LLM-driven Multimodal Target Volume Contouring in Radiation Oncology

Yujin Oh[a,*], Sangjoon Park[b,*], Hwa Kyung Byun[c], Jin Sung Kim[b,**], Jong Chul Ye[a,**]

[a] *Kim Jaechul Graduate School of AI, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea*
[b] *Department of Radiation Oncology, Yonsei College of Medicine, Seoul, Korea*
[c] *Department of Radiation Oncology, Yongin Severance Hospital, Yongin, Gyeonggi-do, Korea*

ARTICLE INFO

ABSTRACT

Target volume contouring for radiation therapy is considered significantly more challenging than the normal organ segmentation tasks as it necessitates the utilization of both image and text-based clinical information. Inspired by the recent advancement of large language models (LLMs) that can facilitate the integration of the textural information and images, here we present a novel LLM-driven multi-modal AI that utilizes the clinical text information and is applicable to the challenging task of target volume contouring for radiation therapy, and validate it within the context of breast cancer radiation therapy target volume contouring. Using external validation and data-insufficient environments, which attributes highly conducive to real-world applications, we demonstrate that the proposed model exhibits markedly improved performance compared to conventional vision-only AI models, particularly exhibiting robust generalization performance and data-efficiency. To our best knowledge, this is the first LLM-driven multimodal AI model that integrates the clinical text information into target volume delineation for radiation oncology.

1. Introduction

Despite the rapid development of Artificial Intelligence (AI) models, there is yet a discernible gap in the realm of medical data processing. Historically, models have predominantly focused on individual data modalities - either visual or linguistic. This approach starkly contrasts with the intrinsic multi-modal practices of physicians, who inherently rely on a confluence of imaging studies and textual electronic medical data for informed decision-making. By understanding diverse data types and their interrelationships, multi-modal AIs would facilitate more accurate diagnoses, personalized treatment development, and a reduction in medical errors by providing a comprehensive view of patient data. For example, in the field of radiation oncology, which is one of the clinical field at the vanguard of potential AI applications and the main focus of this article, the integration of multiple modalities is of paramount importance [1].

For modern intensity-modulated radiation therapy (IMRT) and its inverse planning, two critical components are needed: organs-at-risk (OARs) and the target volume where the dose is prescribed. OARs are defined as the radiosensitive organs susceptible to damage by ionizing radiation during radiation therapy. Traditionally, they were either manually delineated by human experts or

*Co-first authors.
**Co-corresponding authors.
e-mail: jinsung@yuhs.ac (Jin Sung Kim), jong.ye@kaist.ac.kr (Jong Chul Ye)

automatically contoured using atlas-based autocontouring algorithms. However, with the advent of deep learning-based AI models, such tasks have been efficiently accomplished [2, 3]. Notably, these organs, as visualized on planning computed tomography (CT) images, can be directly contoured. This has made the task relatively straightforward, relying solely on imaging.

However, in contrast to OARs segmentation, the task of target volume delineation, which is paramount for treatment planning, has traditionally been the purview of experienced radiation oncologists. This task is perceived as more challenging due to its intrinsic need for the integration of multi-modal knowledge. Although a multitude of segmentation models have been proposed and explored to enhance the precision and efficacy of this task over the last few years [4, 5, 6], a conspicuous gap in research persists, particularly regarding multi-modal target delineation [3]. This is because the delineation of radiation therapy target transcends beyond the mere consideration of visual elements, such as the gross tumor volume [7], and necessitates the incorporation of a myriad of factors, including tumor stage, histological diagnosis, the extent of metastasis, and gene mutation. These factors critically influence the potential for occult metastases, which may compromise the survival outcome of a patient. Areas at elevated risk for such metastatic growth are inclusively treated electively, necessitating a clinical judgment that is deeply rooted in a comprehensive understanding of various data modalities. Furthermore, additional factors, such as a patient's performance status and age, which collectively contribute to the general condition, also exert an impact on treatment target delineation. Given the imperative nature of considering information beyond imaging in target volume delineation, the application of a multi-modal approach in radiation oncology is not merely beneficial but essential for the tasks of the radiation oncology [8]. This is particularly substantiated by the necessity to incorporate textual clinical data, which can significantly influence the identification and subsequent treatment of regions susceptible to occult metastases.

Recently, large language models (LLMs) – AI models proficient in processing and generating text, code, and other data types – have witnessed remarkable advancements [9, 10, 11]. Trained on extensive datasets of text and code, these models discern relationships among varied data types and generate new data, adhering to learned patterns. Furthermore, multimodal data such as images, signals, etc., can be easily integrated into LLMs through adaptors and generative models for image understanding and generation, respectively. Consequently, these models have demonstrated proficiency in a myriad of medical tasks, including multi-modal medical report generation, medical question answering, and multi-modal segmentation with medical images like chest X-rays [12, 13, 14].

Inspired by the multimodal integration capability of LLMs and needs for multimodal information for tumor target delineation, here we present a novel multi-modal clinical target volume (CTV) delineation model by integrating clinical text information through LLM for conditioning a segmentation model. Specifically, by leveraging the textual information from well-trained LLMs through simple prompt tuning, our cross-attention-based segmentation model has adeptly integrated text-based clinical information into the target volume contouring task. More specifically, as illustrated in Fig. 1, we introduce an interactive alignment framework which uses both self-attention and cross-attention mechanisms in a bidirectional manner (text-to-image and image-to-text features), by following the concept of promptable segmentation from Segment Anything Model (SAM) [15]. To further improve the quality of features, we implement the interactive alignment between all the skip-connected image encoder features with the LLM feature. These layer-wise multi-modal features are then combined to jointly predict the target labels through the multi-modal decoder. In this way, we ensure the image encoder to efficiently extract meaningful text-related representations and vice versa. Finally, to transfer the LLM's knowledge while the entire network parameters are kept and achieve superior performance in various downstream

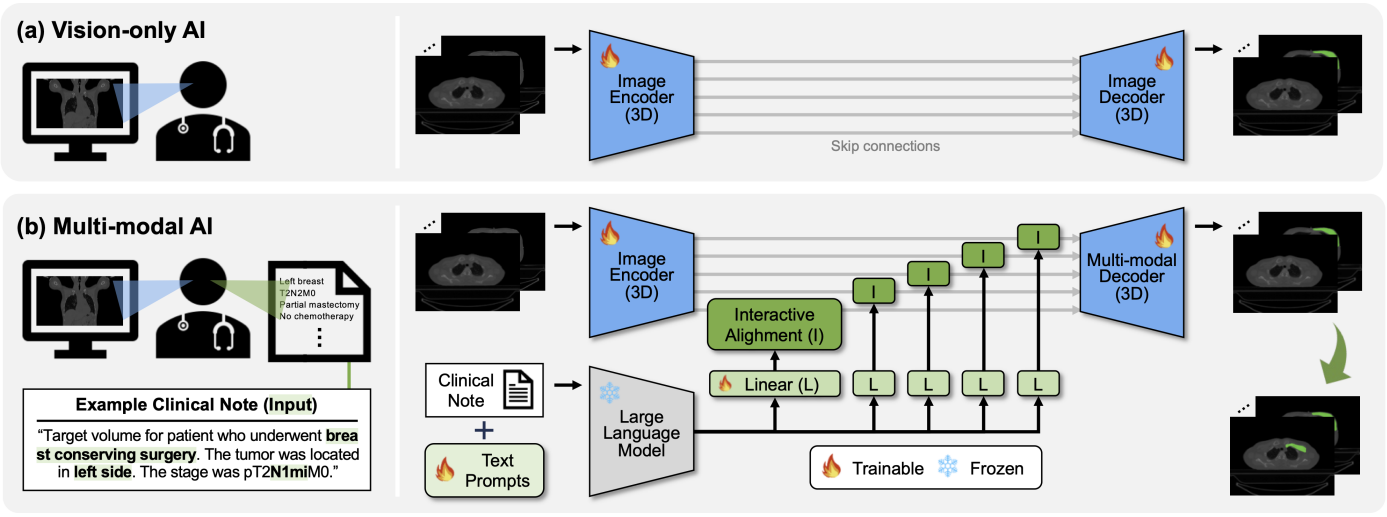

Fig. 1: **Illustration describing a comparative analysis between (a) the vision-only model and (b) the proposed multi-modal large language model in the context of target volume delineation. AI: artificial intelligence.**

tasks [16, 17, 18], we adapt the idea of light-weight text prompts to fully leverage the great linguistic capability of the LLM within the proposed multi-modal AI framework.

As a proof of the concept, the model was applied to a breast cancer target volume delineation task. By utilizing a well-curated, large-scale dataset from two institutions for development and external validation, we verify its capability to integrate pivotal clinical textural information, such as tumor stage, surgery type, and laterality. Experimental results confirm that the model not only demonstrates a significantly enhanced target contouring performance compared to existing vision-only segmentation models but also exhibits behavior that contours targets in accordance with provided clinical information. Notably, the model exhibits superior performance enhancement on an external dataset and shows stable performance gains in data-insufficient settings, demonstrating generalizability and data-efficiency that is not only apt for the characteristics of medical domain data but also resonates with the attributes of clinical experts.

