systems (Wiemker, Opfer, Bulow, Kabus, & Dharaiya, 2008). They argue for the addition of computed conceptual features based upon human appraisal to inform and support the radiologist when using the results of CAD(x) systems.

Towards bridging the semantic gap between medical images and human appraisal, the computer-aided diagnostic characterization (CADc) approach has been proposed. CADc aims to compute these conceptual features by extracting image-based features to predict radiologist-provided opinion of medically-meaningful diagnostic characteristics of focal anomalies. These CADc systems are developed using machine learning and statistical pattern recognition techniques to map quantitative image analysis measurements (features) to expert opinion given by radiologists (ground truth). The challenge for designing and developing CADc models is selecting or developing image feature extraction algorithms that capture relevant visual characteristics as observed by experts and obtaining sufficient, consistent opinion from expert radiologists to train the models. As discussed in this chapter, the most challenging problem is evaluating the prediction performance

of the models when the ground truth is inconsistent due to lack of agreement between radiologists.

This chapter discusses the development and application of a CADc approach towards characterizing shape-related characteristics of the pulmonary nodule, a type of lung lesion that might indicate lung cancer. Understanding the characteristics of nodules aims to help radiologists distinguish cancerous nodules from other types of abnormal tissue caused by infection or other non-cancerous diseases. In clinical usage, the lung nodule CADc scheme would indicate the extent of spiculation, degree of lobulation, and other nodule characteristics, then annotate the nodule with these ratings to provide diagnostic evidence to inform and support the radiologist's diagnostic decision.

In addition to providing quantitative evidence, the CADc ratings can be used to retrieve similar patient cases from medical image databases with known diagnostic and patient outcomes. During the reading of a new patient case, the CADc ratings are computed for a suspicious nodule and used to retrieve similar nodules from an image database. The retrieved nodules will be both medically similar in terms of the ratings for diagnostic characteristics and visually similar since the ratings are based upon direct, quantitative measurements of pulmonary nodules. If the radiologist considers these nodules to be sufficiently similar to the new patient case, then the known diagnoses may be useful to the radiologist during their differential diagnosis (Doi, 2005).

The retrieval of images similar in visual appearance is known as content-based image retrieval (CBIR). Much work has been done using content-based retrieval in the field of mammography. Giger, et al. (2002) pioneered the application of CBIR to mammography was reported and more recent developments are discussed in a review by Zheng (2009). Lam, Disney, Raicu, Furst, & Chanin (2007) developed an initial CBIR framework to explore feature extraction and similarity metrics for the retrieval of lung nodules and later extended the framework (Datterri, Raicu, & Furst, 2008).

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/evaluation-challenges-computer-aided-diagnostic/73523](http://www.igi-global.com/chapter/evaluation-challenges-computer-aided-diagnostic/73523)

## Related Content

---

### Overview of PAKDD Competition 2007

Zhang Junping and Li Guo-Zheng (2010). *Strategic Advancements in Utilizing Data Mining and Warehousing Technologies: New Concepts and Developments* (pp. 277-284).

[www.irma-international.org/chapter/overview-pakdd-competition-2007/40409](http://www.irma-international.org/chapter/overview-pakdd-competition-2007/40409)

### Integrated Model of Inductive-Deductive Inference Based on Finite Predicates and Implicative Regularities

Arkadij Zakrevskij (2013). *Diagnostic Test Approaches to Machine Learning and Commonsense Reasoning Systems* (pp. 1-12).

[www.irma-international.org/chapter/integrated-model-inductive-deductive-inference/69402](http://www.irma-international.org/chapter/integrated-model-inductive-deductive-inference/69402)

### Legal Framework for Data Mining and Privacy

Robert Sprague (2009). *Social Implications of Data Mining and Information Privacy: Interdisciplinary Frameworks and Solutions* (pp. 181-198).

[www.irma-international.org/chapter/legal-framework-data-mining-privacy/29150](http://www.irma-international.org/chapter/legal-framework-data-mining-privacy/29150)

### Semi-Automatic Design of Spatial Data Cubes from Simulation Model Results

Hadj Mahboubi, Sandro Bimonte, Guillaume Deffuant, Jean-Pierre Chanet and François Pinet (2013). *International Journal of Data Warehousing and Mining* (pp. 70-95).

[www.irma-international.org/article/semi-automatic-design-spatial-data/75616](http://www.irma-international.org/article/semi-automatic-design-spatial-data/75616)

### Post-Processing for Rule Reduction Using Closed Set

Huawen Liu, Jigui Sun and Huijie Zhang (2009). *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction* (pp. 81-99).

[www.irma-international.org/chapter/post-processing-rule-reduction-using/8438](http://www.irma-international.org/chapter/post-processing-rule-reduction-using/8438)