

CAD Performance Analysis for Pulmonary Nodule Detection on Thin-slice MDCT Scans

TBD^{a,*}, Matthias Wolf^a, Arun Krishnan^a, Marcos Salganicoff^a, Jinbo Bi^a, Murat Dundar^a, Glenn Fung^a, Jonathan Stoeckel^a, Senthil Periaswamy^a, Hong Shen^b, Peter Herzog^c, David P. Naidich^d

^a Siemens Medical Solutions USA Inc., Malvern, PA

^b Siemens Corporate Research, Princeton, NJ

^c Department of Clinical Radiology, University of Munich

^d Department of Radiology, New York University Medical Center, NY

...

Abstract. Early detection of lung nodules is an important clinical indication for obtaining routine CT studies of the thorax. To date, research has mainly focused on the sensitivity of CAD compared with expert chest radiologists using data obtained from single or multi-detector CT scanners. But, beside sensitivity and specificity it is also important to know how well the CAD system does perform on datasets from different sites with different slice thicknesses and differences in the dosage. The present study focuses on the generalization ability of a CAD system. It describes the architecture of a recent CAD system and assesses the performance on a heterogeneous dataset collected from multiple sites.

Keywords: Computer-Aided Diagnosis, CAD, Lung nodule, Multi-detector CT

1. Introduction

Lung cancer is the leading cause of cancer related death in western countries with a mean 5 year survival rate for all stages of only 14%. The prognosis of stage I cancer is more prosperous with a mean 5 year survival rate of about 49%. Although multi-slice CT scanner allow acquisition of the entire chest with sub-millimetre slice thickness within a breath hold, only 15% of lung cancers are diagnosed at this early stage. The problem is that one CT examination acquires up to 700 axial images whose interpretation is not only tedious but also difficult and may result in increasing false-negative rates. Computer-aided detection (CAD) is considered to be a helpful diagnostic tool to handle the increasing amount of radiological data. It is well recognized that use of CAD not only offers the potential to decrease detection and recognition errors as a second reader, but also to reduce mistakes related to misinterpretation [2]. But when using CAD systems we also need to know and understand how well such a system can

* Corresponding author. *E-mail address:* TBD@siemens.com

generalize and how they will perform on unseen data. This problem is related to the fact that most CAD systems are developed on and tuned for a relatively small set of data. Depending on the quality, source, and variability of a given lesion within this dataset the ability to generalize may be restricted which may lead to a bad sensitivity and specificity on new data. With this in mind, the purpose of the present paper is to describe the design and to assess the generalization ability of a new CAD system. In the following sections we will explain the data used in this study, we will give an overview of the evaluated CAD system, describe the training process and will conclude with results and a conclusion.

2. Material

The dataset of 185 high-resolution CT images was obtained from one North American and three European sites. The CT data was acquired with the following scanners: Siemens VolumeZoom CT, a Sensation 16, and a Sensation 64 scanner, all with 100kV, 120kV, or 140kV, and an exposure from 20mAs to 253mAs. All volumes were reconstructed with a B50f or B60f reconstruction kernel. The slice thickness ranged in 1.0mm and 1.25mm with an axial resolution between 0.5mm and 0.8mm. The ground truth used in this study was established by 3 radiologists (1 expert and 2 experienced readers). Each of the cases was read by 1 reader.

3. Overview of the CAD system

The CAD system is an automatic lung nodule detection system (not commercially available yet) comprising multiple modules to detect solid nodules with or without vascular or pleural attachment.

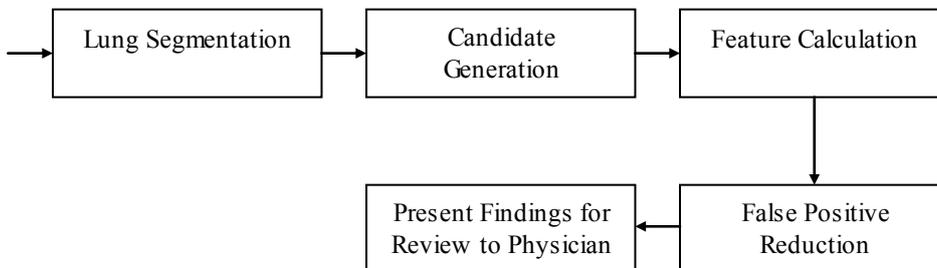


Fig. 1. Graphical overview of the dataflow in the CAD system described

The CAD system, see Fig. 1, comprises the following modules that are processed consecutively: a) lung segmentation; b) candidate generation; c) feature calculation at each candidate location; d) false positive reduction and e) presenting CAD findings to a physician for review. The lung segmentation is used to speed up the processing time by restricting the further analysis to a volume of interest (VOI) that is much smaller than the original volume. Lung segmentation is achieved by first thresholding on the Hounsfield units of the volume and then to use morphological operators to improve the accuracy of the segmentation. For nodule detection it is important to obtain a very accurate

