AI Radiology Assistant: Can Artificial Intelligence Help Radiology Diagnostics?

Abstract

Artificial intelligence (AI) is rapidly transforming radiology by automating image interpretation and workflow tasks. This critical essay examines the current state of AI in radiological diagnostics by analyzing recent scientific studies, focusing on metrics such as sensitivity, specificity, area under the curve (AUC), and comparisons between AI-assisted and human-only diagnosis. The review highlights AI's successes in improving detection rates for abnormalities like pulmonary nodules, pneumothorax, and breast cancer, while also addressing its shortcomings, including limited generalizability, false positives, algorithmic bias, and issues of explainability. Despite promising results—such as increased sensitivity and reduced reading times—challenges persist, particularly regarding real-world applicability and ethical considerations. The essay concludes that while AI holds great promise as a supportive tool for radiologists, caution must be exercised in its implementation. Future research directions emphasize the need for prospective, multi-center studies, improved data diversity, transparent AI development, seamless workflow integration, and continuous performance monitoring. Ultimately, AI should be viewed as an assistant complementing human expertise rather than a replacement, with the potential to significantly improve patient outcomes when used judiciously.

Keywords: artificial intelligence; radiology; diagnostic imaging; deep learning; chest X-ray; CT; performance; mammography; breast imaging; limitations

1. Introduction

Artificial intelligence (AI) technologies have grown in popularity over the years and have started to develop in many fields, such as medical imaging. In radiology, deep learning is poised to revolutionize by automating image analysis and improving clinical decision-making[1][3]. AI algorithms for detection, segmentation, and classification of findings (e.g. tumors, fractures, hemorrhages) have progressed rapidly. These helpful tools are often designed as computer-aided detection/diagnosis (CAD) systems that detects abnormalities so they can be more easily observed and reviewed by the radiologist [3]. To date there is no fully autonomous AI that replaces the human radiologist. Instead, AI is being used as a decision support system where a human accepts or overrides the AI’s suggestions [3]. The first implementations date back to 1980 and were for mammography CAD [2], but since then deep learning has improved image recognition. On large datasets, AI managed to rival or surpass radiologists, showing higher sensitivity or AUC (area under the curve- a measure of how well a model distinguishes between classes) in some experimental tasks [3][5].

However, to integrate AI into radiology requires *large, diverse annotated datasets* for training, which will be difficult to obtain due to GDPR. Many deep learning models act as “black boxes,” so ensuring transparency and explainability is an active concern [2]. The real-world images may be different from the training data, resulting in decreased performance. Ethical and regulatory issues also need to be addressed in order to be able to use AI safely in this field on a large scale [2]. In this critical essay, we review the most recent scientific studies on AI in radiology, focusing on the metrics obtained (diagnostic accuracy, AUC, sensitivity, specificity, etc.) and on comparisons between AI assistant and human radiologist diagnosis. We bring forward both successes (e.g. improved detection and efficiency) and shortcomings (e.g. generalizability, false positives), summarizing the current situation of AI in radiological diagnostics.

2. Relevant Scientific Studies on AI in Radiology

2.1. Chest Radiography and Chest CT

Chest imaging has a very important role in AI research, given the high volume of X-rays and CT scans today. These are used to clearly and quickly detect lung and thoracic diseases. In *Chest radiography*, several studies have shown that artificial intelligence is a good aid to the radiologist to help detect chest abnormalities. In a recent study- "Using AI to Improve Radiologist Performance in Detection of Abnormalities on Chest Radiographs" (Radiology 2023)[6], 12 readers (thoracic radiologists, general radiologists, residents) evaluated 500 X-rays, half of the radiographs without AI and half of the radiographs with AI (ChestView; Gleamer). About 50% of radiographs having abnormal findings and the average age of patients is 54 years ± 19 (261 female, 239 male). It was shown that, for all readers, AI increased sensitivity significantly: sensitivity rose by 26% (from baseline) for pneumothorax detection, 14% for consolidation, and 12% for lung nodules [6]. Specificity also improved modestly (3.9% for pleural effusion, 3.7% for mediastinal and hilar mass , 2.9% for consolidation and 2.1%, for nodule), except a slight 0.2% drop for pneumothorax [6]. Also, reading with the AI assistant reduced the time by 31% on average (81 seconds without AI vs. 56 seconds with AI, $p<0.001$) [6]. We can say, based on this study, that AI can improve both the detection and the efficiency of X-ray interpretation.

