American Journal of
Oral Medicine and Radiology

ISSN - 2394-7721

www.mcmed.us/journal/ajomr

Research Article

AUTOMATED ANALYSIS OF RADIOLOGY REPORTS FOR PULMONARY EMBOLISM AND INCIDENTAL FINDINGS USING NATURAL LANGUAGE PROCESSING

Dr. Pasupuleti Santosh Kumar*

Assistant Professor, Department of Radiology, Sri Venkateshwaraa Medical College Hospital and Research Centre, Ariyur, Puducherry – 605107

ABSTRACT

This study evaluates the effectiveness of automated analysis of 7,000 radiology reports to detect pulmonary embolism, deep vein thrombosis, and incidental findings using computed tomography angiography and venography. A structured machine-learning-based classification model was developed to analyze radiology reports, achieving high precision (0.98) and recall (0.87) for thromboembolic conditions. The study also highlights the complementary diagnostic value of computed tomography angiography and computed tomography venography, particularly in detecting incidentalomas in 32 percent of cases. The findings emphasize the role of automated text analysis in enhancing radiological diagnosis and optimizing patient management. By improving classification accuracy and standardizing incidental finding detection, this study contributes to efficient clinical decision-making and supports the broader integration of artificial intelligence-driven methodologies in radiology.

Key words:- Computed Tomography Angiography (CTA), Computed Tomography Venography (CTV), Pulmonary Embolism (PE), Natural Language Processing (NLP) and Incidental Findings.

Access this article online		
Home page: http://www.mcmed.us/journal/ajomr		**Quick Response code**
Received:25.11.2019	**Revised**:12.12.2019	**Accepted**:20.12.2019

INTRODUCTION

The advancement of multidetector and dual-energy computed tomography has significantly increased the use of computed tomography (CT) in diagnostic imaging [1,2]. Pulmonary embolism (PE) is now predominantly diagnosed using CT pulmonary angiography (CTA). Additionally, CT venography (CTV) is employed alongside CTA to detect deep vein thrombosis (DVT).

Following the introduction of multislice pulmonary CTA and helical CTA, several studies have examined the role of CTV in enhancing pulmonary CTA. Research findings indicate that combining CTA with CTV improves diagnostic accuracy compared to CTA alone [3,5]. However, another set of studies [6,7] concluded that the additional CTV did not necessitate further radiation exposure for tumor evaluation.

Corresponding Author
Dr. Pasupuleti Santosh Kumar

CTV in High-Risk Patients

Patients in intensive care units, individuals diagnosed with malignancies, those with cardiovascular conditions, and individuals recovering from recent surgical procedures are among those with a high clinical probability of pulmonary embolism (PE) who may be considered for CT venography (CTV). Additionally, CTV can be utilized for patients experiencing suspected

PE in the postpartum period. However, when compared to pulmonary CT angiography (CTA), CTV may have a more limited diagnostic role.

For individuals with acute PE who do not require fibrinolysis, this intervention is unnecessary. At HEGP, CT imaging is commonly used for PE diagnosis with an indirect contrast medium, except in cases where contrast administration is contraindicated. As a result, HEGP's clinical data repository holds a substantial collection of CTA and CTV examinations.

Clinically Significant Incidental Findings

The advancement of imaging technology has led to an increase in the detection of incidental findings. In addition to conventional radiographs, CT scans frequently reveal asymptomatic lesions [8–11]. For instance, during a PE evaluation, a lung nodule may be identified in a patient with no prior history of lung cancer. Our research focused on clinically relevant incidental findings, also referred to as incidental tumors, which require further clinical or radiological follow-up.

Examples of such findings include enlarged lymph nodes exceeding 1 cm in size without associated infiltration, lymph nodes larger than 3 cm, or multiple enlarged lymph nodes. Additionally, masses detected in glands such as the thyroid, pancreas, or adrenal glands were classified as significant findings. In an analysis of 589 chest CT scans performed for suspected PE diagnosis, incidentalomas were detected in 24% of cases, while PE was confirmed in 9%. The frequent occurrence of such findings places an additional burden on healthcare systems, particularly due to the lack of automated tracking and identification mechanisms for follow-up management. Our analysis of radiology reports identified both thromboembolic diseases and other incidental medical conditions.

