



18F-FDG PET/CT読影レポート生成システム構築に向けた客観的評価指標の探索

南和孝¹, 吉村高明²⁻⁵, 平田健司^{3,5-7}, 甲木晶枝⁸, 植竹望⁸, 渡邊史郎^{3,6,7}, 工藤與亮^{3,5,6}

1.北海道大学大学院医学院, 2.北海道大学大学院保健科学研究院健康科学分野, 3.北海道大学大学院医学研究院医理工学グローバルセンター, 4.北海道大学病院医学物理部, 5.北海道大学病院医療AI研究開発センター, 6.北海道大学大学院医学研究院画像診断学教室, 7.北海道大学病院核医学診療科, 8.GEヘルスケア・ジャパン研究開発部

COI開示

共同研究者である平田健司は、GEヘルスケア・ジャパンから支援を受けている共同研究者の甲木晶枝と植竹望はGEヘルスケア・ジャパンに属する社員である

背景

- 陽電子放出断層撮影/コンピュータ断層撮影(PET/CT)読影レポートは診断・治療方針決定に重要である
- SUVmaxは臨床的に重要であり、鑑別診断、治療効果判定等に有用である¹
- 検査件数の増加や働き方改革を背景として、画像診断医の負担軽減と業務効率化が求められている²
- 近年、LLMを用いたレポート作成・構造化が注目されており、読影レポート作成支援への応用が期待される³

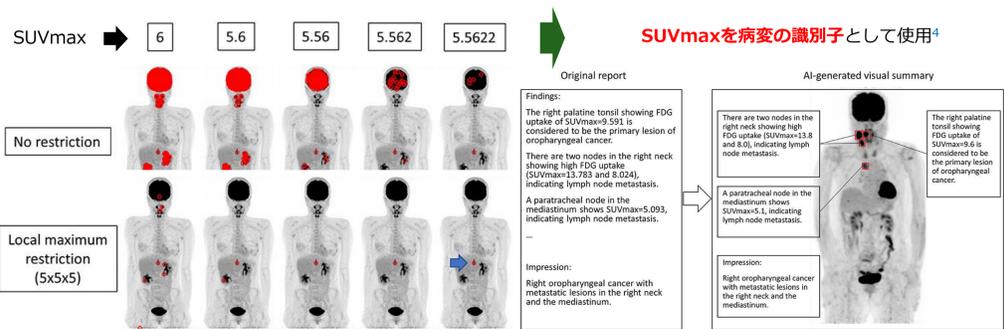


Figure 1 脊椎病変に対してFDG PET-CTを実施した患者の所見⁴ Figure 2 診断レポートを要約するシステムの実現可能性⁴

研究の経緯

- これまでに、SUVmaxおよび集積部位を入力情報としてPET読影レポートをAIで生成する検討を行ってきた
- 生成レポートの活用を進めるためには、出力品質を適切に評価できる枠組みが必要である

目的

- 生成AIにより作成された読影レポートの質評価には専門医による主観的評価が用いられるが、膨大なレポートすべてに対して都度評価を行うことは困難である
- 客観的評価指標を用いて専門医が重点的に確認すべきレポートを絞り込むことができれば、専門医の評価負担を軽減できる可能性がある

従来のレポートの評価 (専門医のみ、専門医に依存)

本研究が目指すもの (客観指標で評価 → 専門医は最終確認)

専門医の主観的評価の代替が可能な客観的指標を探索

対象と方法

対象患者

- 北海道大学病院でPET検査を実施した20名の患者 (IRB: 023-0128)
- 患者ごとに2回分のPET検査の読影レポートを抽出
- 患者を特定可能な情報は含まれていない

方法

- LLM
 - Microsoft Copilot (Microsoft, Redmond, USA)⁵
 - GPT-4(OpenAI)をベースに設計
 - データ保護機能あり(入力情報は学習されない)

