

NTCIR-17 MedNLP-SC Radiology Report Subtask Overview: Dataset and Solutions for Automated Lung Cancer Staging



Yuta Nakamura¹, Shouhei Hanaoka¹, Shuntaro Yada², Shoko Wakamiya², and Eiji Aramaki²
¹The University of Tokyo
²Nara Institute of Science and Technology

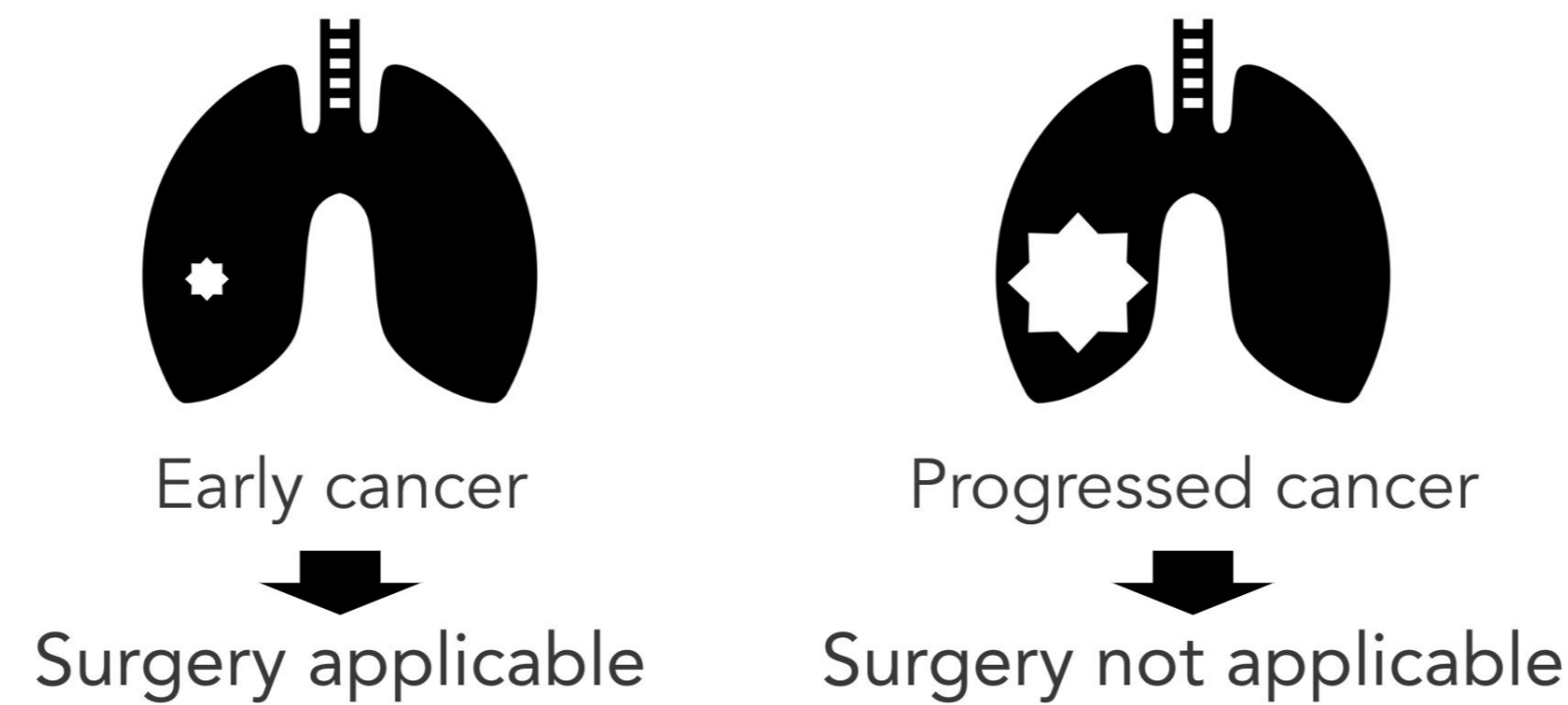
Background

Radiology reports are clinical documents describing CT images



(Created with GPT-4V)

