Pulmonary Nodule Clinical Decision Support System (CDSS)

NLP-Based Automation of Nodule Detection & Follow-Up Recommendations

By Shamiso Mubatsa

SAT 5141-FALL’25

Introduction

- Pulmonary nodules are small, rounded opacities frequently discovered as incidental findings on chest imaging, particularly computed tomography (CT) scan.
- Although many nodules are benign, a small percentage represent early-stage lung cancers, making appropriate follow-up critical for early detection and treatment (MacMahon et al., 2017).
- Radiology workflows rely on manual review of reports, creating opportunities for oversight and variability.
- This project develops an NLP-driven CDSS prototype to improve follow-up compliance, support clinicians, and strengthen patient safety.

PROBLEM STATEMENT

- The Fleischner Society guidelines provide detailed, evidence-based recommendations for managing these nodules based on factors such as size, morphology, and patient risk profile.
- Despite the existence of these standardized guidelines, research has consistently shown low adherence rates, with many patients failing to receive follow-up imaging within the recommended timeframes (Lacson et al., 2012).
- Delayed cancer diagnosis, progression of undetected malignancies, or, unnecessary radiation exposure.(Hammer et al., 2018).

OBJECTIVES

Apply AI/NLP methods to automatically detect pulmonary nodules within unstructured radiology reports.

Use text-mining and extraction techniques to identify clinically relevant attributes such as nodule size.

Develop an AI-assisted CDSS prototype that generates guideline-based follow-up recommendations aligned with Fleischner standards.

Literature Review

- NLP has proven effective in extracting clinical findings from radiology reports, including pulmonary nodules, with high accuracy reported in recent studies (Pons et al., 2016; Zheng et al., 2021).
- Automated nodule extraction and size detection using NLP supports guideline-based management and has shown strong performance in identifying incidental findings (Grolleau et al., 2024).
- Follow-up adherence remains low in clinical practice, but structured workflows and CDSS tools significantly improve communication and timely management (Fu et al., 2023; Schwartz et al., 2021).
- AI and NLP-driven CDSS systems offer a practical, scalable solution for reducing missed follow-ups and enhancing radiology workflow efficiency (Allen et al., 2019).

DATA SET OVERVIEW

- Dataset sourced from the NIH OpenI Chest X-ray Collection, containing 3,900+ radiology reports in XML format.
- Only text reports were used—FINDINGS and IMPRESSION sections extracted and cleaned for NLP analysis.
- Final structured dataset includes:

❖ report_id – unique identifier

❖ report_text – cleaned radiology report

❖ label_nodule – weak label indicating presence of nodule

- Weak labeling generated using nodule-related keywords, following common radiology NLP practices.

Libraries Used

Core NLP & ML	scikit-learn	TF-IDF vectorization, Logistic Regression classifier
	pandas	Data manipulation, CSV creation
	numpy	Numerical operations
Text Processing	re (regex)	Nodule size extraction
	nltk (optional)	Tokenization / stopwords (if present in your code)
XML Parsing	xml.etree, ElementTree	Extract FINDINGS & IMPRESSION sections from XML
Evaluation	sklearn. metrics	Classification report, precision/recall/F1
Colab Utilities	google, colab.files	For downloading output CSV

- The entire project was developed in Google Colab.

```
# 10. TRAIN LOGISTIC REGRESSION CLASSIFIER

clf = LogisticRegression(max_iter=2000)
clf.fit(X_train_vec, y_train)

# Evaluate on held-out test set
y_pred = clf.predict(X_test_vec)
print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.92      1.00      0.96       634
           1       1.00      0.66      0.79       152

    accuracy                           0.93       786
   macro avg       0.96      0.83      0.88       786
weighted avg       0.94      0.93      0.93       786
```

Results summary

- Final test performance:
  - ❖ Accuracy: 0.93
  - ❖ Macro F1: 0.88
  - ❖ High precision for both classes
- High precision - very few false positives
- Lower recall for nodules - expected due to weak labeling
- High precision ensures clinicians won't receive unnecessary alerts.
- Overall, results are strong enough for a prototype CDSS in a real workflow.

