

VLM Multi-Model Comparison Report

Dataset: radiation_treatment_outcomes_v1_1

Subtask: radiation_outcome

Question:

Based on the provided medical history and cancer evidence, what is the patient's expected response to external beam radiation therapy?

Correct Answer: Tumor symptoms controlled

Model: OpenGVLab/InternVL3_5-8B

no radiology report: [INCORRECT] <brief reasoning trace with explanation for label selection>
| <label> The patient's medical history indicates multiple malignancies, including malignant neoplasms of the lung, lymph nodes, and mediastinum. The timeline includes codes for therapeutic radiology simulation-aided field setting, basic radiation dosimetry calculation, and intensity modulated radiotherapy plan, suggesting that the patient has undergone or is undergoing radiation therapy. However, there is no explicit evidence of the patient's response to the treatment, such as imaging results or clinical assessments indicating tumor symptom control or recurrence. Therefore, based on the provided information, the most appropriate label is "Never achieved complete response" as there is no documented evidence of a complete response to the radiation therapy. | <label> "Never achieved complete response"

Model: OctoMed/OctoMed-7B

no radiology report: [INCORRECT] <think>First, I need to analyze the patient's timeline to determine their expected response to external beam radiation therapy (EBRT). The options are "Tumor symptoms controlled," "Tumor symptoms returned," or "Never achieved complete response." Looking at the timeline, the patient has multiple ICD-10 codes indicating various malignancies: C34.90 (lung), C77.0 (lymph nodes of head/neck), C38.1 (anterior mediastinum), and C38.2 (posterior mediastinum). This suggests multiple primary or metastatic sites, which complicates treatment. The STANFORD_OBS/Flowsheet entries provide clinical assessments. Key points include: - Sensory Perception: 4 (normal) - Altered Elimination Patterns: 0 (no issues) - Altered Mental Status: 0 (no issues) - Other Fall Risks: 2 (some risk) - CAM Results: Negative - RASS: Not specified, but other entries like Total Score (21 and 22) might relate to mental status or other assessments. - Fall Interventions: Standard - Delirium Interventions: Manage pain, promote sleep, manage toileting - Mobility: 4 (normal) - Moisture: 4 (normal) - High Risk for Delirium?: No The CPT codes indicate that radiation therapy was planned and delivered (77280, 77300, 77301, 77338, 77427, 77386). This confirms that EBRT was administered. However, the patient's clinical status shows no significant improvement in symptoms. The sensory, elimination, and mental status are normal, but there are other risks (like fall risks). The delirium interventions suggest some concern, but CAM is negative. The

presence of multiple malignancies and the lack of documented symptom improvement or resolution in the timeline suggest that the tumor symptoms were not fully controlled. The options are about tumor symptoms. Since the patient has multiple sites and the timeline doesn't show resolution, "Tumor symptoms controlled" is unlikely. "Tumor symptoms returned" would imply prior control followed by recurrence, but there's no indication of prior improvement. "Never achieved complete response" aligns with persistent symptoms despite treatment. Therefore, the correct label is "Never achieved complete response."</think> <brief reasoning trace with explanation for label selection> | <label> The patient's timeline shows multiple malignancies (lung, lymph nodes, mediastinum) and radiation therapy was planned and delivered (CPT codes). However, there is no documented evidence of tumor symptom control, resolution, or improvement. The STANFORD_OBS/Flowsheet does not indicate symptom relief, and the presence of multiple tumors suggests persistent disease. Thus, the patient likely "Never achieved complete response."</label>

Model: Qwen/Qwen3-VL-8B-Instruct

no radiology report: [INCORRECT] No evidence of tumor response or symptom control documented post-radiation; no imaging, pathology, or clinical outcome data provided to assess efficacy. | Never achieved complete response

Model: google/medgemma-4b-it

no radiology report: [CORRECT] Tumor symptoms controlled

Dataset: tb_v1_1_tb_classification_tasks

Subtask: pneumonitis_infection_answer

Question:

Based on the provided medical history, cancer evidence, and imaging findings, which diagnosis is most consistent with the imaging differential?