In a 2021 systematic review that included articles from EMBASE, PubMed, Cochrane library, SCOPUS, and Web of Science, similar trends were found: in 38 studies of artificial intelligence-assisted radiography/CT, sensitivity increased from 67.8% unaided to 74.6% with AI, specificity from 82.2% to 85.4%, and overall accuracy from 75.4% to 81.7% [3]. Also, the pooled Area Under the ROC Curve (AUC) improved from 0.75 to 0.80. Table 1 shows a comparative sensitivity and specificity before and after the implementation of AI tools, showing the diagnostic performance of human observers when interpreting radiographs/CTs. We can see that they all have positive changes, so these findings imply that when AI is used, radiologists generally perform better than alone.

Table 1-. Sensitivity and specificity for observer tests without and with AI-Based CAD[3]

Author	Without AI-Based CAD		With AI-Based CAD		Change	Statistical Significance between Difference
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)		
			a			
Bai et al. [13]	79	88	88	91	↑	$p < 0.001$
Beyer et al. [19]	56.5	-	61.6	-	↑	$p < 0.001$
de Hoop et al. [20]	56 *	-	56 *	-	↑	-
Dorr et al. [14]	47	79	61	75	↑	$p < 0.007$
Kim et al. [15]	73.9	88.7	82.2	98.1	↑	$p < 0.014$
Koo et al. [21]	92.4	93.1	95.1	97.2	↑	-
Kozuka et al. [22]	68	91.7	85.1	83.3	↑	$p < 0.01$ **
Lee et al. [23]	84	-	88	-	↑	-
Rajpurkar et al. [31]	70	52	73	61	↑	-
Singh et al. [28]	68 *	77.5 *	73 *	74 *	↑	-
Sung et al. [30]	80.1	89.3	88.9	96.6	↑	$p < 0.01$
Yang et al. [17]	89.5	-	94.2	-	↑	$p < 0.05$

In contrast to the above, some evaluations show that artificial intelligence underperforms. For example, Johns *et al*. [8] tested a commercial AI for nodule-detection in chest X-ray on 5,722 radiographs in the UK. Using the radiologist report as reference, the AI had only 54.5% sensitivity and 83.2% specificity. Faced with the final cancer diagnosis, the sensitivity was 60.9% (with 83.3% specificity) [8]. False positives (normal anatomy detected as nodule) were constant, thus risking some unnecessary follow-ups. The authors considered AI to have "considerably underperformed" in this real-world setting [8], probably due to differences between the training and test data (*overfitting*). This highlights an important limitation, which is that an algorithm that excels in one dataset cannot generalize to others without going through careful *validation and support*.

Regarding *chest CT*, the implementation of AI tools for the detection of pulmonary nodules was also studied. In the scientific article "Impact of artificial intelligence assistance on pulmonary nodule detection and localization in chest CT: a comparative study among radiologists of varying experience levels"[7], two commercial artificial intelligence systems were evaluated on 198 chest CT scans (221 nodules) interpreted by radiology residents and experienced radiologists. It was found that AI assistance improved nodule detection by residents: the mean detection rate increasing from 64% to 77% with AI(a significant +13% absolute gain). Senior radiologists, on the other hand, had little change(from 85% to 86%). Improvement was better for residents than for seniors even though both AI tools had similar effects. Detection increased from 80.3% to 85.9% and for the other from 67.1% to 76.9% [7].

2.2. Breast Imaging

Nowadays, breast cancer is one of the few cancers for which screening is effective. Because of this, artificial intelligence has penetrated to the top of this field. Digital mammography and tomosynthesis produce large datasets ideal for deep learning. Recent studies have managed to provide quantitative evidence that AI can improve cancer detection. In the study published in 2025 by Chang et al.[9] a large prospective cohort (24,543 women) was conducted in a Korean screening program. Radiologists using AI-assisted CAD in single-read mammography detected

13.8% more cancers than without AI (cancer detection rate 5.70‰ vs 5.01‰; 140 vs 123 cancers)[9]. Importantly, the recall rates were not significantly different (which means no false alarms). Figure 1 shows that breast radiologists using AI detected 140 cancers 17 more than the 123 cancers detected by BR without AI-CAD. The results of a simulation (exploratory analysis) showed similar trends in diagnostic performance for general radiologists (GR) compared to BR, with an even greater improvement. GR using AI-CAD detected 120 cancers, 25 more than the 95 cancers detected by GR without AI-CAD. AI alone detected 128 cancers. This shows AI-CAD can raise detection of true cancers without unduly increasing recalls.

