UNDERSTANDING EXPECTATIONS, REALITIES AND CHALLENGES OF ARTIFICIAL INTELLIGENCE IN RADIOLOGY

1. Introduction: The Past Expectations

Radiology, a cornerstone of modern medicine, relies on imaging for diagnosis and treatment. However, the field faces a shortage of radiologists, raising concerns as healthcare demands increase [27, 33]. Artificial Intelligence (AI), particularly Convolutional Deep Neural Networks (DNNs), has proven highly effective in image classification, suggesting its transformative potential for radiology [17].

Pioneering work by Wang and Summers [42] underscored AI's promise in radiology, highlighting its role in image segmentation, computer-assisted diagnosis (CAD), and Natural Language Processing (NLP) for standardised report generation [19]. These applications were anticipated to improve speed, accuracy, and cost-effectiveness significantly.

Initially, CAD was seen as a breakthrough in image processing for personalised medicine [19], but its limitations were soon evident [18]. DNN-based AI, however, holds potential where traditional CAD falters, offering powerful support for radiologists [45, 25]. The advantage of AI in identifying patterns and transfer knowledge to unseen data could automate processes, enhance accuracy and efficiency, and boost productivity [13].

Surveys interestingly reflect radiologists' constant optimistic views and needs. Initial surveys indicated radiologists' unfamiliar but positive view towards AI to enhance diagnostic accuracy and reduce risks [43]. Later, radiologists remained optimistic and added workflow enhancement as a need [5]. Surprisingly, few concerned about AI replacing radiologists, but more about potential damage to radiologists' reputations from unregulated AI. More recently, these positive views persisted, with additional expectations around workflow prioritisation, disease grading, visualisation, and auto-generated reports [1]. Concerns now include the need for interpretable AI and transparency.

2. The Current Landscape: Progress and Limitations

Amid high expectations, AI has been applied to many areas in radiology. A DNN trained on 4.8 million head CT scans achieved performance comparable to human radiologists [26]. Another model reduced reaction time of emergency radiographs from minutes to seconds [40], and a DNN for fracture detection cut half error rates among generalists [21], when radiologists unavailable in emergencies.

AI's role in chest radiograph analysis has also been extensively explored [13], with models detecting anomalies on par with radiologists [38]. In more granular disorders, AI surpasses human performance in detecting malignant pulmonary nodules with X-rays, recommended as a second reader [28].

Progress in ontological radiology analysis includes a model achieving a 0.944 ROC (Receiver-Operating Characteristic) AUC (Area Under Curve) in 3D lung cancer screening, surpassing human experts [2]. Moreover, a multi-view residual network effectively predicts breast cancer [45]. Combining AI with CAD and medical expertise has proved effective [8]. In NLP, AI efficiently analyses radiology reports, though less integrated due to a lack of standard frameworks [30].

Despite advancements, challenges persist. Wang and Summers highlight the lack of large-scale annotated data [42], with distribution shifts yet impairing model performance. Patients also often find AI decisions less acceptable than those from doctors.

Translating AI models into clinical practice often reveals suboptimal performance compared to initial claims [25]. Concerns over consensus level of annotated data persist [1], compounding long-standing challenges [14].

Moreover, radiologists do not solely focus on specific diseases like pneumonia during daily tasks, contrasting with the targeted training of AI models, which perform even poorer in some areas [41]. Models frequently lack generalisability outside their training data [47].

Efforts have been made to address these challenges. Few-shot domain adaptation and fine-tuning have been employed to mitigate data scarcity [7, 37], while attempts to counter data acquisition shifts have also been initiated [16]. Unannotated data has been leveraged to build classification and localisation models, demonstrating improved performance with increased data [20]. An extensive international study illustrated generalisability of radiology models across countries [24].

With the initiatives mentioned in place, why does AI adoption in radiology remain so challenging?

3. Major Challenges in AI Radiology

First, many AI models struggle to integrate into clinical workflows because accuracy is a necessary but not sufficient criteria [25, 41]. Many studies function as proof of concept rather than rigorous testing under clinical conditions [4]. Moreover, the metrics used for model evaluation are not always effective. For example, ROC is susceptible to class imbalance despite anomalies are minority class in real world [34].

