



Published in final edited form as:

J Am Coll Radiol. 2019 November ; 16(11): 1587–1594. doi:10.1016/j.jacr.2019.04.026.

Natural Language Processing for Identification of Incidental Pulmonary Nodules in Radiology Reports

Stella K. Kang, M.D., M.Sc.^{1,2,*}, Kira Garry, M.P.H.², Ryan Chung, M.D.¹, William H. Moore, M.D.¹, Eduardo Iturrate, M.D.³, Jordan L. Swartz, M.D.⁴, Danny C. Kim, M.D.¹, Leora I. Horwitz, M.D., M.H.S.^{2,3}, Saul Blecker, M.D., M.H.S.^{2,3}

¹NYU Langone Health, Department of Radiology, 550 First Ave, NY, NY 10016

²NYU Langone Health, Department of Population Health, 227 East 30th St, NY, NY 10016

³NYU Langone Health, Department of Medicine, 550 First Ave, NY, NY 10016

⁴NYU Langone Health, Department of Emergency Medicine, 550 First Ave, NY, NY 10016

Abstract

Purpose: To develop natural language processing (NLP) to identify incidental lung nodules (ILN) in radiology reports for assessment of management recommendations.

Method and Materials: We searched the electronic health record for patients who underwent chest CT during 2014 and 2017, before and after implementation of a department-wide dictation macro of the Fleischner Society recommendations. We randomly selected 950 unstructured chest CT reports and reviewed manually for ILNs. An NLP tool was trained and validated against the manually reviewed set, for the task of automated detection of ILN with exclusion of previously known or definitively benign nodules. For ILN found in the training and validation sets, we assessed whether reported management recommendations agreed with Fleischner Society guidelines. The guideline concordance of management recommendations was compared between 2014 and 2017.

Results: The NLP tool identified ILN with sensitivity and specificity of 91.1% and 82.2%, respectively in the validation set. Positive and negative predictive values were 59.7% and 97.0%. In reports of ILN in the training and validation sets before versus after introduction of a Fleischner reporting macro, there was no difference in the proportion of reports with ILN [108/500 (21.6%) vs. 101/450 (22.4%); $p=0.8$], or in the proportion of reports with ILNs containing follow-up recommendations [75/108 (69.4%) vs. 80/101 (79.2%); $p=0.2$]. Rates of recommendation guideline concordance were not significantly different before and after implementation of the standardized macro [52/75 (69.3%) vs. 60/80 (75.0%); $p=0.43$].

Conclusion: NLP reliably automates identification of ILN in unstructured reports, pertinent to quality improvement efforts for ILN management.

Summary Statement:

*Corresponding Author: Stella Kang, M.D., M.Sc., Assistant Professor, Department of Radiology, Department of Population Health, NYU School of Medicine, 550 First Ave, NY, NY 10016, stella.kang@nyumc.org, Phone: (212) 263-0232, Fax: (212) 263-6634.

Natural language processing can automate searching of unstructured radiology reports for incidental pulmonary nodules, and therefore facilitate assessment of reporting quality.

Keywords

pulmonary nodule; incidental finding; natural language processing; quality improvement

Introduction

In the United States alone, there are an estimated 234,000 new cases of lung cancer each year, and lung cancer deaths represent an estimated 25.3% of all cancer deaths in the US (1). The rate of lung cancer deaths has steadily decreased over the past decade by about 2.7% per year, and detection of small malignant nodules via lung cancer screening may be contributing to this improvement, along with changes in the prevalence of smoking and evolving treatments (1–4). In addition to initiatives for screening and improved public awareness of modifiable risk factors for improved cancer control, there may be further opportunity for improved outcomes through the management of incidentally detected lung nodules (ILN).

An ILN can be defined as a newly identified nodule detected on cross-sectional imaging performed for reasons other than lung cancer screening or cancer follow-up: for example, a nodule detected in a patient without a malignancy undergoing a computed tomography (CT) pulmonary angiogram for a clinically suspected pulmonary embolism. The rate of ILN among chest imaging studies has been shown to be as high as 31%, or an approximate 1.5 million newly diagnosed nodules in the U.S. each year (5, 6). The high prevalence of lung nodules is one driver of the population health impact of appropriate management, as even the small percentage that are cancerous represent substantial absolute numbers. These concerns must be balanced with the potential for over-testing and over-treatment of benign nodules, which have also been described (7). Multiple societies including the American College of Chest Physicians (8) and the Fleischner Society (9) have created evidence-based guidelines for further evaluation of the ILN, based in part on trial data. Since radiologists describe ILN and apply these guidelines as they report recommendations, the radiology report may be a source of valuable data for epidemiological and outcomes studies (5, 8). However, the information contained in radiology reports is typically unstructured or semi-structured, containing free text descriptions that require time-intensive manual review.

