AUTOMATED ASSESSMENT OF TUMOR RESPONSE FROM FREE-TEXT ONCOLOGY REPORTS USING DEEP NATURAL LANGUAGE PROCESSING

Dr Som Biswas, Jefferson Abington Hospital , Pennsylvania

BACKGROUND

- Radiology reports contain critical information about tumor burden and treatment response. However, most reports remain unstructured narrative text, making automated extraction of clinical outcomes difficult.
- Structured oncology reports (SOR) provide standardized data that can be used to train natural language processing (NLP) models. Such models could enable automated tumor response classification, reducing manual workload and supporting large-scale oncology research.

OBJECTIVE

To develop and evaluate a deep NLP model capable of classifying Tumor Response Categories (TRCs) from Free-Text Oncology Reports (FTOR) and compare its performance with:

Conventional NLP algorithms

Human annotators with varying radiology expertise

METHODS

NLP Models

Deep Learning Model

- BERT-based NLP model
- Fine-tuned on structured radiology findings

Conventional NLP Models

- Linear Support Vector Classifier (Linear-SVC)
- k-Nearest Neighbors (kNN)
- Multinomial Naive Bayes
- Bag-of-Words feature representation

Evaluation Metrics

- F1 Score
- Accuracy
- Precision
- Recall

Human Comparison

Model performance was compared with:

- Radiologists
- Medical Students
- Radiology Technologist Students

Additional analyses evaluated:

- **Lexical complexity**
- **Semantic ambiguity**

PERFORMANCE ON STRUCTURED REPORTS

Model	F1 Score
BERT	**0.86**

Structured data significantly improved classification performance.

KEY FINDINGS

- Deep NLP models can automatically classify tumor response from radiology reports.
- BERT achieved near human-level performance.
- Structured reports improve model accuracy.
- Language ambiguity remains a key challenge.

IMPACT OF LANGUAGE COMPLEXITY

Both human annotators and NLP models were affected by:

- **Lexical complexity**
- **Semantic ambiguity**

Maximum observed decrease in F1 score:

- Lexical complexity: **−0.19**
- Semantic ambiguity: **−0.17**

CLINICAL IMPACT

Automated tumor response classification can enable:

- Scalable **oncology outcome extraction**
- **Clinical workflow optimization**
- Improved **longitudinal patient monitoring**
- Large-scale **radiology research datasets**

DATASET
TOTAL REPORTS: 10,455
FROM 3 INDEPENDENT RADIOLOGY DEPARTMENTS

Report Type	Number
Structured Oncology Reports (SOR)	9,653
Free-Text Oncology Reports (FTOR)	802

Tumor response labels were extracted from SOR using Response Evaluation Criteria in Solid Tumors (RECIST).

RESULTS

The **BERT model outperformed conventional NLP methods** and approached the performance of human annotators.

Method	F1 Score
Radiologists	**0.79**
Medical Students	0.73
BERT Model	**0.70**
Technologist Students	0.65
Linear-SVC	0.63

REFERENCES:

Laurent, G., Craynest, F., Thobois, M., & Hajjaji, N. (2023). Automatic classification of tumor response from radiology reports with rule-based natural language processing integrated into the clinical oncology workflow. JCO Clinical Cancer Informatics, 7, e2200139.

Senders, J. T., Karhade, A. V., Cote, D. J., Mehrtash, A., Lamba, N., DiRisio, A., ... & Arnaout, O. (2019). Natural language processing for automated quantification of brain metastases reported in free-text radiology reports. JCO clinical cancer informatics, 3, 1-9.

Chen, P. H., Zafar, H., Galperin-Aizenberg, M., & Cook, T. (2018). Integrating natural language processing and machine learning algorithms to categorize oncologic response in radiology reports. Journal of digital imaging, 31(2), 178-184.

Fink, M. A., Kades, K., Bischoff, A., Moll, M., Schnell, M., Küchler, M., ... & Kleesiek, J. (2022). Deep learning-based assessment of oncologic outcomes from natural language processing of structured radiology reports. Radiology: Artificial Intelligence, 4(5), e220055.

Cheng, L. T., Zheng, J., Savova, G. K., & Erickson, B. J. (2010). Discerning tumor status from unstructured MRI reports–completeness of information in existing reports and utility of automated natural language processing. Journal of digital imaging, 23(2), 119-132.