

VLM Multi-Model Comparison Report

Dataset:

progression_recurrence_survival_1yr_2yr_3yr_4yr_5yr_v1_1

Subtask: died_of_cancer_1_yr

Average unique events: 26.49

Question:

Based on the provided medical history and cancer evidence, is cancer-related mortality expected within 1 year of treatment initiation?

Correct Answer: No

Model: OctoMed/OctoMed-7B

[illegible]

[illegible][illegible]

Subtask: died of cancer 2 yr

Average unique events: 26.49

Question:

Based on the provided medical history and cancer evidence, is cancer-related mortality expected within 2 years of treatment initiation?

Correct Answer: No

Model: OctoMed/OctoMed-7B

[illegible]

[illegible]

radiology report: [INCORRECT]

Subtask: died_of_cancer_3_yr

Average unique events: 26.49

Question:

Based on the provided medical history and cancer evidence, is cancer-related mortality expected within 3 years of treatment initiation?

Correct Answer: No

Model: OctoMed/OctoMed-7B

[illegible]

mediastinal mass, which is likely compressing the SVC. The patient is currently on diltiazem and apixaban. The patient has a history of atrial flutter and a mediastinal mass, which is likely compressing the SVC. The patient is currently on diltiazem and apixaban. The patient has a history of atrial flutter and a mediastinal mass, which is likely compressing the SVC. The patient is currently on diltiazem and apixaban. The patient has a history of atrial flutter and a mediastinal mass, which is likely compressing the SVC. The patient is currently on diltiazem and apixaban. The patient has a history of atrial flutter and a mediastinal mass, which is likely compress

[illegible]

Subtask: died_of_cancer_4_yr

Average unique events: 26.49

Question:

Based on the provided medical history and cancer evidence, is cancer-related mortality expected within 4 years of treatment initiation?

Correct Answer: No

Model: OctoMed/OctoMed-7B

[illegible]

[illegible]

Subtask: died of cancer 5 yr

Average unique events: 26.49

Question:

Based on the provided medical history and cancer evidence, is cancer-related mortality expected within 5 years of treatment initiation?

Correct Answer: No

Model: OctoMed/OctoMed-7B

[illegible]

compressing the SVC, leading to SVC syndrome. The patient is on diltiazem and apixaban, indicating atrial flutter management. The CT scan shows a mediastinal mass, which is compressing the SVC, leading to SVC syndrome. The patient is on diltiazem and apixaban, indicating atrial flutter management. The CT scan shows a mediastinal mass, which is compressing the SVC, leading to SVC syndrome. The patient is on diltiazem and apixaban, indicating atrial flutter management. The CT scan shows a mediastinal mass, which is compressing the SVC, leading to SVC syndrome. The patient is on diltiazem and apixaban, indicating atrial flutter management. The CT scan shows a mediastinal mass, which is compressing the SVC, leading to SVC syndrome. The patient is on diltiazem and apixaban, indicating atrial flutter management.

[illegible]

Subtask: has_recurrence_1_yr

Average unique events: 26.49

Question:

Based on the patient's clinical history and current cancer-related evidence, will the patient achieve an initial treatment response but experience cancer recurrence within 1 year of treatment initiation?

Correct Answer: No

Model: OctoMed/OctoMed-7B

[illegible]

[illegible][illegible]

Subtask: has recurrence 2 yr

Average unique events: 26.49

Question:

Based on the patient's clinical history and current cancer-related evidence, will the patient achieve an initial treatment response but experience cancer recurrence within 2 years of treatment initiation?

Correct Answer: No

Model: OctoMed/OctoMed-7B

[illegible]

[illegible][illegible]

Subtask: has recurrence 3 yr

Average unique events: 26.49

Question:

Based on the patient's clinical history and current cancer-related evidence, will the patient achieve an initial treatment response but experience cancer recurrence within 3 years of treatment initiation?

Correct Answer: No

Model: OctoMed/OctoMed-7B

[illegible]

[illegible]

radiology report: [INCORRECT]

Subtask: has_recurrence_4_yr

Average unique events: 26.49

Question:

Based on the patient's clinical history and current cancer-related evidence, will the patient achieve an initial treatment response but experience cancer recurrence within 4 years of treatment initiation?

Correct Answer: No

Model: OctoMed/OctoMed-7B

[illegible]

[illegible][illegible]

Subtask: has recurrence 5 yr

Average unique events: 26.49

Question:

Based on the patient's clinical history and current cancer-related evidence, will the patient achieve an initial treatment response but experience cancer recurrence within 5 years of treatment initiation?