In order to smoothen the adoption of AI-based tools in radiology context, more efforts on peer reviewed assessment should be conducted. This includes the need for more systematic testing, more reproducible research, and realistic clinical settings [31]. Also, the choice of evaluation metrics should be more carefully considered, such as precision-recall curve [34] that is class-imbalance resistant (reflecting real performance). Finally, more divulgation and exposure of AI models to radiologists and patients are needed, providing more evidence and robust validation to remove scepticism.

Second, interpretability and trustworthiness remain issues even with optimal models. The black-box nature of neural networks has long been a concern [14]. This is a prevalent need among radiologists, given AI's rapid advancement [1]. Furthermore, the EU General Data Protection Regulation (GDPR) emphasises the need of transparency and challenges in interpretability [31].

On interpretability, recent methodologies such as LIME [32], SHAP [23], and Grad-CAM [35], provide degrees of explanation for model decision ground. They provide certain degree of reasoning behind the predictions. Additionally, one could argue that currently, many medical diagnoses and prescriptions are given without full comprehension of underlying mechanisms [31]. Towards future, researchers must be transparent about false positives and negatives rates for patient and radiologists interests [13]. Lastly, the notion of trustworthiness must be reframed: the challenge lies not in blindly trusting AI but in evaluating its appropriateness. Trusting an untrustworthy AI is harmful, but so is distrusting a trustworthy one [4].

The final challenge, notwithstanding the above solutions, is the system's vulnerability to adversarial attacks. Adversarial attacks pose a well-known challenge in AI, where minor perturbations in input data cause misclassification [10]. Certain powerful AI models exploit this paradigm [9]. In radiology and healthcare, adversarial examples can be generated [44], severely reducing model accuracy — for instance, mammographic classifiers' accuracy can drop to below 35% with simple adversarial techniques [46]. Importantly, adversarial attacks can occur in black-box environments, without knowledge of model architecture or parameters, in radiology context [3].

There is no one-size-fits-all solution to this arms race, akin to cybersecurity challenges. However, raising awareness about adversarial attacks is crucial, and studies have begun examining attack and defence strategies for medical imaging systems [15]. Given radiology's sensitivity, attacks could be particularly dangerous; thus, better understanding and protocol development are imperative [39]. Additionally, training more robust models with adversarial examples [22] and enhancing system security [6] can also help.

4. Fairness and Ethics

Fairness and ethical considerations are paramount in AI radiology. Models may display biased predictions throughout their development stages, from problem selection, training data, outcomes definition, to algorithmic development. For example, some proxies for healthcare measures, initially accepted, were later determined inadequate [29]. When models lack fairness, they may exhibit varying accuracy across populations. Some studies have shown less accurate chest X-ray classifiers for underserved populations [36]. This is compounded by model generalisability issues, where accuracy may differ across hospitals [47].

In addition, inequalities in radiology diagnosis quality and AI tool accessibility may be intensified by socioeconomic disparities [31]. Efforts must focus on collecting representative data spanning the population's cross-section. It is vital to consider whether population effects are desirable or not. This is especially relevant in radiology, where populations can display distinct image features.

Additionally, constraints can be imposed on models to ensure fairness, such as demographic parity, equalised odds, or equal opportunity [11]. These can be incorporated through regularisation and post-processing. Careful strategy selection is necessary. Alternatively, models could be created to excel in different populations, or ensemble models employed for combined predictions.

From an ethical standpoint, AI usage in radiology raises significant concerns due to the sensitive nature of radiological images. These images are deeply personal and private, and use of AI possibly facilitates localisation of patients [31]. Unapproved data usage by large companies further complicates this [12]. Is it ethical to employ patient data for model training? Should individual data be sacrificed for the greater good? These are pressing ethical questions requiring further exploration [13]. Additionally, improved privacy and security management should be prioritised too [31].

The most pressing concern remains the lack of clear guidelines for ethical AI implementation, everyone has a word knowing 'what' is ethical but not knowing 'how' to be ethical [4]. Furthermore, AI responsibility in automation mistakes is vague. Thus, AI will likely never fully replace radiologists but may serve as supervised automation [31]. AI should not only aim for accuracy but also adhere to ethical standards, ultimately, improving treatment and patient lives [13].

5. Conclusion: Towards a More Inclusive Future

In conclusion, despite the hype and expectations surrounding AI, it is unlikely to replace radiologists entirely. The tremendous costs associated with algorithmic errors and other challenges impedes full automation. But as suggested above, AI could be complementary to human experts, heralding a future where radiologists are empowered with tools synthesised from millions of samples and expert insights.