Lung cancer is treated differently depending on the cancer stage



However, the evaluation of the cancer stage is complex and NLP can be helpful

Criteria: the 8th edition by the Japan Lung Cancer Society

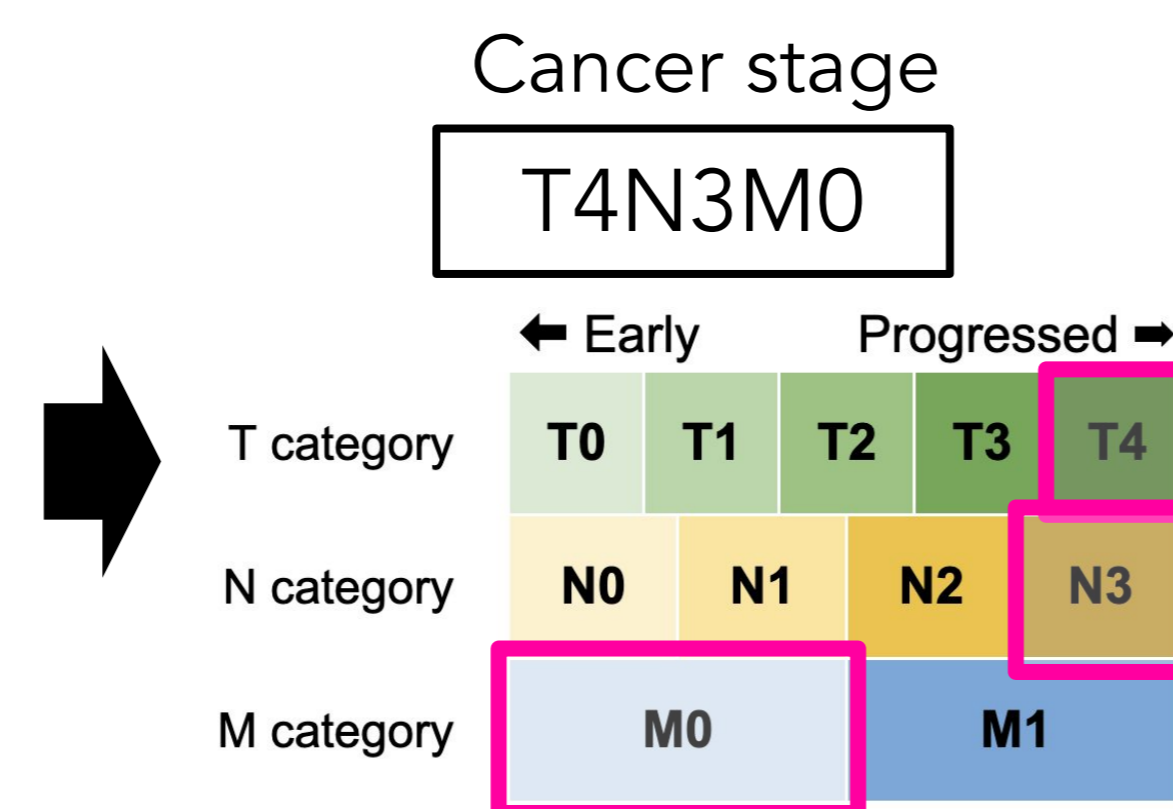
Stage	Definition
T: primary lesion	
TX	原発腫瘍の存在が判定できない、あるいは喀痰または気管支洗浄液細胞診でのみ陽性で画像診断や気管支鏡では観察できない
T0	原発腫瘍を認めない
Tis	上皮内癌 (carcinoma in situ): 肺野型の場合は、充実成分径 0 cm かつ病変全体径 ≤ 3 cm
T1	腫瘍の充実成分径 ≤ 3 cm, 肺または臓側胸膜に覆われている、葉気管支より中樞への浸潤が気管支鏡上認められない (すなわち主気管支に及んでいない)
(T1mi)	微小浸潤性腫瘍: 部分充実型を示し、充実成分径 ≤ 0.5 cm かつ病変全体径 ≤ 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 でも以下のいずれかであるもの
	<ul style="list-style-type: none"> 主気管支に及ぶが気管分枝部には及ばない 臓側胸膜に浸潤 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
(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 でも以下のいずれかであるもの
	<ul style="list-style-type: none"> 壁側胸膜、胸壁 (superior sulcus tumor を含む)、横隔神経、心臓のいずれかに直接浸潤 同一葉内の不連続な副腫瘍結節
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、気管分枝部への浸潤、あるいは同側の異なった肺葉内の副腫瘍結節
N: nodal involvement	
NX	所属リンパ節評価不能
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枝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: distant metastasis	
M0	遠隔転移なし
M1	遠隔転移がある
(M1a)	対側肺内の副腫瘍結節、胸膜または心臓の結節、悪性胸水 (同側・対側)、悪性心嚢水
(M1b)	肺以外の一臓器への単発遠隔転移がある
(M1c)	肺以外の一臓器または多臓器への多発遠隔転移がある