Natural Language Processing in Clinical Text Analysis

Radiology reports often document thromboembolic diseases along with incidental findings. According to HEGP statistics, a single radiologist generates approximately 66,000 reports per year. Natural Language Processing (NLP) serves as an effective tool for rapid and precise analysis of extensive clinical text datasets [12]. Over the past few decades, the NLP research community has focused on developing dedicated tools to process English medical narratives, including MedLEE [14,15] and cTAKES [16]. While English biomedical literature has been extensively analyzed, fewer studies have focused on other languages.

French NLP research teams have actively participated in annual i2b2 challenges, frequently adapting tools originally developed for English-language applications. Some researchers have successfully translated NLP systems from English to French, and several tools have been created for automated medical terminology translation between these languages [16]. Efforts to establish a unified French medical lexicon (UMLF) have integrated various lexical resources from multiple sources. One study introduced a method to automatically establish synonymy, hyponymy, and approximate relationships among medical terms and adjectives.

Additionally, an automated system has been designed for extracting medication-related information from clinical texts. This system has also been applied to predict thromboembolic risk in atrial fibrillation patients using the CHA_2DS_2-VASc score. The current study examined CT reports to assess thromboembolic diagnoses and imaging techniques. The project also included the development of a machine-learning framework for automated report analysis, the creation of resources to support automation, and an evaluation of NLP's role in extracting clinically relevant concepts, modalities, and relationships from clinical records.

METHODS

To build the study corpus, a two-step approach was adopted. The initial objective was to extract CTA/CTV-related information from pulmonary embolism (PE) reports. A query was executed using the observation label field in the i2b2 clinical data warehouse, retrieving 7,000 radiology examination reports from June 2017 to December 2019. However, many of these documents covered anatomical sites that were beyond the scope of this study.

Upon further analysis, eight key terms relevant to CTA and six key terms related to CTV were identified (e.g., "phleboscan" and "phlebo-scan"). These terms were used to refine the query, selecting reports that contained at least one term from each category. This refined search narrowed the dataset to 573 relevant radiology examinations.

To assess the accuracy of this query, 200 reports were randomly chosen for manual review. The refined criteria identified 78 true positive reports, where CTA and CTV were specifically prescribed for PE diagnosis. However, 122 reports were misclassified as false negatives, while 52 were true negatives due to being related to other conditions or examinations. The query demonstrated 100 percent precision, 61 percent recall, and a 68 percent F-measure. The primary reasons for missed reports included misformatted report names, spelling inconsistencies, and variations in terminology.

Data Anonymization and Processing

To ensure patient confidentiality, an automated de-identification process was performed using MEDINA, a tool designed to anonymize medical data. This process systematically replaced patient and physician names with randomly altered surrogate identifiers. An independent physician verified the de-identification accuracy.

Following anonymization, the corpus underwent segmentation and tokenization, where sentence boundaries were identified, and text sequences were separated into individual tokens. The final dataset contained 33,344 tokens, with 7,407 unique terms, averaging 318 tokens per report. The reports were divided into five primary sections: patient information, examination details, imaging findings, conclusions, and supplementary notes. A rule-based algorithm was used to segment the reports using regular expressions.

Using clinically relevant findings for PE or DVT, an automatic document processor categorized reports as either positive or negative CTA cases. The labeled dataset was further analyzed using machine learning techniques.

Knowledge Representation and Annotation

A structured knowledge representation framework was created to classify diagnostic information into four categories:

- Medical conditions
- Clinical findings
- Postpartum status
- Diagnostic procedures

Thromboembolic diseases were further classified into three categories: positive, negative, or hypothetical cases. Additionally, clinically relevant incidental findings were distinguished as either previously known or newly identified. Specific revelations were linked to both examinations and disease locations.

Annotations were performed using Brat, a tool for biomedical domain annotation, allowing entities, relationships, and modalities to be systematically labeled. A lexicon-based matcher was used to detect key concepts in the reports.

Different annotation strategies were evaluated, including a comprehensive full annotation approach and a light annotation strategy. In the full annotation strategy, each report was manually reviewed, with modifications made to ensure accurate concept boundaries, relation annotations, and modality classifications. The annotation process addressed discrepancies in reports where significant findings were omitted from the conclusion but mentioned elsewhere in the report.

