Journal of Digital Imaging

Indexing Thoracic CT Reports Using a Preliminary Version of a Standardized Radiological Lexicon (RadLex)

Dirk Marwede, Thomas Schulz, and Thomas Kahn

Introduction: To validate a preliminary version of a radiological lexicon (RadLex) against terms found in thoracic CT reports and to index report content in RadLex term categories. Material and Methods: Terms from a random sample of 200 thoracic CT reports were extracted using a text processor and matched against RadLex. Report content was manually indexed by two radiologists in consensus in term categories of Anatomic Location, Finding, Modifier, Relationship, Image Quality, and Uncertainty. Descriptive statistics were used and differences between age groups and report types were tested for significance using Kruskal-Wallis and Mann-Whitney Test (significance level <0.05). Results: From 363 terms extracted, 304 (84%) were found and 59 (16%) were not found in RadLex. Report indexing showed a mean of 16.2 encoded items per report and 3.2 Finding per report. Term categories most frequently encoded were Modifier (1,030 of 3,244, 31.8%), Anatomic Location (813, 25.1%), Relationship (702, 21.6%) and Finding (638, 19.7%). Frequency of indexed items per report was higher in older age groups, but no significant difference was found between first study and follow up study reports. Frequency of distinct findings per report increased with patient age (p < 0.05). Conclusion: RadLex already covers most terms present in thoracic CT reports based on a small sample analysis from one institution. Applications for report encoding need to be developed to validate the lexicon against a larger sample of reports and address the issue of automatic relationship encoding.

KEY WORDS: Reporting, classification, chest CT, terminology

INTRODUCTION

T he Radiological Society of North America (RSNA) is developing a lexicon (RadLex) of standardized terms used in radiology.¹ The aim of the lexicon is to unify terms used in radiology and to facilitate indexing and retrieving of images and reports. Validation of the lexicon against terms used in radiological reports and the application of the lexicon to index radiological information sources is a necessary step to guarantee completeness of the lexicon and verify applicability of the lexicon in clinical practice.

In recent years, an increasing demand for standardized terminologies and reporting criteria is arising in radiology.^{2–8} Some lexical resources have been developed for radiology such as the Fleischner Glossary of terms used in thoracic imaging,^{9,10} the Breast imaging reporting and data system (BIRADS) classification system,¹¹ or the American College of Radiology (ACR) Index of Radiological Diagnoses. However, those lexicons represent only a small part of terms used in radiology and are not linked to other medical lexicons. The aim of defining terms in a hierarchical system and linking those terms to other electronic lexical resources is: 1. to use the lexicon for encoding of medical information, and 2. to facilitate the electronic processing of encoded medical information. The use of a standardized terminology in clinical applications brings the prospective of reducing communication

Copyright \bigcirc 2007 by Society for Imaging Informatics in Medicine

Online publication 28 July 2007 doi: 10.1007/s10278-007-9051-6

From the Department of Diagnostic and Interventional Radiology, University Leipzig, Liebigstr. 20, 04103, Leipzig, Germany.

Correspondence to: Dirk Marwede, Tel: +49-0341-9717400; Fax: +49-0341-9717409; e-mail: dirk.marwede@ medizin.uni-leipzig.de

errors in medicine and allowing structured collection and analysis of medical data.¹²

Few studies exist which evaluated terms contained in free-text radiological reports against medical lexicons.^{13–15} The purpose of this study was to validate a preliminary version of a radiological lexicon (RadLex) against terms found in thoracic computed tomography (CT) reports and to encode report content in RadLex term categories.

MATERIALS AND METHODS

This was a retrospective study of thoracic CT reports. Institutional review board approval was granted to retrospectively review reports. Two hundred fifty reports were extracted from the radiology information system of a single teaching hospital. Reports without three report sections (clinical history, main report section, conclusion section) were discarded (n = 50). Only one report per patient was included in the study. Reports had been composed by four board certified radiologists and seven residents (fourth and fifth year) using voice recognition software without use of macros or predefined report phrases. All examinations had been performed on a multislice scanner in 2004 using a standardized imaging protocol (100 mAs, 120 kV, contiguous 3-mm-thick images).