2. Results

Improved CTV Delineation Performance of Multi-modal Model. Table 1 presents a comparative analysis between the vision-only model and our proposed multi-modal model for clinical target volume (CTV) delineation in breast cancer patients. Evidently, the multi-modal model not only exceled the vision-only model in the internal validation data but also significantly outperformed the vision-only model in the external validation data.

Intriguingly, a more pronounced performance improvement was observed in the external validation setting compared to the internal validation. The multi-modal model remained a notably stable performance across the external validation setting. In contrast, the vision-only model showed substantial decrements of around 10% performance drops for Dice and IoU, respectively, thereby highlighting a conspicuous disparity between the two models.

Fig. 2 provides qualitative comparisons of the CTV delineation performance for breast cancer patients. In general, CTV for breast cancer radiation therapy can be categorized into two primary types: one that involves treatment of the breast or chest wall alone, and the other that electively treats the regional lymph nodal area (including axillary, supraclavicular, and internal mammary

Table 1: **Comparison of CTV delineation performance with and without clinical report guidance.**

Model	Internal Validation (N=170)			External Validation (N=98)		
	Dice	IoU	HD-95 ▼	Dice	IoU	HD-95 ▼
Vision-only AI	0.833 (0.816-0.848)	0.726 (0.707-0.746)	6.493 (5.396-7.561)	0.756 (0.723-0.781)	0.625 (0.592-0.653)	18.965 (17.236-20.748)
Multi-modal AI (Ours)	**0.847 (0.830-0.860)**	**0.744 (0.725-0.761)**	**4.436 (3.511-5.491)**	**0.851 (0.837-0.862)**	**0.745 (0.726-0.761)**	**12.650 (9.969-15.242)**

Note. ▼: Lower is better

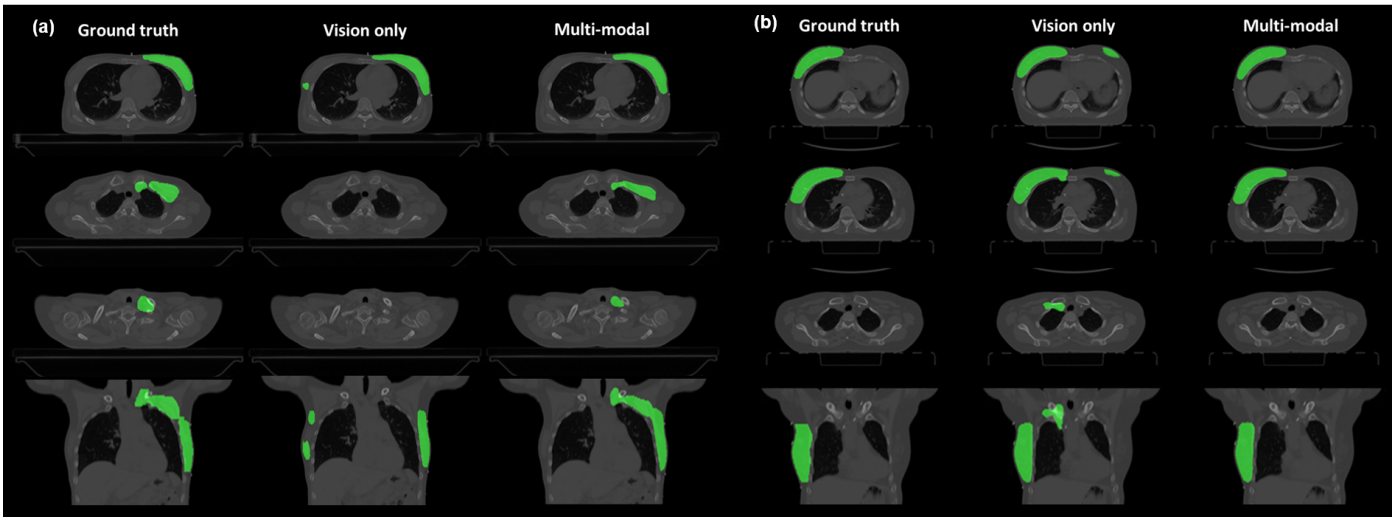

Fig. 2: **Qualitative comparison of target volumes contoured by the vision-only and multi-modal models. The vision-only model (a) fails to incorporate the regional lymph node into the target volume in subjects with positive lymph node metastasis, or (b) erroneously includes the regional lymph node in cases that should be treated with breast-only radiation.**

lymph nodes) in addition to the aforementioned areas, given the frequent metastasis of breast cancer to these regions.

In Fig. 2(a), despite the ground truth label posing CTV on both the breast and regional lymph nodes, the vision-only model only contours the breast alone. Moreover, as the vision-only model lacks information about the laterality of the breast that diagnosed as cancer, partial segmentation masks are observed on the opposite breast. In contrast, the multi-modal model accurately contours the breast and regional lymph nodes that need to be treated as CTV.

In Fig. 2(b), despite early breast cancer case requiring treatment of the breast only, the vision-only model incorrectly includes the regional lymph node as CTV. Moreover, CTVs are extended to the opposite breast. On the other hands, the multi-modal model that integrates the clinical text information accurately contours the requisite treatment areas, encompassing both the breast and the regional lymph nodes, aligning with the ground truth.

Performance Evaluation by Expert Reveals Superiority of Multi-modal Model.. The assessment of the target volume should not be based on mere metric evaluations such as the Dice coefficient, but rather by whether the target volume has been established with appropriate clinical rationale. In the context of breast contouring, this involves considerations such as whether the contouring has been performed on the breast or chest wall, contingent on the type of surgery (breast-conserving surgery or mastectomy), and whether the regional lymph nodal area has been included. If the latter is affirmed, the extent of the included regional lymph nodes becomes important. Therefore, the appropriateness of target contouring should be evaluated by a board-certified radiation oncologist, ensuring a clinically relevant perspective in the assessment. While nuances in details might exist, the evaluation focuses on whether the contouring aptly incorporated the concept of areas that the target should encompass, thereby determining clinical

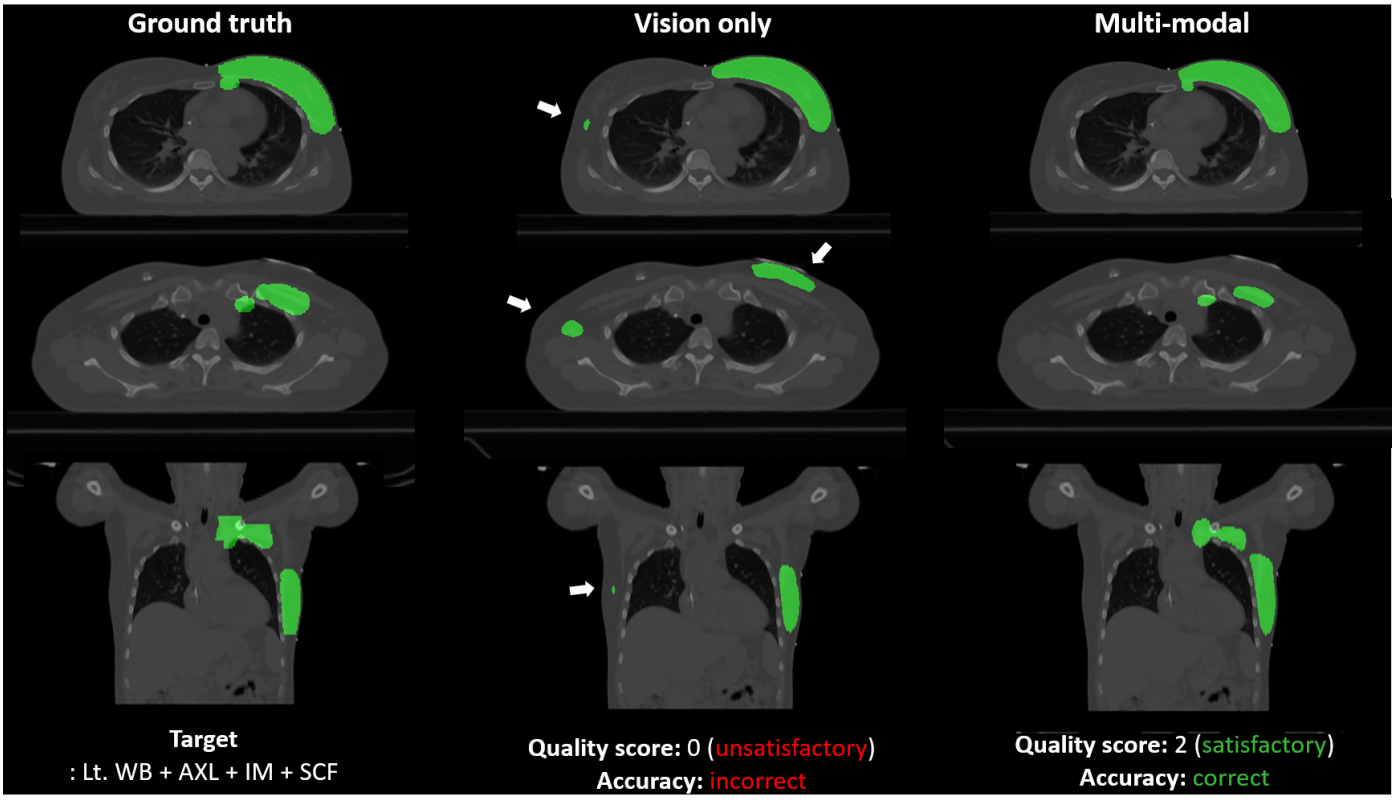

Fig. 3: **An illustrative case demonstrating successful delineation employing a multi-modal model, in contrast to a vision-only model, as evaluated by an expert. Abbreviations: WB, Whole Breast; AXL, Axilla; IM, Internal Mammary; SCF, Supraclavicular Fossa.**

