Environment. Technology. Resources. Rezekne, Latvia
Proceedings of the 16th International Scientific and Practical Conference. 2025. Volume II, 83-92

Comparative Analysis of the Capabilities of Two Artificial Intelligence Models for Chest X-Ray Image Interpretation

Bora Dogan
Faculty of Medicine
Medical University of Varna „Prof. Dr. Paraskev Stoyanov“
Varna, Bulgaria
bdogan008@gmail.com

Yordanka Eneva
Department of Physics and Biophysics, Faculty of Pharmacy
Medical University of Varna „Prof. Dr. Paraskev Stoyanov“
Varna, Bulgaria
stemivada@abv.bg

Yanka Baneva
Department of Physics and Biophysics, Faculty of Pharmacy
Medical University of Varna „Prof. Dr. Paraskev Stoyanov“
Varna, Bulgaria
yanysh@abv.bg

Elena Ivanova
Medical Imaging Clinic
University Hospital St. Marina
Varna, Bulgaria
ed.ivanova@abv.bg

***Abstract*—The automated interpretation of chest X-rays with the help of artificial intelligence (AI) is gaining increasing importance in medical diagnostics. Technologies based on machine learning and deep neural networks demonstrate the potential to assist radiologists by accelerating the diagnostic process and reducing subjective errors. However, the effectiveness of different AI models varies depending on the algorithms, training data used, and methods of analysis. This pilot study conducts a comparative analysis between two AI models – ChatGPT and X-Ray Insight, a GPT model specifically developed to assist in the interpretation of medical images and support doctors in diagnostics. Their ability to interpret chest X-rays is evaluated, along with an examination of their strengths and limitations in real clinical scenarios. A standardized database of radiographs is used, including normal findings, pneumonia, pulmonary infiltrates, and other pathological conditions. The evaluation of the models includes diagnostic accuracy, clarity and correctness of interpretations, the usefulness of the provided explanations, and recommendations for further examinations. The obtained results are compared with expert radiological assessments. A paired t-test revealed no significant differences in Findings Description and Differential Diagnosis, but X-Ray Insight performed significantly better in recommending the Next Steps, suggesting stronger recommendations for further management. The study highlights the challenges and potential of generative and image-based AI models in radiographic diagnostics. Further adaptation and training of the models are necessary to achieve clinical reliability and effectiveness.**

***Keywords*— artificial intelligence, ChatGPT, chest X-rays, X-Ray Insight.**

I. INTRODUCTION

The use of artificial intelligence (AI) in medicine has the potential to revolutionize medical imaging diagnostics. Deep learning algorithms have demonstrated exceptional efficiency in analyzing X-ray images [1]. AI-driven models have demonstrated the potential to assist radiologists in interpreting complex imaging studies, thereby improving diagnostic accuracy and workflow efficiency. The integration of AI into medical imaging has the potential to address challenges such as high workload, diagnostic variability among radiologists, and the need for rapid and accurate assessments [2]. However, despite the growing interest in AI applications, concerns regarding the reliability, clinical applicability, and interpretability of AI-generated analyses remain critical areas of study.

Chest X-ray interpretation is one of the most fundamental tasks in radiology, serving as a primary imaging modality for detecting thoracic diseases [3], including infectious conditions, cardiovascular abnormalities, and malignancies. Traditional diagnostic approaches [4] rely on the expertise of radiologists, who assess imaging findings in conjunction with clinical information to reach an accurate diagnosis and formulate

Online ISSN 2256-070X
https://doi.org/10.17770/etr2025vol2.8573

an appropriate treatment plan. In recent years, AI-based models have been developed to automate and enhance the analysis of chest X-rays, offering promising results in disease detection, differential diagnosis formulation, and clinical decision support [5]. However, a crucial factor in X-ray interpretation is image quality, which directly impacts diagnostic accuracy. Research on computational phantoms and simulation studies, particularly in mammography [6], [7], has demonstrated their ability to enhance image quality and refine key imaging parameters. Integrating these advancements into AI-driven chest X-ray analysis could help minimize interpretation errors and improve overall diagnostic performance.

This study aims to evaluate and compare the performance of two AI models, ChatGPT-4o and X-Ray Insight, in interpreting chest X-ray images. The study specifically assesses their capabilities across three key criteria: Findings Description, Differential Diagnosis [8], [9], and Next Steps. By analyzing the effectiveness of these models, we seek to determine their potential role in assisting radiologists and enhancing patient care. Furthermore, this research explores whether AI-generated recommendations align with those of human experts and whether they can be effectively integrated into clinical workflows. Understanding the strengths and limitations of these AI systems is essential for optimizing their use in practice and ensuring that they complement rather than replace human expertise.

II. MATERIALS AND METHODS

This pilot study is a comparative evaluation of the diagnostic accuracy of two AI models, ChatGPT-4o of OpenAI and X-Ray Insight, developed by AiWebTools.Ai, in analyzing chest X-rays.

The goal was to assess their ability to detect, classify, and provide differential diagnoses for chest abnormalities compared to ground truth diagnoses and expert radiologist interpretations.

TABLE 1 SUMMARY OF CLINICAL CASE CHARACTERISTICS

Category	Details
Imaging View	PA: 16 (80%), AP: 4 (20%)
Patient Gender	Male: 12 (60%), Female: 8 (40%)
Age Range	9 to 90 years
Pathology Type	Infectious: 6 (30%), Cardiovascular: 4 (20%), Neoplastic: 5 (25%), Other: 5 (25%)

Chest X-ray images were obtained from Radiopaedia, a reputable platform for medical imaging cases. [10 – 29] To identify and select images relevant to this study, specific filters were applied within the "Cases" section: the system was set to "chest," and the study modality was restricted to "X-ray." The search results were sorted by case completion percentage, and only cases with a high diagnostic certainty were included. A total of 20 clinical cases, comprising chest X-ray images, were selected. These cases encompassed a wide range of thoracic pathologies, ensuring a diverse and representative dataset for analysis (Table 1).

