

Could ChatGPT-4 be Used to Escalate Abnormal Chest X-Rays to a Radiologist?

Muhammad Abdullah Pirwani¹, Shiraz Imran Syed², Hiba Zafar³ and Muhammad Mudassir Pirwani^{4*}

¹Barts and the London School of Medicine and Dentistry, London, United Kingdom

²Medical University of Sofia, Sofia, Bulgaria

³Jinnah Sindh Medical University, Jinnah Postgraduate Medical Center, Karachi, Pakistan

⁴Southend University Hospital, NHS Foundation Trust, Essex, United Kingdom

ABSTRACT

Background: ChatGPT-4 is a complex neural network primarily used for human interaction. It has been demonstrated that ChatGPT-4 is capable of passing the FRCR 2A examination, thereby showcasing the potential of this artificial intelligence software. With an ever-increasing imaging burden on the healthcare system, AI could be used to escalate abnormal scans for radiographers or radiologists to report, before confirming that normal scans are indeed normal.

Objective: This article studies the potential utility of using ChatGPT-4 to categorize chest X-rays (CXRs) as normal or abnormal, with the potential to prove that AI could be used to escalate abnormal scans and report them as early as possible.

Methodology: The study design used for this study is a retrospective diagnostic study, with the material sourced from various online resources. The study duration is from January 2024 to February 2024. We studied the responses of ChatGPT-4 to 100 CXRs collected from the internet over a period of 1 week. The accuracy of differentiating normal from abnormal X-rays and reaching the correct diagnosis was noted for each radiological image input into the AI tool.

Results: ChatGPT-4 was able to detect normal CXR images with a 100% true negative rate. The true positive rate was 95.6%. ChatGPT-4 was only 35.5% accurate in diagnosing the correct pathology; the most correct diagnoses were those of pulmonary oedema.

Conclusion: Although ChatGPT-4 was able to successfully identify normal CXRs, more sophisticated and refined AI models need to be developed to accurately diagnose them.

Keywords: ChatGPT-4, abnormal chest X-rays, radiologist, radiopedia, CT, MRI.

INTRODUCTION

By incorporating artificial intelligence (AI), the medical industry is undergoing a massive transformation in the way healthcare professionals approach diagnosis and subsequent management of patients [1, 2]. AI, defined as a machine's ability to perform specific tasks by mirroring human behavior and intelligence, has drastically changed over the past few years from a simple concept to a practical innovation with substantial clinical applications [3]. The National Library of Medicine defines AI as "the use of computers to carry out tasks that normally call for objective reasoning and comprehension". This further highlights its capability to boost the limit of human intellectual potential in an intricate medical setting [4].

Healthcare systems have become much better over the past few years due to the capabilities of AI in pattern recognition, data analysis, and decision-making [5-7]. This technological advancement deals with both supervised and unsupervised learning, where AI systems are trained either on distinct datasets or to independently

identify patterns within the data [8]. Concerning medical imaging, a notable example is deep convolutional neural networks (DCNNs), which use sophisticated algorithms to identify medical findings without the need for direct supervision. This leads to unparalleled accuracy when it comes to image interpretation [9].

Multiple medical specialist fields have benefitted from AI integration, ranging from the enhanced decision-making capabilities of an anaesthesiologist to the improved pathological diagnoses and early cancer detection in gastroenterology [10, 11]. These new advances have helped reduce physician workload while also maintaining, or in some cases even improving, the diagnostic accuracy in areas that require extensive pattern recognition and data processing.

However, radiology needs this technological advancement more than ever. Radiologists and trained imaging specialists undergo immense pressure due to the absurd workload they face, and their stress plays an adverse role in both their efficiency and accuracy. The gap between imaging demand and available expertise to read and diagnose those images is growing wider each passing year [12]. As a radiologist, many complex scans (including X-rays, CT scans, and MRI images) have to be visually evaluated, and more often than not, these

*Corresponding Author: Muhammad Mudassir Pirwani, Southend University Hospital, NHS Foundation Trust, Essex, United Kingdom, E-mails: Muhammad.pirwani2@nhs.net; Muhammad.pirwani2@nhs.net
Received: November 19, 2024; Revised: August 23, 2025; Accepted: August 23, 2025
DOI: <https://doi.org/10.37184/nrjp.3007-5181.2.1>

interpretations are subjective and highly dependent on the radiologist's experience and training. This factor becomes particularly troublesome when subtle abnormalities are present in the image, such as radiomic features of the tracheobronchial tree in a CT scan. Even highly experienced practitioners consider it notoriously difficult to detect these minute findings [13].