Natural language processing (NLP) can potentially alleviate the burden of reviewing these reports by converting free text into a computer understandable form, which allows for automated sorting of large volumes of text for key terms (10, 11). One quality improvement application of NLP is the potential to capture instances of lesion discovery, assess the reporting and recommendations provided by the radiologists, and improve appropriate follow up of the lesions (7, 12). Referring providers frequently rely on the report to understand the degree of risk of a lesion, and clarity and guideline concordance of the recommendations can impact management decisions and ultimately, patient health outcomes. Despite the common discovery of the ILN, assessment of radiology reports for

lung nodule incidence and guideline concordance of recommendations is understudied. The purpose of this preparatory study for quality improvement efforts was twofold: 1) train and validate an NLP tool for detection of ILNs that would potentially require follow up evaluation, and 2) given the implementation of a departmental Fleischner Society guideline macro for reporting, assess the reviewed reports for pre- vs. post-implementation agreement of recommendations with Fleischner Society guidelines.

Methods and Materials

Data Source

This retrospective study was HIPAA compliant and approved by the institutional review board, waiving informed consent. All data were collected and analyzed from the electronic health record (EHR) system (Epic Systems Corporation, Verona, Wisconsin) at an urban tertiary care center. We searched the EHR for adult patients who had a CT scan including imaging of the lungs (Supplemental Appendix 1 for full list) between 1/1/2014 and 12/31/2014, prior to implementation of a department-wide macro for reporting Fleischner Society recommendations according to ILN size category, as well as between 6/1/2017 – 12/31/2017, the period immediately following the implementation of this ILN reporting macro.

Study Population and Datasets

We included all patients 35 years of age or older (to conform to the Fleischner Society Guidelines), who underwent a CT chest or thorax during the period of interest (Table 2). Abdominal CT scans were not included. Patients were excluded if they had 1) known primary cancer, 2) immunocompromised state on the basis of HIV positive status or treatment with medications of therapeutic class listed as “immunosuppressants” in the EHR including anti-neoplastics within 6 months before the CT scan; 3) a clinical indication of lung cancer screening as these reports are structured specifically according to American College of Radiology criteria and form a distinct population. Patients who underwent more than one CT during the time frame of interest were analyzed for the first scan describing a new lung nodule. A total of 28,540 reports met the selection criteria, including 13,516 reports from 2014 and 15,024 reports from 2017 (Figures 1a and 1b).

From the dataset, 500 patients’ reports from 2014 were randomly sampled. These reports were manually reviewed to establish a reference standard set for the purposes of training and testing the tool. A random sample of 450 patients’ reports from 2017 was also reviewed for creation of the validation set. An initial training set consisted of 40 patients’ reports from 2014 determined to describe lung nodules, including manual coding for whether or not the nodule was an incidental finding. These cases were manually coded by a board-certified radiologist (SK, with 6 years of experience), followed by a radiology resident who underwent a training session. Inter-reader reliability was calculated using simple kappa coefficients, interpreted as poor agreement when $K < 0.2$, slight agreement when $0.2 < K < 0.4$, fair agreement when $0.4 < K < 0.6$, and as substantial agreement when $K > 0.6$. With sufficient agreement ($K > 0.8$) reached using the final adjudicated training set, the remaining reports from the 2014 and 2017 datasets were

then manually coded by the resident, with all cases considered ambiguous for ILN reviewed by the same attending radiologist. Any cases with disagreement were resolved in consensus by this radiologist and another board-certified chest radiologist (WM, 15 years of experience).

