Developing Prompts from Large Language Model for Extracting Clinical Information from Pathological and Ultrasound Findings in Breast Cancer

Hyeon Seok Choi[1, 2], Jun Yeong Song[1, 2], Kyung Hwan Shin[1, 2,3], Ji Hyun Chang[1, 2], Bum-Sup Jang[1, 2]

[1] Department of Radiation Oncology, Seoul National University Hospital, Seoul, South Korea
[2] Department of Radiation Oncology, Seoul National University College of Medicine, Seoul, South Korea
[3] Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, South Korea

Background & Purpose

- Radiation therapy (RT) for breast cancer patients considers numerous factors derived from pathological and ultrasound findings.
- However, because these information are often scattered within the reports, manual extraction is laborious and time-consuming.
- With an advent of large language models (LLM) like ChatGPT, which demonstrate exceptional natural language processing capabilities.
- Thus, we aimed to develop prompts tailored for LLM to extract clinical factors from pathological and ultrasound findings, to provide individualized RT for breast cancer patients.

Materials & Methods

- Data were collected from surgical pathology reports and ultrasound reports breast cancer patients
- We extracted information in three methods.
  1) 'LLM method': Develop prompts and extract information using LLM
  2) 'LLM-assisted manual method': manually check the information extracted by 'LLM method'
  3) 'Full manual method': extract information manually
- We calculated the time and cost of collecting information for each method.
- To evaluate the accuracy, 60 patients (2% of all patients) were randomly selected, and the information extracted by the 'LLM method' and the information collected by the 'Full manual' method were compared.

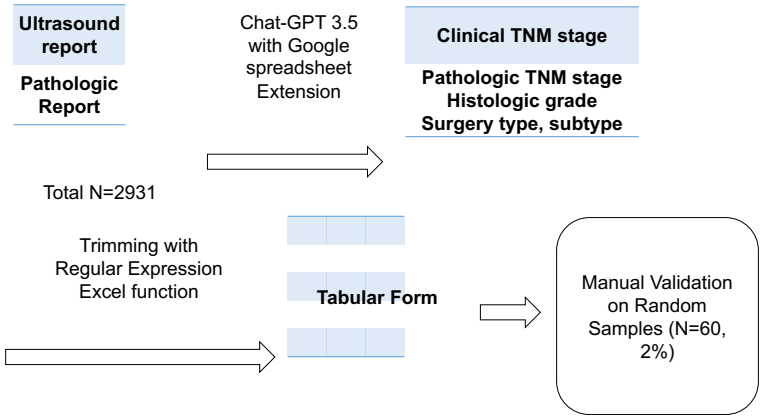

Figure 1. The Schema of the Current Study

Results

- Data from 2,931 patients were collected.
- We developed 12 prompts for Extract function and 12 for Format function to extract and standardize the information.

	Ultrasound Reports
Prompt (Extract Function)	information about largest nodule in diameter or mass with BI-RADS classification of C4 or higher. if there is no C4 or higher nodule, just say 'no cancer'. if not 'no cancer', answer form is 'longest diameter(ex: 1.0cm, 2.5cm by cm), laterality(ex: Rt/Lt), orientation(by clockface(ex: 1H, 11.5H) or by quadrant(ex: SA, center, UO, IL), BI-RADS classification'. all answer is in a single line.
Result	2.3, Lt, 2H, C5/6
Prompt (Format Function)	longest one direction diameter. say just number without unit, not other information. If the unit is mm, change it to cm.
Result	2.3
Prompt (Extract Function)	Pathology Reports the invasive tumor size 또는 종괴의 크기 as a long diameter 또는 the extent of in situ.
Result	1.5 x 0.4 x 2.0 cm (invasive tumor size)
Prompt (Format Function)	T stage. You are breast cancer pathologist interpreting reports with AJCC 8th staging system
Result	T1c

Table 1. The Examples of Prompts Designed, and Their Corresponding Results

- The overall accuracy was 88.3%, and especially for surgery type and lymphovascular invasion, the accuracy was 100%.

Factors	Accuracy (%)	Factors	Accuracy (%)
Clinical T stage	83.3	**Lymph Node Sampling type**	85.0
Clinical N stage	93.3	**Pathologic N stage**	93.3
Tumor Location	63.3	**Metastatic Lymph Node Ratio**	98.3
Surgery Type	100.0	**Extracapsular extension**	86.7
Neoadjuvant Chemotherapy	48.3	**Estrogen receptor**	98.3
Pathologic T stage	76.7	**HER2**	96.7
Histologic grade	98.3	**Ki-67**	93.3
Lymphovascular Invasion	100.0	**Triple-Negative Breast Cancer**	98.3
Resection margin	88.3	**Overall**	**88.3**

Table 2. The Accuracy of the Data Extraction by LLM

- 'LLM-assisted manual' and 'LLM' methods were time- and cost-efficient compared to the 'Full manual' method.

	Full manual		LLM-assisted manual		LLM	
	60 pts	All pts*	60 pts	All pts*	60 pts	All pts
Time spent for data collection	2.6 hrs	127 hrs	1.6 hrs	78.2 hrs	-	-
Time spent for prompt design	-	-	3.5 hrs	3.5 hrs	3.5 hrs	3.5 hrs
Total time spent	2.6 hrs	127 hrs	5.1 hrs	81.7 hrs	3.5 hrs	3.5 hrs
Compared to 'Full manual'	**-**	**-**	**-2.5 hrs**	**45.3 hrs**	**-0.9 hrs**	**123.5 hrs**
Manual data collection wage cost	$19.2	$939.9	$11.8	$578.4	-	-
Prompt design wage cost	-	-	$25.9	$25.9	$25.9	$25.9
GPT API usage Fee	-	-	$1.4	$65.8	$1.4	$65.8
Total estimated cost	$19.2	$939.9	$39.1	$670.1	$27.3	$91.7
Compared to 'Full manual'	**-**	**-**	**-$19.9**	**$269.8**	**-$8.1**	**$848.2**

* extrapolated

Table 3. The Time and Cost Spent on Collecting the Data Using 'Full manual', 'LLM-assisted manual' and 'LLM' Methods

Conclusion

- Developing prompts for LLM to derive clinical factors was efficient and the extracted information was accurate. This study demonstrated the potential of the application of natural language processing using LLM model in breast cancer patients.