Correct Answer: -1

Model: OpenGVLab/InternVL3_5-8B

no radiology report: [INCORRECT] <brief reasoning trace with explanation for label selection>
| <label> The timeline shows multiple radiology visits and imaging studies, particularly focusing on the chest area. The repeated imaging and emergency room visits suggest ongoing monitoring and possible complications. The presence of multiple imaging studies over several years indicates a need for continuous evaluation, which is often seen in cases of cancer progression. The absence of specific evidence pointing to infection or pneumonitis, such as clinical symptoms or specific test results, makes "Cancer Progression" the most consistent diagnosis based on the provided timeline. | Cancer Progression

Model: OctoMed/OctoMed-7B

no radiology report: [INCORRECT] <think>First, I carefully examine the provided CT images to identify any abnormalities or features that might suggest a diagnosis. The images show axial slices progressing from the lower abdomen to the thorax. I note the presence of gas-filled structures in the abdomen, which could be bowel loops, and the lungs appear aerated with no obvious masses or consolidations. The bones, including the spine and pelvis, appear intact without lytic or blastic lesions. There is no visible pleural effusion or pneumothorax. The soft tissues do not show any obvious masses or swelling. Next, I consider the patient's timeline. The patient has a history of medical interventions, including medications (montelukast), and multiple imaging studies over several years. The most recent imaging includes CT scans of the chest and whole body, which do not show any new or progressive lesions. The cytology report and HPV testing are normal, suggesting no active cervical pathology. There is no evidence of cancer progression or new metastatic disease. Given the absence of new findings on imaging, no signs of infection or pneumonitis, and no evidence of cancer progression, the most consistent diagnosis is "Neither." The imaging does not support pneumonitis, infection, or cancer progression at this time.</think> <answer>Neither</answer>

Model: Qwen/Qwen3-VL-8B-Instruct

no radiology report: [INCORRECT] The patient's timeline documents a diagnosis of malignant neoplasm of the right lung (C34.31) and a secondary malignant neoplasm of the brain (C79.31) in 2017. Treatment included 3D radiotherapy planning and stereotactic radiosurgery for the brain lesion. However, the timeline does not contain any subsequent imaging, clinical assessments, or documentation of tumor response (e.g., symptom control, progression, or recurrence) following radiation therapy. Without evidence of clinical outcome or follow-up, it is impossible to determine if the treatment was effective or if symptoms returned. Therefore, the

most accurate option based on the provided evidence is "Never achieved complete response", as there is no documentation of a successful outcome. <brief reasoning trace with explanation for label selection> | "Never achieved complete response"

Model: google/medgemma-4b-it

no radiology report: [INCORRECT] The provided images are chest X-rays. The images show a clear, round, white structure with a central black area. This is consistent with a pneumothorax, which is air in the pleural space. The presence of a "sunburst" pattern in the lung fields is also consistent with pneumothorax. The patient has a history of ER visits, suggesting possible respiratory distress. <Both>

Subtask: pneumonitis_infection_discussed

Question:

Based on the provided medical history, cancer evidence, and imaging findings, is there sufficient diagnostic ambiguity to warrant presentation of this case at a thoracic tumor board to discuss differentiation between radiation pneumonitis and infection?

Correct Answer: No

Model: OpenGVLab/InternVL3_5-8B

no radiology report: [INCORRECT] <brief reasoning trace with explanation for label selection>
| <label> The patient's medical timeline indicates a complex case involving a bone tumor in the right superior pubic rami and a secondary finding of a lung nodule. Imaging and biopsy results suggest a metastatic adenocarcinoma, with the bone tumor likely being a metastasis from the lung primary. The differential diagnosis includes both radiation pneumonitis and infection, given the patient's history and symptoms. The ambiguity in distinguishing between these conditions, especially considering the patient's recent treatments and symptoms, warrants further discussion at a thoracic tumor board to ensure accurate diagnosis and appropriate management. | Yes

Model: OctoMed/OctoMed-7B

no radiology report: [CORRECT] <think>First, I carefully reviewed the provided medical history and imaging findings to understand the patient's condition. The patient is a 49-year-old female with a history of chronic right groin and pelvic pain, leading to wheelchair dependence. Imaging shows a large aggressive bone tumor in the right superior pubic ramus with a 2.5 cm right lung nodule. The biopsy results indicated metastatic adenocarcinoma rather than a primary sarcoma. The patient has a family history of cancer and is a former smoker who quit in 2016. She is currently experiencing significant pain and discomfort, which complicates her clinical management. The key question is whether there is sufficient diagnostic ambiguity to warrant a thoracic tumor board discussion to differentiate between radiation pneumonitis and infection. Radiation pneumonitis typically occurs after radiation therapy and presents with pulmonary infiltrates, cough, and dyspnea, often with elevated inflammatory markers. Infection, on the other hand, would present with fever, leukocytosis, and possibly new infiltrates on imaging. In this case, the patient has a known aggressive bone tumor with metastatic lung involvement, which suggests a primary malignancy rather than an infectious process. The imaging findings do not show typical signs of infection such as consolidation or cavitation. Additionally, the patient's symptoms (chronic pain, no acute fever, and no recent antibiotic use) do not strongly support infection. The biopsy confirmed metastatic adenocarcinoma, which aligns with the clinical picture of a primary malignancy with metastases rather than an infectious etiology. Given the biopsy results, imaging findings, and clinical presentation, there is no significant ambiguity to suggest radiation pneumonitis or infection. The diagnosis is clear as metastatic adenocarcinoma with lung metastases. Therefore, there is no need for a thoracic tumor board discussion to differentiate between these two entities.</think> <answer>No</answer>