segmentation of the lung boundaries. Otherwise pleura attached nodules might not be included in the VOI and would not be considered in subsequent processing steps. In the next module: “candidate generation”, the search space is further reduced by quickly generating a list of suspicious locations for possible nodules with high sensitivity. Each of the candidates points to a structure that might be a nodule and that needs a more exhaustive examination. Shape based characteristics are used during this step to generate the list of candidates. During feature calculation more complex and computationally more expensive feature values are then calculated for each of the candidates. Those features are based on the intensity, 2D and 3D shape descriptors, the curvature, and the location of the corresponding structure. Finally, the goal of the false positive reduction step is to increase the specificity without decreasing the sensitivity by pruning the list of candidates. For this, a classifier is used. The classifier combines all feature values for each candidate to classify it as either 'nodule' or 'no nodule'. The CAD findings are then presented to the physician for further review using the LungCARE software platform.

4. Classification

Given a set of candidates the goal is to construct a classification function to differentiate candidates that are true nodules from the rest of the candidates. Each candidate corresponds to a lung structure that has a potential to be a nodule. Additionally, each candidate is represented by a vector of numerical values or attributes that are commonly referred as features. The features are based on physical attributes and/or geometrical characteristics of the candidate structure such as size, roundness, texture, symmetry, etc.

One of the difficulties in constructing a classifier for this task is the problem of feature selection. It is a well-known fact that a reduction of classifier feature dependences improves the classifier’s generalization capability. However, the problem of selecting an “optimal” small subset of features from a large pool (in the orders of up to hundreds) of potential original features is known to be NP-hard. In order to address this difficulty we use a feature selection greedy approach. Given a subset of features S , our greedy approach consists in finding a new single feature f from the feature pool that improves classification performance when considering the expanded subset of features $S \cup f$.

This procedure is done starting with an empty set of features and it stops when no feature in the feature pool improves classification performance significantly when added to S . At each step, classification performance is measured based on the area under the Receiver Operating Characteristic (ROC) Curve and Leave-One-Patient-Out (LOPO) cross-validation [1].

Another difficulty associated with the task of designing a classifier for nodule classification is the fact that the number of nodules is usually very small compared to the number of non-nodule structures in the candidate list. This is referred in the machine learning community as an unbalanced training set. Error-based classification methods like Support Vector Machines (SVM) are known to have problems with unbalanced datasets since the majority class tends to override the minority class. In order to address this problem we used classifiers where the distribution of each class is estimated separately to avoid inter-class influence.

The operating point for the final system (80.7% sensitivity and 2.7 FP) was chosen based on the training LOPO ROC curve presented in Fig. 2. Note that the slope of the ROC curve at around 2.7 false positives tends to diminish making this point a stable choice.

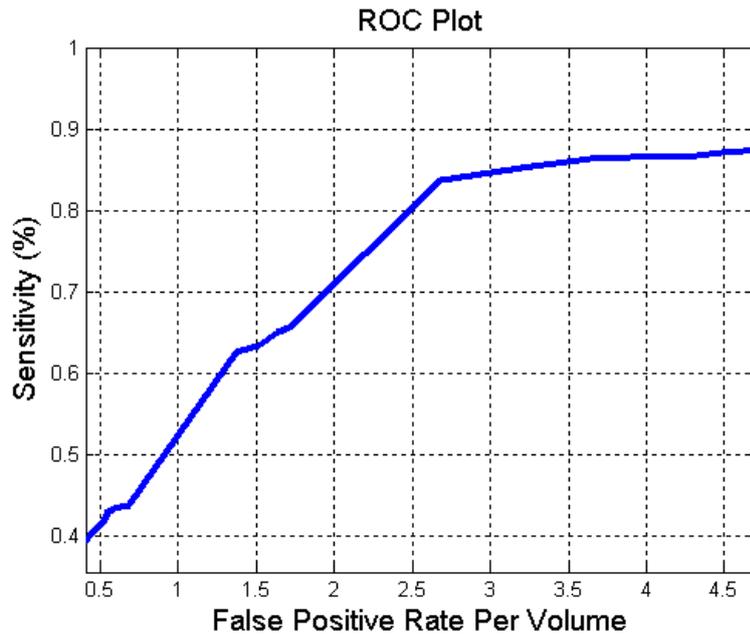


Fig. 2. ROC curve for the training set

5. Results

We tested our approach on a dataset of 185 patients that was randomly partitioned into two groups: training set (n=125) and test set (n=60). The training set was then used to calculate the internal parameters of the classification function.

