NURad at the NTCIR-18 RadNLP Task

Nagoya University Hospital, Department of Radiology, Japan
Marina Higashi, Rintaro Ito, Keita Kato, Ryota Asai, Shingo Iwano, Shinji Naganawa

Introduction

- The TNM classification of lung cancer plays an important role in diagnosis and treatment planning.
- Radiologists classify TNM using mainly CT images.
- In Japan, the number of thoracic radiologists is limited. While the Japanese Radiological Society has approximately 10,500 members [1], the Japanese Society of Thoracic Radiology comprises only around 400 members [2].
- An AI model that automatically applies TNM classifications from radiology reports would be valuable for radiologists and many clinicians.

① Data

From task organizers

108 training data (with TNM 8th)
56 validation data (with TNM 8th)
216 test data (without TNM)

・fictional reports
・both in Japanese and English

From Nagoya University Hospital

876 clinical data (with TNM 8th)

・2017-2024
・classified by a single radiologist

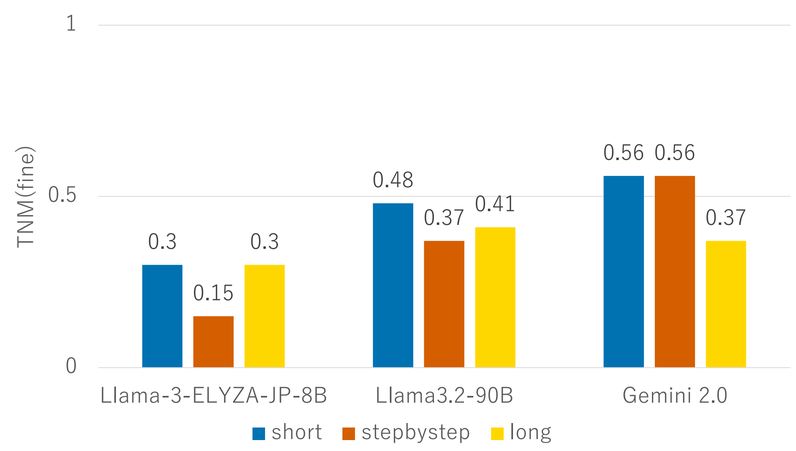
In this study, the accuracy was evaluated using the following indicators.
•TNM (fine): Exact match accuracy for all T, N, and M factors.
•TNM (coarse): Accuracy allowing for differences among Tis/T1mi/T1a/T1b/T1c, T2a/T2b, and M1a/M1b/M1c.

② Make Prompt

Made base Prompt on 30 training data
・3 data use for instruction
・27 data use for compare prompts

Made three types of prompt
(1) Simple prompt
(2) Step-by-step prompts
(3) Detailed prompt

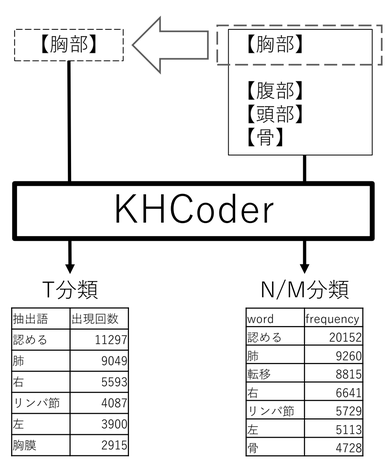
Compare with 3 prompts



Short prompt is better in TNM(fine).

③ Text mining

- 876 clinical cases were used to text-mining analysis using KHCoder.[3]
- Specified the top 50 most frequent words in
  - ・in the chest findings
  - ・in the entire reports
- Made prompts with frequent words.



抽出語	出現回数
認める	11297
肺	9049
右	5593
リンパ節	4087
左	3900
胸膜	2915

word	frequency
認める	20152
肺	9260
転移	8815
右	6641
リンパ節	5729
左	5113
骨	4728

prompt

これらは約2000件のCTレポートを解析した結果、出現頻度が高い単語をまとめたデータです。{頻出単語:出現した回数}の形式でまとめられています。
レポートを解析する際は、これらの単語やその前後を重要視してください。
T分類を評価するときはひとつめのデータセットを、N分類やM分類を評価するときはふたつめのデータセットを参照してください。

###T分類###
[{"word": "認める",
"frequency": 11297},
{"word": "肺",
"frequency": 9049},
……

④ Model comparison

(1) Llama-3-ELYZA-JP-8B[4] with fine-tuning (fine-tuned model)
(2) GPT-O1 pro[5] (commercial model)
(3) Google Notebook LM[6] + source (commercial model)

108 training data
876 clinical data
Use LLAMAFACTORY[7]