Results

Model Performance (from this study):

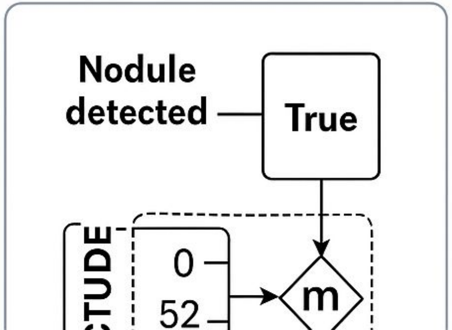
Class	Precision	Recall	F1-score
No Nodule	0.92	1.00	0.96
Nodule (1)	1.00	0.66	0.79

Overall Accuracy: 0.93
Macro F1-score: 0.88

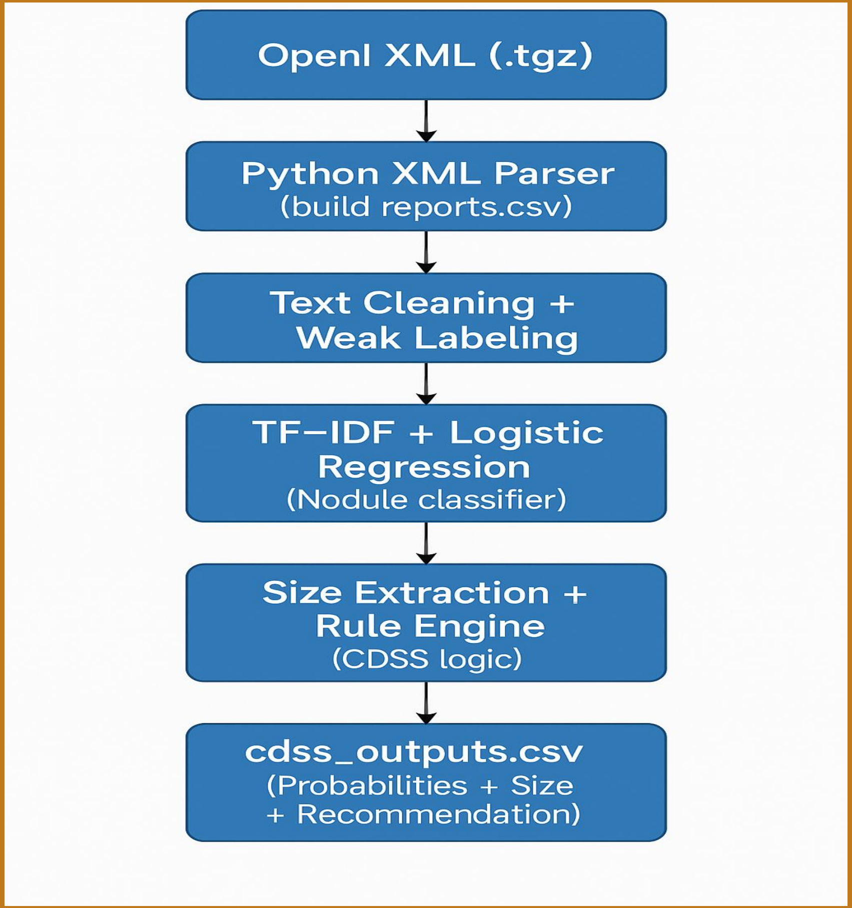
	Predicted No Nodule	Predicted Nodule
Actual No Nodule	634	0
Actual Nodule	100	52

Sample Output
(from real report)

- Nodule detected: True
- Probability: 0.82
- Extracted size: 8 mm
- CDSS Recommendation: CT at 6–12 months



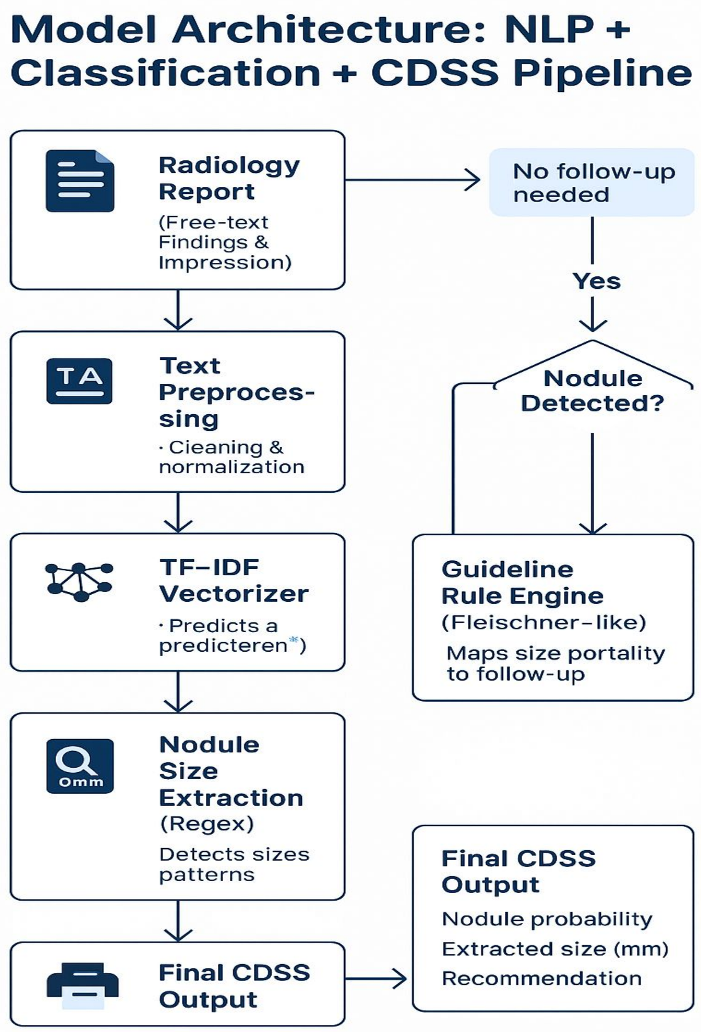
Model Architecture



```python
    for k, v in result.items():
        print(f"  {k}: {v}")
    print()
```

```
================================================================================
Report ID: 3723
Report Text: Lungs are clear bilaterally. Cardiac and mediastinal silhouettes are normal. Pulmonary vasculature is normal. No pneumothorax or

CDSS Output:
  nodule_detected: False
  probability: 0.024840512467023776
  size_mm: None
  recommendation: No pulmonary nodule detected.

================================================================================
Report ID: 3320
Report Text: Chest. Heart size normal lungs are clear. Right knee. Severe osteoarthritis all 3 compartments ...

CDSS Output:
  nodule_detected: False
  probability: 0.08262776360477554
  size_mm: None
  recommendation: No pulmonary nodule detected.

================================================================================
Report ID: 3301
Report Text: The cardiomediastinal silhouette is within normal limits for size and contour. The lungs are normally inflated without evidence
```