Results on the training set: The total number of solid nodules in the training set was 250, with a minimum, maximum, average and median size of 3.0mm, 27.2mm, 4.8mm, and 3.9mm respectively. The sensitivity achieved on the training set was 80.7% with an average of 2.7 FP per volume. The test set was used for validating the approach. Results on the test set: A total number of 109 solid nodules, with a minimum, maximum, average and median size of 3.0mm, 17.8mm, 4.9mm, and 4.0mm respectively. The sensitivity obtained was 77.1% with an average of 2.7 FP per volume.

Examples of some detected and some missed nodules are given in Fig.3 and Fig. 4.

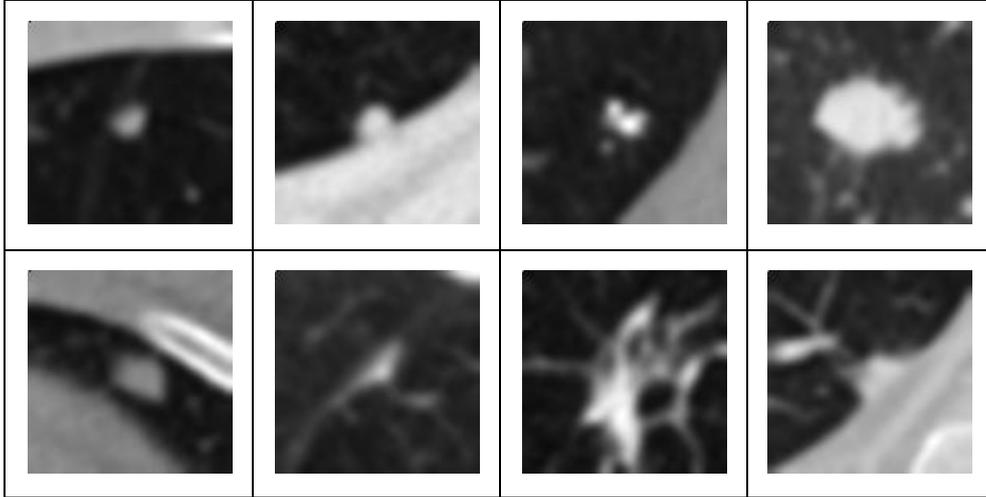


Fig. 3. Shown in the above figure are sample nodules that the algorithm successfully detects. The first row depicts some nodules that are relatively easy to detect because they have a spherical and more regular shape, while the second row shows nodules that are more difficult to detect.

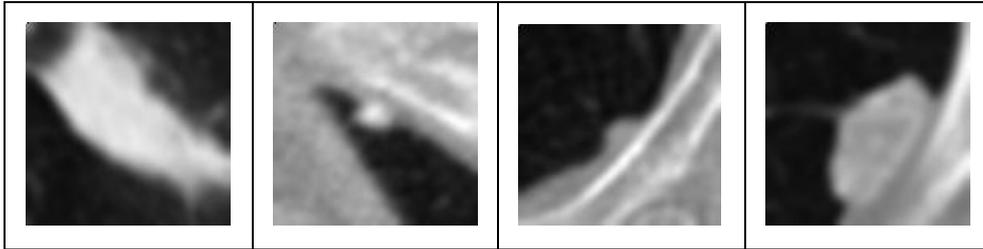


Fig. 4. Shown in the above figure are sample nodules that the algorithm fails to detect.

6. Conclusion

The CAD system developed focuses on the detection of solid pulmonary nodules of sizes from 3mm up to 15mm at a low number of false positives per volume. The results obtained on the training set and on the testing set demonstrate that the CAD algorithm generalizes very well and that the sensitivity for solid nodule detection is close to 80% at a low average of less than 3 false positives per volume.

References

- [1] M. Dunder, G. Fung, L. Bogoni, M. Macari, A. Megibow, B. Rao: A Methodology for Training and Validating a CAD System and Potential Pitfalls. CARS 2004
- [2] D. P. Naidich, J. P. Ko, J. Stoeckel et al.: Computer Aided Diagnosis: Impact on Nodule Detection Amongst Community Level Radiologist. A Multi-Reader Study. CARS 2004