Table 2: **Statistical Analysis of Expert Evaluation.**

Model	**Internal Validation (N=170)**		**External Validation (N=98)**	
	Clinical Accuracy	Clinical Score	Clinical Accuracy	Clinical Score
Vision-only AI	0.741 (0.676-0.806)	0.984 (0.882-1.094)	0.376 (0.286-0.469)	0.457 (0.337-0.582)
Multi-modal AI (Ours)	**0.940 (0.905-0.976)**	**1.506 (1.411-1.601)**	**0.876 (0.796-0.939)**	**1.466 (1.316-1.602)**

Note: Clinical Score 2: *"correct"*; 1: *"minor mistake"*; 0: *"incorrect"*.

accuracy through a binary assessment of correctness. Furthermore, the satisfaction of the target from an expert's perspective is quantified through a clinical score, wherein a score of 2 indicates *"correct"*, 1 denotes *"minor mistake"*, and 0 signifies *"incorrect"*, as shown in Fig. 3.

When evaluated using the proposed scoring method as shown in Table 2, the multi-modal model demonstrated superior performance, with a gain of up to 50% for clinical accuracy and up to 1.0 for clinical score compared to the vision-only model for both the internal and external validation. This performance gain was notably accentuated in the external validation, affirming the robustness and clinical relevance of the multi-modal model across varied datasets and potentially diverse clinical scenarios.

Differential Target Contouring Based on Varied Textual Inputs.. To validate the hypothesis that our multi-modal model genuinely performs CTV delineation based on textual clinical information, we conducted an experiment to assess whether altering the textual clinical information alone would yield different delineation results, even for the same CT, as illustrated in Fig. 4.

As depicted in Fig. 5, the model contours different targets for the same CT, contingent on the provided textual information. In

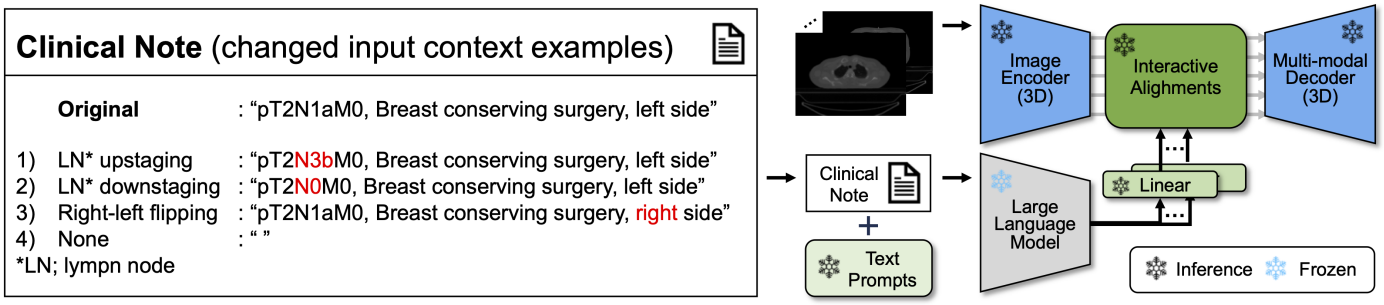

Fig. 4: **Illustration of an experiment involving the modification of a clinical note, given the same computed tomography (CT) image, to ascertain whether the model is reliant on the provided clinical information. Modifications include either the presence of lymph node metastasis or the alteration of right-left orientation. LN, lymph node.**

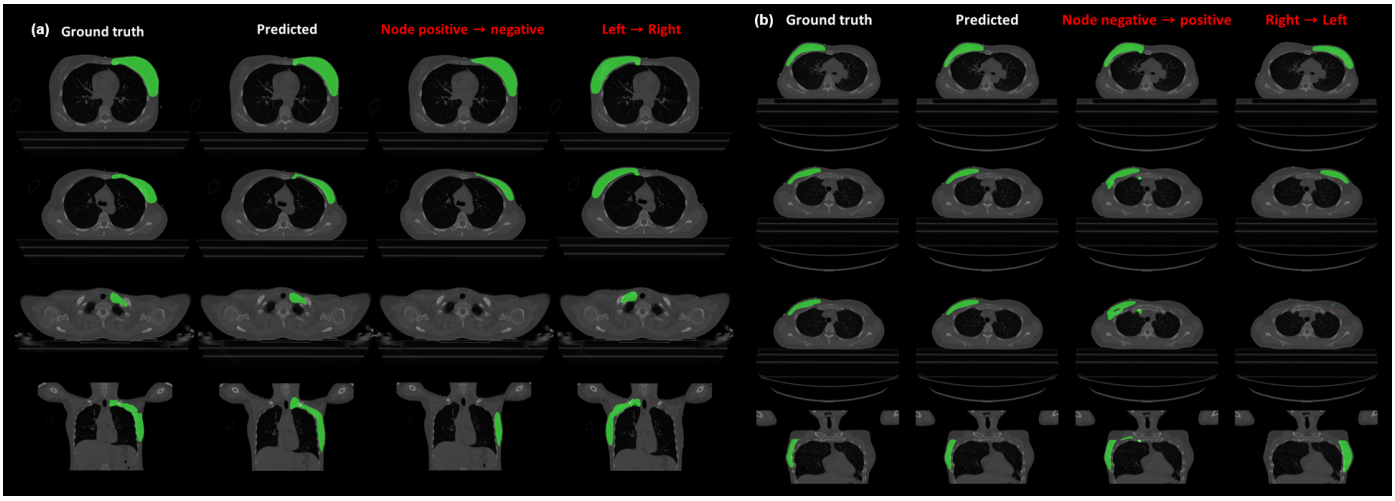

Fig. 5: **Qualitative comparison using modified text input for context-aware target volume delineation. (a) Changing clinical information from positive to negative lymph node metastasis excludes regional lymph nodes from the target volume; altering from left to right breast cancer shifts the contour to the right breast. (b) Conversely, changing from node-negative to node-positive includes the lymph node, and altering laterality shifts the contour from right to left.**

Fig. 5(a), when the textual information for a case, originally positive for lymph node metastasis, was altered to negative, the target volume shifted from encompassing both the breast and regional nodal area to including the breast only. Conversely, as exemplified in Fig. 5(b), when the textual information for a case, initially negative for lymph node metastasis, was changed to positive, the resultant target volume incorporated the regional lymph node. In both cases, when the laterality of the breast cancer was inversely presented in the textual information, delineation occurred on the opposite breast. These experimental outcomes substantiate that our model contours the target volume, referencing not only the imaging but also the textual clinical information.