Data source for Google Notebook LM

Japanese train data：108
Japanese value data：56

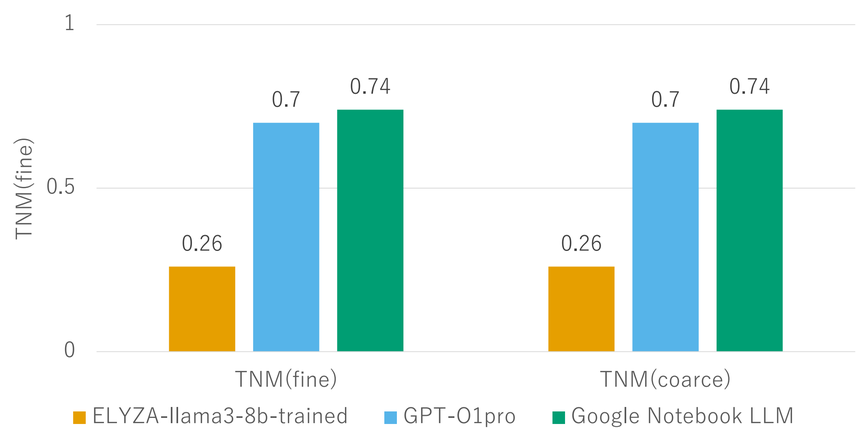
3 cases : Use system prompt
27 cases : Use value model
134 cases : Google Notebook LM

④ Model comparison

The simple prompt (②)

Use 30 training data
・3 data use for instruction
・27 data use for compare models

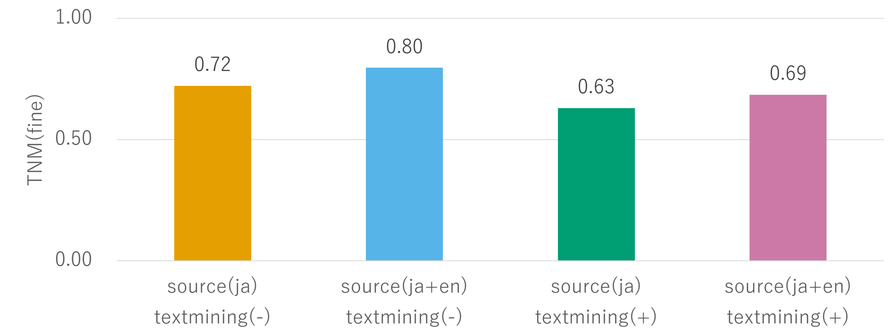
Google Notebook LM is better.



⑤Comparison of Training Source for Google Notebook LM

(1) 108 train data in Japanese
(2) 108 Japanese train data + 164 English data
(3) 108 Japanese train data + text mining (③)
(4) 108 Japanese train data + 164 English data + text mining (③)

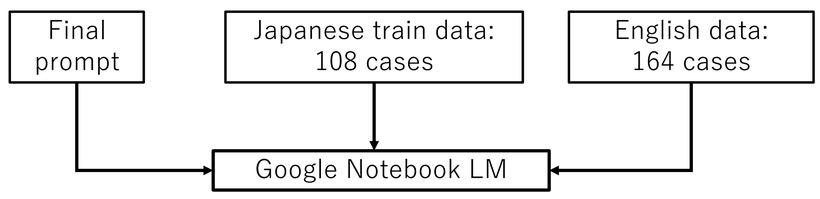
The simple prompt (②)
Use 56 Japanese validation data for compare models.



Japanese and English data without text mining analysis was better.

⑥Final Model

Prompt : the simple prompt (②)
Model : Google Notebook LM (④)
Data : Japanese and English data (⑤)



- If a prediction was not enough, asked only one more time.
- Incomplete predictions were answered as TX or MX.