Diagnostic challenges aren't the only crises that radiologists face; severe workforce constraints and demands for humanly impossible productivity are just a few of the issues that have plagued radiologists around the world. As imaging data continues to grow at rates far exceeding the amount of workforce available to read them, a significant pressure is put on radiologists to increase their productivity while maintaining their diagnostic accuracy. In a study conducted in 2015, the average radiologist was required to interpret one scan every 3-4 seconds during an 8-hour workday [14]. Such a pace will almost inevitably result in an increased number of diagnostic errors and professional burnout.

The situation in the United Kingdom alone is very dire. With approximately 7 million CXRs being requested every year and the Royal College of Radiologists reporting 976,000 awaiting plain film reports for more than 30 days, the healthcare system has reached a bottleneck, and this has directly impacted patient care and disease outcome [15, 16]. These delays can have serious clinical consequences, particularly for patients with conditions that are time-sensitive, where early detection and interventions play a major role in prognosis and outcome.

Amidst this chaos, AI emerges as an answer to both the problems of efficiency and accuracy in image reading. AI's superior pattern recognition software and its ability to automatically provide accurate interpretations make it an indispensable tool for bettering clinical workflow. When integrated properly, AI can be used as a screening tool to provide pre-reviewed images with identified features, therefore increasing productivity and reducing diagnostic errors.

One of the largest advancements in large language model technology, ChatGPT-4, has shown a stunning potential in medical applications [17]. This advanced neural network, mainly designed to imitate human interactions, has demonstrated amazing results in the field of radiology [18-20]. To top it all off, ChatGPT-4 successfully passed the Fellowship of the Royal College of Radiologists (FRCR) part 2A examination. However, it narrowly failed FRCR part 1, hinting at the untapped potential but current limitations of ChatGPT-4 in specialised radiological applications [21].

The current radiological crisis can be addressed by utilizing the concept of AI in CXR triage. Healthcare systems may prioritize abnormal scans for immediate assessment and put the normal ones on hold during triage by implementing AI systems that can identify whether a CXR is normal or not. This may significantly limit reporting delays, and as a result, enhance patient outcomes through improved earlier detection and optimize resource allocation within an already overwhelmed department.

The addition of a successful AI system in CXR triage helps the healthcare systems beyond just workflow improvements. These systems could later encompass more complex imaging modalities like CT scans and MRIs, potentially smoothing out the entire workflow in radiology. Patients with acute conditions could benefit from early detection and treatment of abnormal findings, and such improvements can prove to be life-saving.

Primarily, this study aims to assess whether ChatGPT-4 can accurately differentiate between normal and abnormal CXRs, thereby optimizing radiological workflow by prioritizing abnormal scans for review. This capability could establish a foundation for future AI models to extend similar functionality to CT and MRI scans, potentially enabling faster reporting for acutely ill patients whose management could benefit from earlier intervention, with potentially life-saving implications.

Here, we also aim to analyze the advantages and disadvantages of implementing AI-powered workflows in fluid organizational workflows in healthcare. We will also look into ChatGPT-4's diagnostic capabilities and its clinical usefulness and shortcomings by providing abnormal CXR diagnoses.

MATERIALS AND METHODOLOGY

This study is a retrospective diagnostic accuracy study, with the material sourced from various online resources. The study duration is from January 2024 to February 2024. No patient, participant, or institute was involved in the study, and all radiological scans were taken from the internet. Hence, no Institutional Review Board's approval was taken, and consent was not sought for this study.