Data Efficiency and Robustness of the Multi-modal Model.. During the training process of clinical specialists, learning is expedited when textual clinical information is integrated alongside imaging studies, as opposed to focusing on target volume in images alone. This approach facilitates a more rapid assimilation of tendencies and principles of target volume contouring. Consequently, in actual clinical settings, specialists undergo training that amalgamates not only imaging but also pertinent clinical information through text, enabling effective learning even with fewer cases. We sought to determine whether this principle of enhanced learning through the integration of textual clinical information could be applied to our multi-modal LLM. This was evaluated by progressively reducing

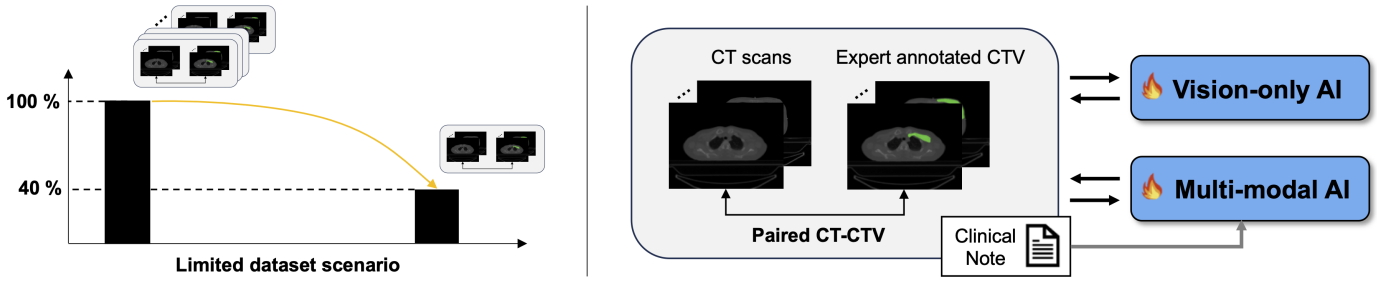

Fig. 6: **An illustration describing an experiment conducted to assess the data efficiency of the multi-modal model, particularly when confronted with a progressively decreasing volume of available data.**

Table 3: **Comparison of CTV delineation performance with and without clinical report guidance on different training dataset size.**

Dataset Size	AI Model	Internal Validation (N=170)			External Validation (N=98)		
		Dice	IoU	HD-95 ▼	Dice	IoU	HD-95 ▼
40%	Vision-only	0.756 (0.737-0.775)	0.623 (0.602-0.645)	12.991 (11.779-14.253)	0.615 (0.577-0.650)	0.468 (0.431-0.502)	20.264 (18.562-21.861)
	Multi-modal	**0.832 (0.816-0.845)**	**0.721 (0.703-0.737)**	**6.138 (5.093-7.235)**	**0.815 (0.800-0.828)**	**0.693 (0.673-0.709)**	**9.187 (7.580-10.891)**
Performance Gain		**10%**	**16%**	53%	**33%**	**48%**	54%
60%	Vision-only	0.798 (0.780-0.815)	0.677 (0.656-0.698)	11.055 (9.838-12.320)	0.662 (0.627-0.694)	0.516 (0.483-0.547)	25.761 (23.877-27.577)
	Multi-modal	**0.847 (0.829-0.860)**	**0.744 (0.725-0.760)**	**4.893 (3.955-5.877)**	**0.829 (0.815-0.840)**	**0.712 (0.694-0.728)**	**4.728 (3.990-5.575)**
Performance Gain		3%	5%	**56%**	25%	38%	**77%**
80%	Vision-only	0.819 (0.802-0.836)	0.707 (0.685-0.728)	8.039 (6.837-9.199)	0.681 (0.656-0.701)	0.526 (0.502-0.547)	20.548 (19.276-21.746)
	Multi-modal	**0.832 (0.816-0.845)**	**0.721 (0.703-0.737)**	**6.138 (5.093-7.235)**	**0.815 (0.800-0.828)**	**0.693 (0.673-0.709)**	**9.187 (7.580-10.891)**
Performance Gain		4%	7%	44%	23%	39%	53%
100%	Vision-only	0.833 (0.816-0.848)	0.726 (0.707-0.746)	6.493 (5.396-7.561)	0.756 (0.723-0.781)	0.625 (0.592-0.653)	18.965 (17.236-20.748)
	Multi-modal	**0.847 (0.830-0.860)**	**0.744 (0.725-0.761)**	**4.436 (3.511-5.491)**	**0.851 (0.837-0.862)**	**0.745 (0.726-0.761)**	**12.650 (9.969-15.242)**
Performance Gain		2%	2%	31%	13%	19%	34%

Note. ▼: Lower is better

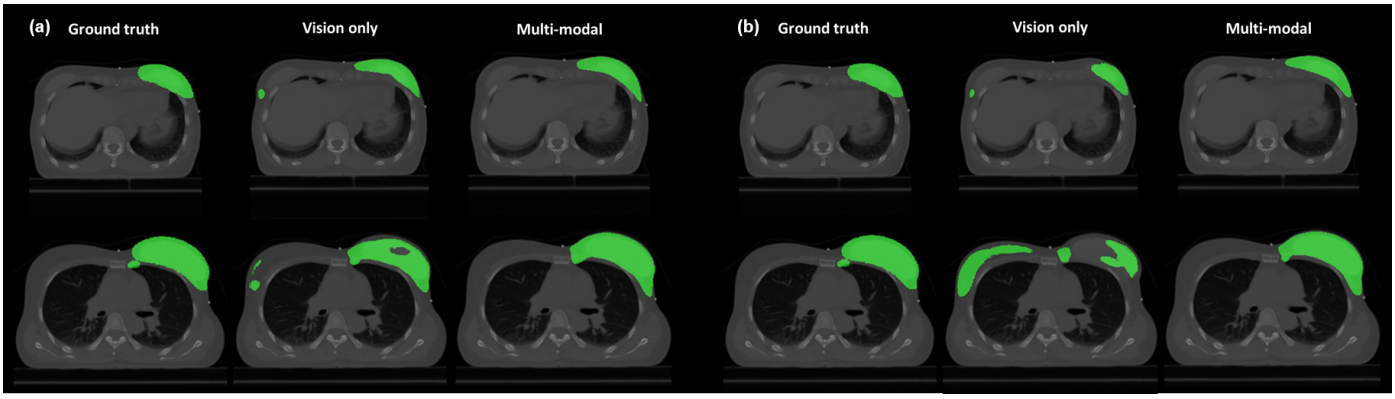

Fig. 7: **Comparison of target contouring by the vision-only and multi-modal models when trained with different amounts of training data: (a) 100% of the data and (b) 40% of the data. The multi-modal model maintains adequate target contouring performance, while the vision-only model does not.**

the amount of data available for learning, as illustrated in Fig. 6.

As presented in Table 3, the multi-modal model demonstrated its data efficiency by maintaining stable performance above 80% of Dice and around 70% of IoU even when data availability was limited, in stark contrast to the vision-only model with performance drop below 65% of Dice and 50% of IoU. This difference was particularly pronounced in the external validation data.

We further analyzed performance gain of the multi-modal model compared to the vision-only model with regard to training dataset size in Table 3. In detail, as depicted in Fig. 7, even when 100% of the data was available for training, the multi-modal model

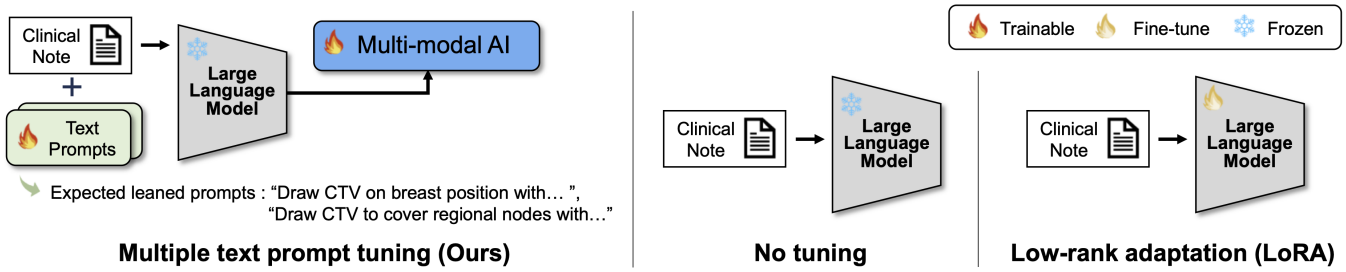

Fig. 8: **Illustration delineating a comparative analysis among diverse fine-tuning methodologies applied to the large language model. AI refers to Artificial Intelligence, and CTV denotes Clinical Target Volume.**

Table 4: **Ablation studies on network components.**

	Text Encoder	Fine-tuning Method	Internal Validation (N=170)			External Validation (N=98)		
			Dice	IoU	HD-95 ▼	Dice	IoU	HD-95 ▼
Ours	Llama2-7B-chat	Multiple text prompts	0.847 (0.830-0.860)	0.744 (0.725-0.761)	4.436 (3.511-5.491)	**0.851 (0.837-0.862)**	**0.745 (0.726-0.761)**	12.650 (9.969-15.242)
(a)	Llama2-7B-chat	One text prompt	0.846 (0.829-0.861)	0.745 (0.726-0.762)	5.248 (4.134-6.490)	0.835 (0.819-0.849)	0.723 (0.703-0.740)	18.579 (16.394-20.728)
		LoRA	0.842 (0.824-0.856)	0.738 (0.718-0.756)	6.625 (5.287-8.004)	0.823 (0.796-0.844)	0.712 (0.684-0.736)	16.185 (13.910-18.509)
		No tuning	0.840 (0.824-0.855)	0.736 (0.716-0.755)	4.621 (3.771-5.552)	0.751 (0.718-0.776)	0.619 (0.587-0.648)	16.734 (15.071-18.450)
(b)	Llama2-7B	Multiple text prompts	0.847 (0.831-0.860)	0.744 (0.726-0.761)	**4.151 (3.397-5.005)**	0.827 (0.811-0.839)	0.710 (0.690-0.727)	**5.191 (3.912-6.734)**
	CLIP text encoder		**0.851 (0.839-0.863)**	**0.749 (0.732-0.765)**	4.678 (3.715-5.733)	0.822 (0.803-0.838)	0.706 (0.683-0.727)	21.793 (19.484-23.986)

Note. ▼: Lower is better

suggested a more accurate target volume compared to the vision-only model with a margin of 2% and 13% of Dice for internal and external validation, respectively. While the vision-only model did produce generally accurate target volume contouring results, the disparity between the models became glaring when only 40% of the data was utilized for training. In this scenario, the multi-modal model continued to suggest target volumes analogous to those suggested with full data availability. In contrast, the vision-only model exhibited a complete failure, providing inappropriate delineation outputs on both breasts and thereby revealing a significant performance gap of over 10% and 33% of Dice for internal and external validation, respectively. These results demonstrate that our multi-modal model exhibit stable performance in data-insufficient settings.

