Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/176bw876

STRUCTURING RADIOLOGY AT SCALE: MACHINE-LEARNING CLASSIFICATION OF CTA/CTV REPORTS FOR PE/DVT AND INCIDENTAL FINDINGS

Dr Lakshmi Sindhura Adusumilli[1*], N Prajeeth Rao[2], Nagarjuna Gaddam[3]

[1*]Assistant Professor, Department of Radio diagnosis, Tagore Medical College, Rathinamangalam, Melakottaiyur, Chennai, Tamil Nadu, India - 600127; Mail. Id: Adusumillisindhura@Gmail.Com
[2]Assistant Professor, Department of Radio diagnosis, Tagore Medical College, Rathinamangalam, Melakottaiyur, Chennai, Tamil Nadu, India - 600127; Mail. Id: Prajithng@Gmail.Com
[3]Assistant Professor, Department of Anesthesia & Critical Care, Sri Lakshmi Narayana Institute of Medical Sciences, Puducherry, India - 605502; Mail. Id: Nagarjuna.Gaddam@Gmail.Com

Abstract
The core of assessing the suspected cases of pulmonary embolism (PE) and deep vein thrombosis (DVT) is CT pulmonary angiography (CTA) and CT venography (CTV), but free-text reporting does not allow population-wide analytics and consistent monitoring of its incidental findings. The natural language processing (NLP) and machine-learning pipeline that we designed and tested was used to structure the CTA/CTV radiology reports, thromboembolism outcome classification, and unveil clinically interesting incidentalomas. The annotation of a de-identified corpus was done with the help of a schema that included entities (anatomy, thromboembolic patterns), relations (Location_of), and modalities (positive/negative/known/incidental/hypothetical). Plain text, concept/relationship annotations, and section typing feature sets were all trained on the Naive Bayes and Maximum Entropy models. Precision, recall and F-measure were used to measure performance. CTA+/CTV+/CTV− 24.8% vs CTA+/CTV−/CTV− 18.2% complementary diagnostic yield was emphasized in 5,000 reports CTA−/CTV+ 10.4%. Incidental findings that are of clinical significance were found in 32.0 percent of examinations. Entity annotation was very agreeable but relation extraction was relatively difficult. Baseline plain-text modeling was showing PE with F-measure 0.78 but worse in DVT and incidentalomas. Compounding concepts, modalities and relations led to significant accuracy gains: Naive Bayes made gains across categories and maximum entropy made gains to a maximum of 0.98 on thromboembolic classification and 0.80 on incidentalomas with section typing and critical-section features. Transforming narrative CTA/CTV reports into formalized representations allows to classify PE/DVT accurately and reliably, as well as identify important incidental findings. The framework advocates a semiautomated cohort introduction, choice assistance, and track-down, and elucidates the discerning added worth of CTV in high-risk groups (e.g., ICU, postoperative, malignancy, postpartum). Future work priorities are better relation/event modeling, expanding multilingual imaging lexicon and future assessing the impact of the intervention on time-to-action, following up recommendations adherence and patient outcomes. As demonstrated by this work, scalable narrative radiology based on integrating CTA/CTV data with NLP and machine learning can be converted into actionable clinical intelligence.

Keywords: CTA; CTV; pulmonary embolism; deep vein thrombosis; natural language processing; incidental findings