Model: Qwen/Qwen3-VL-8B-Instruct

no radiology report: [CORRECT] The patient has a known primary pelvic bone tumor, which was biopsied and later confirmed to be metastatic adenocarcinoma. The imaging (CT scan)

shows a 2.5 cm spiculated mass in the right lung, which is consistent with metastatic disease. The question asks whether there is diagnostic ambiguity to differentiate between radiation pneumonitis and infection. However, the patient's history and imaging findings do not suggest a prior radiation therapy to the chest, which is a key factor in radiation pneumonitis. The CT scan shows a spiculated mass, which is more characteristic of a malignant tumor than radiation pneumonitis or infection. The timeline shows the lung nodule was identified after the pelvic tumor was diagnosed, and the biopsy results confirmed metastatic adenocarcinoma. There is no evidence of prior radiation to the chest, and the imaging findings are consistent with metastatic disease, not radiation pneumonitis or infection. Therefore, there is no diagnostic ambiguity to warrant a thoracic tumor board discussion for differentiating between radiation pneumonitis and infection. No

Model: google/medgemma-4b-it

no radiology report: [INCORRECT] The provided medical history indicates a 49-year-old female patient with a large, aggressive-appearing bone tumor in the right superior pubic ramus and pelvic bones, along with a lung nodule. The patient has a history of chronic pain and has been wheelchair-bound for a significant period. The imaging findings show a destructive lesion in the right superior pubic ramus, extending into the acetabulum and soft tissues. The patient has been referred to thoracic oncology for evaluation of the lung nodule. The final pathology report revealed metastatic adenocarcinoma. Given the presence of a known primary malignancy (lung cancer) and the metastatic disease in the bone, the question is whether the imaging findings are consistent with radiation pneumonitis or infection. While radiation pneumonitis is a possibility in patients who have received radiation therapy to the chest, the patient's history does not indicate prior radiation exposure. Infection is also a possibility, but the imaging findings are more suggestive of metastatic disease. Therefore, the answer is "No".

Subtask: progression_assessment_discussed

Question:

Based on the provided medical history and cancer evidence, is there ambiguity in radiologic interpretation that warrants presentation of this case at a thoracic tumor board for progression assessment?

Correct Answer: No

Model: OpenGVLab/InternVL3_5-8B

no radiology report: [INCORRECT] <The patient has a history of nasopharyngeal carcinoma with evidence of recurrent disease and the presence of solitary pulmonary nodules. Imaging and biopsy results are pending, and there is suspicion of metastatic disease or infection. The complexity and ambiguity of the radiologic findings, along with the need for further diagnostic workup, suggest that a thoracic tumor board discussion is warranted to assess progression and guide management. | Yes>

Model: OctoMed/OctoMed-7B

no radiology report: [INCORRECT] <think>First, I will carefully analyze the provided CT scan images and clinical notes to understand the patient's condition and the current clinical scenario. The patient is a 44-year-old female with a history of nasopharyngeal carcinoma treated with chemoradiation in 2001. She now presents with a right nasal mass and bilateral upper lobe nodules. The CT scan shows a 1.5 cm right upper lobe (RUL) spiculated nodule that is hypermetabolic, and a 1.6 cm left apical nodule that is not hypermetabolic. The Quantiferon test is positive, indicating latent or active tuberculosis infection. The patient is scheduled for a CT-guided biopsy of the hypermetabolic RUL nodule. The clinical team is considering the possibility of recurrence of nasopharyngeal carcinoma or metastatic disease, and the biopsy is planned to confirm the diagnosis. The patient is also undergoing evaluation for potential treatment options, including surgery or systemic therapy. The key question is whether there is ambiguity in the radiologic interpretation that necessitates a thoracic tumor board review for progression assessment.</think> <answer>Yes</answer>

Model: Qwen/Qwen3-VL-8B-Instruct

no radiology report: [INCORRECT] The provided timeline documents a complex clinical scenario involving a patient with a history of nasopharyngeal carcinoma (NPC) who presented with a new right nasal obstruction. Imaging revealed a destructive mass in the nasopharynx and two lung nodules. The radiologic interpretation of these findings is explicitly described as ambiguous. Specifically, the note from Dr. Mendenhall on 09/14/2012 states: "The nodules are also necrotic appearing, also probably more consistent with an infection such as reactivation TB, but not entirely eliminating the possibility these could represent metastatic lesions." This directly indicates that the radiologic findings are not definitive and could represent either recurrent NPC metastases or an infectious process (reactivation TB), which was later confirmed by a positive Quantiferon test. This ambiguity in differentiating between metastatic disease and infection is a key reason for the recommendation to present the case at a tumor board. The tumor board is the appropriate forum to discuss such complex, uncertain findings and to determine the optimal diagnostic and therapeutic pathway. Therefore, the radiologic interpretation is ambiguous and warrants presentation at a tumor board. Yes