RadLex

RadLex is currently available as a preliminary online version¹⁶ and consists of 7,466 terms (including synonyms) grouped in nine major term categories: *Treatment*, *Image Acquisition/Processing/ Display*, *Modifier*, *Finding*, *Anatomic Location*, *Uncertainty*, *Teaching Attribute*, *Relationship*, and *Image Quality*.

The majority of terms is defined hierarchically in term categories of *Anatomic Location* (3,290 [terms]), *Finding* (1,925), and *Modifier* (762). For example, the category of *Anatomic Location* has subterms "abdomen", "lower extremity", "nervous system", "thorax", "trunk", "upper extremity", and "blood vessel". Those subterms have further subterms defined, for example, "thorax" has subterms like 'airways", "lung", and "mediastinum", which in turn have further subterms (Fig. 1). The term category of *Finding* contains terms describing processes and diagnoses frequently given in radiology reports. Subterms of *Finding* include "infectious or inflammatory disease", "body substance", "cardiovascular disease", "disorder caused by drugs or toxins", "foreign body", and "growth disorder", which have further subterms defined (Fig. 2). The category of *Modifier* defines the wide range of attributes used in radiology. Those modifiers are *finding modifier* (eg, subterm 'disc composition' with subterm modifiers 'cartilaginous', collagenous', desiccated', 'gaseous', liquefied', 'nuclear', and 'osseous') and other types of modifiers including *general modifier*, *anatomyspecific modifier*, *modality-related modifier*, *patient modifer* and *surgery modifier*.

RadLex defines eight main relationships to link terms used in radiology to each other. Those relationships are *member of* and the converse *member* (eg, 'liver' *member of* "set of viscera of abdomen"), *is a* (to link a child term to a parent term, eg, 'cardiac tamponade' *is a* 'heart disease'), *contains* and *contained in* (eg, 'thrombus' *contained in* 'pulmonary artery'), *part* and *part of* (eg, 'upper lobe of right lung' *part of* 'right lung'), and *continous with* (eg, 'ileum' *continous with* 'jejunum').

Term Matching

A text processor was used (Word Smith Tools Version 4.0) to extract single-word terms from the reports. From the resulting term list stop words were removed automatically (eg, the, a, at). The resulting term list was manually reviewed by two radiologists in consensus to identify single-word terms and multiple-word terms (eg, "anterior mediastinum"). Identification of multiple-word terms was based on correlation analysis between single-word terms. Terms were then matched against terms in RadLex categories of Modifier, Finding, Anatomic Location using the text processor. If a term had an exact match, the term category was noted. If no exact match was found, the different hierarchies were browsed manually to verify if a synonym could be found. Terms not found in the lexicon were classified in a RadLex term category.

Indexing of Reports

All reports were indexed manually by two radiologists in consensus to verify applicability of the lexicon to index report information. Manual

```
Anatomic Location
```

```
🛨 thorax
```

```
🛨 airway
```

```
🛨 diaphragm
```

- **H** diaphragmatic sulcus
- 🛨 lung
 - 🛨 left lung
 - right lung
 - + lower lobe of right lung
 - + middle lobe of right lung
 - + upper lobe of right lung
 - anterior segment of upper lobe of right lung
 - + apical segment of upper lobe of right lung
 - posterior segment of upper lobe of right lung
 - 🛨 lung zone
 - + small lung component
- 🛨 mediastinum
 - 🛨 heart
 - 🛨 cardiac chamber
 - + heart valve
 - + mvocardium
 - + pericardium
 - 🛨 mediastinal spaces
 - thymus
- 🛨 pleura
- + thoracic lymph nodes

Fig. 1. RadLex term category of Anatomic Location (extract).

indexing was performed to take account for encoding terms by decomposition and to detect information implicitly contained in the reports.

For indexing of terms in categories of Modifier, Finding, and Anatomic Location exact matches and term decomposition matches were allowed. For example, the term "mediastinal shift" has no direct match in RadLex, but can be decomposed in the terms "mediastinum" and "shifted", which are present in RadLex. Terms with a partial match were not indexed (for example, the term "miliary tuberculosis" is not contained and cannot be composed, but the term "tuberculosis" is contained in RadLex). If the exact quantity was not described (eg, "multiple nodules"), a single finding was recorded. Findings reported in the main report section and repeated in the conclusion section were noted only once. The number of distinct findings (term category Finding) per report and the number of modifiers (term category *Modifier*) used for the description of findings were quantified. Negation of findings and description of normal anatomical conditions were not noted.