References

[1] Anjali Agrawal, Garvit D. Khatri, Bharti Khurana, Aaron D. Sodickson, Yuanyuan Liang, and David Dreizin. A survey of ASER members on artificial intelligence in emergency radiology: trends, perceptions, and expectations. *Emergency Radiology*, 30(3):267–277, June 2023.

[2] Diego Ardila, Atilla P. Kiraly, Sujeeth Bharadwaj, Bokyung Choi, Joshua J. Reicher, Lily Peng, Daniel Tse, Mozziyar Etemadi, Wenxing Ye, Greg Corrado, David P. Naidich, and Shravya Shetty. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine*, 25(6):954–961, June 2019. Publisher: Nature Publishing Group.

[3] Gerda Bortsova, Cristina González-Gonzalo, Suzanne C. Wetstein, Florian Dubost, Ioannis Katramados, Laurens Hogeweg, Bart Liefers, Bram van Ginneken, Josien P. W. Pluim, Mitko Veta, Clara I. Sánchez, and Marleen de Bruijne. Adversarial attack vulnerability of medical image analysis systems: Unexplored factors. *Medical Image Analysis*, 73:102141, October 2021.

[4] Valerie K. Bürger, Julia Amann, Cathrine K. T. Bui, Jana Fehr, and Vince I. Madai. The unmet promise of trustworthy AI in healthcare: why we fail at clinical translation. *Frontiers in Digital Health*, 6, April 2024. Publisher: Frontiers.

[5] Francesca Coppola, Lorenzo Faggioni, Daniele Regge, Andrea Giovagnoni, Rita Golfieri, Corrado Bibbolino, Vittorio Miele, Emanuele Neri, and Roberto Grassi. Artificial intelligence: radiologists' expectations and opinions gleaned from a nationwide online survey. *La radiologia medica*, 126(1):63–71, January 2021.

[6] Benoit Desjardins, Yisroel Mirsky, Markel Picado Ortiz, Zeev Glozman, Lawrence Tarbox, Robert Horn, and Steven C. Horii. DICOM Images Have Been Hacked! Now What? *American Journal of Roentgenology*, 214(4):727–735, April 2020. Publisher: American Roentgen Ray Society.

[7] Mohsen Ghafoorian, Alireza Mehrtash, Tina Kapur, Nico Karssemeijer, Elena Marchiori, Mehran Pesteie, Charles R. G. Guttmann, Frank-Erik de Leeuw, Clare M. Tempany, Bram van Ginneken, Andriy Fedorov,

Purang Abolmaesumi, Bram Platel, and William M. Wells. Transfer Learning for Domain Adaptation in MRI: Application in Brain Lesion Segmentation. In Maxime Descoteaux, Lena Maier-Hein, Alfred Franz, Pierre Jannin, D. Louis Collins, and Simon Duchesne, editors, *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*, pages 516–524, Cham, 2017. Springer International Publishing.

[8] Jeremy Goecks, Vahid Jalili, Laura M. Heiser, and Joe W. Gray. How Machine Learning Will Transform Biomedicine. *Cell*, 181(1):92–101, April 2020.

[9] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[10] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and Harnessing Adversarial Examples, March 2015. arXiv:1412.6572 [stat].

[11] Moritz Hardt, Eric Price, Eric Price, and Nati Srebro. Equality of Opportunity in Supervised Learning. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

[12] Hal Hodson. Revealed: Google AI has access to huge haul of NHS patient data, April 2016.

[13] Sabeena Jalal, Savvas Nicolaou, and William Parker. Artificial Intelligence, Radiology, and the Way Forward. *Canadian Association of Radiologists Journal*, 70(1):10–12, February 2019. Publisher: SAGE Publications Inc.

[14] C E Kahn. Artificial intelligence in radiology: decision support systems. *RadioGraphics*, 14(4):849–861, July 1994.

[15] Sara Kaviani, Ki Jin Han, and Insoo Sohn. Adversarial attacks and defenses on AI in medical imaging informatics: A survey. *Expert Systems with Applications*, 198:116815, July 2022.

[16] Wouter M. Kouw and Marco Loog. An introduction to domain adaptation and transfer learning, January 2019. arXiv:1812.11806 [cs].