100 CXR images were collected from various online resources for this study. The criteria for the images were that only those images would be chosen that were taken at adequate inspiration (8-10 posterior ribs at the midclavicular line). The scans were selected on a random basis with a mix of subtle and obvious findings in a ratio of 1:1. The images were downloaded onto a

computer in .img format. Out of 100 X-rays, 10 were normal CXRs and 90 were pathological CXRs. The pathologies consisted of pneumonia, pleural effusion, pneumothorax, congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), pulmonary oedema, tension pneumothorax, atelectasis, and lung carcinoma/metastasis. One of our researchers randomly selected a mix of subtle and obvious findings as abnormalities. First, an individual CXR was submitted to ChatGPT-4 with the prompt, “Please analyze the following chest X-ray and report it as either normal or abnormal,” and the response from the AI tool was recorded. ChatGPT-4 was then opened on an alternative account, the same image was uploaded again, and ChatGPT-4 was given the following prompt: “Please analyze this chest X-ray and report any pathological findings, or report it as normal,” and the response was recorded. The corresponding diagnosis, whether normal or abnormal, was then confirmed by a radiologist.

Data collected was analyzed using SPSS software version 25.0. Frequencies and percentages were computed to summarize ChatGPT-4 results. Differences in ChatGPT-4’s capacity to recognize normal from abnormal scans and pathological diagnoses among CXRs were evaluated using Chi-square testing. The results were presented in a clear and concise manner with figures and text to illustrate key findings. The statistical analysis used a p-value of <0.05 as the threshold for determining statistical significance.

RESULTS

100 CXRs were submitted to ChatGPT-4. Out of the 10 CXRs that were normal, ChatGPT-4 demonstrated a 100% accuracy rate; 10 out of 10 CXRs were reported as normal. Among the 90 abnormal CXR images, ChatGPT-4 correctly identified 86 CXRs, showcasing a 96% positive result. The remaining 4 images, that is, 4% of the dataset, were false negatives. No false positives were observed in this dataset. Table 1 shows a two-by-two confusion matrix depicting these results.

When asked to identify the pathology, ChatGPT-4 correctly diagnosed 32 CXRs, which amounts to 35.5%.

Table 1: A confusion matrix that shows the performance of ChatGPT-4 in interpreting 100 CXRs. (TN= True negative, FN= False negative, FP= False positive, TP= True positive).

ChatGPT Prediction	Normal (Actual)	Abnormal (Actual)	Total
Normal (Predicted)	10 (TN)	4 (FN)	14
Abnormal (Predicted)	0 (FP)	86 (TP)	86
Total	10	90	100

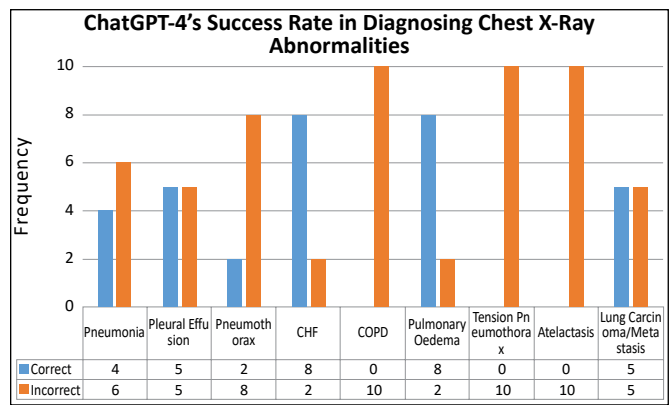


Fig. (1): Figure showing a cluster bar chart representing ChatGPT-4’s correct and incorrect responses in diagnosing chest X-ray abnormalities. (CHF= congestive heart failure, COPD= chronic obstructive pulmonary disorder).

Among the abnormal images, ChatGPT-4’s performance for each pathology is given as follows: ChatGPT-4 correctly diagnosed 4 out of 10 images of pneumonia (40%), 5 out of 10 images of pleural effusion (50%), 2 out of 10 images of pneumothorax (20%), 8 out of 10 images of CHF (80%), 0 out of 10 images of COPD (0%), 8 out of 10 images of pulmonary oedema (80%), 0 out of 10 images of tension pneumothorax (0%), 0 out of 10 images of atelectasis (0%), and 5 out of 10 images of lung carcinoma/metastasis (50%). These preliminary findings indicate ChatGPT-4 is able to identify normal chest X-ray images but is not 100% successful at identifying abnormal chest X-rays. The diseases that were correctly diagnosed the most were pulmonary oedema (80%) and CHF (80%), while COPD, tension pneumothorax, and atelectasis were not identified correctly by ChatGPT-4. Fig. (1) shows a cluster bar chart representing ChatGPT-4’s response to identifying abnormal CXRs.