Spatial relationships were indexed if a term from RadLex *Relationship* hierarchy was present (eg, *contains*, *located*, or *branch of*) or if an implicit relationship between terms from different RadLex term categories was present in the reports (eg, "hyperdensity in the upper mediastinum" or "subpleural nodule segment 6 of right lung"). In the latter case, a spatial relationship was noted, but no distinction was made between the different types of spatial relationships. *Image Quality* was evaluated on a two-point scale ("limited", "diagnostic") on the basis of reported reduced image quality (eg, "artifacts"). If no statement about reduced image quality was present, it was assumed to be "diagnostic". *Uncertainty* was evaluated on the + finding + body substance • body fluid ascites bile blood cerebrospinal fluid chyle fat tissue + cardiovascular disease • disorder caused by drugs or toxins • foreign body + growth factor + aplasia dysgenesis + dysplasia • ectopic tissue + enlargement distension goiter + hyperplasia hypertrophy lymphadenopathy splenic sequestration • fusion + growth rate disorder + iatrogenic disease or disorder the infectious or inflammatory disease • injury

+ physiologic or pathologic process

Fig. 2. RadLex term category of *Finding* (extract).

presence of qualifiers like "possibly", "probably", "suggestive for", "unlikely".

Statistical Analysis

Term-matching results and indexing results of reports were analyzed descriptively. Means and standard deviations were calculated for indexing results and percentages were reported for matching and indexing results. Indexing results were stratified according to patient age (age group I–V; <40, 40–49, 50–59, 60–69, and \geq 70) and study type (first study, follow-up) and tested for significant differences using Kruskal–Wallis and Mann– Whitney test. *P* values of 0.05 or less were considered statistically significant. Statistics were calculated using statistical software (SPSS for Windows, Version 11.0.1).

RESULTS

Eighty-two of 200 (41%) reports were firststudy reports and 118 (59%) were follow-ups. Mean age of patients was 54 years (range 6– 89 years), 121 (60%) were men and 79 (40%) were women. One hundred seven of 200 patients (54%) were inpatients, 71 (35%) were outpatients, and 11 patients (11%) were referred to by the emergency department.

Term Matching

Three hundred sixty-three distinct terms were extracted from the reports. Most terms were from category of *Finding* (137 of 363, 38%), followed by *Modifier* (126, 35%), and *Anatomic Location* (100, 27%, Table 1). Three hundred four of 363 (84%) terms were found in RadLex and 59 (16%) were not found. Matches across all term categories were between 78% and 90%, with highest percentage for terms from the category of *Finding*.

Indexing Results

During the indexing process, term categories were encoded 3,244 times (Table 1). Most terms found were from the category of *Modifier* (1,030 of 3,244, 31.8%) followed by the category of *Anatomic Location* (813, 25.1%) and *Relationship* (702, 21.6%). Terms from the category of *Finding* were found 638 times (19.7%) and a mean of 2.5 distinct findings per report was found.

Most anatomic locations were from the term hierarchy of "lungs" (509 of 813, 62.6%), "thoracic lymph nodes" (86, 10.6%) and "mediastinum" (63, 7.7%). Subterms of "lungs" most frequently encoded were "lung" (207, 25.5%), "segment of lung" (166, 20.4%), and "lobe of lung" (75, 9.2%). Subterms of "mediastinum" most frequent found were 'thyroid gland' (24, 3%), subterms of "mediastinal spaces" (25, 3%), and "heart" (14, 1.7%). Other frequently reported anatomic locations were "pleura" (38, 4.7%), "spine" (40, 4.9%), subterms of "blood vessel" (23, 2.8%), and "rib" (18, 2.2%). Most frequent findings reported were "fibrosis", "nodule", "pneumonia", "effusion", and "metastasis" (Table 2). Modifier most frequently found were from the term subcategories of Position (245 of 1,030, 23.8%), Temporal (192, 18.6%), Morphology (166, 16.1%), and Size (142, 13.8%) accounting for 72.3% (745 of 1.030) of all reported Modifiers. Other frequently encoded Modifier subcategories included Shape (56, 5.4%), Amount (54, 5.2%), Composition (33, 3.2%), and Extent (23, 2.2%). Most frequently found terms were "consolidated" (108, term subcategory Morphology), "stable" (72, Trend), "small" (66, Size), "enlarged" (51, Size), "dorsal" (45, Position), "basilar" (39, Position) and "apical" (36, Position). Terms from the subcategory of *Modifier* used for the description of terms form the category of Finding were found 794 times (1.3 [mean]). Findings most frequently described by *modifiers* were "fibrosis" (2 [mean], 0-5 [range]), "tumor" (2.1, 0-5), "pneumonia" (1.9, 0-6), and "metastasis" (1.8, 0-5).