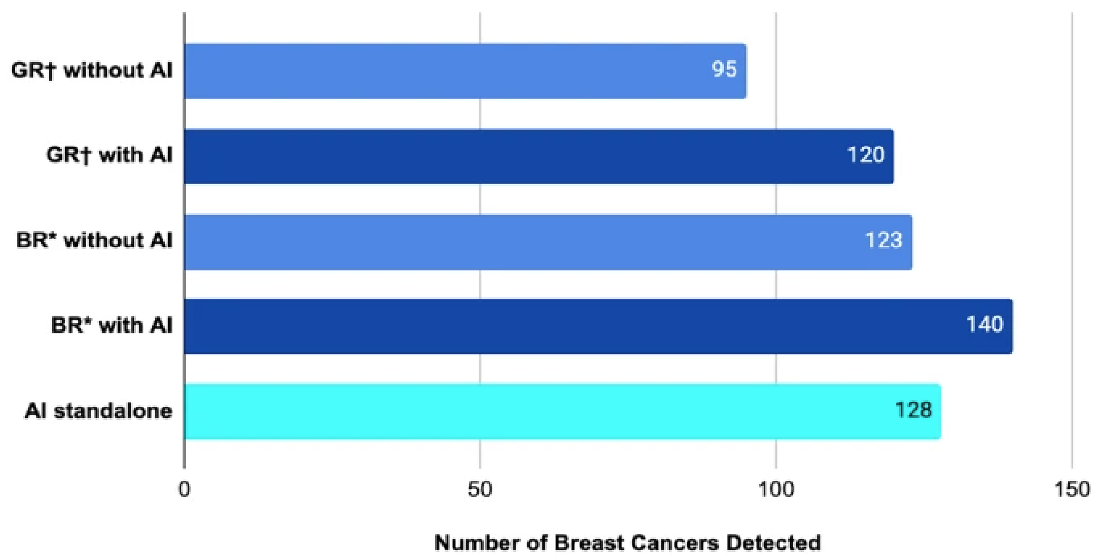

Figure 1-Breast Cancer Detection by Reading Strategy [9]

The research study created by Yoon et al. (Radiology 2023)[5] confirmed the performance of artificial intelligence in breast cancer detection. Across the 16 studies, standalone AI obtained a significantly higher AUC than radiologists in the reader studies: 0.87 for AI vs 0.81 for humans (p=0.002) [5]. AI also tended to have higher sensitivity and lower specificity than radiologists. In the few studies of digital breast tomosynthesis, AI again outperformed human AUC (0.90 vs 0.79, p<0.001) [5]. It was finally concluded that "AI performed as well or better than radiologists" for mammography interpretation. These findings are supported by other reports that deep learning models can identify breast abnormalities much faster and easier.

The main quantitative conclusions for breast imaging seem to be the following:

- **AI-CAD augmenting radiologists:** +13.8% cancer detection rate with AI [9];
- **Standalone AI:** AUC ~0.87 vs 0.81 for radiologists [5];
- **Sensitivity/specificity:** In meta-analysis, AI had higher sensitivity but somewhat lower specificity (leading to more recalls) [5].

So, AI demonstrates once again that it is powerful in detecting many types of abnormalities, which is essential in screening.

3. Other Modalities and General Performance

The AI radiology assistant's help goes beyond chest and breast imaging. For example, in *neuroimaging*, AI algorithms have been shown to have high accuracy. Several models to detect intracranial hemorrhage or aneurysm have sensitivities of ~90-95% and specificities of ~90-96% [10]. The meta-study [10] on aneurysms (DLM for MRI angiography) found a pooled sensitivity of 0.92 and specificity of 0.96. In stroke imaging, artificial intelligence-based software (e.g. detection of large vessel occlusions on CT angiography) has achieved performance comparable to that of neuroradiologists experienced in such tests (sensitivity ~88-90%) [10].

In *cardiovascular imaging*, artificial intelligence is used for calcium scoring and quantification of ventricles. Deep learning tools for calcium assessment in coronary CT have been reported to achieve approximately 95% sensitivity and specificity for identifying high-risk patients. Similarly, in MRI, AI models have equaled cardiology-wide accuracy for estimating variances within clinically acceptable limits [2].

Overall, grouping all studies and modalities presented, Ai algorithms report sensitivities in the range 85-95% and specificities in the range 80-97% [3] [5] [10]. Receiver-operating characteristic (ROC) AUC values are often above 0.85 for well-studied tasks (lung nodule detection, breast cancer screening, intracranial hemorrhage, etc.). However, caution must be exercised as many published results come from retrospective studies with high quality labeled data. A recent article showed that most studies on artificial intelligence have a high risk of bias and limited real-world testing [11]. The impressive in-study metrics cannot always translate directly into clinical practice.