Of the 20 cases, 80% (16/20) were posterior-anterior (PA) views, while 20% (4/20) were anterior-posterior (AP) views. The cases included 60% (12/20) male patients and 40% (8/20) female patients, with an age range of 9 to 90 years. The pathologies represented infectious diseases (e.g., pneumonia, tuberculosis), cardiovascular conditions (e.g., aortic valve stenosis, pericardial effusion), neoplastic processes (e.g., adenocarcinoma, thymoma), and other thoracic abnormalities (e.g., hiatal hernia, Chilaiditi sign).

Next, the following study prompt was carefully crafted to be entered into the AI models:

“This is a chest X-ray in the [PA/AP/lateral view] of a [age]-year-old [gender] patient, presenting with [clinical presentation, e.g., cough, shortness of breath, chest pain, fever, hemoptysis, or routine screening]. Please describe the findings in the image, including abnormalities in the lungs (e.g., opacities, nodules, pneumothorax, pleural effusion), mediastinum (e.g., widening, masses, lymphadenopathy), cardiac silhouette (e.g., size, shape, chamber enlargement), bony structures (e.g., fractures, lesions), and any other findings (e.g., diaphragmatic elevation, free air). Provide a differential diagnosis, listing the most likely diagnosis first, and justify your reasoning with a confidence level (low, intermediate, high) for each. Recommend next steps for the physician, such as additional imaging (e.g., CT scan, ultrasound), laboratory tests, biopsy, follow-up intervals, or urgent interventions (e.g., thoracentesis, chest tube placement) if life-threatening conditions are suspected.”

The prompt was designed to guide the AI through a systematic and detailed analysis of the chest X-ray, ensuring comprehensive and structured responses. It also encouraged the AI to provide rationales and evidence for its assessments, which was critical for evaluating the model’s reasoning capabilities. To maintain consistency and enable fair comparisons across different AI models and datasets, the prompt was standardized.

The final step involved inputting the prompt, along with a single chest X-ray image from each case, into a new chat session within the AI models. This approach was essential for several reasons. First, it ensured that the AI evaluated each image independently, without influence or bias from previous cases, closely mimicking real-world clinical practice where radiologists assess each case on its own merits. Second, starting a new session prevented the AI from retaining context or information from prior interactions, which could distort results and compromise the fairness of the evaluation. Finally, this method maintained consistency across the study, enabling a more accurate comparison of the AI’s performance across different cases and models. By isolating each case, the study upheld scientific rigor and ensured the reliability of its findings.

The responses generated by the AI were recorded in a shared document, and the relevant portions were organized into a table to facilitate comparison and evaluation. The radiologist's remarks for each case were also included in the comparison to ensure a balanced assessment.

The paired t-test was conducted, using R software (version 4.3.2; 2023 The R Foundation for Statistical Computing) to compare the performance of ChatGPT-4o and X-Ray Insight in interpreting chest X-ray images across three criteria: Findings Description, Differential Diagnosis, and Next Steps. The paired t-test was chosen because both models evaluated each case, allowing for a direct comparison of their performance. This method is ideal for detecting statistically significant differences in means while accounting for the paired nature of the data.

III. Results and discussion

A board-certified radiologist analyzed the AI models' responses for each clinical case across three components: Findings Description, Differential Diagnosis, and Next Steps. The radiologist evaluated each component across all cases, assigning scores on a scale from 2 to 6, with 2 representing the lowest grade and 6 the highest. The table with AI-generated answers to the relative clinical cases together with scores from the radiologist are given in Appendix.

Fig. 1 illustrates the percentage distribution of individual scores assigned to ChatGPT-4o and X-Ray Insight across all components of the clinical cases. The score distribution indicates that X-Ray Insight received a higher proportion of top scores (6) at 41.67%, compared to 33.33% for ChatGPT-4o, suggesting stronger overall performance. Mid-range scores (4 and 5) were more evenly distributed between the two models, though ChatGPT-4o had a slightly higher percentage of scores in the 4 range (23.33% vs. 16.67%). Lower scores (3 and 2) were relatively uncommon for both models, with ChatGPT-4o receiving slightly lower scores than X-Ray Insight.

Fig. 2 presents the average scores assigned to AI models by the radiologist for each clinical case. X-Ray Insight generally performed slightly better in key diagnostic cases. It received higher scores in conditions such as Aortic Valve Stenosis, Giant Hiatal Hernia, and Bronchial Carcinoid. Both models performed equally in cases like Chilaiditi Sign, Oesophageal-Pleural Fistula, Sarcoidosis, Right Lobe Consolidation, Mitral Valve Regurgitation, Adenocarcinoma, Pulmonary Metastases, and COVID-19 Pneumonia, indicating comparable diagnostic accuracy. ChatGPT-4o demonstrated a slight advantage in Phrenic Nerve Palsy but received lower scores for Pneumothorax and Thymoma.