Task scheme

Three-label document classification to assign T, N, and M categories

Radiology report (Japanese)

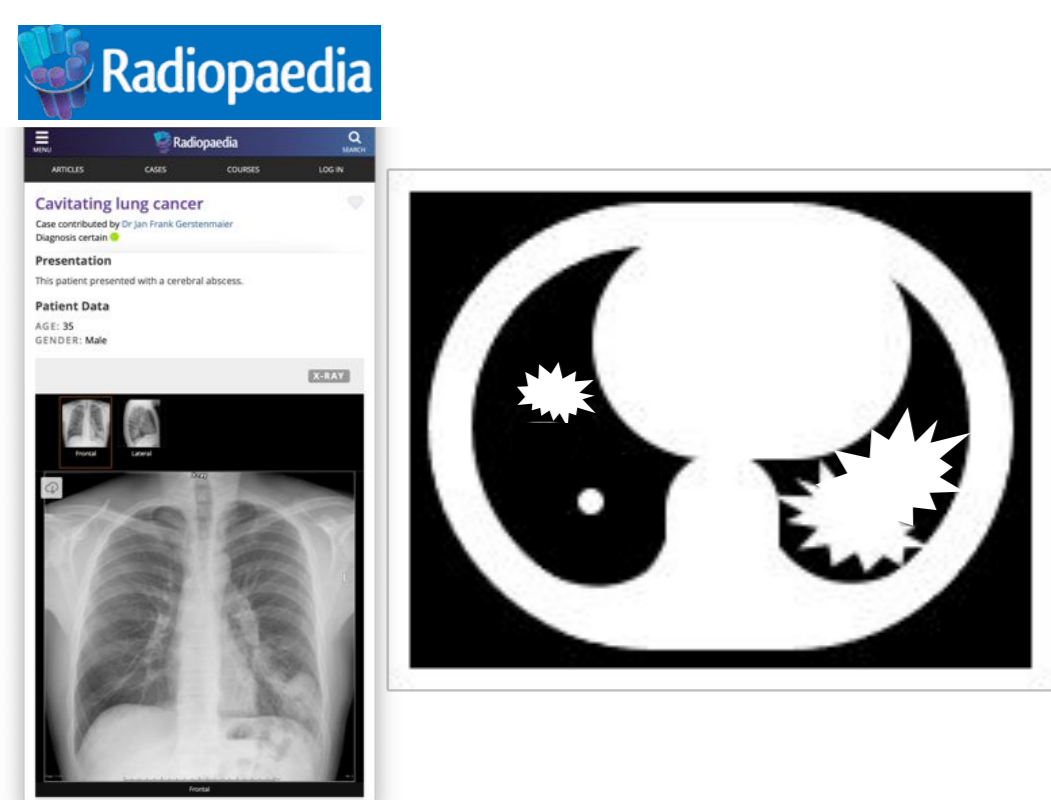
左上下葉に広がる長径 12cm の腫瘍を認めます。既知肺癌が示唆されます。胸膜に広範囲に接しており左第 3 肋骨の破壊を伴っています。肋骨、壁側胸膜浸潤を疑います。左上葉に小結節あり、副腫瘍結節を疑います。左縦隔、両側肺門部リンパ節が腫大、転移を疑います。胸水は認めません。撮像範囲の上腹部臓器に明らかな異常は認めません。



Dataset

Created a Japanese corpus of 243 radiology reports in the courtesy of a teleradiology company (Y's Reading, Inc.)