Exploring Textual Clinical Information Provision Methods in the Multi-modal Model.. Various methodologies could be employed to introduce textual clinical information into the multi-modal model, as illustrated in Fig. 8. These include utilizing a single or multiple text prompts through prompt tuning, directly employing a pre-trained LLM without tuning, and employing LLM with low-rank adaptation (LoRA) fine-tuning [19]. We conducted experiments to evaluate which of these tuning methods proves most effective in enhancing the performance of the multi-modal model in the context of integrating textual clinical information for delineation tasks. We further altered the text encoder module with the traditional CLIP text encoder or a variant of LLM to confirm the proposed version of LLM archives the best performance.

As indicated in Table 4(a), instances utilizing prompt tuning consistently demonstrated superior performance relative to those employing LoRA fine-tuning. Additionally, the use of multiple text prompts revealed a marked improvement in performance compared to scenarios where a single prompt was utilized. Surprisingly, the performance gain was emphasized in the external validation setting. The use of multiple text prompts was the only method which avoided overfitting issues during fine-tuning. The results indicate that the introduced multiple prompts were optimized to efficiently extract the text-related image features regardless of the input data distribution. In Table 4(b), by altering the backbone structure of the text encoder, we confirmed the used version of LLM archived the best performance for both the internal and external validation. Consequently, our entire experimental procedure was

conducted employing Llama2-7B-chat with prompt tuning method, specifically utilizing multiple text prompts.

3. Discussion

Despite the promising outcomes demonstrated by AI models in various studies, a notable limitation prevalent in the field of medical AI has been the predominant development of models tailored for singular, specialized tasks [20]. For instance, models have been specifically designed and trained to excel in a singular task, such as segmentation [6, 4], diagnosis [21, 22], or prognosis prediction [23, 24], without the adaptability to transition across various tasks. While these specialized models perform commendably within their designated task, they lack the flexibility to navigate the complex challenges in the medical domain, where the ability to integrate, and concurrently process diverse tasks is crucial.

The horizon of medical AI is envisaged to be dominated by the era of multi-modal medical foundation models, as they promise to bridge the gap between specialized task-specific models and a more holistic, multi-tasking AI paradigm [12, 14, 13]. The concept of a multi-modal foundation model, as defined and reviewed in a recent review [25], emphasize the imperative of multi-modality, with a particular emphasis on the integration of text and image through vision-language models as a key component.

In the nascent stages of applying vision-language models to the medical domain, initial research endeavors have predominantly focused on the most simple form of vision-text paired data, such as chest radiographs [26]. These studies have explored various tasks, including zero-shot classification [27], report generation [28, 29], and text-guided segmentation [30, 31]. However, the field of radiation oncology emerges as a particularly potent application area for such models [8].

Radiation oncology exemplifies a robust case for the adoption of multi-modality, underpinned by two fundamental factors [1]. Firstly, decision-making in Radiation Oncology, especially in determining treatment scope and dose, extends beyond imaging to include a plethora of clinical information, such as surgical notes, pathology reports, and electronic medical records, which can be conveyed textually. Secondly, the integration of prior knowledge, including standard treatment guidelines and radiation oncology textbooks, is vital for informed treatment decision-making, with these guidelines also being expressible in textual formats. Consequently, the necessity for multi-modality is markedly emphasized in Radiation Oncology (Fig. S1).

Consequently, we have applied LLMs in our research. Our model introduces several novel aspects and has demonstrated commendable results by accurately segmenting radiation therapy target volume based on clinical information, thereby achieving absolute performance where the multi-modal model surpasses the vision-only model. It also exhibits a pronounced performance differential in external validation settings and demonstrates data-efficiency in data-insufficient settings.

This resonates intriguingly with the clinical implications, especially mirroring the learning trajectory and characteristics of clinical experts. In the clinical training of experts, reliance is placed on multi-modality information; learning is not confined to either images or text but is rather a confluence of both, facilitating the inference of text-image relationships and enabling effective learning even with relatively fewer cases. This aspect of the clinical learning paradigm, being data-efficient, aligns seamlessly with our proposed multi-modal model.

The decrement in classical AI-driven delineation generalization performance is often attributed to variations in image acquisition settings and characteristics of devices from different vendors, among other factors. Nonetheless, the ability of clinical experts to perform target contouring is scarcely influenced by external factors such as CT scanning conditions. This is because linguistical concepts embodied in textual clinical information, are independent of such acquisition settings. Therefore, it is plausible that our

model, which learns in conjunction with such textual clinical information by leveraging the great linguistic capability of LLMs, demonstrates particularly commendable performance in external validation settings. This characteristic is particularly optimal for the medical domain, where training data is often limited and stable generalization performance is a prerequisite across varied external settings, thereby heralding a promising future for the application of multi-modal models in medical AI.

Our study has several limitations. First, our evaluation was confined to patients at their initial diagnosis, leaving a scope for further exploration into varied patient scenarios and treatment stages, which could potentially influence the model's applicability and performance. Second, the model does not incorporate considerations for radiation therapy doses in target volume contouring, presenting an opportunity to explore how dose-related variables could be integrated to enhance delineation and treatment planning in future studies. Lastly, while the model utilized refined, rather than raw, clinical data, future research could explore mechanisms for automating the data refinement process or developing capabilities to process raw clinical data, thereby reducing the need for manual intervention and potentially uncovering additional insights from unprocessed data.

Despite aforementioned limitations, our research serves as a pivotal step towards the multi-modal models in the field of radiation oncology, verifying the clinical utility and emphasizing the significance of intertwining textual clinical data with medical imaging. The model illuminates a pathway for crafting more adaptable and clinically pertinent AI models in medical imaging and treatment planning. Future research will likely refine and broaden such models, closer to harnessing the full potential of multi-modal medical foundation models in elevating clinical decision-making and patient care.

4. Methods

4.1. Definition of task

In contrast to traditional segmentation tasks, where the primary objective is to segment visible portions in an image, clinical target volume (CTV) delineation in radiation therapy necessitates the consideration of additional clinical information. In specific, when defining the target volume for radiation therapy, one must consider not only the anatomy visible on the CT scan but also various factors such as the type of primary tumor, histology type, cancer stage (TNM stage), patient age, and performance status.

Taking breast cancer as an example, in early-stage cases (e.g., stage I) where there is no regional lymph node metastasis, only the whole breast is included in the radiation therapy target volume. However, in advanced stages (e.g., stage IIIB) where regional lymph node metastasis is identified during surgery, there is a need for elective nodal irradiation across all regional nodal areas. However, such distinctions are not discernible during the CT simulation for post-operative radiation therapy planning and require acquisition through other forms of information. Consequently, we aimed to develop a model that can consider clinical information such as primary tumor type, stage, age, and performance status in a manner akin to an experienced radiation oncologist by providing such data in the form of textual information to a multi-modal model.

Among the primary cancer types, we targeted breast cancer. This was predicated on the fact that breast cancer presents with relatively uniform guidelines for target delineation according to the clinical information including primary tumor location, size, and the presence of nodal metastasis, etc. Furthermore, the inter-observer variability in target delineation for breast cancer is also expected to be small compared with other cancer types. Within the task of radiation therapy target delineation for breast cancer, we exclusively incorporated cases of patients at their initial diagnosis of breast cancer. This decision was based on the understanding

that treatments with aims such as salvage or palliative often exhibit significant variability according to the preferences of the physicians as well as the patients, and other circumstances.