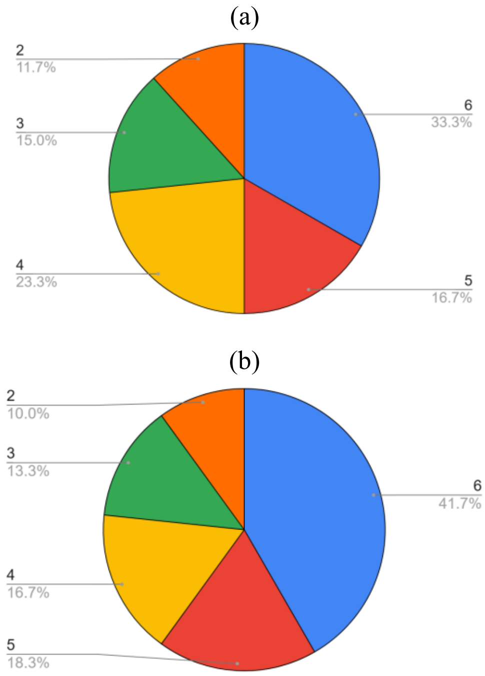


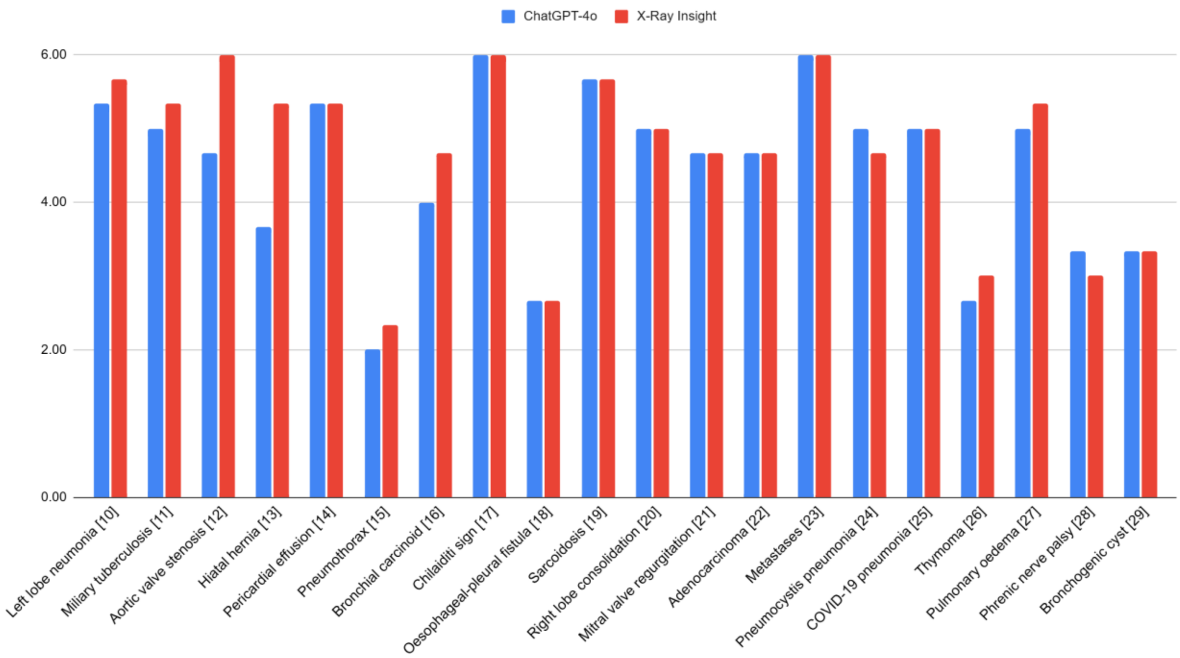
Fig. 1. Percentage distribution of scores assigned to ChatGPT-4o (a) and X-Ray Insight (b).

Table 2 presents the results of the paired t-test, including the mean difference (MD), standard deviation of differences (SD), t-value, degrees of freedom (df), and p-value.

TABLE 2 THE RESULTS OF THE PAIRED T-TEST

Criteria	**MD**	**SD**	**t-value**	**df**	**p-value**
Findings Description	-0.2	0.894	-1	19	0.33
Differential Diagnosis	-0.1	0.852	-0.525	19	0.606
Next Steps	-0.35	0.671	-2.33	19	0.031

The statistical analysis revealed that both AI models performed similarly in Findings Description and Differential Diagnosis, with no significant differences observed ($p = 0.330$ and $p = 0.606$, respectively). These results indicate that ChatGPT-4o and X-Ray Insight possess comparable abilities in identifying radiographic abnormalities and generating differential diagnoses. This suggests that both models could serve as valuable tools in supporting radiologists by providing initial impressions of chest X-ray findings.



Radiologist-assigned average scores, ranging from 2 to 6, for ChatGPT-4o and X-Ray Insight for each clinical case.

However, a notable discrepancy was observed in the Next Steps category, where X-Ray Insight outperformed ChatGPT-4o ($p = 0.031$). This indicates that X-Ray Insight is more effective at recommending appropriate clinical actions based on the interpreted findings. The superior performance in this category may stem from differences in training data, algorithmic design, or model optimization strategies. This finding is particularly relevant in clinical settings, as appropriate follow-up actions are crucial for patient management and treatment planning.

Despite these findings, it is essential to consider the study's limitations. The dataset consisted of only 20 clinical cases, which may not fully represent the broad spectrum of thoracic pathologies encountered in real-world practice. Additionally, while efforts were made to standardize prompts and minimize biases, AI models inherently rely on training data that may introduce certain limitations. Future research should explore larger datasets and incorporate diverse imaging conditions to further validate these results.

IV. Conclusions

This study provides valuable insights into the capabilities of AI models in chest X-ray interpretation. The findings suggest that ChatGPT-4o and X-Ray Insight perform comparably in describing radiographic findings and formulating differential diagnoses. However, X-Ray Insight demonstrates superior performance in recommending the next steps, highlighting its potential as a more clinically actionable tool. While AI-assisted radiology holds promise, further research is necessary to refine these models, expand their training datasets, and ensure their safe and effective integration into clinical workflows. Ultimately, AI should be viewed as a complementary tool to human expertise, enhancing diagnostic accuracy and patient care rather than replacing radiologists.

V. References

[1] M. A. Mazurowski, M. Buda, A. Saha, and M. R. Bashir, "Deep learning in radiology: An overview of the concepts and a survey of the state of the art with a focus on MRI," J. Magn. Reson. Imaging, vol. 49, no. 4, pp. 939–954, 2019, doi: 10.1002/jmri.26534.

[2] J. G. Lee, S. Jun, Y. W. Cho, et al., "Deep learning in medical imaging: general overview," Korean J. Radiol., vol. 18, pp. 570–584, 2017, doi: 10.3348/kjr.2017.18.4.570.