For the purposes of identifying nodules that would potentially need follow up evaluation, an incidental pulmonary nodule was defined as a pulmonary nodule previously not described and without a clinical context of known malignancy, pneumonia, or small airways disease. A clinical indication in the report of lung cancer screening or mycobacterial infection, mycobacterium avium complex (MAC), mycobacterial avium intracellulare (MAI) or tuberculosis led to categorization of any nodules as non-incidental. Reported nodules were also not incidental if the description of the nodule included words such as “stable,” “unchanged,” or “minimal change” — or other similar phrases indicating the lesion had been first discovered on a prior CT scan. Nodules with definitively benign description on CT (e.g. benign patterns of calcification), or described in association with definitive features of inflammatory processes without equivocation (e.g. tree-in-bud or small airways disease, tubular configuration compatible with mucoid impaction, scarring, diffuse centrilobular nodules) were deemed benign and were not included.

Development of NLP Algorithm Rules

Simple Natural Language Processing, or “SimpleNLP,” is an open-source, rule-based natural language processing tool developed at our institution (<http://iturrate.com/simpleNLP/>) (10). The tool can be trained to detect any specific clinical entity or diagnosis, and relies on pattern matching to determine whether a condition of interest is reported as present or absent in a report. The application takes text as input from a file formatted as comma separated values (CSV) and the report narrative is stored in the first column of the CSV file, with one radiology report per row. Next, the user defines the rules for analysis by specifying arguments to the following parameters: target phrases, skip phrases, start phrases, absolute assertions, and negative absolute assertions.

Target phrases are the list of terms that describe the clinical diagnosis of interest, such as “nodule” or “groundglass nodule.” When the target phrase is found within a report sentence, the tool also checks to see whether the target phrase is negated using a list of predefined commonly used clinical negation phrases (e.g., “no nodule” or “no pulmonary nodule”). If the tool finds more sentences that contain the target phrase than negate the target phrase, the report is scored as “present;” otherwise it is scored as “absent.” In cases of a tie, the report is scored as present. Skip phrases are used to skip entire sentences in the report containing words that do not define the clinical entity and may confuse the tool. For example, “thyroid” and “breast” were included as skip phrases in our application to skip any sentences discussing nodules in other organ systems. Absolute assertions are phrases that are used to definitively mark the report as present or absent. If an absolute positive assertion is found in the report, the report is always marked as positive; similarly, if the absolute negative assertion phrase is found anywhere in the report, it will always be marked negative.

The research team, working with the chest radiologist (WM) noted the various phrases that radiologists use to describe and define lung nodules, and also identified the phrases that

should negate the presence of an ILN. These phrases were developed iteratively in several additional cycles using the 2014 set of patients' CT reports, broken into the initial set of 44 reports with lung nodules, followed by the remaining 456 reports with and without nodules for the training set. Skip phrases excluded cases of diffuse, geographic, or multifocal disease where the Fleischner Society guidelines do not apply (see Supplemental Appendix 2 for full list of applied phrases). In the tool training process, false positive and negative results revealed further skip phrases needed to eliminate non-incidental lung nodules (e.g. "unchanged nodule"). We evaluated the performance characteristics of the final version of the SimpleNLP tool, including sensitivity, specificity, positive predictive value, and negative predictive value in identifying ILNs in the training set, and then without further changes to the tool for the validation set.

Assessment of Recommendation Concordance with Guidelines in the Dataset

We then compared the contents of recommendations in the reports with the Fleischner Society guidelines to establish guideline-concordance according to the nodule size, and compared the proportion of reports that had guideline concordant ILN recommendations in 2014 vs. 2017. The Fleischner Society guidelines were updated in 2017 (9), and therefore, the previous guidelines were used to evaluate 2014 cases, while the updated guidelines were used to evaluate the 2017 cases. For the ILNs, we evaluated for the presence of specific high risk imaging features as described by the Fleischner Society and by the American College of Chest Physicians (8, 9). The published low risk follow-up recommendations apply to those patients with minimal or absent smoking history, and no other risk factors (e.g. radon, asbestos, uranium, lung cancer in 1st-degree relative). For the analysis, it was assumed high or low risk category information was not available for each patient at the time of interpretation, as is common in practice, and therefore in order to be a concordant recommendation, the report needed to contain recommendations for both high and low-risk patients in the applicable ILN size categories. The concordance of management recommendations was also assessed for ILN with versus without any high imaging risk features described. We analyzed recommendation concordance using a Chi-squared test and binomial exact confidence intervals. All analyses were performed using Microsoft Excel (Redmond, Washington). A p-value less than 0.05 was considered statistically significant.