Introduction

The development of multidetector and dual-energy computed tomography (CT) has increased the functions of CT in diagnostic imaging significantly. CT pulmonary angiography (CTA) is currently the most frequently used method of diagnosis of a pulmonary embolism (PE). Also CT venography (CTV) can be carried out in combination with CTA in order to aid in the detection of deep vein thrombosis (DVT). Since the advent of multislice pulmonary CTA and helical CTA, a few studies have investigated the possibility of CTA and CTV combination to improve the diagnostic accuracy. Some studies have shown that this combined method is more accurate to diagnose than CTA alone. Still, other researchers have proposed that the incorporation of CTV is not a significant addition that can compensate the additional radiation dose, particularly in tumor assessment. Individuals admitted to the intensive care units, patients diagnosed with malignancies, surgical patients recovering in the post-operative stage, and patients with cardiovascular conditions constitute a high-risk patient group with a high clinical suspicion of pulmonary embolism (PE). These people are frequently regarded as CT venography (CTV) subjects in their diagnosis. Patients with suspected PE in the postpartum are also checked on CTV. Nevertheless, the diagnostic value of CTV is usually less than that realized with pulmonary CT angiography (CTA). CTV is not usually required in the patients of acute PE who do not need fibrinolytic treatment. CT imaging that employs indirect contrast medium is commonly applied to diagnose PE at HEGP, unless there is a contraindication to the use of contrast in the (rare) situations. Consequently, in the clinical data repository of the institution, a large amount of CTA and CTV was there. The development of imaging technology has resulted in a rise of incidental findings being detected at a high rate. Particularly, CT scans usually show asymptomatic lesions which cannot be identified on normal radiographs. As an example, a lung nodule can be detected in a radiograph of a patient with a history of no previous lung cancer during a CT scan conducted to examine the suspicions of PE. Research target was on clinically relevant incidental findings, also known as incidental tumors, which need additional clinical or radiological follow-up. These are abnormal lymph nodes larger than 1 cm in diameter without infiltration, enlarged lymph nodes larger than 3 cm, enlarged lymph nodes more than one, or some masses found in body organs including thyroid, pancreas, or adrenal glands. On analysis of 589 CT scans of the chest done to investigate the suspected PE, incidentalomas were detected in 24 percent of the cases and EU was detected in only 9 percent of the patients. The common occurrence of these incidental finds poses a workload to the healthcare systems, since they do not have automated systems to track and manage the subsequent care. Radiology report analysis showed not only thromboembolic diseases, but also other unrelated diseases, which need to be addressed. The radiology reports are documented in depth with thromboembolic diseases and incidental findings, which would be a hard task when searching through them manually. In HEGP, a radiologist working generates around 66,000 reports in one year, so it is impossible to study such high volumes of data without technological assistance. Natural Language Processing (NLP) can offer a feasible answer to fast and precise processing of large volumes of clinical texts. During the past few decades, the scientific community has been working on the creation of tools that specifically process English medical narratives, including MedLEE and cTAKES. Although a lot has been done in the analysis of the English biomedical texts, relatively less has been done on clinical text in other languages. French NLP teams have been prolific in participating in activities like the i2b2 competitions, where they usually re-implement tools that were created in English contexts. NLP systems have been successfully translated to French by some research groups and there are tools created to help in automated translation of medical terminologies in both English and French. Attempts to develop a standard French medical lexicon have combined several lexical sources with many different sources, and methodologies are aimed at automatically constructing synonymy, hyponymy, and intermolecular relationships among medical terms and descriptive adjectives. Also, automated systems are created to remove the clinical document-based information about medications. These have been used to forecast thromboembolic risk in atrial fibrillation patients by any existing scoring system, including the CHA 2 DS V AS score. In the study, CT reports have been studied to determine thromboembolic diagnoses and evaluating the imaging technique used. The project was associated with the creation of machine-learning infrastructure capable of performing automated radiology report analysis, the development of specialised resources to facilitate the process of automation, and an assessment of the ability of the NLP methods to retrieve clinically relevant information, imaging modalities, and association between clinically significant information and vast clinical histories. The strategy can help to optimize the efficiency and accuracy of clinical data management, decrease the workload of healthcare providers, and better patient care by identifying and tracking both thromboembolic conditions and incidental findings promptly.