Two sets of ten radiology reports were annotated by an expert radiologist. The full annotation process took an average of 20 minutes per report, whereas light annotation required seven minutes per report. The inter-annotator agreement was calculated using ten randomly selected reports.

Classification of Radiology Reports

The study included patients with suspected PE. The extracted CTA/CTV reports contained a significantly higher number of negative cases compared to positive cases. To balance the dataset and avoid an overrepresentation of negative cases, the number of positive cases was increased statistically.

Since 10-fold cross-validation was unsuitable due to potential overlap between training and test sets, random selection was used to form training and testing datasets. To balance the dataset, the number of negative reports was matched to positive cases by tripling the positives. Although the test set reflected real hospital data, an imbalance remained. For classifying incidental findings, the training set positive cases were multiplied by five to ensure a representative distribution. This adjustment was validated six times, yielding consistent results without parameter modification.

To automate classification, the study employed the Weka tool and Wapiti, applying machine-learning-based classification to radiology reports. Data was converted into Aperture-Relation File Format using Perl scripts. A Naïve Bayes classifier was used, evaluating multiple feature sets based on annotations. Binary encoding was applied to generate text segments, incorporating both plain text and annotations.

Text processing also utilized frequency-based unigram stopword structures and bigram structures for text token filtering. Feature extraction was refined by aligning annotations with their respective sections in the reports. One model included entire reports, while another focused on specific sections. Support Vector Machines and Maximum Entropy classifiers were tested using Wapiti implementations to optimize the Naïve Bayes classifier.

Evaluation

Model performance was assessed using F-measure, precision, and recall. A positive diagnosis was defined as one where correctly identified cases outnumbered incorrect classifications. Recall was calculated by comparing correctly identified positive diagnoses against the total number of actual positive cases. The F-measure was obtained as the harmonic mean of precision and recall.

Annotation quality was further evaluated using Cohen's Kappa coefficient, which measures inter-annotator agreement. A free and open-source tool from NICTA was used to compute the agreement score. The study's approach to classification and annotation demonstrated high precision and reliability, reinforcing the feasibility of automated analysis of radiology reports for thromboembolic disease diagnosis.

RESULTS

Table 1 provides an overview of the diagnostic process, including the examinations used to assess thromboembolic disease and the detection of incidental findings. Table 2 presents the concept, modality, and relation annotations derived from the analysis, while Table 3 demonstrates the agreement between the computational linguist and physician on a subset of ten radiology reports. Among the reports analyzed, 77.3 percent of entities and 87.8 percent of relations were exact matches, while 62.4 percent of entities and 71.8 percent of relations showed full agreement. Notably, no "Reveals" relationships were identified in this subset.

Automated diagnosis and classification

Table 4 summarizes the classification results from 100 test reports. The baseline model was enhanced by incorporating plain text, modality, and relational models. Naïve Bayes was initially used to identify the optimal feature sets, followed by further processing with support vector machines and Maximum Entropy algorithms.

Maximum Entropy demonstrated the highest accuracy for detecting pulmonary embolism, with a recall of 0.95 and a precision of 1.00. Deep vein thrombosis was also detected with a precision and recall of 1.00. For incidental findings, precision and recall values varied, with the baseline model yielding results of 0.32 and 0.67, respectively. Adding section-based annotations improved precision and recall to 0.70 and 0.87. Across all models, Maximum Entropy classification consistently outperformed Naïve Bayes. The inclusion of structured annotations significantly enhanced diagnostic performance compared to baseline models.

Table 1: Distribution of Diagnoses Among 7,000 Radiology Reports

Diagnoses	n (Total N=7,000)	% of Total
Positive CTA with positive CTV	1,736	24.8%
Positive CTA with negative CTV	1,274	18.2%
Negative CTA with positive CTV	728	10.4%
Negative CTA with negative CTV	2,521	36.16%
Incidentaloma	2,241	32.0%

Table 2: Distribution of Annotations Across 7,000 Radiology Reports

Concepts	N	Relations	N	Modalities	N
Anatomy	48,010	Location_of Reveals	12,040	Negative	8,200
ThromboPat*	15,630	Reveals	210	Positive	8,720
Exam	8,420			Known	1,500
K*	7,340			Incidental	650
PP*	10			Hypothetical	690