4. Limitations and Challenges

The studies presented above have highlighted the strengths of AI, but they also reveal some limitations and challenges:

- **Generalizability**: As mentioned, artificial intelligence models often underperform on data that differ from their training set. The Johns et al. [8] study on chest radiography showed some drastic changes in sensitivity when moving from selected test sets to routine clinical images. Similarly, a study on artificial intelligence for mammography found that stand-alone artificial intelligence only performs well on high-quality screening images and poorly on low-quality images. This means rigorous external validation (on diverse, real-world populations) is needed before adoption.

- **False positives and specificity:** Many artificial intelligence systems tend to trade sensitivity for specificity, in the sense that higher sensitivity often comes with lower specificity, thus leading to more false alarms. For instance, in screening mammography, AI tended to recall more cases (lower specificity) even as it caught more cancers [5]. In chest X-rays, Johns et al. reported a positive predictive value of only ~5.5% for cancer detection [8], meaning most flagged nodules were false.

- **Inherit biases:** AI models can inherit biases from their training data. If certain patient groups (by age, race, comorbidity) are underrepresented, AI may underperform in those groups and exacerbate health disparities [2]. Algorithmic bias is a known risk: for example, an AI chest X-ray system might be less accurate on images from older machines.
- **Explainability:** Deep learning networks do not usually provide interpretability. Their "black-box" nature makes it difficult for radiologists to trust AI results, especially in critical diagnoses [2].

In sum, although many studies report exceptional results, with high accuracy and high efficiency, these results must be viewed in context. A balanced perspective recognizes that artificial intelligence can enhance the work of radiologists by highlighting abnormalities and reducing labor time [2] [3], but it is not ideal and currently serves best as a second reader rather than a standalone diagnostician.

5. Conclusions and Future Directions for Research

Artificial intelligence has shown that it can help radiologists to improve accuracy and efficiency in many contexts. Previously reviewed studies have shown that AI assistance improves the performance of human doctors: pooled analyses indicate absolute sensitivity gains of 7–26% and AUC increases of 0.05–0.10 across tasks [3] [6]. In thoracic imaging, Ai has been able to detect pulmonary nodules and pneumothoraxes with greater accuracy than humans, and AI-assisted reading can produce even higher detection rates [6]. In breast cancer screening, AI augmentation significantly raised cancer detection rates with minimal harm to specificity [5] [9]. All of the above can translate into *earlier diagnoses and more concrete outcomes for patients*. AI also *shortens reading times* and can *simplify workflow*, solving the shortage of radiologists and the high volume of cases [3] [6].

However, there are some valid issues. In terms of *generalizability*, several studies have shown that real-world AI models perform worse than their counterparts in studies due to differences in patient populations [8]. Specificity is often sacrificed in favor of sensitivity, leading to many false positives, which can negate efficiency gains. Algorithmic bias and data privacy challenges loom large [2]. The current evidence (research studies) has been conducted in silico, meaning few large-scale prospective or randomized studies have been conducted.

Therefore, the adoption of artificial intelligence in radiology *must remain cautious*. People should view AI as an assistant to complement human expertise, not replace it [2] [3]. The strongest evidence to date has suggested that when AI is used judiciously, diagnostic accuracy improves. However, using AI without oversight can lead to errors, which we cannot afford because a human life is at stake.

In conclusion, we can say that AI is extremely promising for radiological diagnostics, demonstrating its ability to increase detection rates and reduce the workload of the human radiologist. With a balanced assessment, we can recognize that these gains are real, but rigorous testing must continue, taking into account the limitations. In this way, AI can become a reliable and safe ally in radiology.

Regarding future work, the following research directions are essential to advance AI in radiology:

- *Conduct prospective, multi-center studies*: conducting prospective, multicenter studies: for AI to thrive across diverse populations, imaging equipment, and clinical environments, thus surpassing single-center retrospective studies
- *Improving data diversity and quality:* creating federated learning networks can allow institutions to collaboratively train models without sharing sensitive data [2]. Federated approaches preserve patient privacy while leveraging larger, more representative datasets – a promising solution to current data limitations.
- *Developing explainable and transparent AI:* research into XAI techniques (e.g., attention maps, decision trees) can help radiologists understand AI suggestions and identify times when AI might be wrong [2]. Regulatory science also needs to keep pace with clear standards (akin to STARD for diagnostics) for reporting AI performance [11].
- *Integration into the radiologists' workflow:* Human factors research can design AI interfaces that truly complement radiologists' work. For instance, choosing how AI alerts are presented, or deciding threshold settings for triage, can significantly affect utility. Studies like Bennani et al.[6] (which measured time savings) should be expanded to other modalities and tasks.
- *Performance monitoring:* as imaging practices and populations change, AI algorithms can drift. Thus, it is important to establish continuous monitoring systems (with periodic recalibration or retraining).