The mean number of indexing results per age group for term categories of *Finding, Anatomic Location*, and *Relationship* was higher for older age groups (III–V) than for younger age groups (I–II, p < 0.05), but great variation of indexing results between different age groups was found. Lowest indexing results were found in age group I (11.9 [mean per report]) and highest in age group V (18.7), but there were no significant differences between all age groups (p = 0.88). Findings were more frequently in the older (III–V) than in the

Table 1. Automatic Term Matching and Hand-Curated Indexing Results

Term Category	Matching Results			Indexing Results			
	No. of terms found (%)	No. of terms not found (%)	Sum	No. of terms in RadLex	Indexed items	Term categories encoded (%)	Mean no. of indexed items per report
Anatomic							
location	83 (83%)	17 (17%)	100	3,290	813	25.1	4.1
Finding	123 (90%)	14 (10%)	137	1,925	638	19.7	3.2
Modifier	98 (78%)	28 (22%)	126	762	1,030	31.8	5.2
Relationship	_	-	_	6 ^a	702	21.6	3.5
Uncertainty	_	_	-	7	61	1.9	0.3
Sum	304 (84%)	59 (16%)	363	5,990	3,244	100	16.2

^aOnly spatial relationships

Table 2. Most Frequent Findings in the Reports, Number of Patients with Findings and Mean Number of Findings Per Patient

Findings	No. of Findings in the Reports	No. of Patients with Findings	Mean No. of Findings Per Patient (Range)	
Fibrosis	82	63	1.3 (1–4)	
Nodule	74	45	1.6 (1–8)	
Pneumonia	62	45	1.4 (1–3)	
Effusion	60	41	1.5 (1-4)	
Metastasis	43	26	1.7 (1–3)	
Degeneration	27	27	1	
Tumor	20	16	1.3 (1–3)	
Granuloma	17	13	1.3 (1–3)	
Tube/catheter	16	10	1.6 (1–4)	
Pneumothorax	11	11	1	
Postoperative				
change	11	8	1.4 (1–3)	
Interstitial				
pneumonia	11	8	1.4 (1–3)	
Calcification	8	7	1.1 (1–2)	
Tuberculosis	7	7	1	
Surgical clip	6	6	1	

younger age groups (I–II, p < 0.05), but no continuous increase from younger to older age groups was found (Fig. 3). Greatest variation of findings frequency was found in the category of *Modifiers* (age group II, 3.8 [mean]; age group III and V, 5.8). Frequency of distinct findings showed a stepwise increase from age group I to age group V, with significant differences between all age groups (p < 0.05). Follow-up study reports had more terms encoded per report (12.6 [mean]) than first study reports (12.4), but no significant differences were found.

DISCUSSION

The RSNA is developing a lexicon aiming at unifying terms in radiology. The lexicon is intended to be used for the indexing of images and reports and to facilitate the electronic processing of encoded medical information. The lexicon is continuously extended and, as the study was carried out, a new version has been published.¹⁶ In our study, we evaluated which quantity of terms contained in a random sample of thoracic CT reports was contained in a preliminary version of RadLex. Our results show that 84% of terms used in thoracic CT reports are contained in RadLex. Indexing report content showed that most terms contained in thoracic CT reports are from the category of *Modifier*. For indexing of report content, the preliminary version of the lexicon was complete enough to demonstrate a continuous increase of distinct findings from younger to older age groups.