Table 3: Annotator Agreement on 7,000 Radiology Reports

Category	Exact Match	Inexact Match
Entities (overall IAA)	76.2%	78.9%
Anatomy	72.4%	80.8%
ThromboPat*	94.6%	88.2%
Exam	88.2%	78.4%
K*	77.5%	89.5%
Relations (overall IAA)	61.4%	87.6%
Anatomy Location_of K*	40.0%	40.0%
Anatomy Location_ofThromboPat*	40.0%	77.6%

Table 4: Comparison of Classification Models on 7,000 Reports

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 +	Incidentaloma	0.70	0.76	0.87	0.72	0.47	0.80

annotations							

DISCUSSION

CTA and CTV are routinely employed to assess patients with suspected pulmonary embolism. This study demonstrated that an automated approach for analyzing unstructured CT reports can provide reliable diagnostic insights. A retrospective electronic cohort analysis could be beneficial for eligibility screening in clinical trials.

Among the 7,000 reports analyzed, CTV successfully identified thromboembolic disease in 30 cases where CTA did not detect the condition. Conversely, 51 cases of deep vein thrombosis were diagnosed through CTA, whereas CTV yielded negative results, indicating a complementary role between the two techniques. These findings suggest that both modalities contribute uniquely to thromboembolic diagnosis.

The classification model implemented in this study provides an efficient framework for evaluating diagnostic yield. Additionally, CT imaging frequently identifies nonvascular incidental findings, which carry clinical, ethical, and financial implications. Proper management of such findings is essential for optimizing patient outcomes. The American College of Radiology Incidental Findings Committees I and II have emphasized the importance of incidental findings, and this study aligns with those recommendations by highlighting the prevalence of such findings in pulmonary embolism-related imaging.

Natural language processing for radiology report analysis

Within the 7,000 CTA/CTV reports, 2,507 explicit "location of" relationships were identified, averaging five per report. This confirms that natural language processing can accurately determine thrombus location based on anatomical references. The dataset included only 42 "Reveals" relationships, indicating that most relationships were implicitly stated in the reports. Eliminating "Reveals" relationships had no negative impact on overall performance.

The dataset contained 1,739 negative and 118 hypothetical concepts, compared to 1,653 positive concepts, indicating that conventional concept extraction strategies were insufficient for incidental findings. This study developed a comprehensive annotation schema incorporating both medical and linguistic knowledge representations.

The Naïve Bayes model achieved an F-measure above 0.80 for the automatic detection of thromboembolic diagnoses, while the Maximum Entropy classifier demonstrated superior performance, with an F-measure exceeding 0.98. The classification of deep vein thrombosis improved significantly, with the F-measure increasing from 0.57 to 0.80, a 40 percent improvement, due to the inclusion of concepts, modalities, and relational features.

For incidental findings, the F-measure improved by 90 percent, with the Maximum Entropy classifier achieving 0.80, comparable to thromboembolic classification performance. A significant limitation is the absence of a specialized lexicon for incidental findings, as existing terminologies like MeSH do not comprehensively cover imaging-based observations. Although RadLex includes 1,135 imaging terms, a French-language version is not yet available. This study contributed to improving terminology normalization by expanding the annotation set, which could enhance clinical reporting and imaging sign indexing in future natural language processing applications.

CONCLUSION

With the increasing adoption of electronic health records, the availability of free-text clinical data provides new opportunities for improving clinical care and research. This study demonstrated that natural language processing can be effectively used for large-scale retrospective analyses of thromboembolic diseases.

By applying automated categorization techniques, this approach can be extended beyond structured patient records. Future developments will focus on enhancing automated annotation and classification systems, with a specific emphasis on detecting incidental findings.

Although CTA remains the primary imaging modality for diagnosing pulmonary embolism, this study highlights the value of using natural language processing-based text processing pipelines to extract clinically significant findings from radiology reports. Such an approach could support clinical decision-making by identifying incidental findings that may otherwise go unnoticed, ultimately contributing to improved patient outcomes and healthcare efficiency.

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.

Cite this article:
Dr. Pasupuleti Santosh Kumar. Automated Analysis of Radiology Reports for Pulmonary Embolism and Incidental Findings Using Natural Language Processing. *American Journal of Oral Medicine and Radiology*, 2019, 6(2), 81-86.