[3] G. Litjens, T. Kooi, B. E. Bejnordi, et al., "A survey on deep learning in medical image analysis," Med. Image Anal., vol. 42, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.

[4] S. Waite, A. Grigorian, R. G. Alexander, S. L. Macknik, M. Carrasco, D. J. Heeger, and S. Martinez-Conde, "Analysis of perceptual expertise in radiology - current knowledge and a new perspective," Front. Hum. Neurosci., vol. 13, p. 213, Jun. 2019, doi: 10.3389/fnhum.2019.00213. Erratum in: Front. Hum. Neurosci., vol. 13, p. 272, Aug. 2019, doi: 10.3389/fnhum.2019.00272.

[5] S. Nazir, D. M. Dickson, and M. U. Akram, "Survey of explainable artificial intelligence techniques for biomedical imaging with deep neural networks," Comput. Biol. Med., vol. 156, p. 106668, 2023, doi: 10.1016/j.compbiomed.2023.106668.

Online ISSN 2256-070X
https://doi.org/10.17770/etr2025vol2.8573

[6] Y. Baneva, T. Georgiev, and K. Bliznakova, "Design and use of computational models for contrast-enhanced spectral mammography," Scripta Scientifica Medica, vol. 54, no. 3, pp. 27-33, 2022, doi: 10.14748/ssm.v54i2.8361.

[7] Y. Baneva and K. Bliznakova, "Influence of incident mammography spectra and breast phantom characteristics on mammography features," in Contemporary Methods in Bioinformatics and Biomedicine and Their Applications, S. S. Sotirov, T. Pencheva, J. Kacprzyk, K. T. Atanassov, E. Sotirova, and G. Staneva, Eds., BioInfoMed 2020, Lecture Notes in Networks and Systems, vol. 374, Springer, Cham, 2022, pp. 573-581, doi: 10.1007/978-3-030-96638-6_41.

[8] Y. Eneva and B. Dogan, "A Qualitative Assessment of Medical Diagnosis Capabilities of Three Artificial Intelligence Models: ChatGPT-4o, CodyMD, and Dr. Gupta," J. Biosci. Med., vol. 12, pp. 243–254, 2024, doi: 10.4236/jbm.2024.1211021.

[9] Y. Eneva and B. Dogan, "Evaluation of Medical Diagnosis Capabilities of Three Artificial Intelligence Models – ChatGPT-3.5, Google Gemini, Microsoft Copilot: Sustainable Development Goals (SDGs)," J. Lifestyle and SDGs Review, vol. 5, no. 2, e03545, 2025, doi: 10.47172/2965-730X.SDGsReview.v5.n02.pe03545.

[10] R. Kulkarni, "Left upper lobe pneumonia," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-27378. [Accessed: Feb. 12, 2025].

[11] M. Holland, "Miliary tuberculosis," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-20025. [Accessed: Feb. 12, 2025].

[12] F. Gaillard, "Aortic valve stenosis," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-15391. [Accessed: Feb. 12, 2025].

[13] K. Dreżewski, "Giant hiatal hernia," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-95741. [Accessed: Feb. 24, 2025].

[14] H. Knipe, "Pericardial effusion," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-42975. [Accessed: Feb. 12, 2025].

[15] M. Mohammed, "Traumatic pneumothorax," Case study, Radiopaedia.org, [Online]. Available: https://radiopaedia.org/cases/204575. [Accessed: Feb. 13, 2025].

[16] L. Silverstone, "Bronchial carcinoid," Case study, Radiopaedia.org, [Online]. Available: https://radiopaedia.org/cases/202984. [Accessed: Feb. 13, 2025].

[17] A. Shetty, "Chilaiditi sign," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-26874. [Accessed: Feb. 14, 2025].

[18] L. Hartley, "Oesophageal-pleural fistula (spontaneous)," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-52346. [Accessed: Feb. 14, 2025].

[19] M. Hall, "Sarcoidosis," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-44870. [Accessed: Feb. 14, 2025].

[20] M. Mohammed, "Right lower lobe consolidation," Case study, Radiopaedia.org, [Online]. Available: https://radiopaedia.org/cases/205135. [Accessed: Feb. 15, 2025].

[21] M. Holland, "Mitral valve regurgitation," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-18796. [Accessed: Feb. 15, 2025].

[22] J. Ren, "Adenocarcinoma of lung," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-24509. [Accessed: Feb. 15, 2025].

[23] A. Shetty, "Pulmonary metastases - cannonball," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-27024. [Accessed: Feb. 15, 2025].

[24] A. Dixon, "Pneumocystis pneumonia," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-9613. [Accessed: Feb. 15, 2025].

[25] M. Al-Tibi, "COVID-19 pneumonia," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-75305. [Accessed: Feb. 15, 2025].

[26] A. Alomari, "Thymoma," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-181667. [Accessed: Feb. 15, 2025].

[27] A. Alhelali, "Pulmonary oedema," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-51448. [Accessed: Feb. 15, 2025].

[28] D. Smith, "Left phrenic nerve palsy," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-37404. [Accessed: Feb. 15, 2025].

[29] A. Johnstone, "Bronchogenic cyst," Case study, Radiopaedia.org, [Online]. Available: https://doi.org/10.53347/rID-77801. [Accessed: Feb. 15, 2025].