6. References

1. Sharma P., Suehling M., Flohr T., Comaniciu D. *Artificial Intelligence in Diagnostic Imaging: Status Quo, Challenges, and Future Opportunities.* J. Thorac. Imaging. 2020;35:S11–S16. doi: 10.1097/RTI.0000000000000499, [Google Scholar]

2. Abhiyan B. Alexander M. , John R A., *Revolutionizing Radiology With Artificial Intelligence*, 2024 Oct 29;16(10):e72646. doi: 10.7759/cureus.72646, [NIH]

3. Li D., Pehrson L., Carsten A. L., Lea T, Fraccaro M., Desmond E., Hubert D., Sune D., Carlsen J., Bachmann M. , *The Added Effect of Artificial Intelligence on Physicians' Performance in Detecting Thoracic Pathologies on CT and Chest X-ray: A Systematic Review*, 2021 Nov 26;11(12):2206. doi: 10.3390/diagnostics11122206 [MDPI]

4. Aidoc. [(accessed on 28 April 2025)]. Available online: https://www.aidoc.com/

5. Yoon J., Strand F., Baltzer P., Conant E., Gilbert F., Lehman C., Morris E., Mullen L., Nishikwa R., Sharma N., *Standalone AI for Breast Cancer Detection at Screening Digital Mammography and Digital Breast Tomosynthesis: A Systematic Review and Meta-Analysis.* 023 Jun;307(5):e222639. doi: 10.1148/radiol.222639, [NIH]

6. Bennani S, Regnard NE, Ventre J, Lassalle L, Nguyen T, Ducarouge A, Dargent L, Guillo E, Gouhier E, Zaimi SH, Canniff E, Malandrin C, Khafagy P, Koulakian H, Revel MP, Chassagnon G., *Using AI to Improve Radiologist Performance in Detection of Abnormalities on Chest Radiographs*. Radiology. 2023 Dec;309(3):e230860. doi: 10.1148/radiol.230860. Erratum in: Radiology. 2024 Apr;311(1):e249015. doi: 10.1148/radiol.249015. PMID: 38085079, [RSNA]

7. Peters, A.A., Wiescholek, N., Müller, M. et al*., Impact of artificial intelligence assistance on pulmonary nodule detection and localization in chest CT: a comparative study among radiologists of varying experience levels*. Sci Rep 14, 22447 (2024). https://doi.org/10.1038/s41598-024-73435-3

8. Maiter A, Hocking K, Matthews S, Taylor J, Sharkey M, Metherall P, Alabed S, Dwivedi K, Shahin Y, Anderson E, Holt S, Rowbotham C, Kamil MA, Hoggard N, Balasubramanian SP, Swift A, Johns CS., *Evaluating the performance of artificial intelligence software for lung nodule detection on chest radiographs in a retrospective real-world UK population*. BMJ Open. 2023 Nov 8;13(11):e077348. doi: 10.1136/bmjopen-2023-077348. PMID: 37940155; PMCID: PMC10632826., [PMC]

9. Chang, YW., Ryu, J.K., An, J.K. et al. *Artificial intelligence for breast cancer screening in mammography (AI-STREAM): preliminary analysis of a prospective multicenter cohort study*. Nat Commun 16, 2248 (2025). https://doi.org/10.1038/s41467-025-57469-3

10. Gu F, Wu X, Wu W, Wang Z, Yang X, Chen Z, Wang Z, Chen G., *Performance of deep learning in the detection of intracranial aneurysm: A systematic review and meta-analysis*. Eur J Radiol. 2022 Oct;155:110457. doi: 10.1016/j.ejrad.2022.110457. Epub 2022 Jul 29. PMID: 35961157. https://www.ejradiology.com/article/S0720-048X(22)00307-2/abstract

11. Aggarwal, R., Sounderajah, V., Martin, G. et al., *Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis*. npj Digit. Med. 4, 65 (2021). https://doi.org/10.1038/s41746-021-00438-z