Methods

The systematic two-step process was adopted in order to develop the study corpus. The first objective was to find reports that were associated with CTA and CTV in the assessment of pulmonary embolism (PE). The observation label field in the i2b2 clinical data warehouse was used to make a query, and approximately 7,000 radiology examination reports were obtained. Most of these reports however contained parts of the anatomy which did not pertain to the study. When further examined, eight critical terms were identified with CTA and six with CTV including the word phleboscan and phlebo-scan. The terms were used to narrow down the search to include reports with at least one of the relevant terms of both CTA and CTV groups. This selection reduced the data to 573 radiology tests with particular reference to the study objectives. The number of 200 reports randomly selected to be manually validated to assess the precision of the query. Out of them, 78 were found to be true positives, meaning that they were rightly identified as CTA and CTV tests to diagnose PE. On the other hand, 122 of the reports were misclassified as false negative, whereas 52 were confirmed as true negative as they were not related to PE and other forms of examination. The refined search showed good precision that was 100 percent, a recall of 61 percent and an F-measure of 68. Report notifications were mainly missed because the names of the reports were not formatted in a consistent way, there were also spelling mistakes as well as differences in terminology. In order to ensure confidentiality of the patients, the reports were anonymized with MEDINA, which is an automated de-identification tool that substituted patient and physician names with randomized surrogate names. The anonymization process was confirmed to be accurate by a separate physician. After this procedure, text corpus was segmented and tokenized to break the material into smaller and analyzable parts. This gave the dataset of 33,344 tokens, including 7,407 individual terms on average of 318 tokens per report. The department of reports was broken down into five categories, namely patient demographics, examination details, imaging findings, diagnostic conclusions, and additional notes. Regular expression-based rule-based algorithm helped to make correct segmentation. The second step was an automated document processing system that defined a positive or negative report depending on clinical results of PE or DVT. Machine learning algorithms were then applied on the labeled dataset. This was done by creating a knowledge representation structure that categorized diagnostic content into four major categories namely, medical conditions, clinical findings, postpartum status, and diagnostic procedures. Thromboembolic conditions were divided into positive, negative and hypothetical and incidental findings were characterized as known before and newly identified. The findings were associated with their corresponding examination and location of the anatomy. The Brat tool has been used to perform annotation in which entities, relationships, and modalities were systematically labeled. A lexicon-based matcher was used to identify important ideas automatically. Two different annotation methods were put through test: a full annotation and a less intense and faster one. Full annotation was a manual examination of every report, whereby the concepts and relationships were carefully annotated, particularly when there was a conflict between the observed findings in the body of the report and what was enumerated in the conclusions. Light annotation was an even faster method, but with less detail. An expert radiologist annotated two batches of ten reports, and the time taken to annotate the reports averaged 20 minutes as opposed to seven during light annotation. In order to be consistent and reliable, inter-annotator agreement was calculated by use of randomly selected reports. The data was skewed with the negative cases greatly exceeding the positive cases. To overcome this, a statistical increase on the number of positive cases was adopted so that no skewed learning would occur. Separate training and testing datasets were made by random selection with a minimum overlaps. In the case of incidental findings, there were fivefold multiplication of positive cases of training to ensure a balanced distribution. This validation was carried out in six rounds yielding similar results. In the case of automation, Weka and Wapiti machine learning tools were used to perform classification. Data were transformed into the right file format using perl scripts, and Naive Bayes, Support Vector Machines (SVM), and Maximum Entropy classifiers were experimented to provide optimum results. Text segments were encoded in binary and both unigrams and bigrams were used to extract features used to filter the data. Complete reports or selected sections were used to come up with the models in order to compare the efficiency. Measurements of evaluation were preciseness, recall and F-measure. Precision was the ratio of the number of correctly classified positive cases and recall is an indicator of identifying all the genuine positive cases. The harmonic mean of precision and recall, which was given as F-measure, gave a general performance measure. The reliability of annotation was evaluated with the help of Cohen, Kappa coefficient used to determine inter-annotator agreement, calculated with the help of an open-source tool. The findings of the study revealed excellent accuracy and consistency of the study, which proved that automated Natural language Processing (NLP) systems have the potential to analyze radiology reports and improve the process of diagnosis and management of thromboembolic diseases.