A Chi-square test of Independence was conducted to evaluate whether ChatGPT-4’s success rates in detecting chest X-ray abnormalities differed significantly across various pathological conditions. The analysis revealed a statistically significant difference in diagnostic performance across conditions, $\chi^2(8) = 391.23$, $p < 0.0001$. This suggests that the variability in success rates is unlikely to be due to chance and indicates that ChatGPT-4’s ability to identify abnormalities is condition-dependent.

DISCUSSION

The results of this research demonstrate both the endless potential and the limitations of ChatGPT-4 in CXR interpretation, revealing a complex landscape of AI performance that requires careful consideration for clinical implementation. While ChatGPT-4 achieved perfect accuracy (100%) in identifying normal CXRs, its

inconsistent performance across different pathological conditions raises significant concerns about patient safety and clinical utility.

The 96% sensitivity for detecting abnormal CXRs, while initially promising, must be interpreted within the context of the 4% false-negative rate. Recent studies have shown that AI tools like qXR, with computer-aided detection (CAD) software, can achieve 99.7% sensitivity for abnormal CXR detection, highlighting the suboptimal performance of ChatGPT-4 in this critical application [22]. Recent studies have shown that comprehensive deep-learning models can achieve superior performance in chest X-ray interpretation across a large breadth of clinical practice [23]. The false-negative rate observed in our study is particularly concerning, given that missed diagnoses can have severe clinical consequences, especially for life-threatening conditions such as tension pneumothorax, which ChatGPT-4 failed to detect in all cases.

The condition-specific performance variations observed in our study align with recent literature documenting the challenges of AI generalizability across different pathologies [24-26]. Contemporary AI systems have demonstrated superior sensitivity for lung lesions (0.83 *versus* 0.52), consolidations (0.88 *versus* 0.78), and atelectasis (0.54 *versus* 0.43) compared to written reports, yet these improvements often come with higher false-detection rates [27]. Our findings of 80% accuracy for both CHF and pulmonary oedema, contrasted with 0% accuracy for COPD, tension pneumothorax, and atelectasis, underscore the need for condition-specific AI training and validation.

The limitations of ChatGPT-4 become more apparent when compared to dedicated chest X-ray AI systems. While ChatGPT-4 demonstrated only 40% accuracy for pneumonia detection in our study, recent foundation auto-refined models specifically trained on chest radiography have shown significant improvements in diagnostic scope, generalizability, and robustness [28]. Advanced AI models have achieved 93.5% accuracy in pneumothorax detection, compared to ChatGPT-4's 20% detection rate for pneumothorax and 0% for tension pneumothorax in our study [29].

Real-world implementations of AI triaging systems have shown more promising results. Studies evaluating AI-assisted chest radiograph interpretation have demonstrated that AI engines can improve reader performance and efficiency when used concurrently with radiologist interpretation [30]. External validations of AI algorithms in clinical practice have shown that AI-

aided interpretation provides significant advantages in the detection and localization of lung lesions [31].

The safety implications of implementing ChatGPT-4 in clinical workflows cannot be overstated. The 4% false-negative rate represents a significant patient safety risk, particularly for critical conditions. Recent systematic reviews have identified that commercially available chest radiograph AI tools, while effective for detecting various pathologies, can produce more false-positive findings than radiology reports, and their performance decreases for smaller-sized target findings and when multiple findings are present [32]. The complete failure to detect tension pneumothorax in our study exemplifies these safety concerns, as this condition requires immediate intervention to prevent fatal outcomes. Specialized AI models for pneumothorax detection have shown much better performance, with studies demonstrating accurate detection of both pneumothorax and tension pneumothorax in chest radiographs [33]. However, factors such as patient positioning and imaging quality can significantly influence AI performance [34].

The integration of AI assistance with radiologist expertise has shown particularly promising results. Studies comparing radiologist performance with and without AI assistance have demonstrated that deep learning algorithms can help radiologists achieve improved efficiency and accuracy in chest radiograph diagnosis [35]. This collaborative approach may represent the optimal path forward for clinical implementation.

We acknowledge the numerous limitations of our study. The retrospective design using internet-sourced images does not reflect the complexity and variability of real clinical practice. Online radiological images are substantial for educational purposes, but often do not reflect the true clinical dataset. Using non-verified online datasets and unblinded radiologist review introduces significant bias and lacks clinical validity. Additionally, the relatively small sample size of 100 images, while adequate for preliminary assessment, may not capture the full spectrum of diagnostic challenges encountered in routine clinical practice. Recent multicenter studies have shown that AI algorithm performance can vary significantly across different institutions and patient populations [31].