Results

The study population, including both training and validation sets, consisted of 362 men and 588 women with mean age of 73 years (range 36–99 years). Most patients were white, non-Hispanic and either former or never smokers as reported in the EHR (Table 1). After all definitions, inclusions, and exclusions were defined, coders for the training dataset had substantial agreement as indicated by $K = 0.88$ before coding the remainder of the training/testing and validation sets. ILN type and morphologic features were mostly not specified in the radiology report text, and among those with specific descriptors in radiology reports, most were solid (Table 2). One hundred eight total incidental nodules were identified among the 500 randomly selected charts reviewed from 2014, while 101 incidental nodules were identified among the 450 charts reviewed in 2017.

Performance of NLP Tool for Incidental Lung Nodule Detection

The sensitivity and specificity of the NLP tool for identification of ILN in the 2014 training set were 90.5% (95% CI, 82.78% – 95.6%) and 86.1% (95% CI, 82.2% – 89.5%), respectively (Table 3). The positive and negative predictive values were 63.2% (95% CI, 56.9 – 69.2%) and 97.2% (95% CI, 94.9–98.5%), respectively. The sensitivity and specificity of the NLP tool for identification of ILN in the 2017 validation set were: 91.1% (95% CI, 83.8% - 95.8%) and 82.2% (95% CI, 77.8% – 86.1%), respectively, and positive and negative predictive values were also similar to those for the training set. False negative and positive cases largely resulted from incorrect or uncommon, ambiguous descriptors for nodules (appendix).

Assessment of Recommendation Concordance

We then explored the guideline concordance of the reported recommendations for ILN discovered in the 2014 and 2017 datasets (Table 4). In 2014, 69.4% of ILN (75/108) were accompanied by recommendations, compared with 79.2% of ILN (80/101) in 2017 ($p=0.13$). The proportion of recommendations that were provided using a macro was significantly lower in 2014 than in 2017, or 24.0% (18/75) vs. 47.5% (38/80), respectively ($p = 0.002$). However, there was no significant difference in the concordance of recommendations provided when examining various risk categorizations, including for ILNs that were described with intrinsically high risk features on imaging.

Discussion

We trained an open-source NLP tool developed at our institution for the task of analyzing chest/thorax CT reports for ILN, for which Fleischner Society guidelines for follow up evaluation are applicable. This NLP tool performed with excellent sensitivity and specificity of 91% and 82%, respectively, in the validation dataset and enabled exclusion of nodules that were described with definitively benign features that do not require follow up. The tool was developed to facilitate assessment of large-scale datasets before and after interventions on the quality of reporting, including guideline concordance of management recommendations. In this case, rates of guideline concordance were compared before and after the mid-2017 implementation of a reporting macro for ILN management recommendations according to Fleischner Society guidelines, which was encouraged for use by all departmental radiologists. The macro contents automatically transmit to our EHR as a trackable finding. However, leveraging this capability depends upon use of the radiologist using the macro.

Having established reasonable performance for these purposes, the NLP tool will be used to rapidly identify all potential ILNs in large datasets, assess whether or not the macro was used to provide a recommendation, and then facilitate monitoring of quality metrics of standardized macro use and guideline-concordance of ILN recommendations. When ILN are reported without recommendations for follow up despite warranting further imaging according to Fleischner Society guidelines, we also have a potential means of identifying such cases to communicate to providers the need for follow-up, and eventually track whether appropriate follow up was completed.

Few studies to date have evaluated the guideline concordance of lung nodule recommendations longitudinally. One prior study assessed agreement of recommendations for incidental lesions with society recommendations before and after an educational intervention with point-of-care resources made available, including reporting macros (13), but all imaging studies originated in the Emergency Department. Presenting an initiative to use department-wide point of care resources might be expected to increase the overall proportion of reports with ILNs containing recommendations, in contrast with our examination of an implemented macro only. With regard to previously developed NLP tools for the ILN, a previous method demonstrated automated detection of all types of lung nodules, without discrimination of definitively benign nodules as described in radiology reports (5, 14). Typically, such definitive imaging features would not result in further testing or treatment recommendations. We believe estimates of lung nodule prevalence including these definitively benign nodules would overestimate the economic, health-related, and psychological consequences of incidental lung nodules. Another strength of our tool is that we did not include nodules described in context of known immunosuppressed status, or known infection or small airways disease, or imaging findings that were of high likelihood to represent a geographic or diffuse airways process. Such nodules would not be evaluated according to the Fleischner Society guidelines, and may be followed or treated differently depending upon clinical signs or symptoms.