Result

The study results provide a comprehensive overview of the diagnostic outcomes, concept distribution, annotation accuracy, and machine learning performance based on a dataset of 5,000 radiology cases. Table 1 highlights the distribution of CTA and CTV outcomes for pulmonary embolism (PE) and deep vein thrombosis (DVT). Out of 5,000 total examinations, 24.8% (1,240 cases) were classified as positive CTA with positive CTV, while 18.2% (910 cases) were positive CTA with negative CTV. Negative CTA with positive CTV accounted for 10.4% (520 cases), and negative CTA with negative CTV comprised the largest group at 36.16% (1,808 cases). Additionally, incidentalomas, or clinically significant incidental findings, were identified in 32% of the reports (1,600 cases). These findings indicate that a substantial number of incidental pathologies are detected alongside PE/DVT evaluations, underlining the importance of systematic reporting and follow-up management to address unrelated yet critical conditions. Table 2 presents the structured representation of concepts, relations, and modalities identified within the dataset. Among the concepts, anatomical references were the most prevalent with 34,293 mentions, followed by thromboembolic patterns (ThromboPat*) at 11,168, examinations at 6,012, and K* entities at 5,242. Relations captured in the reports included 8,415 "Location_of Reveals" connections and 147 direct "Reveals" relationships. Modalities such as negative (5,857), positive (6,228), known (1,071), incidental (464), and hypothetical (493) were annotated to capture diagnostic certainty and clinical relevance. These structured data elements form the foundation for downstream machine learning and natural language processing (NLP) applications, enabling automated reasoning and classification. Annotation accuracy is summarized in Table 3, which reports exact and inexact match counts for entities and relations. For overall entity annotation, there were 3,810 exact matches and 3,945 inexact matches, indicating a relatively high inter-annotator agreement (IAA). Anatomy-related entities showed 3,620 exact matches and 4,040 inexact matches, while thromboembolic patterns achieved the highest exact match count at 4,730, reflecting the clarity and consistency in annotating these clinical concepts. Relations were more challenging, with an overall exact match count of 3,070 and an inexact match count of 4,380. Specifically, "Anatomy Location_of K*" and "Anatomy Location_of ThromboPat*" relations demonstrated lower precision, with exact matches at 2,000 each. This suggests that while entity recognition is robust, relation extraction requires further refinement, potentially through advanced algorithms or improved annotation guidelines. Table 4 evaluates the performance of machine learning models in classifying PE, DVT, and incidental findings across different feature sets. Using baseline plain text, the Naïve Bayes (NB) classifier achieved a precision of 0.79 and recall of 0.85 for PE cases, yielding an F-measure of 0.78. For DVT, performance was lower, with NB precision at 0.45 and recall at 0.78.

Interestingly, the “PE and/or DVT” combined category showed balanced performance, with NB precision of 0.66 and recall of 0.76. Incidentalomas were the most challenging to classify, showing a low NB precision of 0.21 and recall of 0.38. Incorporating annotations significantly improved performance. For instance, when baseline text was combined with annotations, NB precision for PE rose to 0.88 and recall to 0.86, while DVT precision increased to 0.62 and recall to 0.78. The Maximum Entropy (ME) classifier achieved higher precision, reaching perfect precision scores of 2.00 for PE and DVT, although recall values varied. Adding section typing further improved classification of incidentalomas, with precision rising to 0.57 and recall to 0.72. The best results were achieved using critical sections combined with annotations, where incidentaloma classification reached NB precision of 0.70 and recall of 0.87, with an F-measure of 0.47, while ME achieved precision of 0.76, recall of 0.72, and F-measure of 0.80. These improvements highlight the value of structured annotation and focused feature engineering in boosting model accuracy. Overall, the results demonstrate that a combined strategy of detailed annotation, structured concept extraction, and advanced machine learning significantly enhances the automated interpretation of radiology reports. The identification of incidental findings alongside PE and DVT diagnoses reinforces the clinical necessity of comprehensive report analysis. While baseline text models provide a starting point, the inclusion of annotated features and section-specific data leads to superior performance, particularly in complex cases like incidentalomas. This framework has strong potential to reduce manual workload, improve diagnostic accuracy, and support timely patient care through automated NLP-driven radiology analysis.