CONCLUSION

While ChatGPT-4 demonstrates potential as a screening tool for identifying normal chest X-rays, its current diagnostic capabilities are insufficient for safe clinical

implementation. The 4% false-negative rate and complete failure to detect critical conditions like tension pneumothorax represent unacceptable risks to patient safety.

However, the field of AI-assisted chest X-ray interpretation continues to evolve rapidly. Purpose-built AI systems designed specifically for radiological applications show greater promise for clinical integration. The implementation of AI-assisted chest X-ray interpretation is about time, but it must be done with comprehensive validation, appropriate safeguards, and continuous monitoring to ensure patient safety.

The path forward requires a balanced approach that harnesses AI's potential while maintaining rigorous safety standards. As AI technology advances, the goal should be to develop systems that enhance rather than replace human expertise, ultimately improving patient outcomes through more efficient and accurate diagnostic workflows.

LIST OF ABBREVIATIONS

AI : Artificial Intelligence
 CXR : Chest X-Ray
 CT : Computed Tomography
 MRI : Magnetic Resonance Imaging
 DCNN : Deep Convolutional Neural Networks
 FRCR : Fellowship of the Royal College of Radiologists

ETHICS APPROVAL

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

AUTHORS' CONTRIBUTION

All authors made significant contributions to the scientific content of this manuscript. M.A.P. conceptualized and designed the study, analyzed and interpreted the data, wrote and revised the manuscript. M.M.P., H.Z., and S.S. were involved in data collection

and manuscript revision. All authors read and approved the final manuscript.

REFERENCES

- Zhang P, Zhang Q, Li S. Advancing cancer prevention through an AI-based integration of traditional and Western medicine. *Cancer Discov* 2024; 14(11): 2033-6. DOI: <https://doi.org/10.1158/2159-8290.CD-24-0832>
- Gruson D, Bernardini S, Dabla PK, Gouget B, Stankovic S. Collaborative AI and laboratory medicine integration in precision cardiovascular medicine. *Clin Chim Acta* 2020; 509: 67-71. DOI: <https://doi.org/10.1016/j.cca.2020.06.001>
- Bini SA. Artificial intelligence, machine learning, deep learning, and cognitive computing: What do these terms mean and how will they impact health care? *J Arthroplasty* 2018; 33(8): 2358-61. DOI: <https://doi.org/10.1016/j.arth.2018.02.067>
- Artificial Intelligence [Internet]. NNLM. Available from: <https://www.nlm.gov/guides/data-glossary/artificialintelligence>
- Zhang Q, Zhou D. Machine learning electrocardiogram for mobile cardiac pattern extraction. *Sensors (Basel)* 2023; 23(12): 5723. DOI: <https://doi.org/10.3390/s23125723>
- Lin S, Wei C, Wei Y, Fan J. Construction and verification of an endoplasmic reticulum stress-related prognostic model for endometrial cancer based on WGCNA and machine learning algorithms. *Front Oncol* 2024; 14: 1362891. DOI: <https://doi.org/10.3389/fonc.2024.1362891>
- Elhaddad M, Hamam S. AI-driven clinical decision support systems: An ongoing pursuit of potential. *Cureus* 2024; 16(4): e57728. DOI: <https://doi.org/10.7759/cureus.57728>
- Tack C. Artificial intelligence and machine learning | applications in musculoskeletal physiotherapy. *Musculoskelet Sci Pract* 2019; 39: 164-9. DOI: <https://doi.org/10.1016/j.msksp.2018.11.012>
- Vakalopoulou M, Christodoulidis S, Burgos N, Colliot O, Lepetit V. Deep learning: Basics and convolutional neural networks (CNNs). In: Colliot O, Eds. *Machine Learning for Brain Disorders*. New York, NY: Humana; 2023; pp. 77-15.
- Çelik E, Turgut MA, Aydoğan M, Kılınç M, Toktaş İ, Akelma H. Comparison of AI applications and anesthesiologist's anesthesia method choices. *BMC Anesthesiol* 2025; 25(1): 2. DOI: <https://doi.org/10.1186/s12871-024-02882-2>
- Uche-Anyan E, Anyane-Yebo A, Berzin TM, Ghassemi M, May FP. Artificial intelligence in gastroenterology and hepatology: How to advance clinical practice while ensuring health equity. *Gut* 2022; 71(9): 1909-15. DOI: <https://doi.org/10.1136/gutjnl-2021-326271>
- Hillman BJ, Pandya BJ. Radiologists' burden of inefficiency using conventional imaging workstations. *J Am Coll Radiol* 2013; 10(11): 875-7. DOI: <https://doi.org/10.1016/j.jacr.2013.04.007>
- Shroff GS, Ocazionez D, Vargas D, Carter BW, Wu CC, Nachiappan AC, et al. Pathology of the trachea and central bronchi. *Semin Ultrasound CT MR* 2016; 37(3): 177-89. DOI: <https://doi.org/10.1053/j.sult.2015.11.003>