APPENDIX

TABLE. AI-GENERATED ANSWERS TO THE RELATIVE CLINICAL CASES TOGETHER WITH SCORES FROM THE RADIOLOGIST

To visually highlight the performance of the AI models relative to the correct answers, colors to the cells in the table were assigned based on the closeness of the answers to the ground truth.	
Green	The AI model's answer is correct or very close to the correct answer
Yellow	The AI model's answer is partially correct or has minor inaccuracies
Red	The AI model's answer is incorrect or significantly deviates from the correct answer

Case	Criteria	ChatGPT-4o	X-Ray Insight	Ground Truth	Specialist
1	**Findings Description**	Left upper lobe consolidation noted, but air bronchograms not clearly visible. No pleural effusion or pneumothorax.	Left upper lobe consolidation with air bronchogram sign noted. No pleural effusion or pneumothorax	Left upper zone consolidation with air bronchograms typical of pneumonia. No pleural effusion, pneumothorax, or other abnormalities	5; 6
	Differential Diagnosis	Left upper lobe pneumonia (most likely), but includes less likely diagnoses like lung abscess and	Community-acquired pneumonia (most likely), with appropriate consideration of TB and pulmonary	Left upper lobe pneumonia	5; 6

		malignancy without strong justification	embolism. Less emphasis on unlikely diagnoses like malignancy		
	Next Steps	Recommends antibiotics, CT scan, sputum analysis, and blood cultures. Suggests follow-up in 6 weeks, which is less urgent than needed	Recommends immediate antibiotics, CT scan, sputum analysis, blood cultures, and repeat X-ray in 48-72 hours. Considers hospitalization if severe, aligning well with clinical urgency	Immediate antibiotics, follow-up imaging, sputum analysis, and blood cultures. Consider hospitalization if severe	6; 5
2	**Findings Description**	Multiple small, diffuse nodular opacities throughout both lung fields. No pleural effusion, pneumothorax, or lymphadenopathy	Diffuse micronodular opacities throughout both lung fields. No pleural effusion, pneumothorax, or obvious lymphadenopathy	Widespread, randomly distributed innumerable tiny nodules in both lungs. No consolidation, pleural fluid, or lymphadenopathy	4; 4
	Differential Diagnosis	Miliary TB (most likely), disseminated fungal infection, metastatic disease, sarcoidosis, pneumoconiosis	Miliary TB (most likely), disseminated fungal infection, metastatic disease, pneumoconiosis, sarcoidosis	Miliary TB (most likely), miliary metastases, atypical infections, hypersensitivity, pneumoconioses (if history fits)	6; 6
	Next Steps	Recommends HRCT, sputum AFB, GeneXpert, blood cultures, and HIV testing. Suggests bronchoscopy if sputum is negative, but lacks emphasis on urgency of anti-TB therapy	Recommends immediate sputum AFB, GeneXpert, HRCT, HIV testing, and urgent anti-TB therapy. Highlights need for hospitalization if worsening	Immediate TB confirmation (sputum AFB, GeneXpert), HRCT chest, HIV testing, and urgent anti-TB therapy	5; 6
3	**Findings Description**	Mild aortic valve calcification noted, but misses the prominent ascending aorta. Correctly identifies no lung/pleural abnormalities	Prominent aortic arch and widened mediastinum noted, consistent with ascending aorta enlargement. Correctly identifies no lung/pleural abnormalities	Prominent right mediastinal border (ascending aorta), normal descending aorta, normal heart size, no lung/pleural abnormalities	4; 6
	Differential Diagnosis	Aortic valve stenosis (most likely), left ventricular hypertrophy, early heart failure, coronary artery disease (less likely)	Aortic valve stenosis (most likely), hypertensive heart disease, aortic aneurysm (less likely), mild pulmonary congestion	Aortic valve stenosis (most likely), hypertensive heart disease, aortic aneurysm (less likely)	6; 6
	Next Steps	Recommends echocardiography, CT chest, BNP, and cardiology referral. Lacks emphasis on urgency if severe aortic stenosis is confirmed	Recommends echocardiography, CT angiography, BNP, ECG, and cardiology referral. Highlights urgency if severe aortic stenosis is confirmed	Echocardiography (TTE) to confirm aortic stenosis, CT angiography if aortic aneurysm suspected, BNP, ECG, cardiology referral	4; 6
4	**Findings Description**	Correctly identifies pacemaker placement and no pneumothorax. Misses the large air-fluid level but notes elevated left hemidiaphragm and abnormal gastric position	Correctly identifies pacemaker placement, no pneumothorax, and notes a large retrocardiac radiolucency (giant hiatal hernia). Also notes elevated diaphragm	Pacemaker leads correctly placed. Single large air-fluid level (giant hiatal hernia). No pneumothorax or pleural effusion	4; 6
	Differential Diagnosis	Giant hiatal hernia (most likely), congestive heart failure, COPD, aortic aneurysm. Misses retrocardiac lung abscess and pneumopericardium	Giant hiatal hernia (most likely), congestive heart failure, COPD. Focuses on the correct primary diagnosis and provides relevant alternatives	Giant hiatal hernia (most likely), retrocardiac lung abscess, pneumopericardium	2; 4
	Next Steps	Recommends CT chest/abdomen, echocardiogram, BNP, and barium swallow study. Lacks emphasis on surgical consultation for symptomatic hiatal hernia	Recommends CT chest, barium swallow study, echocardiogram, BNP, and surgical consultation for symptomatic hiatal hernia. Aligns well with clinical urgency	CT scan to confirm hiatal hernia, barium swallow study, echocardiogram, BNP, and surgical consultation if symptomatic	5; 6
5	**Findings Description**	Correctly identifies enlarged cardiac silhouette and bilateral pleural effusions but misses the small left pleural effusion specifically	Correctly identifies enlarged cardiac silhouette, bilateral pleural effusions, and notes the globular shape suggestive of pericardial effusion	Enlarged cardiac silhouette, small left pleural effusion, no focal consolidation	4; 4
	Differential Diagnosis	Pericardial effusion (most likely), congestive heart failure, pneumonia, malignancy, constrictive pericarditis	Pericardial effusion (most likely), congestive heart failure, pneumonia, malignancy	Pericardial effusion (most likely), cardiomegaly, pericardial effusion, cardiac tumors (rare)	6; 6
	Next Steps	Recommends echocardiogram (STAT), ECG, CT scan, BNP, cardiac enzymes, and pericardiocentesis if tamponade is present	Recommends echocardiogram (STAT), CT scan, BNP, cardiac enzymes, and pericardiocentesis if tamponade is present. Also suggests	Echocardiogram (STAT), CT scan, BNP, cardiac enzymes, pericardiocentesis if tamponade suspected	6; 6