[17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[18] Constance D. Lehman, Robert D. Wellman, Diana S. M. Buist, Karla Kerlikowske, Anna N. A. Tosteson, Diana L. Miglioretti, and for the Breast Cancer Surveillance Consortium. Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. *JAMA Internal Medicine*, 175(11):1828–1837, November 2015.

[19] King C. Li, Peter Marcovici, Andrew Phelps, Christopher Potter, Allison Tillack, Jennifer Tomich, and Srini Tridandapani. Digitization of Medicine: How Radiology Can Take Advantage of the Digital Revolution. *Academic Radiology*, 20(12):1479–1494, December 2013.

[20] Zhe Li, Chong Wang, Mei Han, Yuan Xue, Wei Wei, Li-Jia Li, and Li Fei-Fei. Thoracic Disease Identification and Localization With Limited Supervision. pages 8290–8299, 2018.

[21] Robert Lindsey, Aaron Daluiski, Sumit Chopra, Alexander Lachapelle, Michael Mozer, Serge Sicular, Douglas Hanel, Michael Gardner, Anurag Gupta, Robert Hotchkiss, and Hollis Potter. Deep neural network improves fracture detection by clinicians. *Proceedings of the National Academy of Sciences*, 115(45):11591–11596, November 2018. Publisher: Proceedings of the National Academy of Sciences.

[22] Siqi Liu, Arnaud Arindra Adiyoso Setio, Florin C. Ghesu, Eli Gibson, Sasa Grbic, Bogdan Georgescu, and Dorin Comaniciu. No Surprises: Training Robust Lung Nodule Detection for Low-Dose CT Scans by Augmenting With Adversarial Attacks. *IEEE Transactions on Medical Imaging*, 40(1):335–345, January 2021. Conference Name: IEEE Transactions on Medical Imaging.

[23] Scott Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions, November 2017. arXiv:1705.07874 [cs].

[24] Scott Mayer McKinney, Marcin Sieniek, Varun Godbole, Jonathan Godwin, Natasha Antropova, Hutan Ashrafian, Trevor Back, Mary Chesus, Greg S. Corrado, Ara Darzi, Mozziyar Etemadi, Florencia Garcia-Vicente, Fiona J. Gilbert, Mark Halling-Brown, Demis Hassabis, Sunny Jansen, Alan Karthikesalingam, Christopher J. Kelly, Dominic King, Joseph R. Ledsam, David Melnick, Hormuz Mostofi, Lily Peng, Joshua Jay Reicher, Bernardino Romera-Paredes, Richard Sidebottom, Mustafa Suleyman, Daniel Tse, Kenneth C. Young, Jeffrey De Fauw, and Shravya Shetty. International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788):89–94, January 2020. Publisher: Nature Publishing Group.

[25] Frederick J. A. Meijer. Managing expectations and challenges of AI in radiology. *European Radiology*, 34(11):7347–7348, November 2024.

[26] Jameson Merkow, Robert Lufkin, Kim Nguyen, Stefano Soatto, Zhuowen Tu, and Andrea Vedaldi. DeepRadiologyNet: Radiologist Level Pathology Detection in CT Head Images, December 2017. arXiv:1711.09313 [cs].

[27] Yasuo Nakajima, Kei Yamada, Keiko Imamura, and Kazuko Kobayashi. Radiologist supply and workload: international comparison. *Radiation Medicine*, 26(8):455–465, October 2008.

[28] Ju Gang Nam, Sunggyun Park, Eui Jin Hwang, Jong Hyuk Lee, Kwang-Nam Jin, Kun Young Lim, Thienkai Huy Vu, Jae Ho Sohn, Sangheum Hwang, Jin Mo Goo, and Chang Min Park. Development and Validation of Deep Learning–based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. *Radiology*, 290(1):218–228, January 2019. Publisher: Radiological Society of North America.

[29] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, October 2019. Publisher: American Association for the Advancement of Science.

[30] Ewoud Pons, Loes M. M. Braun, M. G. Myriam Hunink, and Jan A. Kors. Natural Language Processing in Radiology: A Systematic Review. *Radiology*, 279(2):329–343, May 2016. Publisher: Radiological Society of North America.