14. McDonald RJ, Schwartz KM, Eckel LJ, Diehn FE, Hunt CH, Bartholmai BJ, *et al.* The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. *Acad Radiol* 2015; 22(9): 1191-8. DOI: <https://doi.org/10.1016/j.acra.2015.05.007>
15. Standards for the education, training and preceptorship of reporting practitioners in adult chest X-ray | The Royal College of Radiologists [Internet]. 2023; Available from: <https://www.rcr.ac.uk/our-services/all-our-publications/clinical-radiology-publications/standards-for-the-education-training-and-preceptorship-of-reporting-practitioners-in-adult-chest-x-ray/>
16. Radiology delays worst on record despite spend on private providers soaring | The Royal College of Radiologists [Internet]. 2025; Available from: <https://www.rcr.ac.uk/news-policy/latest-updates/radiology-delays-worst-on-record-despite-spend-on-private-providers-soaring/>
17. Günay S, Öztürk A, Özerol H, Yiğit Y, Erenler AK. Comparison of emergency medicine specialist, cardiologist, and chat-GPT in electrocardiography assessment. *Am J Emerg Med* 2024; 80: 51-60. DOI: <https://doi.org/10.1016/j.ajem.2024.03.017>
18. Mao Y, Xu N, Wu Y, Wang L, Wang H, He Q, *et al.* Assessments of lung nodules by an artificial intelligence chatbot using longitudinal CT images. *Cell Rep Med* 2025; 6(3): 101988. DOI: <https://doi.org/10.1016/j.xcrim.2025.101988>
19. Illimoottil M, Ginat D. Recent advances in deep learning and medical imaging for head and neck cancer treatment: MRI, CT, and PET Scans. *Cancers (Basel)* 2023; 15(13): 3267. DOI: <https://doi.org/10.3390/cancers15133267>
20. Mitsuyama Y, Tatekawa H, Takita H, Sasaki F, Tashiro A, Oue S, *et al.* Comparative analysis of GPT-4-based ChatGPT's diagnostic performance with radiologists using real-world radiology reports of brain tumors. *Eur Radiol* 2025; 35(4): 1938-47. DOI: <https://doi.org/10.1007/s00330-024-11032-8>
21. Sood A, Mansoor N, Memmi C, Lynch M, Lynch J. Generative pretrained transformer-4, an artificial intelligence text predictive model, has a high capability for passing novel written radiology exam questions. *Int J Comput Assist Radiol Surg* 2024; 19(4): 645-53. DOI: <https://doi.org/10.1007/s11548-024-03071-9>
22. Blake SR, Das N, Tadeballi M, Reddy B, Singh A, Agrawal R, *et al.* Using artificial intelligence to stratify normal *versus* abnormal chest X-rays: External validation of a deep learning algorithm at East Kent Hospitals University NHS Foundation Trust. *Diagnostics (Basel)* 2023; 13(22): 3408. DOI: <https://doi.org/10.3390/diagnostics13223408>
23. Seah JCY, Tang CHM, Buchlak QD, Holt XG, Wardman JB, Aimoldin A, *et al.* Effect of a comprehensive deep-learning model on the accuracy of chest X-ray interpretation by radiologists: A retrospective, multireader multicase study. *Lancet Digit Health* 2021; 3(8): e496-506. DOI: [https://doi.org/10.1016/S2589-7500\(21\)00106-0](https://doi.org/10.1016/S2589-7500(21)00106-0)
24. Wang Q, Liu Q, Luo G, Liu Z, Huang J, Zhou Y, *et al.* Automated segmentation and diagnosis of pneumothorax on chest X-rays with fully convolutional multi-scale ScSE-DenseNet: A retrospective study. *BMC Med Inform Decis Mak* 2020; 20(Suppl 14): 317. DOI: <https://doi.org/10.1186/s12911-020-01325-5>