Correct Answer: No

Model: OctoMed/OctoMed-7B

[illegible]

history of atrial flutter and atrial fibrillation. The patient has a history of atrial flutter and atrial fibrillation. The patient has a history of atrial flutter and atrial fibrillation. The patient has a history of atrial flutter and atrial fibrillation. The patient has a history of atrial flutter and atrial fibrillation. The patient has a history of atrial flutter and atrial fibrillation. Okay, let's analyze the patient's clinical course and imaging findings to determine the most likely diagnosis and whether the patient has a mediastinal mass causing SVC syndrome. The CT chest from outside hospital shows a mediastinal mass compressing the SVC, which is likely the cause of SVC syndrome. The patient has a history of atrial flutter and atrial fibrillation. The patient has a history of atrial flutter and atrial fibrillation. The patient has a history of atrial flutter and atrial fibrillation. The patient has a history of atrial flutter and atrial fibrillation.

[illegible]

Dataset: tb_v1_1_tb_classification_tasks

Subtask: early_stage_management_answer

Average unique events: 25.72

Question:

Based on the provided medical history and cancer evidence, what is optimal curative-intent treatment for stage I-III NSCLC?

Correct Answer: Radiation alone

Model: OctoMed/OctoMed-7B

no radiology report: [INCORRECT] <think>First, I need to carefully analyze the provided CT images and the patient's timeline to determine the most appropriate treatment for stage I-III non-small cell lung cancer (NSCLC). The patient is a 77-year-old female with a history of aortic valve replacement, chronic heart failure, atrial fibrillation, and other comorbidities. She has a solitary pulmonary nodule and additional suspicious lung nodules seen on imaging. The CT scans show a spiculated right upper lobe mass and a new right lower lobe nodule, raising concern for malignancy. The patient has undergone PET/CT and CT head imaging, which are critical for staging and assessing resectability. The options for treatment include systemic therapy followed by surgery with or without adjuvant therapy, surgery followed by systemic therapy, surgery alone, radiation alone, or definitive chemoradiation with or without adjuvant systemic therapy. Given the patient's age, comorbidities, and the presence of multiple lung nodules, surgery alone may not be the best option due to the risk of incomplete resection or recurrence. Radiation alone is less likely to be curative for stage I-III NSCLC. Definitive chemoradiation is typically reserved for patients who are not surgical candidates or have locally advanced disease. However, the patient's history of heart surgery and current cardiac status must be considered. The most appropriate curative-intent treatment for stage I-III NSCLC in a patient with resectable disease is surgery followed by systemic therapy if needed. This approach allows for complete tumor removal and addresses potential micrometastatic disease. The patient's cardiac history and the need for surgical risk assessment are important, but the imaging findings suggest resectable disease. Therefore, the optimal treatment is surgery followed by systemic therapy.</think> <answer>Systemic therapy followed by surgery +/- adjuvant therapy</answer>

radiology report: [INCORRECT] <think>First, I need to understand the patient's medical history and current condition based on the timeline provided. The patient is a 77-year-old female with a history of aortic valve endocarditis and replacement in 2013. She has chronic diastolic heart failure, atrial fibrillation, and is on multiple medications including diuretics and anticoagulants. She presents with a right upper lobe spiculated nodule and a right middle lobe subsolid nodule, both concerning for malignancy. The CT scan shows these nodules with increased SUV max values, indicating metabolic activity. Additionally, there are hypermetabolic mediastinal and hilar lymph nodes, which could be granulomatous disease or malignancy. The PET/CT report notes that the right upper lobe nodule has grown compared to a prior scan, and the middle lobe nodule has a solid component. The lymph nodes have varying SUV max values but are calcified, which may suggest granulomatous disease, but malignancy cannot be excluded. The patient is a lifelong non-smoker, which is important because lung