4.2. Details of datasets.

For model development and internal validation, we acquired data from 844 patients treated at the Department of Radiation Oncology at Yonsei Cancer Center between January 2009 and December 2022. These patients had been initially diagnosed with the breast cancer and underwent radiation therapy post-curative surgery with the primary objective of preventing recurrence. We not only utilized their simulation CT images and CTVs for radiation therapy, but also incorporated text-based clinical information that is essential for precise target delineation. This additional information included the location of the primary cancer, type of surgery undertaken, disease stage, and the status of nodal metastasis. Example of clinical information as for textual input used in our study is shown in Table 5. The clinical information was prepared by the tabular format derived from raw clinical data. The resulting clinical note was then structured into text prompt according to custom criteria. Initially, these criteria were devised by a board-certified radiation oncologist. Subsequent refinement was achieved through ablation studies on the components of the raw clinical data to construct the most effective text prompt, incorporating selected clinical contexts. The resulting examples of text prompts are illustrated in the right-most column of Table 5.

Table 5: Examples of the input clinical information

	Example Raw Clinical Data								**Example Input Clinical Information**
	Age	**Location**	**T stage**	**N stage**	**M stage**	**Surgery type**	**Chemotherapy**	**Pathology**	
Criteria		✓		✓		✓			
Example #1	61	Left breast	2	1mi	0	BCS	Adjuvant	IDC	N1mi, breast conserving surgery, left side
Example #2	73	Right breast	4d	2	0	Total	Neoadjuvant	ILC	N2, total mastectomy surgery, left side
Example #3	42	Both breast	Left - 1c Right - is	0	0	BCS	None	IDC	N0, breast conserving surgery, left side; N0, breast conserving surgery, right side

Note. BCS: breast conserving surgery; Total: total mastectomy surgery; IDC: Invasive ductal carcinoma; ILC: Invasive lobular carcinoma

To better reflect real clinical application, the ideal approach for external validation needs the use of patient data acquired under different conditions and with equipment from a different vendor. Therefore, we utilized data from 98 patients treated at the Department of Radiation Oncology at Yongin Severance Hospital between January 2018 and December 2022. These patients, like the previous cohort, were initially diagnosed with breast cancer and underwent radiation therapy following curative surgery to prevent recurrence. We confirmed that the external cohort was non-overlapping with those included in the model development nor internal validation. Alongside the CT simulation images and CTVs, we also incorporated text-based clinical information.

Table 6 provides a detailed description of patient characteristics. Across the train, internal, and external validation sets, distributions of factors such as location and T stage were observed to be consistent. The proportion of patients with lymph node metastasis and those undergoing total mastectomy was higher in the train and internal validation sets than the external validation set. Consequently, the percentage of patients receiving irradiation to the chest wall and regional lymph nodes was also higher in the train and internal validation sets compared to the external validation set.

4.3. Details of Implementation.

The schematic of our multi-modal AI is illustrated in Fig. 1. For the image encoder/decoder and the large language model (LLM), we employed the 3D U-Net [32] and the pre-trained Llama2-7B-chat [10] model, respectively. For the interactive alignment

Table 6: Details of data partitioning and patient characteristics

Characteristics	Training and Internal Validation		External Validation
	Train (N=674)	**Validation (N=170)**	**Validation(N=98)**
Age	53.6 (27-87)	54.3 (28-83)	53.2 (31-86)
Location			
Left	338 (50.1%)	86 (50.6%)	49 (50.0%)
Right	312 (46.3%)	80 (47.1%)	49 (50.0%)
Both	24 (3.6%)	4 (2.4%)	0 (0.0%)
T stage			
Tis	71 (10.2%)	28 (16.1%)	19 (19.4%)
T1	261 (37.4%)	63 (36.2%)	41 (41.8%)
T2	264 (37.8%)	63 (36.2%)	32 (32.7%)
T3	64 (9.2%)	13 (7.5%)	5 (5.1%)
T4	38 (5.4%)	7 (4.0%)	1 (1.0%)
N stage			
N0	324 (46.4%)	84 (48.3%)	69 (70.4%)
N1	220 (31.5%)	56 (32.2%)	17 (17.3%)
N2	101 (14.5%)	21 (12.1%)	9 (9.2%)
N3	53 (7.6%)	13 (7.5%)	3 (3.1%)
Surgery type			
Partial mastectomy	458 (65.6%)	128 (73.6%)	89 (90.8%)
Total mastectomy	240 (34.4%)	46 (26.4%)	9 (9.2%)
Neoadjuvant chemotherapy			
Yes	304 (43.6%)	69 (39.7%)	25 (25.5%)
No	394 (56.4%)	105 (60.3%)	73 (74.5%)
Target			
Breast only	249 (35.7%)	77 (44.3%)	62 (63.3%)
Breast + RNI	209 (29.9%)	51 (29.3%)	27 (27.6%)
CW only	31 (4.4%)	2 (1.1%)	1 (1.0%)
CW + RNI	210 (30.1%)	44 (25.3%)	8 (8.2%)

Note: RNI; regional node irradiation, CW; chest wall

modules, we utilized the two-way transformer modules of SAM [15]. During training, we let the entire LLM frozen, while made the image encoder/decoder modules, the interactive alignment modules and their corresponding linear layers, and the text prompts as trainable parameters. Details of the interactive alignment modules and the text prompts for performing CTV delineation are deferred to Supplementary Section 4.5 and 4.5. Details of the network training complexity is further specified in Supplementary Table S3.

When pre-processing the data, all the chest CT images and CTVs were initially re-sampled to have an identical voxel spacing of $1.0 \times 1.0 \times 3.0\ mm^3$. The image intensity values were truncated between -1,000 and 1,000 of Hounsfield unit (HU), and linearly normalized within a range between 0 and 1.0. When training the network, a 3D patch with a size of $384 \times 384 \times 128$ pixels was randomly cropped to cover the entire breast alongside with its paired clinical note with batch size of 2. When evaluating the trained network, the entire 3D CT image was tested using sliding windows with a 3D patch with a size of $384 \times 384 \times 128$ pixels. As baseline methods, we utilized the pre-trained Llama2-7B [10] model and the pre-trained CLIP ViT-B/16 [33] model as the text encoder.

As the loss function, we computed both the binary cross-entropy loss and the Dice loss, with the weight value for each loss as 1.0, respectively. The network parameters were optimized using AdamW [34] optimizer with a learning rate of 0.0001, until the training epoch reaching 100. We implemented the network using the open-source library MONAI[1]. All the experiments were conducted using the PyTorch [35] in Python using CUDA 11.4 on either NVIDIA RTX A6000 48GB or NVIDIA Tesla A100 40GB.

4.4. Details of Evaluation.

To quantitatively evaluate the CTV delineation performance, we calculated Dice coefficient (Dice), Intersection over Union (IoU), and the 95th percentile of Hausdorff Distance (95-HD) [36] to measure spatial distances between the ground-truth and the predicted contours. When calculating the 95-HD, all the measured distances in the pixel unit are converted with respect to the original pixel resolution, and the results are expressed in centimeters (cm). For statistical analysis, we used the non-parametric bootstrap method to calculate the 95th percentile of confidence intervals for each metric. We randomly sampled the total size of dataset from the original dataset while allowing replacement for 1,000 times, repeatedly. Then, the 95th percentile of confidence intervals were estimated from the relative frequency distribution of each trial.

4.5. Ethic committee approval.

The hospital data deliberately collected for this study were ethically approved by the Institutional Review Board of each hospital (approval numbers of 2023-1460-001 and 2023-0364-001) and the requirement for informed consent was waived due to the retrospective nature of the study.

Correspondence

Correspondence and requests for materials should be addressed to Jin Sung Kim (email: jinsung@yuhs.ac) and Jong Chul Ye (email: jong.ye@kaist.ac.kr).

[1] https://monai.io/

Acknowledgement

This research was supported by Basic Science Research Program through the NRF funded by the Ministry of Education under Grant RS-2023-00242164.

Author Contributions

Y.O. conducted all experiments, extended the code, and contributed to manuscript preparation. S.P. conceptualized the study, gathered and labeled the data, and also contributed to manuscript preparation. H.K.B. was responsible for data collection. J.S.K. and J.C.Y. provided supervision throughout the project, from conception to discussion, and assisted in preparing the manuscript.

Competing Interests

The authors declare that they have no competing financial interests.

Data Availability

The data utilized for this study are not publicly available due to patient privacy obligations. Interested researchers may request access for research purposes by contacting the corresponding author, J.C.Y., at (jong.ye@kaist.ac.kr). Data sharing is permissible following IRB approval and de-identification, accompanied by a signed data transfer and usage agreement. Initial request responses will be provided within 10 working days. Data usage is restricted to research purposes only, and redistribution is prohibited.

Code availability

The Pytorch codes for the proposed Multi-modal AI used in this study is available at the following Github repository at https://github.com/tvseg/MM-LLM-RO.

References

[1] Huynh, E. *et al.* Artificial intelligence in radiation oncology. *Nature Reviews Clinical Oncology* **17**, 771–781 (2020).