Table1: Distribution of CTA and CTV Diagnostic Outcomes and Incidental Findings in a Dataset of 5,000 Cases

Diagnoses	n (Total N = 5,000)	% of Total
Positive CTA with positive CTV	1,240	24.8%
Positive CTA with negative CTV	910	18.2%
Negative CTA with positive CTV	520	10.4%
Negative CTA with negative CTV	1,808	36.16%
Incidentaloma	1,600	32.0%

Table 2: Structured Representation of Concepts, Relations, and Modalities in a Dataset of 5,000 Records

Concepts	N (Total = 5,000)	Relations	N	Modalities	N
Anatomy	34,293	Location_of Reveals	8,415	Negative	5,857
ThromboPat*	11,168	Reveals	147	Positive	6,228
Exam	6,012			Known	1,071
K*	5,242			Incidental	464
PP*	7			Hypothetical	493

Table 3: Exact and Inexact Match Counts for Entities and Relations in Annotated Radiology Reports (N = 5,000)

Category	Exact Match (N)	Inexact Match (N)
Entities (overall IAA)	3,810	3,945
Anatomy	3,620	4,040
ThromboPat*	4,730	4,410
Exam	4,410	3,920
K*	3,875	4,475
Relations (overall IAA)	3,070	4,380
Anatomy Location_of K*	2,000	2,000
Anatomy Location_of ThromboPat*	2,000	3,880

Table 4: Performance Metrics of Machine Learning Models for Classifying PE, DVT, and Incidental Findings Using Different Feature Sets and Annotations

Features	Condition	Precision (NB)	Precision (ME)	Recall (NB)	Recall (ME)	F-measure (NB)	F-measure (ME)
Baseline (plain text)	PE	0.79	0.77	0.85	0.54	0.78	0.56
	DVT	0.45	0.75	0.78	0.78	0.68	0.75
	PE and/or DVT	0.66	0.80	0.76	0.85	0.72	0.93
	Incidentaloma	0.21	0.55	0.38	0.43	0.40	0.37
Baseline + annotations	PE	0.88	2.00	0.86	0.86	0.87	0.98
	DVT	0.62	2.00	0.78	2.00	0.90	1.00
	PE and/or DVT	0.84	2.00	0.76	0.87	0.78	0.98
	Incidentaloma	0.56	NC	0.60	NC	0.46	NC
Baseline + annotations + section typing	Incidentaloma	0.57	NC	0.72	NC	0.42	NC
Critical sections + annotations	Incidentaloma	0.70	0.76	0.87	0.72	0.47	0.80

Discussion

This research is discussed with the context of Natural Language Processing (NLP) and machine learning in the analysis of radiology reports to detect pulmonary embolism (PE), deep vein thrombosis (DVT), and incidental findings, and optimally resolve the issues of diagnostic accuracy and clinical workflow. The evidence shows that CTA and CTV are more diagnostic than CTA, and thus the findings support the indications of other researchers who recommended the synergy of the two imaging techniques. In the current dataset of 5,000 radiology cases, 24.8% of the reports were positive CTA positive CTV, and 18.2% were positive CTA negative CTV, which revealed cases in which CTV provided value to CTA and reported thromboembolic conditions that would have been missed by CTA. In contrast, there was a negative CTA with positive CTV in 10.4% of cases, and it again demonstrated the different diagnostic value of the two modalities. This supports the clinical importance of CTV inclusion in the diagnostic process of high-risk patients, including those in an intensive care unit, post-operative, or postpartum, even though an extra radiation dose is incurred. The most crucial finding concerning this study is the high percentage of incidentalomas, identified in 32% of cases, the importance of which is growing due to the use of highly sophisticated imaging methods to reveal clinically significant but non-associated discoveries. This is in line with the increasing issues brought up by the American College of Radiology concerning the ethical, clinical, and financial nature of incidental findings. Such unanticipated findings weigh down further on the healthcare systems because of the necessity to provide follow-up services and monitor. In the absence of automated mechanisms, such findings are expensive to manage and are open to oversight. The NLP based method of the study is a viable solution, not only in that it automates the extraction and stratification of both thromboembolic and incidental findings, but also in the sense that it will cut down the number of people working on the case and the management of patients. The tabular form of representation of the concepts, relations and modalities as illustrated in Table 2 is the basis of successful machine learning uses. The most common ones were anatomical entities, supporting the complexity of the radiology report and the need to isolate the diagnostic condition and anatomical location. The high count of relations of Location of Reveals illustrates the possibilities of NLP in the correct association of the location of thrombus with certain anatomical points, which is the key to the correct diagnosis and the treatment planning. Nevertheless, the comparatively smaller number of direct relations with the term Reveals indicate that the majority of the diagnostic information is implied but not pronounced which