adenocarcinoma is more common in non-smokers, especially in Asian women, and EGFR mutations are a consideration. The patient's cardiac history complicates surgical options, as she has significant heart failure and underwent valve replacement. The cardiologist evaluation is recommended to assess surgical risk. The question is about the optimal curative-intent treatment for stage I-III NSCLC. The options are various combinations of surgery, systemic therapy, and radiation. For stage I-III NSCLC, the standard approach is typically surgery followed by adjuvant therapy if indicated, or definitive chemoradiation for inoperable cases. However, the presence of multiple nodules and lymph node involvement complicates staging. The PET/CT shows a primary tumor with possible second nodule and hypermetabolic lymph nodes. The lymph nodes are calcified, which may suggest benign etiology, but malignancy cannot be ruled out. Given the patient's cardiac comorbidities and the need for surgical risk assessment, the decision must balance the potential benefits of surgery against her overall health. The presence of multiple nodules and lymph node involvement may suggest stage III disease. However, the exact stage is not definitively clear from the timeline. The patient's history of heart failure and valve replacement increases surgical risk, so a multidisciplinary approach is necessary. The options include surgery followed by systemic therapy, which is standard for resectable stage II-III NSCLC. However, if the patient is not a surgical candidate due to cardiac issues, definitive chemoradiation would be appropriate. The timeline does not provide clear evidence of resectability or surgical feasibility, but the recommendation from the note is to see a cardiologist for surgical risk assessment. Therefore, the optimal treatment would depend on that assessment. Since the question asks for the optimal curative-intent treatment based on the timeline, and the timeline does not confirm resectability, the safest choice would be to consider definitive chemoradiation with or without adjuvant therapy, especially if surgery is deemed high risk. However, if the patient is deemed a candidate for surgery, then surgery followed by systemic therapy would be appropriate. The timeline does not provide enough information to confirm resectability, but the presence of multiple nodules and lymph node involvement may suggest that surgery is not the first choice. Given the complexity and the need for multidisciplinary evaluation, the answer should reflect the standard approach for stage I-III NSCLC when surgery is feasible. However, the timeline does not confirm surgery is feasible, so definitive chemoradiation is a reasonable option. But the options include "Surgery followed by systemic therapy" and "Definitive chemoradiation with or without adjuvant systemic therapy." The patient's cardiac status and the need for surgical risk assessment suggest that definitive chemoradiation might be preferred if surgery is not feasible. However, the timeline does not explicitly state that surgery is not feasible, only that it requires risk assessment. In the absence of definitive evidence of inoperability, the standard approach for stage I-III NSCLC is surgery followed by adjuvant therapy if indicated. The presence of lymph node involvement (stage III) would typically require adjuvant chemotherapy after surgery. However, the patient's cardiac status complicates this. The timeline does not provide clear evidence of resectability, but the recommendation to see a cardiologist implies that surgery is a possibility but needs evaluation. Given the information, the most appropriate answer based on standard guidelines for stage I-III NSCLC with no clear contraindication to surgery would be "Surgery followed by systemic therapy." However, the patient's cardiac history and the need for surgical risk assessment might lead to definitive chemoradiation. But since the question is based solely on the timeline and the timeline does not confirm inoperability, the answer should be "Surgery followed by systemic

Subtask: pneumonitis_infection_discussed

Average unique events: 28.53

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: OctoMed/OctoMed-7B

no radiology report: [CORRECT] <think>First, I carefully review the provided patient timeline and imaging findings. The patient has a history of a right upper lobe mass identified on chest x-ray and confirmed by CT scan. The patient is scheduled for a CT-guided biopsy, and a PET scan is planned for staging. The patient's symptoms include chronic cough, occasional productive cough, and minimal dyspnea on exertion. There is no significant weight loss or other systemic symptoms reported. The patient's physical exam is unremarkable with clear lungs bilaterally. Laboratory values show normal hemoglobin, platelets, and coagulation parameters, indicating no immediate contraindications for biopsy. The imaging findings show a right upper lobe mass with mediastinal nodal involvement, which is highly suspicious for malignancy. The patient's history of smoking (though distant) and family history of cancer further support the suspicion of malignancy. There is no mention of prior radiation therapy or infection in the timeline. The question asks whether there is sufficient diagnostic ambiguity to warrant presentation at a thoracic tumor board to differentiate between radiation pneumonitis and infection. Radiation pneumonitis typically occurs after radiation therapy, which is not mentioned in the timeline. Infection would require clinical signs such as fever, leukocytosis, or infiltrates on imaging, none of which are documented here. The imaging shows a mass rather than diffuse infiltrates or consolidation typical of infection. The patient's presentation aligns more with a neoplastic process rather than infection or radiation pneumonitis. Therefore, there is no sufficient ambiguity to warrant tumor board discussion for differentiation between these entities.</think> <answer>No</answer>