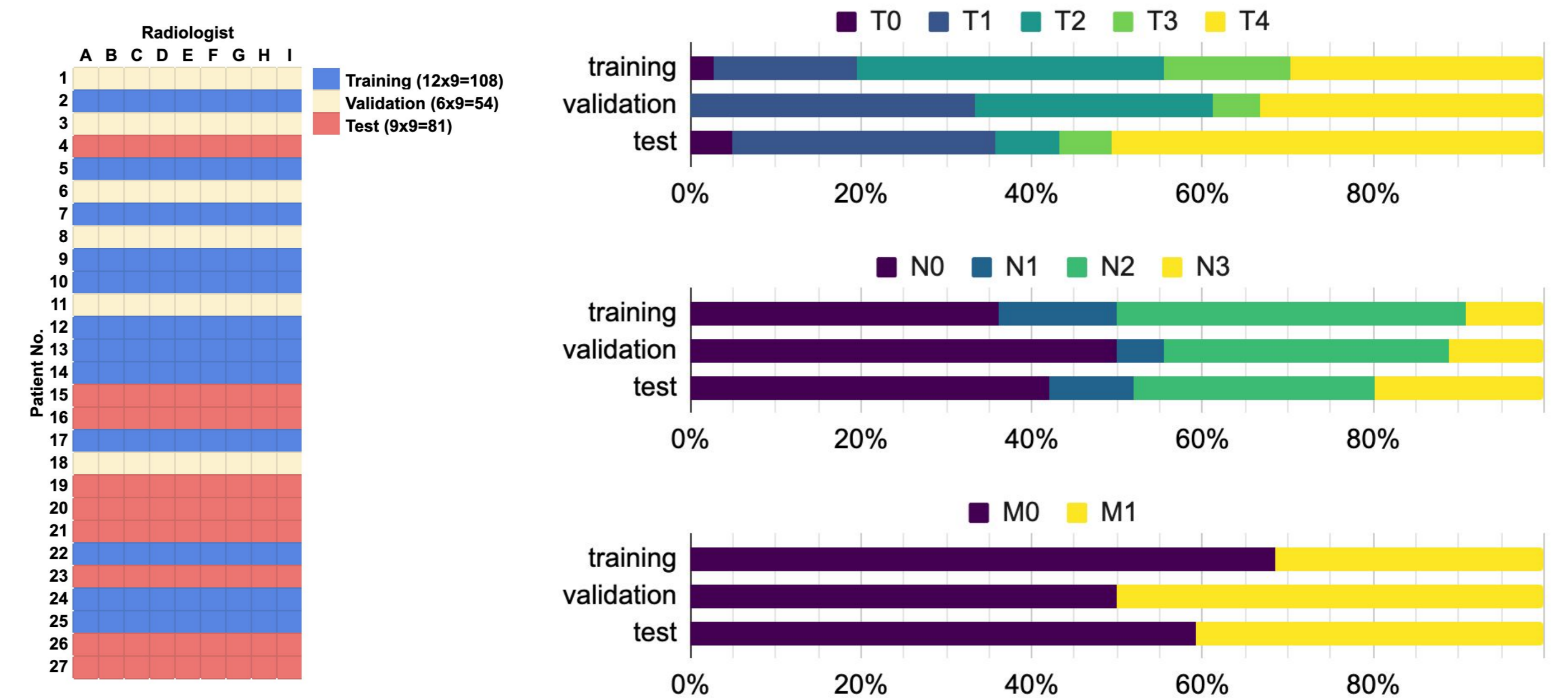
De-identified lung cancer CT images on Radiopaedia



243 documents (27 cases × 9 radiologists)



The split of the dataset was case-based so that radiology reports diagnosing the same case did not overlap beyond partition



Evaluation metrics

- (i) Separate evaluation
 - T accuracy: accuracy score for T category
 - N accuracy: accuracy score for N category
 - M accuracy: accuracy score for M category
- (ii) Joint evaluation
 - Joint accuracy: fraction of cases whose T, N, and M categories were all predicted correctly