			thoracic ultrasound for pleural effusion		
6	**Findings Description**	Misses pneumothorax entirely. Incorrectly suggests a possible rib fracture	Misses pneumothorax but correctly notes no rib fractures or mediastinal shift. Suggests mild interstitial changes, which could be reactive	Left visceral pleural edge with no lung markings (pneumothorax). No rib fracture, no mediastinal shift, no soft tissue emphysema	2; 2
	Differential Diagnosis	Misses pneumothorax entirely. Suggests rib fracture (incorrect) and pulmonary contusion	Misses pneumothorax but suggests pulmonary contusion (most likely), rib microfracture, and occult pneumothorax	Traumatic pneumothorax (most likely), pulmonary contusion, rib fracture (if present)	2; 2
	Next Steps	Misses pneumothorax entirely. Recommends CT scan and follow-up X-ray but does not address pneumothorax management	Misses pneumothorax but recommends CT scan if symptoms worsen and suggests chest tube placement if pneumothorax develops	Immediate chest tube placement if pneumothorax is large or symptomatic. Follow-up imaging if needed	2; 3
7	**Findings Description**	Identifies increased opacity in the right lung and possible pleural effusion but misses hypertransradiancy and indistinct right heart border	Identifies increased opacity in the right lower lung field and blunted costophrenic angle (effusion). Misses hypertransradiancy but notes elevated right hemidiaphragm	Patchy consolidation in the right lower zone, indistinct right heart border, hypertransradiancy of the right upper/mid zones, no effusions	4; 5
	Differential Diagnosis	Suggests bronchial carcinoid tumor (most likely), pneumonia, pulmonary embolism, and malignancy	Suggests pulmonary carcinoid tumor (most likely), atypical pneumonia, pleural effusion, and lung mass. Focuses on carcinoid as the primary diagnosis	Atypical bronchial carcinoid tumor (most likely), pneumonia, pulmonary embolism, malignancy	4; 4
	Next Steps	Recommends CT chest, ultrasound, thoracentesis, and bronchoscopy with biopsy. Misses tumor markers and surgical referral	Recommends CT chest, ultrasound, thoracentesis, bronchoscopy with biopsy, and tumor markers. Emphasizes referral to pulmonology and oncology, aligning with surgical referral	CT chest with contrast, bronchoscopy with biopsy, thoracentesis if effusion present, tumor markers, and surgical referral	4; 5
8	**Findings Description**	Correctly identifies bowel gas interposition (Chilaiditi sign) and rules out pneumoperitoneum. No other abnormalities noted	Correctly identifies bowel gas interposition (Chilaiditi sign) and rules out pneumoperitoneum. No other abnormalities noted	Interposition of bowel loop (hepatic flexure of colon) between the right hemidiaphragm and liver. No other abnormalities	6 6
	Differential Diagnosis	Chilaiditi sign (most likely), pneumoperitoneum (less likely), right hemidiaphragmatic paralysis (less likely)	Chilaiditi sign (most likely), Chilaiditi syndrome, pneumoperitoneum (less likely)	Chilaiditi sign (most likely), pneumoperitoneum (less likely)	6 6
	Next Steps	Recommends no further intervention if asymptomatic. Suggests abdominal X-ray or CT scan if perforation is suspected	Recommends no further intervention if asymptomatic. Suggests abdominal X-ray, CT scan, or ultrasound if perforation or obstruction is suspected	No further intervention if asymptomatic. CT scan if perforation is suspected	6 6
9	**Findings Description**	Identifies mediastinal shift but misunderstnads large left-sided pleural effusion for right-sided and misses the gas-containing collection and air-fluid level	Identifies left-sided pleural effusion but misses the gas-containing collection, air-fluid level, and mediastinal shift	Large gas-containing collection in the left hemithorax with air-fluid level, loss of left heart and diaphragm silhouette, mediastinal shift to the right, and overlying atelectasis	2 2
	Differential Diagnosis	Suggests malignant pleural effusion, parapneumonic effusion, TB, CHF, and esophageal-pleural fistula. Misses hiatus hernia and pyopneumothorax	Suggests malignant pleural effusion, parapneumonic effusion, CHF, and esophageal-pleural fistula. Misses hiatus hernia and pyopneumothorax	Esophageal-pleural fistula (most likely), hiatus hernia, intrapulmonary cavitating lesion, hydro- or pyopneumothorax	3 3
	Next Steps	Recommends thoracentesis, CT chest, pleural fluid analysis, and oncology referral. Misses water-soluble contrast swallow and intercostal drain placement	Recommends thoracentesis, CT chest, pleural fluid analysis, and esophagogram. Misses intercostal drain placement and surgical referral	CT chest, water-soluble contrast swallow, pleural fluid analysis, intercostal drain placement, and surgical referral	3 3
10	**Findings Description**	Correctly identifies bilateral reticulonodular opacities and hilar	Correctly identifies bilateral reticulonodular opacities and hilar	Widespread reticulonodular pattern, bilateral hilar and paratracheal	5 5