Our study had several limitations. Although the tool substantially reduces the manual chart review to identify incidental pulmonary nodules in radiology reports, some manual review is still required to determine whether nodules detected by the NLP are true cases of ILN, given its imperfect specificity. In addition, some nonspecific or ambiguous terminology in reports may have led to missed incidental nodules since no images were reviewed to establish the reference standard sets. We also did not evaluate CT scans for nodules that were not detected by radiologists. However, the purpose was not to aid diagnosis of nodules or lung nodule etiology, and the tool performance is sufficient to enhance but not entirely supplant manual process of reviewing semi-structured reports in which ILN may be described in a number of ways. Machine learning-based approaches could potentially improve performance, for example, by better separating inflammatory lesions from true ILN. NLP will facilitate longitudinal analyses of large numbers of reports, to gauge the rates of reported management recommendations and concordance of these recommendations with the Fleischner Society guidelines.

In conclusion, we trained an open source NLP tool that may be useful to other departments for automated searching of unstructured radiology reports for incidental pulmonary nodules, which excluded definitively benign imaging features and retained nodules likely to require follow up evaluation. Baseline analyses showed that the rates of providing specific management recommendations overall, and also for nodules with imaging features associated with elevated risk of cancer, did not yet improve in the time frame immediately following the implementation of a department-wide standardized macro for incidental lung nodules. Application of the SimpleNLP ILN tool to broader datasets of thoracic imaging reports will facilitate evaluation of potential longitudinal trends in recommendation concordance, identification of nodule or study types susceptible to suboptimal reporting, and study of downstream resource use and health outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Research reported in this publication was supported by the Agency for Healthcare Research and Quality (5P30HS024376-03). The findings and conclusions in this document are those of the authors, who are responsible for its content, and do not necessarily represent the views of AHRQ. Authors had full access to all of the data in this study and the authors take complete responsibility for the integrity of the data and the accuracy of the data analysis.

Stella Kang receives financial support from the National Cancer Institute/NIH for unrelated work, and royalties from Wolters Kluwer for unrelated work. William Moore receives personal fees as a consultant for unrelated work from Merck, Pfizer, and BTG. The remaining authors report no financial support for unrelated work.

References

1. Cancer Stat Facts: Lung and Bronchus Cancer 2018 [cited 2018 July 9]. Available from: <https://seer.cancer.gov/statfacts/html/lungb.html>.
2. National Lung Screening Trial Research T, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395–409. [PubMed: 21714641]
3. Moyer VA, Force USPST. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;160(5):330–8. [PubMed: 24378917]
4. Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. *Transl Lung Cancer Res*. 2016;5(3):288–300. [PubMed: 27413711]
5. Farjah F, Halgrim S, Buist DS, Gould MK, Zeliadt SB, Loggers ET, et al. An Automated Method for Identifying Individuals with a Lung Nodule Can Be Feasibly Implemented Across Health Systems. *EGEMS (Wash DC)*. 2016;4(1):1254.
6. Gould MK, Tang T, Liu IL, Lee J, Zheng C, Danforth KN, et al. Recent Trends in the Identification of Incidental Pulmonary Nodules. *Am J Respir Crit Care Med*. 2015;192(10):1208–14. [PubMed: 26214244]
7. Wiener RS, Gould MK, Slatore CG, Fincke BG, Schwartz LM, Woloshin S. Resource use and guideline concordance in evaluation of pulmonary nodules for cancer: too much and too little care. *JAMA Intern Med*. 2014;174(6):871–80. [PubMed: 24710850]
8. Gould MK, Donington J, Lynch WR, Mazzone PJ, Midhun DE, Naidich DP, et al. Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*. 2013;143(5 Suppl):e93S–e120S. [PubMed: 23649456]
9. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. *Radiology*. 2017;284(1):228–43. [PubMed: 28240562]
10. Swartz J, Koziatsek C, Theobald J, Smith S, Iturrate E. Creation of a simple natural language processing tool to support an imaging utilization quality dashboard. *Int J Med Inform*. 2017;101:93–9. [PubMed: 28347453]
11. Pons E, Braun LM, Hunink MG, Kors JA. Natural Language Processing in Radiology: A Systematic Review. *Radiology*. 2016;279(2):329–43. [PubMed: 27089187]
12. Wadia R, Akgun K, Brandt C, Fenton B, Levin W, Marple A, et al. Comparison of Natural Language Processing and Manual Coding for the Identification of Cross-Sectional Imaging Reports Suspicious for Lung Cancer. *JCO Clinical Cancer Informatics*. 2018;2:1–7.
13. Zygmunt ME, Shekhani H, Kerchberger JM, Johnson JO, Hanna TN. Point-of-Care Reference Materials Increase Practice Compliance With Societal Guidelines for Incidental Findings in Emergency Imaging. *J Am Coll Radiol*. 2016;13(12 Pt A):1494–500. [PubMed: 27652569]
14. Danforth KN, Early MI, Ngan S, Kosco AE, Zheng C, Gould MK. Automated identification of patients with pulmonary nodules in an integrated health system using administrative health plan