[2] Shi, F. *et al.* Deep learning empowered volume delineation of whole-body organs-at-risk for accelerated radiotherapy. *Nature Communications* **13**, 6566 (2022).

[3] Zhang, L. *et al.* Segment anything model (sam) for radiation oncology. *arXiv preprint arXiv:2306.11730* (2023).

[4] Chung, S. Y. *et al.* Clinical feasibility of deep learning-based auto-segmentation of target volumes and organs-at-risk in breast cancer patients after breast-conserving surgery. *Radiation Oncology* **16**, 1–10 (2021).

[5] Offersen, B. V. *et al.* Estro consensus guideline on target volume delineation for elective radiation therapy of early stage breast cancer. *Radiotherapy and oncology* **114**, 3–10 (2015).

[6] Choi, M. S. *et al.* Clinical evaluation of atlas-and deep learning-based automatic segmentation of multiple organs and clinical target volumes for breast cancer. *Radiotherapy and Oncology* **153**, 139–145 (2020).

[7] Guo, Z., Guo, N., Gong, K., Li, Q. *et al.* Gross tumor volume segmentation for head and neck cancer radiotherapy using deep dense multi-modality network. *Physics in Medicine & Biology* **64**, 205015 (2019).

[8] Liu, C. *et al.* Artificial general intelligence for radiation oncology (2023). 2309.02590.

[9] Bubeck, S. *et al.* Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712* (2023).

[10] Touvron, H. *et al.* Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288* (2023).

[11] Liu, Z. *et al.* Radiology-gpt: A large language model for radiology. *arXiv preprint arXiv:2306.08666* (2023).

[12] Moor, M. *et al.* Foundation models for generalist medical artificial intelligence. *Nature* **616**, 259–265 (2023).

[13] Singhal, K. *et al.* Large language models encode clinical knowledge. *arXiv preprint arXiv:2212.13138* (2022).

[14] Tu, T. *et al.* Towards generalist biomedical ai. *arXiv preprint arXiv:2307.14334* (2023).

[15] Kirillov, A. *et al.* Segment anything. *arXiv preprint arXiv:2304.02643* (2023).

[16] Kim, K., Oh, Y. & Ye, J. C. Zegot: Zero-shot segmentation through optimal transport of text prompts. *arXiv preprint arXiv:2301.12171* (2023).

[17] Jia, M. *et al.* Visual prompt tuning. *arXiv preprint arXiv:2203.12119* (2022).

[18] Zhou, K., Yang, J., Loy, C. C. & Liu, Z. Conditional prompt learning for vision-language models (2022). 2203.05557.

[19] Hu, E. J. *et al.* Lora: Low-rank adaptation of large language models (2021). 2106.09685.

[20] Shen, D., Wu, G. & Suk, H.-I. Deep learning in medical image analysis. *Annual review of biomedical engineering* **19**, 221–248 (2017).

[21] De Fauw, J. *et al.* Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature medicine* **24**, 1342–1350 (2018).

[22] Rajpurkar, P. *et al.* Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225* (2017).

[23] Choi, B. G. *et al.* Machine learning for the prediction of new-onset diabetes mellitus during 5-year follow-up in non-diabetic patients with cardiovascular risks. *Yonsei medical journal* **60**, 191–199 (2019).

[24] Yoo, T. K. *et al.* Osteoporosis risk prediction for bone mineral density assessment of postmenopausal women using machine learning. *Yonsei medical journal* **54**, 1321–1330 (2013).

[25] Rajpurkar, P. & Lungren, M. P. The current and future state of ai interpretation of medical images. *New England Journal of Medicine* **388**, 1981–1990 (2023).

[26] Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H. & Aerts, H. J. Artificial intelligence in radiology. *Nature Reviews Cancer* **18**, 500–510 (2018).

[27] Tiu, E. *et al.* Expert-level detection of pathologies from unannotated chest x-ray images via self-supervised learning. *Nature Biomedical Engineering* **6**, 1399–1406 (2022).

[28] Moon, J. H., Lee, H., Shin, W., Kim, Y.-H. & Choi, E. Multi-modal understanding and generation for medical images and text via vision-language pre-training. *IEEE Journal of Biomedical and Health Informatics* **26**, 6070–6080 (2022).

[29] Huang, Z., Zhang, X. & Zhang, S. Kiut: Knowledge-injected u-transformer for radiology report generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19809–19818 (2023).

[30] Huemann, Z., Hu, J. & Bradshaw, T. Contextual net: A multimodal vision-language model for segmentation of pneumothorax. *arXiv preprint arXiv:2303.01615* (2023).

[31] Lee, S., Kim, W. J. & Ye, J. C. Llm itself can read and generate cxr images. *arXiv preprint arXiv:2305.11490* (2023).

[32] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical*

Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19, 424–432 (Springer, 2016).

[33] Radford, A. *et al.* Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, 8748–8763 (PMLR, 2021).

[34] Kingma, D. & Ba, J. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)* (San Diega, CA, USA, 2015).

[35] Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* **32** (2019).

[36] Crum, W. R., Camara, O. & Hill, D. L. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE transactions on medical imaging* **25**, 1451–1461 (2006).

Supplementary Information

Schematic Comparison of the Workflows of Radiology and Radiation Oncology.

Fig. S1 delineates the clinical workflows in Radiology and Radiation Oncology. Radiology primarily formulates diagnostic outcomes based on imaging findings. In contrast, Radiation Oncology not only incorporates imaging but also considers various clinically relevant textual information, such as surgery notes, pathology reports, and electronic medical records, in decision-making processes like determining treatment scope and dose decisions. Additionally, the integration of prior knowledge, including standard treatment guidelines and radiation oncology textbooks, is crucial for informed treatment decision-making and can also be expressed in textual formats. Therefore, the significance of multi-modality is notably enhanced in Radiation Oncology compared to Radiology.

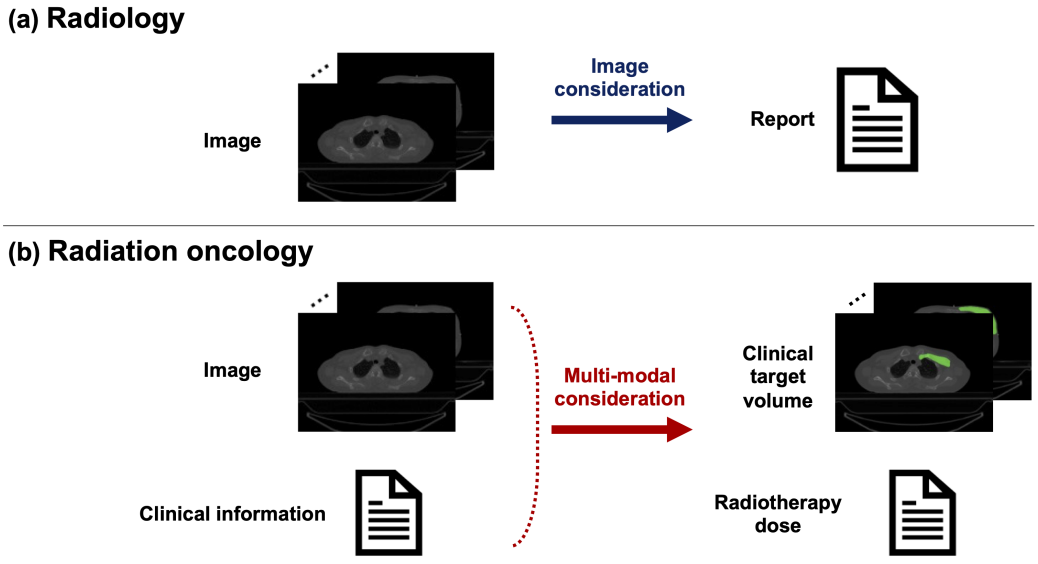

Fig. S1: **Comparison of workflows in Radiology (upper) and Radiation Oncology (lower). While Radiology primarily employs imaging information for interpretations, Radiation Oncology integrates not only imaging but also considers clinical information available from pathology results and patient's electronic medical records to inform treatment decisions.**

Details of Multi-modal AI Framework.

In this section, we further explain our proposed multi-modal AI framework for performing clinical target volume (CTV) delineation tasks as illustrated in Fig. S2. We introduce three key components: (a) text prompt tuning, (b) multi-modal interactive alignment, and (c) CTV delineation.