Most distinct terms extracted from the reports were from the category of *Modifier*, which shows the wide range of terms even in a small sample size. Across all term categories of RadLex between 78% and 90% of terms from the reports were matched. Best results for matching terms were obtained in the category of *Finding* (90%) and *Anatomic Location* (83%). Based on our results from a small sample size, those categories seem to be the most complete. In contrast, 22% of terms from the term category of *Modifier* were not found, which indicates that for encoding of report

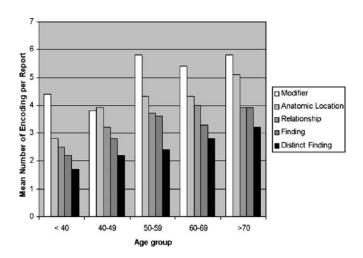


Fig. 3. Mean Number of Modifier, Anatomic Location, Relationship, Finding, and Distinct Finding per report according to age group.

content this term category needs to be extended. Terms not found include those frequently found in reports evaluating observations (eg, "normal", "pathological", "relapse").

Terms from the category of *Modifier* were most frequently found in the reports (mean 5.1/report) and within this term category most terms were from the subcategory of *Position*. In addition, spatial relationships were found frequently in the reports (mean of 3.5/report), which indicates that the description and location of features (eg, modifiers) and findings represent a substantial part of radiological report content.

Overall frequency analysis of findings, anatomic locations, and spatial relationships showed no stepwise increase between age groups. Frequency analysis of reported findings depends on the type of findings and the detail of findings reported. For example, we found that some findings were reported frequently more than once in the reports (eg, *nodule*, *metastasis*) and some only once (eg, degeneration). Furthermore, explicit information about the frequency of findings may be hidden (eg, multiple metastasis) and detail of the description of findings varied (eg, the use of modifiers for the description of findings). However, analysis of distinct findings reported showed a stepwise increase from younger to older age groups, which was recently reported for a screening population.¹⁷ We assume that frequency analysis of distinct findings in reports is the most accurate measure for the incidence of findings reducing the effect of individual reporting patterns of radiologists.

Results of prior studies have shown that existing medical lexicons do not represent terms used in clinical radiology sufficiently. For example, one study matched terms from chest x-ray reports against the Unified Medical Language System (UMLS) and found only 33% of terms.¹⁸ In another study, terms from ultrasound reports were matched against the UMLS, the Systemized Nomenclature of Medicine (SNOMED), and the ACR Index, but no lexicon contained more than 25% of terms.¹⁹ In a recent study, the completeness of the UMLS, ICD 10, and SNOMED were examined by matching terms used in different subdomains of clinical radiology, but none of the lexicons provided matches for more than 50% of terms examined.¹²

Some studies have examined radiology report content and classified terms used in radiological reports in term categories^{20–22}. One study has quantified terms extracted from reports into categories of *Rad-Finding*, *Body Part*, *Body Regions*, and *Modifiers* (eg, degree, change, status).¹⁸ Results are difficult to compare as reports examined were from different imaging examinations, and different term categories were used. However, for the term categories of *Rad-Findings* 26% (RadLex 20%) and *Modifier* 39% (RadLex 34% including *Uncertainty*) were found, which suggests similar results.

In a recent study, reports of patients who underwent whole-body CT screening were examined and a mean of 1.5 distinct thoracic findings per patient was found.¹⁷ Our indexing results showed a mean of 2.5 distinct thoracic findings. The authors of the study reported that most thoracic findings were found in the "lung" (59%) followed by the "mediastinum" (35%), which approximately was confirmed in our study ("lung" 62.6%, "mediastinum" 21.1% including blood vessel and thoracic lymph nodes subcategories). However, the type of findings reported in our study differed: most frequent findings were "fibrosis', "nodule", "effusion", and "pneumonia", in contrast to "parenchymal scars", "nodule", "granuloma" and "emphysema" in the cited study. Even if comparison of those results is limited because distinct terminologies were used for indexing, differences in frequency and types of findings are most likely caused by the different study populations.

Indexing results between different report types showed no significant differences, but those results are biased because reports not containing three separate report sections were excluded (eg, reports stating only "no change"). The results of our study show that according to RadLex term categories thoracic CT reports contain a mean of 3.2 findings per report from a random sample of patients referred to our department. In our opinion, this shows that the lexicon can be already applied in an experimental setting to index radiological reports. However, further analysis is necessary to prove applicability of the lexicon for standardized reporting tasks and indexing of teaching files.