data, radiology reports, and natural language processing. *J Thorac Oncol.* 2012;7(8):1257–62.
[PubMed: 22627647]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Take Home Points:

1. SimpleNLP is an open-source, rule-based natural language processing tool that was trained to detect lung nodules and exclude nodules with definitively benign features as compared with a reference standard set of 950 cases.
2. In the validation set, natural language processing identified ILN with sensitivity and specificity of 91.1% (95% CI, 83.8% - 95.8%) and 82.2% (95% CI, 77.8% – 86.1%), respectively.
3. Before and immediately after implementation of a dictation macro for the Fleischner Society Guidelines, there was no significant difference in the proportion of reports with ILNs containing follow-up recommendations [75/108 (69.4%) vs. 80/101 (79.2%); p=0.2].

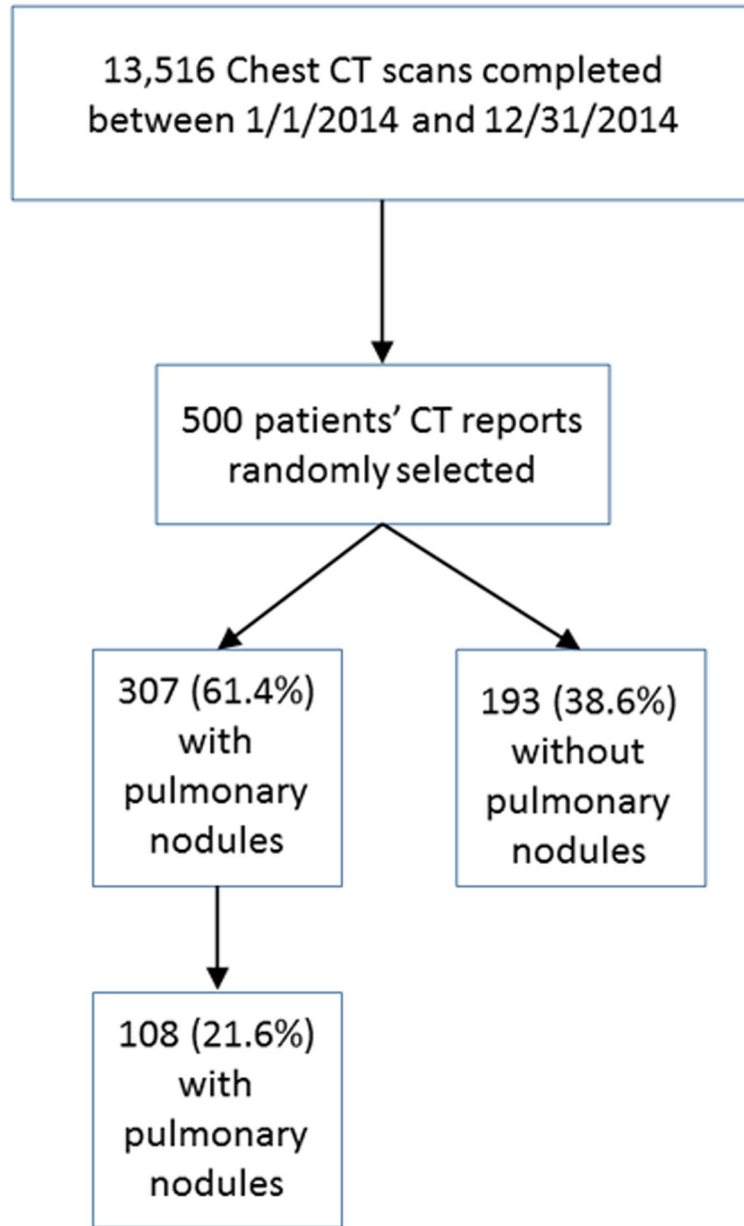


Figure 1a. Selection of cases from 2014 dataset of CT scans of the thorax. Presence of pulmonary nodules and incidental pulmonary nodules were determined by manual review to establish a reference standard set.