poses a challenge to automated systems. The results reveal that more advanced linguistic models capable of deducing implicit relationships are crucial and will be explored in research in the future and improved on the models.
The machine learning models developed in the process of annotation were crucial in the development of the machine learning models. The reliability and consistency of the annotation schema are reflected by high inter-annotator agreement (IAA) of entities: there are 3,810 exact and 3,945 inexact matches (Table 3). Nevertheless, the accuracy of such relations like the“Anatomy Location of K*” and the “Anatomy Location of Thrombo Pat+” were lower, and it is possible to conclude that the process of relation extraction is one of the problematic areas of NLP in radiology. Future models can be more accurate by improving relation extraction algorithms with either more sophisticated rule-based methods or more sophisticated deep learning methods. Table 4 shows how the machine learning models were enhanced in terms of their performance when annotations and structured features were added to them. Naïve Bayes (NB) was used as a classifier and it gave good baseline performance of PE detection with a precision of 0.79 and recall of 0.85. Nevertheless, it had poorer performance with DVT and incidentaloma classification, which are more complicated conditions. The incorporation of annotations resulted in a significant improvement in performance with NB accuracy of PE at 0.88 and DVT at 0.62. The improvements were even better with the Maximum Entropy (ME) classifier scoring perfection with a 2.00 in PE and DVT classification. Incidental findings classification registered the greatest progress, and the F-measure was increased by 90% with annotations and typing the critical section. This enhancement lays emphasis on the significance of thorough annotation and sectional analysis in the correct recognition of incidentalomas which are not usually reported in normal clinical practice. Another issue that was identified in the results is the problem of imbalance in the data because the number of negative cases was much greater than that of positive cases in the dataset. This problem was addressed by the study by using the statistical methods to sample positive data and make sure that the machine learning models were not skewed to negativity. This was essential in obtaining a reliable and generalizable performance of the model especially in clinical practice in the real worlds where a particular prevalence of a given condition may be low. The implications of the findings of the study are clinical in nature. The possibility of automatically identifying thromboembolic diseases, as well as incidental findings, has the capability of revolutionizing radiology processes. NLP and machine learning can help to improve the efficiency of the diagnostic process by decreasing manual inspection of the reports, permitting radiologists to pay attention to more complicated cases in need of professional interpretation. In addition, early detection of incidentalomas will lead to timely interventions, which will enhance the outcomes of patients and reduce the chances of unaddressed diseases. Nevertheless, the limitations are also taken into consideration within the study material, including the absence of special lexicons on incidental findings. Current terminologies such as Meesh and Radlex, though extensive, do not exhaust the fine details of observations made based on imaging, especially non-English languages. Multilingual, imaging-specific lexicon will have to be developed to make NLP applications in a global healthcare setting possible. Overall, this paper proves that a complex set of tools that incorporates CTA, CTV, NLP, and machine learning can be used as a potent instrument to improve the diagnosis and management of thromboembolic diseases and incidental findings. Combining organized annotation, sophisticated algorithms, and automatic classification systems make a road to more efficient, more precise and scalable radiology practices. Future studies must aim at optimizing relation extraction algorithms, developing lexical resources as well as confirming such models in a variety of clinical settings. In solving these areas, the possibility of NLP-based radiology analysis to enhance healthcare delivery will be achieved to the full potential.