評価

- 客観的評価指標の算出
- 専門医2名による主観的評価

解析

- Spearmanの順位相関係数を算出
- 95%信頼区間をブートストラップ1万回で推定

客観的評価指標の分類

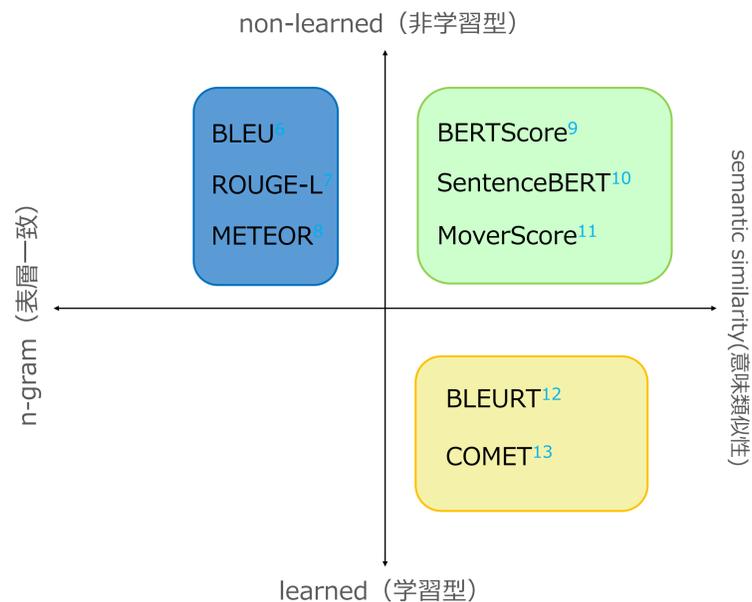


Figure 3 客観的評価指標の分類図

主観的評価

- 評価方法
 - 核医学専門医2名(経験年数:10年、20年)が20症例を5段階で評価
 - 1カ月の間隔を空けて2回の評価を実施

結果

Table.1 客観的評価指標ごとの平均相関係数と範囲

指標	Mean p	Range
BLEU	0.518	0.104-0.825
ROUGE-L	0.532	0.037-0.885
METEOR	-0.234	-0.572-0.187
BERTScore	0.686	0.353-0.857
SentenceBERT	0.018	-0.488-0.497
MoverScore	0.565	0.127-0.847
BLEURT	0.11	-0.301-0.455
COMET	0.7	0.403-0.858

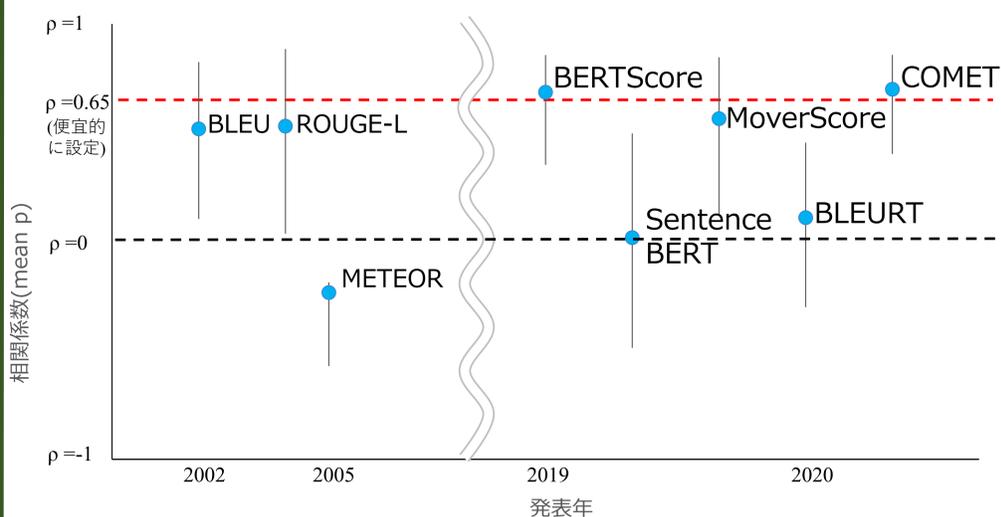


Figure 4 各評価指標の発表年と平均相関係数の関係

COMET及びBERTScoreが他指標と比べ高い相関を示した(mean p>0.65)

考察と結論

- COMET/BERTScoreが比較的高い相関を示したことから、表層一致よりも意味的整合性を拾いやすく、主観評価に近い評価指標と考えられる
- 使用した指標は一般的自然言語生成(NLG)評価であり、SUVmax・部位などタスク固有の重要項目を直接評価できない
- 症例数が少なく、推定の不確実性が大きいいため、症例数を増やす必要がある

COMET/BERTScoreは一次スクリーニングとして、専門医の主観的評価の代替候補となる可能性が示唆された

1. Krak NC et al. Eur J Nucl Med Mol Imaging. 2005;32:294-301. 5. Tepe M et al. Patient Educ Couns.

2. Nishiyama Y et al. Ann Nucl Med. 2019;33:725-32.

3. Hirata K et al. Ann Nucl Med. 2024.

4. Hirata K et al. Front Med. 2021;8:647562.

6. Papineni K et al. ACL. 2002.

7. Lin C-Y. Text Summarization Branches Out. 2004.

8. Banerjee S et al. ACL Workshop. 2005.

9. Zhang T et al. ICLR. 2020.

10. Reimers N et al. EMNLP-IJCNLP. 2019.

11. Zhao W et al. EMNLP-IJCNLP. 2019.

12. Sellam T et al. ACL. 2020.

13. Rei R et al. EMNLP. 2020.