Architecture for the CDSS



Clinical Impact

- Reduces missed follow-ups by automatically flagging nodules
- Standardizes recommendations using Fleischner guidelines
- Converts unstructured text structured actionable data
- Demonstrates feasibility of AI-assisted radiology workflow improvement

Challenges Encountered

- Lack of expert-annotated labels → required weak labeling using keyword rules.
- Inconsistent radiology terminology made it difficult for simple NLP patterns to capture all nodule mentions.
- Extracting nodule size was challenging due to variation in how measurements appear (e.g., “6 mm”, “1.2 cm”, “6-mm”).
- Balancing model performance — high precision but lower recall for nodules due to data imbalance and weak labels.
- Integration of guideline logic required careful mapping of size ranges to

Solutions Applied

- Developed a weak supervision strategy using curated nodule-related keywords.
- Applied text cleaning + TF–IDF to normalize report language and improve signal quality.
- Used regex-based extraction to reliably detect size patterns in multiple formats.
- Tuned Logistic Regression with max_iter adjustment for convergence and stable performance.
- Built a modular CDSS rule engine replicating Fleischner guidance for follow-up recommendations.

Recommendations/Future Work

- Incorporate advanced language models (BioBERT, ClinicalBERT) for improved nodule detection and contextual understanding.
- Expand labels using radiologist validation or semi-supervised learning for stronger ground truth.
- Integrate additional clinical risk factors (age, smoking status, cancer history) into CDSS logic.
- Extend system to analyze imaging data (CT scans) using deep learning for combined text + image decision support.
- Deploy as a FHIR-enabled module within EHR systems to automate notifications and follow-up tracking.

CONCLUSIONS

- Successfully built an NLP-driven CDSS prototype that identifies pulmonary nodules from free-text radiology reports.
- System achieved 93% accuracy and produced meaningful follow-up recommendations aligned with Fleischner guidelines.
- Demonstrated how AI/NLP can transform unstructured clinical text into actionable decision support.
- Highlights the feasibility and potential impact of integrating informatics tools into radiology workflows to reduce missed follow-ups.

References

- Fu, T., Liu, B., & Zhang, J., et al. (2023). Implementing a streamlined radiology workflow to close the loop on actionable incidental findings. Frontiers in Radiology, 3, 1162332.
- Lacson, R., Prevedello, L. M., & Andriole, K. P., et al. (2012). Factors associated with radiologists' adherence to Fleischner Society guidelines for management of pulmonary nodules. Journal of the American College of Radiology, 9(7).
- MacMahon, H., Naidich, D. P., & Goo, J. M., et al. (2017). Guidelines for management of incidental pulmonary nodules detected on CT images: From the Fleischner Society 2017. Radiology, 284(1), 228–243.
- Pons, E., Braun, L. M. M., & Hunink, M. G. M., et al. (2016). Natural language processing in radiology: A systematic review. Radiology, 279(2), 329–343.
- Allen, B., Seltzer, S. E., & Langlotz, C. P., et al. (2019). A road map for translational research on artificial intelligence in medical imaging: From the 2018 NIH/RSNA/ACR workshop. Journal of the American College of Radiology, 16(9), 1179–1189.

Thank You