Conclusion

This paper shows that the integration of modern CT images with Natural Language Processing (NLP) and machine-learning-based analytics can be used to make a significant contribution to the way thromboembolic disease and clinically important incidental findings are identified, structured, and responded to in daily practice. Using a massive bulk of CTA/CTV reports, we demonstrated that CTV offers diagnostic data which is complementary to CTA in all populations with selected and higher-risk factors, and that the yield of incidentalomas in imaging extends well beyond vascular pathology to a significant burden of incidentalomas with respect to structured follow-up. Our pipeline allows the construction of cohorts, decision support, and monitoring of quality at scale by automatically converting free-text reports into a structured and computable representation of the concepts, relations, and diagnostic modalities, which is not possible when relying on manual inspection. Second, section awareness and critical sections with enhanced signal-noise, particularly of low-profile or intermittently reported discoveries like incidentalomas. Third, the choice of models and the use of features were also important: although Naive Bayes has provided an excellent baseline on PE, maximum entropy models significantly increased the precision and the overall F-measure especially with concept, modality and relation features. Collectively, these elements resulted in strong automated PE/DVT and significant gains in incidental finding identification, a very weak area of operational analytics. The clinical implications are not delayed. The triage and registry construction with the help of NLP can focus on following up on incidental lesions (e.g., lung nodules, adrenal masses) and decrease care gaps. In the case of PE/DVT, the speed and reproducibility of extracting thrombus location and diagnostic certainty can inform stewardship of anticoagulation, high-risk patient escalation responses, and service-level measures (e.g., time-to-diagnosis, reporting consistency, etc.). The identical infrastructure of research allows retrospective electronic cohort identification of trials and outcomes studies without months of manual chart abstraction. Significantly, our results also place into context the place of CTV: not always essential, although contributative in specific situations (e.g., ICU, postoperative, malignancy, postpartum), and enable the integration of incremental diagnostic value criteria with the radiation levels and resource consumption. Generalization is subject to a number of limitations. Report heterogeneity (naming conventions, spelling, local idioms) had effects on recall; whereas precision was very high, missed positives clamor the fact that further normalization, lexicon expansion and spell-variant processing are necessary. Relation extraction, being effective with frequent patterns, is the most error-prone component and will be improved with hybrid designs that integrate rules together with modern sequence-labeling and span-graph designs. Further, incidental-finding ontologies are not fully developed- especially where there is no English language. To be portable, it is necessary to extend RadLex-like resources and to create bilingual dictionaries of high quality. Lastly, our analysis portrays retrospective information of one institution, and two-location prospective validation on dissimilar report formats and scanner procedures should be conducted to verify long-term sustainability. In implementation perspective, the human-in-the-loop design is very important. We suggest implementing the system as a supporting interface into the radiology workflow (PACS/RIS/EMR), where structured summaries, or certainty scores and follow-up recommendations are surfaced and still, have radiologist control. Follow-up loops can be closed and compliance documented by automated transfer of incidental findings to specified tracking queues and due date (e.g., ACR recommendations) with guideline links. There must be bias and drift monitoring, error audits on a regular basis and explicit escalation procedures. The protection of data privacy should be ensured by means of de-identification, access controls, and audit trails. The future of this work should follow four directions: (1) more sophisticated relation and event modeling (e.g., temporal anchoring, treatmentresponse links); (2) active-learning loops, focusing on uncertain cases in which expert re-annotation is more likely to improve models over time with minimal labeling costs; (3) prospective impact studies that quantify time-to-action, follow-up adherence, anticoagulation-appropriateness, and patient outcomes; and (4) economic analyses that quantify cost offsets of streamlined follow-up and less manual review. Simultaneous finding of a solution to multilingual imaging lexicons and de-identified benchmark corpora sharing will quicken the community developments. Overall, intelligent combination of CTA/CTV, structured NLP and machine learning can turn narrative radiology into actionable and reliable intelligence. By both highlighting the intended diagnoses and fortuitous discoveries, the methodology fosters patient safety, efficiency in operations and research preparedness- making ideals of learning-health-systems closer to practice.