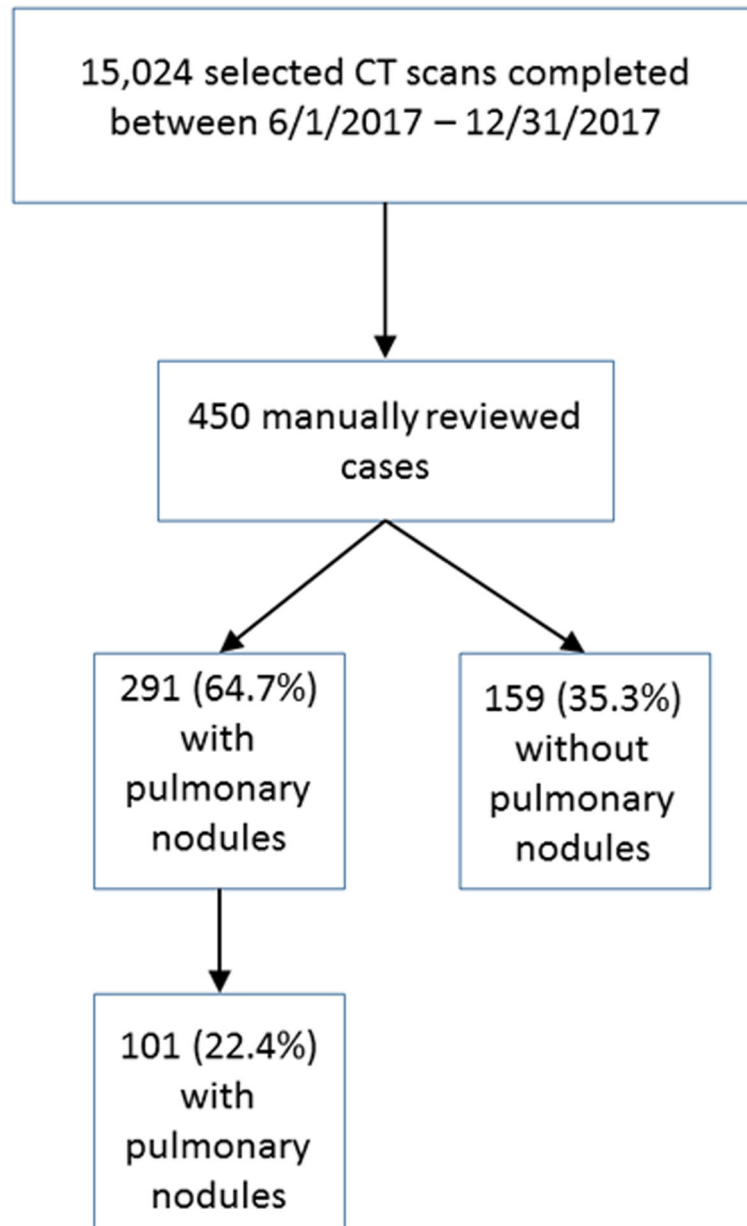


Figure 1b.

Selection of cases from 2017 dataset of CT scans of the thorax. Presence of pulmonary nodules and incidental pulmonary nodules were determined by manual review to establish a reference standard set.

Table 1.

Patient demographics at the time of index CT scan (n=950 patients).

Characteristic	Number of Encounters
General Demographics	
Mean age in years (range)	73 (36–99)
Sex	
Men	362 (38.1%)
Women	588 (61.9%)
Race	
African American	35 (3.7%)
Asian	23 (2.4%)
White	803 (84.5%)
Native Hawaiian or Other Pacific Islander	2 (0.2%)
Other	52 (5.5%)
Unknown	35 (3.7%)
Ethnicity	
Not of Spanish/Hispanic Origin	586 (61.7%)
Spanish/Hispanic Origin	21 (2.2%)
Unknown	343 (36.1%)
Smoking Status*	
Current every day/ Heavy smoker	42 (4.4%)
Current some day/ Light smoker	18 (1.9%)
Former smoker	473 (49.8%)
Smoker, current status unknown	1 (0.1%)
Passive smoke exposure	4 (0.4%)
Never smoker	350 (36.8%)
Unknown	62 (6.5%)

* Represented categories are the available choices in the electronic health record for specifying smoking history

Table 2.