radiology report: [CORRECT] <think>First, I need to determine if there's sufficient diagnostic ambiguity to warrant presenting this case at a thoracic tumor board to differentiate between radiation pneumonitis and infection. The key here is to look for evidence in the patient timeline that might suggest either condition or if the findings are ambiguous. Looking at the patient's history: He has a right upper lobe mass confirmed by CT and PET scans. The PET scan shows intense hypermetabolism in the mass and lymphadenopathy, which is suspicious for malignancy. The patient has a history of smoking (quit 30 years ago) and family history of cancer, which increases suspicion for malignancy. The physical exam shows no signs of infection like fever, wheezing, or rales. The lungs are clear bilaterally. He has a chronic cough but no recent weight loss. His vital signs are stable, and oxygen saturation is normal. Lab results show normal hemoglobin, platelets, and coagulation studies, which don't indicate infection or bleeding issues. There's no mention of recent radiation therapy, which would be necessary for radiation pneumonitis. Radiation pneumonitis typically occurs after radiation treatment, and the timeline here doesn't show any such treatment. The patient was referred for CT-guided biopsy, indicating the working diagnosis is malignancy, not infection or radiation pneumonitis. The imaging findings (hypermetabolic mass and lymphadenopathy) strongly suggest malignancy rather than infection or radiation pneumonitis. Radiation pneumonitis would

not typically show such intense FDG uptake. Infections might have variable uptake but usually not as intense as malignancy. Also, the absence of systemic symptoms like fever or elevated inflammatory markers supports malignancy over infection. Therefore, there is no ambiguity; the findings point towards malignancy, and the case should proceed with biopsy and staging. No need to present at a tumor board for differentiation between radiation pneumonitis and infection since radiation pneumonitis is not a consideration here.</think> <answer>No</answer>

Subtask: progression_assessment_discussed

Average unique events: 28.53

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: OctoMed/OctoMed-7B

no radiology report: [CORRECT] <think>First, I carefully reviewed the provided patient timeline and medical records. The patient is a 58-year-old male with a history of chronic cough and a right upper lobe mass identified on chest x-ray and confirmed by CT scan. The CT scan shows a right upper lobe mass and mediastinal nodules, which are highly suspicious for malignancy. The patient has undergone multiple evaluations including blood work and imaging, but no definitive diagnosis or treatment plan is mentioned. The patient is scheduled for a CT-guided biopsy and a PET scan for staging. The timeline does not indicate any prior thoracic tumor board review or multidisciplinary discussion. There is no mention of progression assessment or ambiguity in radiologic interpretation. The imaging findings are consistent with a suspicious mass, but the timeline does not show any conflicting or unclear data that would necessitate a tumor board review at this stage. Therefore, based on the provided information, there is no ambiguity in radiologic interpretation that warrants presentation at a thoracic tumor board for progression assessment.</think> <answer>No</answer>

radiology report: [INCORRECT] <think>First, I need to determine if there is ambiguity in the radiologic interpretation that would warrant presenting this case at a thoracic tumor board for progression assessment. The key here is to look for any conflicting or unclear findings in the imaging reports that might make the interpretation uncertain. Looking at the patient timeline, the initial CT scan on 05/09/2005 showed a right upper lobe mass and multiple mediastinal nodules. The assessment noted a right upper lung mass highly suspicious for malignancy. The plan included a CT-guided biopsy and a PET scan for staging. The PET scan report on 08/21/2005 (dated June 9th, 2005, but likely a typo) shows a large intensely hypermetabolic right lung mass with SUV 19.8, which is highly suspicious for malignancy. It also shows hypermetabolic right peritracheal and bilateral hilar lymphadenopathy, with SUVs of 3.9 and 4.4 respectively. The left lung is negative for abnormal FDG uptake. There is a small focus of borderline increased activity in the right axillary soft tissues with SUV 1.9. Now, the question is whether there is ambiguity in the radiologic interpretation. The PET scan findings are consistent with malignancy given the high SUV values in the primary mass and lymph nodes. However, the borderline increased activity in the right axillary region (SUV 1.9) might be ambiguous. This could represent a benign process or early metastasis, but it's not definitively malignant. The report mentions it as "borderline increased activity," which suggests uncertainty. Additionally, the initial CT showed mediastinal nodules, which were confirmed on PET as hypermetabolic lymphadenopathy. The presence of hypermetabolic lymph nodes supports malignancy, but the axillary finding adds a point of ambiguity. The tumor board would likely want to assess whether this axillary focus is significant or if further evaluation is needed. Therefore, the radiologic interpretation is not entirely clear due to the borderline axillary finding, which warrants a multidisciplinary discussion to determine the next steps and whether this represents progression or a benign condition.</think> <brief reasoning trace with explanation for answer selection> |

<answer> The PET scan report identifies a hypermetabolic right lung mass and hypermetabolic right peritracheal and bilateral hilar lymphadenopathy, strongly suggesting malignancy. However, there is a "small focus of borderline increased activity" in the right axillary region (SUV 1.9), which is not definitively malignant and could represent a benign process or early metastasis. This ambiguity in the axillary finding necessitates a thoracic tumor board review to clarify progression and guide further management. <answer> </answer></think>
<answer>Yes</answer>