⑦Final Prompt

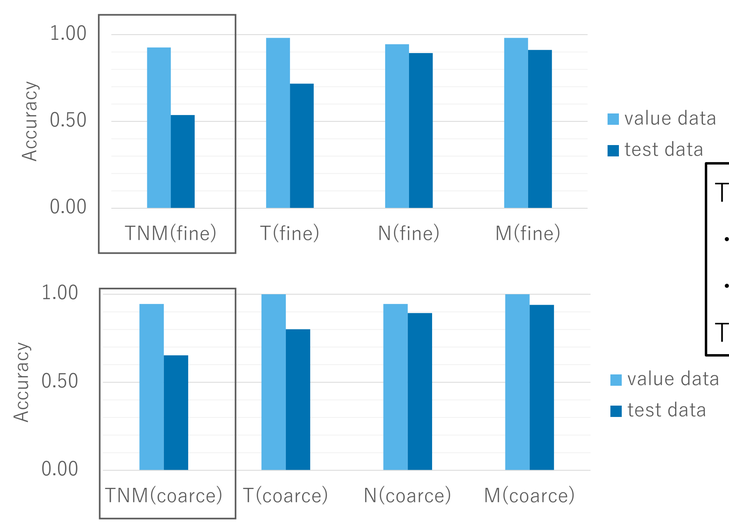
###肺癌のTNM分類についての説明###
==T分類==
T0：{原発腫瘍を認めない}
Tis：{上皮内癌（carcinoma in situ）やAIS：
肺野型の場合は，充実成分径0 cmかつ病変全体径≦3 cm}
T1：[腫瘍の充実成分径≦3 cm，肺または臓側胸膜に覆われている，葉気管支より中枢への浸潤が気管支鏡上認められない
（すなわち主気管支に及んでいない）
さらに、腫瘍の大きさにより以下の4つに分類される。
T1mi：{微少浸潤性腺癌：部分充実型を示し，
充実成分径≦0.5 cmかつ病変全体径≦3 cm}
T1a：{充実成分径≦1 cmでかつTis・T1miには相当しない}
T1b：{充実成分径＞1 cmでかつ≦2 cm}
T1c：{充実成分径＞2 cmでかつ≦3 cm}]
T2：[充実成分径＞3 cmでかつ≦5 cm
または充実成分径≦3 cmでも以下2つのいずれかであるもの :{
a)主気管支に及ぶが気管分岐部には及ばない
b)臓側胸膜や葉間胸膜に浸潤、肺門まで連続する部分的
または一側全体の無気肺か閉塞性肺炎がある}
さらに、腫瘍の大きさにより以下の2つに分類される。:{
T2a：充実成分径＞3 cmでかつ≦4 cm
T2b：充実成分径＞4 cmでかつ≦5 cm}]
T3：[充実成分径＞5 cmでかつ≦7 cm
または充実成分径≦5 cmでも以下2つのいずれかであるもの:{
a)壁側胸膜，胸壁（superior sulcus tumorを含む），
横隔神経，心膜のいずれかに直接浸潤。
b)同一葉内の不連続な副腫瘍結節(肺転移、肺内転移)を持つ}]
T4：{充実成分径＞7 cm，または大きさを問わず横隔膜，縦隔，心臓，大血管，気管，反回神経，食道，椎体，気管分岐部への浸潤，あるいは同側の異なった肺葉内に副腫瘍結節がある}

==N分類==
肺癌の所属リンパ節は以下の通りです。[
鎖骨上窩リンパ節　前斜角筋リンパ節
縦郭リンパ節：[上部気管傍リンパ節　血管前リンパ節、気管後リンパ節　下部気管傍リンパ節　大動脈下リンパ節　大動脈傍リンパ節
気管分岐下リンパ節　食道傍リンパ節
肺靭帯リンパ節}
肺門リンパ節：{主気管支周囲リンパ節　葉気管支間リンパ節}
これらのリンパ節について、以下のように分類します。
N0：{所属リンパ節転移なし}
N1：{同側の気管支周囲かつ/または同側肺門，肺内リンパ節への転移で原発腫瘍の直接浸潤を含める}
N2：{同側縦隔かつ/または気管分岐下リンパ節への転移}
N3：{対側縦隔，対側肺門，同側あるいは対側の前斜角筋，鎖骨上窩リンパ節への転移}

==M分類==
M0：{遠隔転移なし}
M1a：{対側肺内の副腫瘍結節，胸膜または心膜の結節，悪性胸水(同側・対側)，悪性心嚢水}
M1b：{肝臓や副腎、骨など肺以外の一臓器への単発遠隔転移}
M1c：{肺以外の一臓器または多臓器への多発遠隔転移}

###解答の形式###
以下は、あなたの回答が従うべき注意点です。{
・"T*N*M*"のように答えをまとめて出力してください。
・最も可能性が高い候補が複数ある場合は、
最も低いステージの分類をひとつだけを答えてください。
・Inputされた文章以外の情報をTNM分類に使用しないでください。
・ソースで与えられた文章、思考過程、Inputされた文章を
繰り返し出力しないでください。
}

⑧Official Result



The TNM (fine) score
・0.93 for the validation data
・0.54 for the test data.
T score was not good.

Discussion

- T stage of invasive mucinous adenocarcinoma (IMA) was difficult for our model. We need to update prompt for consideration of difficulty of IMA.
- Our model partly failed to classify TNM when diagnostic reports lack a description of TNM precisely(e.g. “T2”).
- The fine-tuning model with our clinical data resulted in decreased performance. It may be because our clinical data differ from the task organizer’s data in their description of findings suggestive of invasion or metastasis.

APPENDIX

1) Japan Radiological Society https://www.radiology.jp/jrs_about/outline.html Accessed February 25, 2025.
2) Japanese Society of Thoracic Radiology https://www.jstr.jp/english Accessed February 25, 2025.
3) KH Coder. Kouichi Higuchi. https://khcoder.net/ Accessed February 11, 2025.
4) elyza/Llama-3-ELYZA-JP-8B. Masato Hirakawa and Shintaro Horie and Tomoaki Nakamura and Daisuke Oba and Sam Passaglia and Akira Sasaki. https://huggingface.co/elyza/Llama-3-ELYZA-JP-8B Accessed February 11, 2025.
5) GPT-O1 pro. Open AI. https://openai.com/ Accessed February 11, 2025.
6) Google Notebook LM. Google https://notebooklm.google/. Accessed February 11, 2025.
7) Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, Zheyan Luo, Zhangchi Feng, Yongqiang Ma. 2024. LLAMAFACTORY: Unified Efficient Fine-Tuning of 100+ Language Models. arXiv:2403.13372