Model: google/medgemma-4b-it

no radiology report: [INCORRECT] The patient has a history of nasopharyngeal carcinoma (NPC) treated with chemoradiation in 2001. She presented with a right nasal obstruction in 2012, and subsequent imaging revealed a destructive mass eroding the pterygoid plate and inferior sphenoid bone, extending into the orbit and posterior sphenoid sinus. The patient is currently being evaluated for a possible recurrence of the NPC. The imaging findings are concerning for local recurrence, and a biopsy is planned. The patient's anxiety and the complexity of her case warrant a discussion at a thoracic tumor board to determine the optimal management strategy. Therefore, the answer is: Yes

Subtask: radiation_oligoprogression_answer

Question:

TODO: Based on the provided medical history, cancer evidence, and imaging findings, is focal radiotherapy recommended for the management of this patient's oligopressive disease?

Correct Answer: Yes

Model: OpenGVLab/InternVL3_5-8B

no radiology report: [INCORRECT] <Multiple myeloma and lung cancer are present in the timeline. The patient has undergone various treatments including chemotherapy and bone marrow biopsy. Imaging studies like CT and MRI have been conducted, but there is no specific mention of focal progression or oligopression. Given the complexity and the lack of clear evidence of oligopression, focal radiotherapy might not be the most appropriate next step without further detailed assessment of progression.> | No

Model: OctoMed/OctoMed-7B

no radiology report: [CORRECT] <think>Okay, let's analyze the question about whether focal radiotherapy is appropriate for the management of oligopression. The patient has a history of multiple myeloma and lung cancer, with recent imaging and lab work indicating disease progression. The key here is to determine if focal radiotherapy is suitable given the patient's cancer status and treatment history. First, oligopression refers to the presence of a limited number of new lesions in a patient with previously controlled disease. In this case, the patient has multiple myeloma and lung cancer, and the imaging findings would need to show isolated progression in a limited number of sites to qualify for oligopression. The provided imaging includes CT scans of the chest and pelvis, showing progression of disease in specific areas. The decision for focal radiotherapy depends on the number, size, and location of lesions, as well as the patient's overall status and prior treatments. Focal radiotherapy is generally considered when there are 1-5 new lesions, and the patient is otherwise stable with no widespread progression. The patient's history includes a bone marrow biopsy showing progression, and the imaging shows localized progression in the pelvis and chest. The patient is stable otherwise, with no evidence of widespread metastatic disease. This aligns with the criteria for oligopression. Focal radiotherapy could be appropriate to target these isolated lesions while sparing the rest of the body from systemic therapy side effects. However, the decision must also consider the patient's prior treatments, response to previous therapies, and overall prognosis. If the patient has already received extensive systemic therapy and is stable, focal radiotherapy could be a good option. If the disease is more widespread or the patient is not a candidate for further treatment, other options may be considered. Given the information provided, the patient has oligopression with localized disease and is a candidate for focal radiotherapy.</think> <answer>Yes</answer>

Model: Qwen/Qwen3-VL-8B-Instruct

no radiology report: [INCORRECT] No | The patient's medical history and imaging findings do not support the use of focal radiotherapy for oligopression. The timeline indicates a diagnosis of multiple myeloma (ICD10CM/C90.00) and lung adenocarcinoma (ICD10CM/C34.92), with a documented stage of 2A for the lung cancer. The imaging series (CT

scans of chest, abdomen, pelvis, and spine) and lab results (including bone marrow biopsy) are consistent with a systemic disease process, not a localized, oligopressive lesion that would be a candidate for focal radiotherapy. The patient's condition is managed with systemic therapies (lenalidomide, dexamethasone, etc.), and there is no evidence of isolated, measurable disease progression that would warrant focal radiation. Therefore, focal radiotherapy is not appropriate for this patient's management.

Model: google/medgemma-4b-it

no radiology report: [CORRECT] The provided medical history indicates a patient with a history of lung cancer (C34.90 and C34.92) and multiple myeloma (C90.00). The patient is currently being treated for multiple myeloma with lenalidomide and aspirin. The imaging findings show a large, well-defined mass in the chest, likely representing the lung cancer. Focal radiotherapy is a reasonable option for the management of oligoprogression in this patient. Therefore, the answer is: Yes