Participant solutions and performance

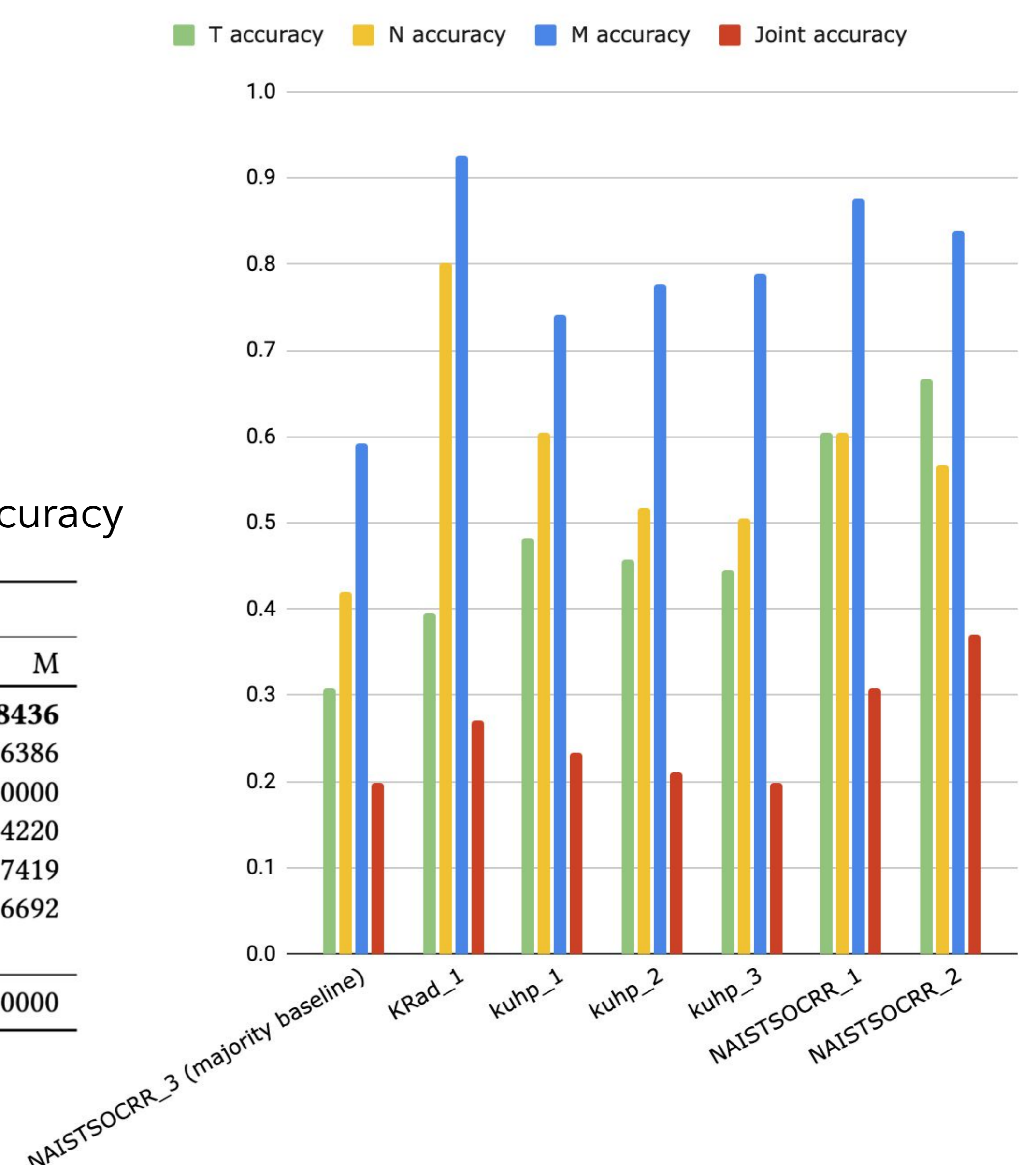
We received 7 solutions from 3 teams

- 2 teams used LLMs (KRad, kuhp)
- 1 team used BERT-based supervised learning (NAISTSOCRR)
- Baseline was provided by NAISTSOCRR

All solutions showed improvement from the baseline

- +10–35% T accuracy, +10–40% N accuracy, +15–30% M accuracy, and +0–20% joint accuracy

System	Model		Accuracy				Weighted kappa		
	Name(s)	Type	T	N	M	Joint	T	N	M
KRad-1	gpt-3.5-turbo	Causal	0.3951	0.8025	0.9259	0.2716	0.3660	0.8422	0.8436
kuhp-1	open-ca1m-7b	Causal	0.4815	0.6049	0.7407	0.2346	0.6536	0.5699	0.6386
kuhp-2	open-ca1m-7b	Causal	0.4568	0.5185	0.7778	0.2099	0.0000	0.0000	0.0000
kuhp-3	open-ca1m-7b	Causal	0.4444	0.5062	0.7901	0.1975	0.5016	0.6453	0.4220
NAISTSOCRR-1	JMedRoBERTa	Bidirectional	0.6049	0.6049	0.8765	0.3086	0.6018	0.6577	0.7419
NAISTSOCRR-2	JMedRoBERTa, Tohoku-BERT-v3, UTH-BERT	Bidirectional	0.6667	0.5679	0.8395	0.3704	0.6556	0.6933	0.6692
NAISTSOCRR-3	(Majority baseline)	-	0.3086	0.4198	0.5926	0.1975	0.0000	0.0000	0.0000



Remaining Issues

- Further exploration for better performance is valuable for real-world applications
- How to extend the process to create open radiology report datasets to non-Japanese languages, especially in English?