		lymphadenopathy. Misses paratracheal lymphadenopathy and sparing of lung bases	lymphadenopathy. Misses paratracheal lymphadenopathy and sparing of lung bases	lymphadenopathy, relative sparing of lung bases	
	Differential Diagnosis	Sarcoidosis (most likely), tuberculosis, lymphoma, hypersensitivity pneumonitis, pneumoconiosis	Sarcoidosis (most likely), tuberculosis, lymphoma, hypersensitivity pneumonitis	Sarcoidosis (most likely), tuberculosis, lymphoma, hypersensitivity pneumonitis, pneumoconiosis	6 6
	Next Steps	Recommends CT chest, serum ACE, TB workup, bronchoscopy with biopsy, calcium levels, and referral to pulmonology	Recommends CT chest, serum ACE, TB workup, bronchoscopy with biopsy, calcium levels, and referral to pulmonology	CT chest, serum ACE, TB workup, bronchoscopy with biopsy, calcium levels, and referral to pulmonology	6 6
11	**Findings Description**	Correctly identifies right lower lobe consolidation and silhouetting of the right hemidiaphragm. Misses air bronchograms and loss of lucency over the lower thoracic spine	Correctly identifies right lower lobe consolidation and silhouetting of the right hemidiaphragm. Misses air bronchograms and loss of lucency over the lower thoracic spine	Homogeneous opacification of the right lower lung zone with air bronchograms, silhouetting of the right hemidiaphragm, and loss of lucency over the lower thoracic spine	4 4
	Differential Diagnosis	Bacterial pneumonia (most likely), atypical pneumonia, tuberculosis, aspiration pneumonia, lung abscess/empyema	Bacterial pneumonia (most likely), atypical pneumonia, tuberculosis, aspiration pneumonia	Bacterial pneumonia (most likely), atypical pneumonia, tuberculosis, aspiration pneumonia, lung abscess/empyema	5 5
	Next Steps	Recommends empiric antibiotics, sputum culture, chest X-ray follow-up, CT scan if no improvement, and TB workup if indicated	Recommends empiric antibiotics, sputum culture, chest X-ray follow-up, CT scan if no improvement, and TB workup if indicated	Empiric antibiotics, sputum culture, chest X-ray follow-up, CT scan if no improvement, and TB workup if indicated	6 6
12	**Findings Description**	Correctly identifies cardiomegaly, left atrial enlargement, and pulmonary venous congestion. Misses specific features like double contour and splayed carina	Correctly identifies cardiomegaly, left atrial enlargement, and pulmonary venous congestion. Misses specific features like double contour and splayed carina	Cardiomegaly, left atrial enlargement (double contour, splayed carina, enlarged left atrial auricle), pulmonary venous congestion	4 4
	Differential Diagnosis	Mitral valve disease (most likely), congestive heart failure, hypertensive heart disease, pericardial effusion	Mitral regurgitation (most likely), congestive heart failure, hypertensive heart disease, pericardial effusion	Mitral regurgitation (most likely), congestive heart failure, hypertensive heart disease, pericardial effusion	5 5
	Next Steps	Recommends echocardiography, BNP, ECG, cardiology referral, and consideration of valve repair/replacement	Recommends echocardiography, BNP, ECG, cardiology referral, and consideration of valve repair/replacement	Echocardiography (urgent), BNP, ECG, cardiology referral, and consideration of valve repair/replacement	5 5
13	**Findings Description**	Identifies increased opacity in the central/lower lung zones but misses the specific right lower lobe lesion and hilum overlay sign	Identifies an ill-defined opacity in the left mid-lung field (incorrect side) and misses the right lower lobe lesion and hilum overlay sign	Round opacity lesion in the right lower lobe (hilum overlay sign), no pneumothorax or pleural effusion	2 2
	Differential Diagnosis	Primary lung malignancy, pulmonary metastasis, post-radiotherapy fibrosis, infectious process	Primary lung adenocarcinoma, pulmonary metastasis, post-radiotherapy fibrosis, infectious process	Primary lung adenocarcinoma (most likely), pulmonary metastasis, post-radiotherapy fibrosis, infectious process	6 6
	Next Steps	Recommends CT chest, PET-CT, biopsy, tumor markers, and follow-up imaging	Recommends CT chest, PET-CT, biopsy, tumor markers, and follow-up imaging	CT chest, PET-CT, biopsy, tumor markers, and follow-up imaging	6 6
14	**Findings Description**	Correctly identifies multiple well-defined nodules, no pleural effusion, no pneumothorax, and normal bony structures	Correctly identifies multiple well-defined nodules, no pleural effusion, no pneumothorax, and normal bony structures	Multiple well-defined nodules in both lungs, no pleural effusion, no pneumothorax, normal bony structures	6; 6
	Differential Diagnosis	Pulmonary metastases (most likely), septic emboli, granulomatous diseases (e.g., sarcoidosis, fungal infections)	Pulmonary metastases (most likely), septic emboli, granulomatous diseases (e.g., TB, fungal infections)	Pulmonary metastases (most likely), septic emboli, granulomatous diseases (e.g., TB, fungal infections)	6 6
	Next Steps	Recommends CT chest, PET-CT, tumor markers, biopsy if needed, oncology consultation, and follow-up imaging	Recommends CT chest, PET-CT, tumor markers, biopsy if needed, oncology consultation, and follow-up imaging	CT chest, PET-CT, tumor markers, biopsy if needed, oncology consultation, and follow-up imaging	6 6