[31] Erich Prem. From ethical AI frameworks to tools: a review of approaches. *AI and Ethics*, 3(3):699–716, August 2023.

[32] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1135–1144, New York, NY, USA, 2016. Association for Computing Machinery.

[33] Abi Rimmer. Radiologist shortage leaves patient care at risk, warns royal college. *BMJ*, page j4683, October 2017.

[34] Takaya Saito and Marc Rehmsmeier. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*, 10(3):e0118432, March 2015.

[35] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *International Journal of Computer Vision*, 128(2):336–359, February 2020. arXiv:1610.02391 [cs].

[36] Laleh Seyyed-Kalantari, Haoran Zhang, Matthew B. A. McDermott, Irene Y. Chen, and Marzyeh Ghassemi. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature Medicine*, 27(12):2176–2182, December 2021. Publisher: Nature Publishing Group.

[37] Hoo-Chang Shin, Holger R. Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M. Summers. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298, May 2016. Conference Name: IEEE Transactions on Medical Imaging.

[38] Ramandeep Singh, Mannudeep K. Kalra, Chayanin Nitiwarangkul, John A. Patti, Fatemeh Homayounieh, Atul Padole, Pooja Rao, Preetham Putha, Victorine V. Muse, Amita Sharma, and Subba R. Digumarthy. Deep learning in chest radiography: Detection of findings and presence of change. *PLOS ONE*, 13(10):e0204155, October 2018. Publisher: Public Library of Science.

[39] Vera Sorin, Shelly Soffer, Benjamin S. Glicksberg, Yiftach Barash, Eli Konen, and Eyal Klang. Adversarial attacks in radiology – A systematic review. *European Journal of Radiology*, 167:111085, October 2023.

[40] Joseph J. Titano, Marcus Badgeley, Javin Schefflein, Margaret Pain, Andres Su, Michael Cai, Nathaniel Swinburne, John Zech, Jun Kim, Joshua Bederson, J. Mocco, Burton Drayer, Joseph Lehar, Samuel Cho, Anthony Costa, and Eric K. Oermann. Automated deep-neural-network surveillance of cranial images for acute neurologic events. *Nature Medicine*, 24(9):1337–1341, September 2018. Publisher: Nature Publishing Group.

[41] Eric J. Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1):44–56, January 2019. Publisher: Nature Publishing Group.

[42] Shijun Wang and Ronald M. Summers. Machine learning and radiology. *Medical Image Analysis*, 16(5):933–951, July 2012.

[43] Q. Waymel, S. Badr, X. Demondion, A. Cotten, and T. Jacques. Impact of the rise of artificial intelligence in radiology: What do radiologists think? *Diagnostic and Interventional Imaging*, 100(6):327–336, June 2019.

[44] Jelmer M. Wolterink, Anirban Mukhopadhyay, Tim Leiner, Thomas J. Vogl, Andreas M. Bucher, and Ivana Išgum. Generative Adversarial Networks: A Primer for Radiologists. *RadioGraphics*, 41(3):840–857, May 2021. Publisher: Radiological Society of North America.

[45] Nan Wu, Jason Phang, Jungkyu Park, Yiqiu Shen, Zhe Huang, Masha Zorin, Stanislaw Jastrzebski, Thibault Fevry, Joe Katsnelson, Eric Kim, Stacey Wolfson, Ujas Parikh, Sushma Gaddam, Leng Leng Young Lin, Kara Ho, Joshua D. Weinstein, Beatriu Reig, Yiming Gao, Hildegard Toth, Kristine Pysarenko, Alana Lewin, Jiyon Lee, Krystal Airola, Eralda Mema, Stephanie Chung, Esther Hwang, Naziya Samreen, S. Gene Kim, Laura Heacock, Linda Moy, Kyunghyun Cho, and Krzysztof J. Geras. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. *IEEE Transactions on Medical Imaging*, 39(4):1184–1194, April 2020. Conference Name: IEEE Transactions on Medical Imaging.

[46] Ibrahim Yilmaz. Practical Fast Gradient Sign Attack against Mammographic Image Classifier, January 2020. arXiv:2001.09610 [cs].

[47] John R. Zech, Marcus A. Badgeley, Manway Liu, Anthony B. Costa, Joseph J. Titano, and Eric Karl Oermann. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLOS Medicine*, 15(11):e1002683, November 2018. Publisher: Public Library of Science.