One limitation of our study was that it was a retrospective analysis of radiological reports. Images were not reviewed for confirmation of reported findings or for gaining additional information in cases where textual reports were not explicit (eg, multiple nodules). Differences in reporting patterns between residents and board certified radiologists were not assessed. Another limitation of our study was the small sample size analyzing reports from only one institution. Report indexing was not performed automatically by a text processor as some report information was not explicit. In the matching procedure, we recognized only exact matches underestimating the completeness of the lexicon. During the indexing approach term decomposition was allowed, which demonstrated the power of the lexicon to index report information; however, comparison between those two approaches was not assessed. Another limitation of the study is that interobserver agreement for term identification and indexing results was not assessed to verify variability among radiologists.

CONCLUSIONS

RadLex covers 84% of terms present in thoracic CT reports based on a small sample analysis from one institution. We believe that further analysis of report content is necessary to learn more about common reporting and interpretation patterns by radiologists and relations between findings reported. Applications for indexing of report content need to be developed to validate the lexicon against a larger sample of reports and to address the issue of automatic relationship encoding.

REFERENCES

1. Langlotz CP: RadLex: a new method for indexing online educational materials. Radiographics 26(6):1595–7, 2006

2. Itai Y: "Peripheral washout" sign: terminology does not reflect the exact mechanism of enhancement. Radiology 197(1):317–9, 1995

3. Itai Y: CT findings of fulminant hepatitis: terminology and distribution of massive necrosis. Radiology 200(3):872, 1996

4. Kopans DB: Terminology and imaging sequencing in breast imaging. Radiology 196(2):579, 1995

5. Krinsky G: Terminology of hepatocellular nodules in cirrhosis: plea for consistency. Radiology 224(3):638, 2002

6. Nour SG: Standardization of terms and reporting criteria for image-guided tumor ablation. Radiology 232(2):626–7, 2004

7. Talner LB, Davidson AJ, Lebowitz RL et al: Acute pyelonephritis: can we agree on terminology? Radiology 192(2):297–305, 1994

8. Goldberg SN, Charboneau JW, Dodd GD, III et al.: Image-guided tumor ablation: proposal for standardization of terms and reporting criteria. Radiology 228(2):335–45, 2003

9. Austin JH, Muller NL, Friedman PJ et al.: Glossary of terms for CT of the lungs: recommendations of the Nomenclature Committee of the Fleischner Society. Radiology 200(2):327–31, 1996

10. Tuddenham WJ: Glossary of terms for thoracic radiology: recommendations of the Nomenclature Committee of the Fleischner Society. AJR Am J Roentgenol 143(3):509–17, 1984

11. Liberman L, Menell JH: Breast imaging reporting and data system (BI-RADS). Radiol Clin North Am 40(3):409–30, 2002

12. Langlotz CP, Caldwell SA: The completeness of existing lexicons for representing radiology report information. J Digit Imaging 15 Suppl 1:201–5, 2002

13. Friedman C: The UMLS coverage of clinical radiology. Proc Annu Symp Comput Appl Med Care:309–13, 1992

14. Hripcsak G, Friedman C, Alderson PO et al.: Unlocking clinical data from narrative reports: a study of natural language processing. Ann Intern Med 122(9):681–8, 1995

15. Huang Y, Lowe HJ, Klein D et al.: Improved identification of noun phrases in clinical radiology reports using a high-performance statistical natural language parser augmented with the UMLS specialist lexicon. J Am Med Inform Assoc 12(3):275–85, 2005

16. RadLex, Radiology Society of North America (RSNA). Available at http://www.radlex.com. Accessed January 2007

17. Furtado CD, Aguirre DA, Sirlin CB et al.: Whole-body CT screening: spectrum of findings and recommendations in 1192 patients. Radiology 237(2):385–94, 2005

18. Friedman C: The UMLS coverage of clinical radiology. Proc Annu Symp Comput Appl Med Care:309–13, 1992

19. Bell DS, Greenes RA: Evaluation of UltraSTAR: performance of a collaborative structured data entry system. Proc Annu Symp Comput Appl Med Care:216–22, 1994

20. Friedman C, Cimino JJ, Johnson SB: A schema for representing medical language applied to clinical radiology. J Am Med Inform Assoc 1(3):233–48, 1994

21. Ranum DL: Knowledge-based understanding of radiology text. Comput Methods Programs Biomed 30(2–3):209–15, 1989

22. Taira RK, Johnson DB, Bhushan V et al.: A conceptbased retrieval system for thoracic radiology. J Digit Imaging 9(1):25–36, 1996