(a) Text Prompt Tuning. To efficiently fine-tune the large language model (LLM), we introduce N-text prompts $\{v^n|_{n=1}^N\}$ as illustrated in Fig. S2(a), where each $v^n \in \mathbb{R}^{M\times D}$ consists of M vectors with the dimension D, which is same embedding dimension as the LLM. These learnable vectors are randomly initialized, and then consistently prepended to each of tokenized clinical note, which denoted as [TEXT] tokens. We additionally append a token, denoted as [SEG], which is intended to attend to all the aforementioned vectors

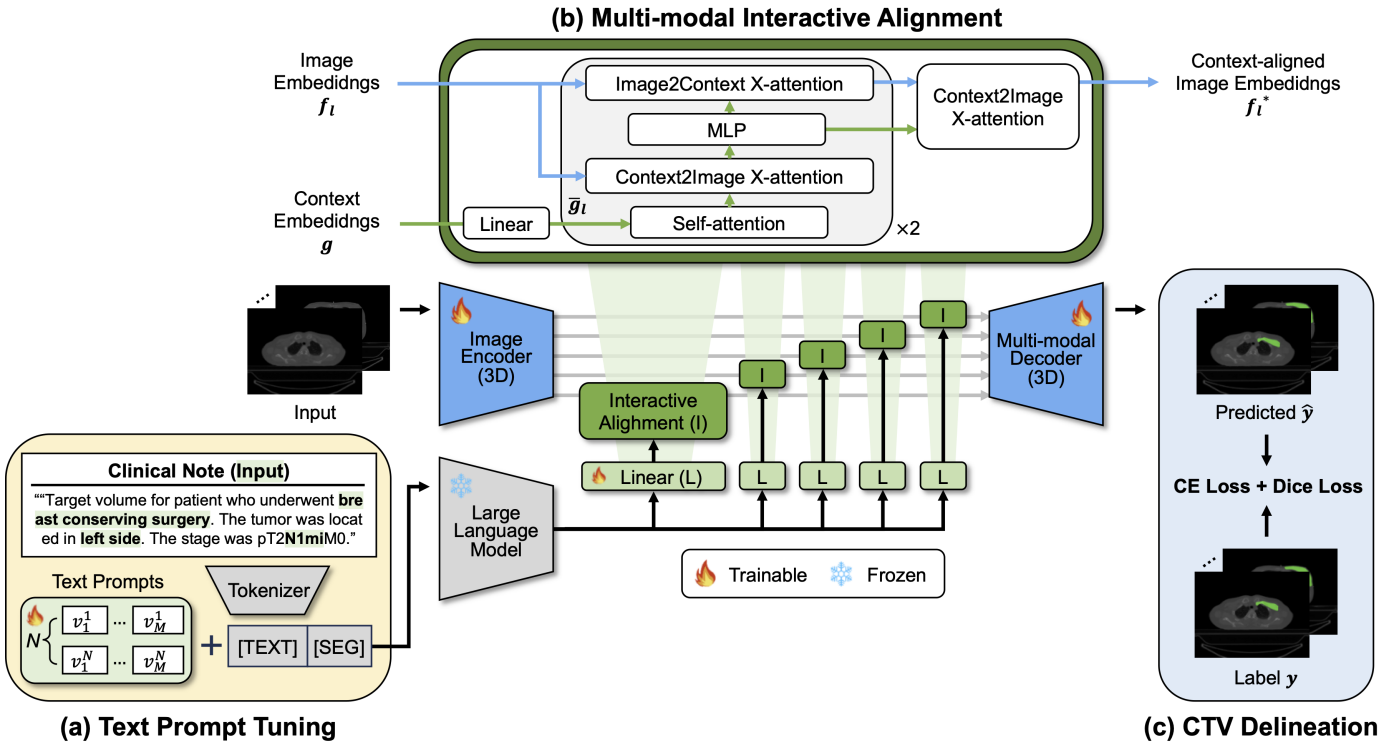

Fig. S2: **Details of our proposed multi-modal AI framework. (a) Text Prompt tuning: we tune the large language model by using the light-weight text prompts. (b) Multi-modal Interactive Alignment: we adapt interactive alignment modules between image and context embeddings. (c) CTV Delineation: we adapt both the cross-entropy and dice loss given ground truth CTV labels.**

and tokens. Here, the final prompted text input t can be formulated as follows:

$$t = \{v_1^n, v_2^n, ..., v_M^n, [\text{TEXT}], [\text{SEG}]\}. \tag{1}$$

Then, using the prompted text input t, the frozen LLM results the context embeddings $g \in \mathbb{R}^{N \times D}$ as output embeddings as for the inputted [SEG] token.

(b) Multi-modal Interactive Alignment . To align the context embeddings g with the image embeddings $f_l \in \mathbb{R}^{H_l W_l S_l \times C_l}$, where f_l is the l-th layer output of the 3D image encoder, H_l, W_l, and S_l correspond to height, width, and slice of the image embeddings, and C_l is the intermediate channel dimension of each l-th layer output, we first project g to have the identical dimension with that of each f_l through layer-wise linear layer. As illustrated in Fig. S2(b), the linearly projected context embeddings $\bar{g}_l$ are then self-attended and crossly-attended with the image embedding f_l to result context-aligned image embeddings f_l^*. Detailed specifications of each l-th layer embeddings and the interactive alignment module are listed in Table S1.

(c) CTV Delineation . After the multi-modal interactive alignment, the context-aligned image embeddings f_l^* become inputs for the 3D image decoder. As illustrated in Fig. S2(c), for the final predicted output $\hat{y}$, we calculated the combination of the Cross-entropy (CE) loss and the Dice coefficient (Dice) loss by using the ground-truth label y as follows:

$$\mathcal{L} = \lambda_{\text{ce}} \mathcal{L}_{\text{ce}}(\hat{y}, y) + \lambda_{\text{dice}} \mathcal{L}_{\text{dice}}(\hat{y}, y), \tag{2}$$

Table S1: **Detailed specifications of the multi-modal AI framework. Linear$_{D,C_l}$ denotes the linear layer which convert dimension of context embedding g from D to C_l, Conv$_{\mathbf{Ch}_{in},\mathbf{Ch}_{out}}$ denotes the convolution layer, Norm, Act, MLP denote Normalization layer, Activation layer, Multilayer Perceptron, respectively.**

Layer (l)	**Ch**$_{in}$	**Multi-modal AI model**						**Ch**$_{out}$	H_l, W_l, S_l
		Image Encoder	Interactive Alignment			Image Decoder			
1	3	ResBlock$_{3,48}$	Linear$_{4096,48}$	CrossBlock$_{48}$	↘		TransConv$_{48,1}$	2	384, 384, 128
2	48	ResBlock$_{48,48}$	Linear$_{4096,48}$	CrossBlock$_{48}$	↘	UpResBlock$_{48,48}$		48	192, 192, 64
3	48	ResBlock$_{48,96}$	Linear$_{4096,96}$	CrossBlock$_{96}$	↘	UpResBlock$_{96,48}$		48	96, 96, 32
4	96	ResBlock$_{96,192}$	Linear$_{4096,192}$	CrossBlock$_{192}$	↘	UpResBlock$_{192,96}$		96	48, 48, 16
5	192	ResBlock$_{192,384}$	Linear$_{4096,384}$	CrossBlock$_{384}$	→	UpResBlock$_{384,192}$		192	24, 24, 8

Note: ResBlock$_{\mathrm{Ch}_{in},\mathrm{Ch}_{out}}$ = [Conv$_{\mathrm{Ch}_{in},\mathrm{Ch}_{out}}$ - Norm - Act - Conv$_{\mathrm{Ch}_{out},\mathrm{Ch}_{out}}$ - Norm - Residual Shortcut - Act]
CrossBlock$_{\mathrm{Ch}}$ = [Self Attention - Context2Image X-attention - MLP$_{\mathrm{Ch}}$ - Image2Context X-attention]
UpResBlock$_{\mathrm{Ch}_{in},\mathrm{Ch}_{out}}$ = [TransConv$_{\mathrm{Ch}_{in},\mathrm{Ch}_{out}}$ - Skip Connection Shortcut (↘) - ResBlock$_{\mathrm{Ch}_{in},\mathrm{Ch}_{out}}$]

where λ_{ce} and λ_{dice} are hyper-parameters for each CE loss and Dice loss, respectively.

Details of Hyper-parameter Settings.

We set the optimal hyper-parameters as listed in Table S2 by varying experimental conditions.

Table S2: **Hyper-parameter settings for experiments.**

Hyper-parameter	**Symbol**	**Value**
Text prompt tuning		
# of multiple text prompts	N	2
length of each text prompt	M	8
Embedding dimensions		
Text encoder	D	4096
Image model	C_l	{48, 48, 96, 192, 384}
Loss function balance factors		
for the CE loss	λ_{ce}	1
for the Dice loss	λ_{dice}	1

Details of Network Training Complexity.

Table S3: **Training complexity.**

	Backbone	**Parameters (M)**		**Training Duration**		**Complexity (GFlops)**
		Total	Trainable	(s/input)	(m/epoch)	
Vision-only AI						
Image encoder/decoder	3D residual U-Net	13.10	13.10	3.16	27.00	1,516
Multi-modal AI						
Text encoder	Llama2-7B-chat	7,000	0			
Text prompts	-	0.07	0.07			
Interactive alignment	-	8.32	8.32			
Total	-	7,021.49	21.49	3.61	30.81	15,120