Nodule Characteristics as specified in radiology report text (n=950 patients), determined using manual review of reference datasets.

Nodule Features	2014 Dataset (total n =108)	2017 Dataset (n=101)
<i>Nodule Size</i>		
Mean nodule diameter (range)	7.8 mm (2 – 52 mm)	8.2 mm (2 – 36 mm)
<i>Nodule Morphology/Type, count</i>		
Solid	6	0
Spiculated	9	4
Ground-Glass	12	23
Not Specified	81	72
<i>Single vs. Multiple Nodules (count)</i>		
Multiple *	52 (48%)	60 (59%)
Single	56 (52%)	41 (41%)

* Non-significant difference in proportion of multiple vs. solitary nodules between 2014 and 2017, p = 0.11.

Table 3.

Performance of NLP in Training and Validation sets of data.

Training Set	Value	95% CI
Sensitivity	90.53% (86/95)	82.78% – 95.58%
Specificity	86.15% (311/361)	82.15% – 89.54%
Positive Predictive Value	63.24% (86/136)	56.88% – 69.16%
Negative Predictive Value	97.19% (311/320)	94.88% – 98.47%
Validation Set		
Sensitivity	91.09% (92/101)	83.76% – 95.84%
Specificity	82.23% (287/349)	77.81% – 86.10%
Positive Predictive Value	59.74% (92/154)	54.01% – 65.21%
Negative Predictive Value	96.96% (287/296)	94.46% – 98.35%

CI = Confidence Interval CI = Confidence Interval

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4.

Concordance of Report recommendations with Fleischner Society guidelines in 2014 cases and 2017 validation set cases (as determined by manual review).

Comparison between 2014 and 2017 reports					
	2014		2017		p value
	n/total	Percentage (95% CI)	n/total	percentage	
Reports with ILN	108/500	21.6% (18.0 – 25.2%)	101/450	22.4% (18.7 – 26.4%)	0.754
Reports with ILN and follow-up recommendation	75/108	69.4% (60.2 – 77.8%)	80/101	79.2% (71.3 – 87.1%)	0.107
Reports with ILN and follow-up recommendation using a macro*	18/75	24.0% (14.7 – 33.3%)	38/80	47.5% (36.3 – 58.8%)	0.002
Reports with ILN and guideline concordant follow up recommendation	53/75	70.7% (60.0 – 80.0%)	60/80	75.0% (65.0 – 83.8%)	0.544
ILN Warranting Both Low and High Risk Fleischner Society Recommendations					
ILN in category warranting both low and high clinical risk recommendations (if/then statement)	39/75	52.0% (41.3 – 62.7%)	39/80	48.8% (37.5 – 60.0%)	0.686
Report reflected guideline concordant recommendation	18/39	46.2% (30.8 – 61.5%)	24/39	61.5% (46.2 – 76.9%)	0.173
ILN with Imaging Features Associated with Elevated Risk for Malignancy[#]					
ILN with imaging features suggestive of elevated risk	45/75	60.0% (49.3 – 70.7%)	39/80	48.8% (37.5 – 60.0%)	0.160
Report reflected guideline concordant recommendation	37/45	82.2% (71.1 – 93.3%)	30/39	76.9% (64.1 – 89.7%)	0.547
ILN without Imaging Features Associated with Elevated Risk for Malignancy					
ILN without imaging features suggestive of elevated risk	30/75	40.0% (29.3 – 50.7%)	41/80	51.3% (40.0 – 62.5%)	0.160
Report reflected guideline concordant recommendation	16/30	55.1% (36.7 – 70.0%)	30/41	73.2% (58.5 – 85.4%)	0.084

ILN = Incidental lung nodules

* In 2014, individual radiologists may have used personal macros while in 2017 a department-wide standardized macro was implemented.

[#] elevated risk features according to American College of Chest Physicians, Fleischner Society.