25. Li Y, Zhang Z, Dai C, Dong Q, Badrigilan S. Accuracy of deep learning for automated detection of pneumonia using chest X-Ray images: A systematic review and meta-analysis. *Comput Biol Med* 2020; 123: 103898. DOI: <https://doi.org/10.1016/j.compbiomed.2020.103898>
26. Tam MD, Dyer T, Dissez G, Morgan TN, Hughes M, Illes J, *et al.* Augmenting lung cancer diagnosis on chest radiographs: Positioning artificial intelligence to improve radiologist performance. *Clin Radiol* 2021; 76(8): 607-14. DOI: <https://doi.org/10.1016/j.crad.2021.03.021>
27. Niehoff JH, Kalaitzidis J, Kroeger JR, Schoenbeck D, Borggrefe J, Michael AE. Evaluation of the clinical performance of an AI-based application for the automated analysis of chest X-rays. *Sci Rep* 2023; 13(1): 3680. DOI: <https://doi.org/10.1038/s41598-023-30521-2>
28. Ma D, Pang J, Gotway MB, Liang J. A fully open AI foundation model applied to chest radiography. *Nature* 2025; 643: 488-98. DOI: <https://doi.org/10.1038/s41586-025-09079-8>
29. Wang Q, Liu Q, Luo G, Liu Z, Huang J, Zhou Y, *et al.* Automated segmentation and diagnosis of pneumothorax on chest X-rays with fully convolutional multi-scale ScSE-DenseNet: A retrospective study. *BMC Med Inform Decis Mak* 2020; 20(Suppl 14): 317. DOI: <https://doi.org/10.1186/s12911-020-01325-5>
30. Ahn JS, Ebrahimian S, McDermott S, Lee S, Naccarato L, Di Capua JF, *et al.* Association of artificial intelligence-aided chest radiograph interpretation with reader performance and efficiency. *JAMA Netw Open* 2022; 5(8): e2229289. DOI: <https://doi.org/10.1001/jamanetworkopen.2022.29289>
31. Homayounieh F, Digumarthy S, Ebrahimian S, Rueckel J, Hoppe BF, Sabel BO, *et al.* An artificial intelligence-based chest X-ray model on human nodule detection accuracy from a multicenter study. *JAMA Netw Open* 2021; 4(12): e2141096. DOI: <https://doi.org/10.1001/jamanetworkopen.2021.41096>
32. Lind Plesner L, Müller FC, Brejnbøl MW, Lastrup LC, Rasmussen F, Nielsen OW, *et al.* Commercially available chest radiograph ai tools for detecting airspace disease, pneumothorax, and pleural effusion. *Radiology* 2023; 308(3): e231236. DOI: <https://doi.org/10.1148/radiol.231236>
33. Hillis JM, Bizzo BC, Mercaldo S, Chin JK, Newbury-Chaet I, Digumarthy SR, *et al.* Evaluation of an artificial intelligence model for detection of pneumothorax and tension pneumothorax in chest radiographs. *JAMA Netw Open* 2022; 5(12): e2247172. DOI: <https://doi.org/10.1001/jamanetworkopen.2022.47172>
34. Monti CB, Bianchi LMG, Rizzetto F, Carbonaro LA, Vanzulli A. Diagnostic performance of an artificial intelligence model for the detection of pneumothorax at chest X-ray. *Clin Imaging* 2025; 117: 110355. DOI: <https://doi.org/10.1016/j.clinimag.2024.110355>
35. Guo L, Zhou C, Xu J, Huang C, Yu Y, Lu G. Deep learning for chest X-ray diagnosis: Competition between radiologists with or without artificial intelligence assistance. *J Imaging Inform Med* 2024; 37(3): 922-34. DOI: <https://doi.org/10.1007/s10278-024-00990-6>