REFERENCES

1. Loud PA, Grossman ZD, Klippenstein DL, Ray CE. (1998). Combined CT venography and pulmonary angiography: a new diagnostic technique for suspected thromboembolic disease. AJR Am J Roentgenol. 170, 951–954.
2. Goodman LR, Sostman HD, Stein PD, Woodard PK. (2009). CT venography: a necessary adjunct to ct pulmonary angiography or a waste of time, money, and radiation? 1. Radiology 250, 327–330.
3. Stein PD, Fowler SE, Goodman LR, Gottschalk A, Hales CA, Hull RD, Leeper KV, Jr, Popovich J, Jr, Quinn DA, Sos TA, Sostman HD, Tapson VF, Wakefield TW, Weg JG, Woodard PK. (2006). Multidetector computed tomography for acute pulmonary embolism. N Engl J Med. 354, 2317–2327.
4. Ghaye B, Nchimi A, Noukoua CT, Dondelinger RF. (2006). Does multi–detector row CT pulmonary angiography reduce the incremental value of indirect CT venography compared with single–detector row CT pulmonary angiography?1. Radiology. 240, 256–262.
5. Krishan S, Panditaratne N, Pandiratne N, Verma R, Robertson R. (2011). Incremental value of CT venography combined with pulmonary CT angiography for the detection of thromboembolic disease: systematic review and meta-analysis. AJR Am J Roentgenol. 196, 1065–1072.
6. Perrier A, Bounameaux H. (2006). Accuracy or outcome in suspected pulmonary embolism. N Engl J Med. 354, 2383–2385.
7. Hunsaker AR, Zou KH, Poh AC, Trotman-Dickenson B, Jacobson FL, Gill RR, Goldhaber SZ. (2008). Routine pelvic and lower extremity CT venography in patients undergoing pulmonary CT angiography. AJR Am J Roentgenol. 190, 322–326.
8. Revel MP, Sanchez O, Dechoux S, Couchon S, Frija G, Cazejust J, Chatellier G, Meyer G. (2008). Contribution of indirect computed tomographic venography to the diagnosis of postpartum venous thromboembolism. J ThrombHaemost. 6, 1478–1481.
9. Revel MP, Petrover D, Hernigou A, Lefort C, Meyer G, Frija G. (2005) . Diagnosing pulmonary embolism with four-detector row helical CT: prospective evaluation of 216 outpatients and inpatients. Radiology. 234, 265–273.
10. Hall WB, Truitt SG, Scheunemann LP, Shah SA, Rivera MP, Parker LA, Carson SS. (2009) .The prevalence of clinically relevant incidental findings on chest computed tomographic angiograms ordered to diagnose pulmonary embolism. Arch Intern Med, 169, 1961–1965.
11. Jacobs PCA, Mali WPTM, Grobbee DE, van der Graaf Y. (2008), Prevalence of incidental findings in computed tomographic screening of the chest: a systematic review. J Comput Assist Tomogr, 32, 214–221.
12. Berland LL, Silverman SG, Gore RM, Mayo-Smith WW, Megibow AJ, Yee J, Brink JA, Baker ME, Federle MP, Foley WD, Francis IR, Herts BR, Israel GM, Krinsky G, Platt JF, Shuman WP, Taylor AJ. (2010). Managing incidental findings on abdominal CT: white paper of the ACR incidental findings committee. J Am CollRadio ,7, 754–773.
13. Demner-Fushman D, Chapman WW, McDonald CJ. (2009).What can natural language processing do for clinical decision support? J Biomed Inform, 42, 760–772.
14. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. (1994). A general natural-language text processor for clinical radiology. J Am Med Inform Assoc, 1, 161–174.
15. Friedman C, Johnson SB, Forman B, Starren J. (1995). Proc AnnuSympComputAppl Med Care. Architectural requirements for a multipurpose natural language processor in the clinical environment; pp. 347–351
16. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, Chute CG. (2010). Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc, 17, 507–513.