15	**Findings Description**	Correctly identifies diffuse, bilateral, hazy opacities with a ground-glass appearance. Misses pneumatocoeles	Identifies diffuse bilateral reticulonodular opacities but misses hazy perihilar opacification and pneumatocoeles	Hazy perihilar mid and upper zone opacification, interstitial prominence, pneumatocoeles, no pleural effusion or nodal enlargement	4 5
	Differential Diagnosis	Pneumocystis pneumonia (most likely), atypical pneumonia, miliary TB, pulmonary edema	Pulmonary TB (most likely), Pneumocystis pneumonia, fungal infection, lymphoma	Pneumocystis pneumonia (most likely), TB, fungal infection, viral pneumonia	5 5
	Next Steps	Recommends CT chest, sputum/BAL PCR, LDH, beta-D-glucan, empiric TMP-SMX, corticosteroids if hypoxic	Recommends CT chest, sputum AFB smear/culture, GeneXpert, beta-D-glucan, and empiric anti-TB therapy. Misses empiric TMP-SMX for Pneumocystis pneumonia	CT chest, sputum/BAL PCR for Pneumocystis, LDH, beta-D-glucan, empiric TMP-SMX, corticosteroids if hypoxic	6 4
16	**Findings Description**	Identifies bilateral diffuse patchy opacities but misses specific consolidations and air bronchograms	Identifies bilateral diffuse opacities but misses specific consolidations and air bronchograms	Bilateral peripheral consolidations with air bronchograms, right upper lobe pneumonia, left lower lobe pneumonia, support lines in situ	3 3
	Differential Diagnosis	Severe COVID-19 pneumonia (most likely), bacterial pneumonia, ARDS, cardiogenic pulmonary edema	Severe COVID-19 pneumonia (most likely), bacterial pneumonia, ARDS, cardiogenic pulmonary edema	Severe COVID-19 pneumonia, bacterial pneumonia, ARDS, cardiogenic pulmonary edema	6 6
	Next Steps	Recommends CT chest, COVID-19 PCR, blood cultures, ABG, broad-spectrum antibiotics, corticosteroids, and ventilatory support	Recommends CT chest, blood cultures, ABG, broad-spectrum antibiotics, corticosteroids, and ventilatory support. Misses COVID-19 PCR	CT chest, COVID-19 PCR, blood cultures, ABG, broad-spectrum antibiotics, corticosteroids, and ventilatory support	6 6
17	**Findings Description**	Identifies increased opacification in the right lower lung field but misses the specific lesion and hilum overlay sign	Identifies increased opacity in the right middle/lower lung zones but misses the specific lesion and hilum overlay sign	Well-defined, smoothly marginated opaque lesion in the right cardiophrenic angle, hilum overlay sign, no pleural effusion or pneumothorax	3 3
	Differential Diagnosis	Pulmonary fibrosis, lung malignancy, atypical pneumonia, aortic aneurysm. Misses thymoma and other mediastinal masses	Pulmonary malignancy, pneumonia, tuberculosis, sarcoidosis. Misses thymoma and other mediastinal masses	Thymoma (most likely), lymphoma, teratoma, pericardial cyst	2 2
	Next Steps	Recommends CT chest, sputum culture, bronchoscopy, and oncology referral. Misses biopsy and tumor markers	Recommends CT chest, CBC, tumor markers, and biopsy. Misses referral to thoracic surgery	CT chest, biopsy, tumor markers, and referral to oncology or thoracic surgery	3 4
18	**Findings Description**	Correctly identifies bilateral interstitial and alveolar opacities, cardiomegaly, and pacemaker in situ. Misses upper lobe pulmonary venous diversion and Kerley B lines	Correctly identifies bilateral interstitial and alveolar opacities, cardiomegaly, and pacemaker in situ. Misses upper lobe pulmonary venous diversion and Kerley B lines	Bilateral lung congestion, upper lobe pulmonary venous diversion, Kerley B lines, cardiomegaly, pacemaker in situ	4 5
	Differential Diagnosis	Congestive heart failure (most likely), volume overload, pneumonitis, pulmonary venous hypertension	Congestive heart failure (most likely), pneumonia, interstitial lung disease	Cardiogenic pulmonary edema (most likely), pneumonia, interstitial lung disease	5 5
	Next Steps	Recommends echocardiogram, BNP, diuretics, oxygen therapy, and cardiology referra	Recommends chest CT, BNP, diuretics, oxygen therapy, and pacemaker evaluation. Misses echocardiogram and cardiology referral	Echocardiogram, BNP, diuretics, oxygen therapy, and cardiology referral	6 6
19	**Findings Description**	Identifies elevated left hemidiaphragm and possible left lower lung opacity but misses surgical clips and specific findings like no consolidation or atelectasis	Identifies elevated left hemidiaphragm and left lower lung opacity but misses surgical clips and specific findings like no consolidation or atelectasis	Elevated left hemidiaphragm, surgical clips at base of left neck, no consolidation, atelectasis, or pneumothorax	4 4
	Differential Diagnosis	Post-surgical changes, postoperative pneumonia, atelectasis, pleural effusion. Misses phrenic nerve palsy as the primary diagnosis	Post-operative atelectasis, pleural effusion, pulmonary edema, aspiration pneumonia. Misses phrenic nerve palsy as the primary diagnosis	Phrenic nerve palsy (most likely), atelectasis, pleural effusion, aspiration pneumonia	3 2

	Next Steps	Recommends CT chest, ultrasound, sputum culture, CBC, and follow-up imaging. Misses specific focus on phrenic nerve palsy	Recommends repeat chest X-ray, ultrasound, CBC, sputum culture, and supportive management. Misses CT chest and specific focus on phrenic nerve palsy	CT chest, ultrasound, sputum culture, CBC, and follow-up imaging	3 3
20	**Findings Description**	Identifies bilateral diffuse airspace opacities but misses splaying of the carina, subcarinal mass, and retrocardiac opacification	Identifies bilateral diffuse interstitial and alveolar opacities but misses splaying of the carina, subcarinal mass, and retrocardiac opacification	Splaying of the carina, large subcarinal mass, prominent pulmonary vasculature, retrocardiac opacification, bilateral lung opacities	3 3
	Differential Diagnosis	ARDS, peripartum cardiomyopathy, sepsis, atypical pneumonia, pulmonary embolism. Misses bronchogenic cyst	Severe pneumonia, ARDS, peripartum cardiomyopathy, pulmonary embolism. Misses bronchogenic cyst	Bronchogenic cyst (most likely), pneumonia, ARDS, pulmonary embolism	3 3
	Next Steps	Recommends CT chest, echocardiography, ABG, blood cultures, and antibiotics. Misses urgent surgical consultation	Recommends CT chest, ABG, blood cultures, and antibiotics. Misses echocardiography and urgent surgical consultation	CT chest, echocardiography, ABG, blood cultures, and urgent surgical consultation	4 4