

# The Rebirth of CAD: How Is Modern AI Different from the CAD We Know?

Luke Oakden-Rayner, MBBS, FRANZCR

From the Department of Radiology, Royal Adelaide Hospital, North Terrace, Adelaide, SA, Australia 5000; School of Public Health, University of Adelaide, Adelaide, Australia; and Australian Institute for Machine Learning, Adelaide, Australia. Received December 14, 2018; revision requested January 14, 2019; revision received April 24; accepted May 2. Address correspondence to the author (e-mail: [luke.oakden-rayner@adelaide.edu.au](mailto:luke.oakden-rayner@adelaide.edu.au)).

Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2019; 1(3):e180089 • <https://doi.org/10.1148/ryai.2019180089> • Content code: 

Computer-aided diagnosis is a phrase that inspires strong opinions among radiologists, and many of those opinions are negative.

The term *computer-aided detection* (CAD) arose during the 1980s and 1990s, during the second era of artificial intelligence (AI). Like the first era in the 1950s and 1960s, and the third era through which we are living currently, the second era was built around new and exciting technologies paired with the promise that computers would soon solve all of our problems. The technology of the second era fell far short of these lofty expectations and led to a period of extreme disillusionment, often described as an “AI winter.” Given this history, it is no wonder that many consider CAD to be a disappointment.

Radiologists have more reason than most to be disappointed, because CAD in medical imaging was more than an unrealized promise. Almost uniquely across the world of technology, medical or otherwise, the hype and optimism around second-era AI led to the widespread utilization of CAD in clinical imaging. This use was most obvious in screening mammography, where it has been estimated that by 2010 more than 74% of mammograms in the United States were read with CAD assistance (1).

Unfortunately, CAD’s benefit has been questionable. Several large trials came to the conclusion that CAD has at best delivered no benefit (2) and at worst has actually reduced radiologist accuracy (3), resulting in higher recall and biopsy rates (4).

Not only were the medical outcomes disappointing, but they came with an estimated 20% increase in the time needed to interpret each study, a result of the radiologist needing to dismiss the many false alarms that these systems produced (5). It was not unusual for a radiologist to be asked to second-guess a half dozen spurious CAD-detected lesions per study, which research has shown can bias the radiologist’s interpretation (6), for example making them less likely to detect a real cancer when the CAD system does not flag any potential lesions.

Given this experience of increased costs without improved performance, and of poor user experiences and unfulfilled promises, why are radiologists now expressing renewed interest in CAD?

The answer is that a new technology has been developed that vastly outperforms the methods of historical CAD. This technology is known as deep learning, which has rapidly spread across the technology sector and appears

entirely capable of fulfilling the promises that second-era CAD could not.

This leads us to the first difference between deep learning and traditional computer vision systems: at least outside of medicine, deep learning actually works. As an example, in the well-known ImageNet image analysis competition, the best traditional CAD methods produced five times as many errors as a practiced human when asked to identify everyday objects in photographs, such as bicycles, dogs, and airplanes (7). In the space of a few short years, deep learning has surpassed human performance in this challenge and now makes around half as many errors as humans (8).

Similarly, before deep learning, autonomous vehicles could not get out of the parking lot. Now hundreds of cars are being tested on real roads all over the world, with over 10 million miles driven (9). The same technology is already part of your life, being used to categorize your photographs, understand your speech, and suggest sensible responses to your e-mails.

Although these applications are unrelated to medical imaging, they represent a massive change in the perceptual abilities of our computer systems. Computer vision in the era of traditional CAD was unable to perform visual tasks that a toddler would find trivial, but modern AI is succeeding in tasks that have previously been the domain of human experts. The difference in capability is difficult to overstate.

Although we do not yet have the level of evidence required to prove that deep learning can do the work of human doctors, early results show systems that appear to perform at a human level in common medical tasks such as retinal assessment (10) and skin lesion analysis (11). We have seen similar results in radiology; for example, Chilamkurthy et al used a large dataset to produce a system that can detect a variety of critical findings on CT head scans (12), which may be useful for the triage of reporting worklists.

Further testing is required, particularly given our previous experiences with traditional CAD, but the success of deep learning in so many “human” domains is unprecedented and a degree of optimism is justified.

This ability of deep learning to succeed across a range of perceptual tasks also distinguishes the technology from traditional CAD. Second-era AI was task-specific; each system had to be crafted specifically for each task. Deep

learning is task-agnostic; it simply learns from the data it is given, whether that data are clinical photographs, radiographs, or pathologic slides. The ability is incredibly powerful, because deep learning systems often can perform well on multiple similar tasks and can be fine-tuned on new tasks with much less effort than with traditional CAD (13). This ability to generalize across tasks may mean that the way we think about CAD will need to change.

We have historically thought of CAD as a solution to a variety of independent tasks and have even defined specific subgroups of CAD based on the intended use of the systems. CADx (computer-aided diagnosis) and CADE (computer-aided detection) are the most commonly used terms, but the list of CAD variants also includes CADq (computer-aided quantification) and CAST (computer-aided simple triage), among others. These subgroups have even been used to define regulatory frameworks (14), which made sense when the previous technology could only do the single task for which it was designed.

Deep learning systems discover patterns that are useful beyond the task they are designed to perform. Just as a human can only diagnose a disease if he or she can identify the image features that inform the diagnosis, a diagnostic deep learning system also contains information that can help to localize and explain its decisions. Similarly, a triage system that identifies cases with possible cerebral artery occlusion for urgent review (15) is inherently performing a diagnostic task; it is recognizing some of the same features a neuroradiologist would look for to diagnose a stroke. With this flexibility in mind, we need to consider if the boundaries we previously drew between CADx, CADE, and the other CAD variants are still relevant in the third era of AI. Although these categories are undoubtedly useful descriptions, they have been used historically to define risk. For example, the U.S. Food and Drug Administration (FDA) has treated CADx as higher risk than CADE (14).

In the current third era of AI, where computer systems approach human-level capability and are able to perform human tasks, it may be far more important to define risk based on the level of human supervision required over an algorithm. A diagnostic algorithm that is backed up by radiologist review may engender lower risk than a system that only quantifies the size of a lung nodule, if a clinician could make a treatment decision based on that measurement with no human ever looking at the images themselves. While this idea of autonomous AI might seem too futuristic to some readers, we should remember that the FDA already has approved an autonomous system to review retinal photographs for signs of diabetic eye disease (16). It is promising that the Therapeutic Goods Administration in Australia has proposed new regulations that take the degree of autonomy into account when determining the risk of AI systems (17), building on earlier work from the International Medical Device Regulators Forum (18).

These developments highlight the most important way that modern AI differs from traditional CAD; we have learned from the past. At a basic level, CAD failed because there was a difference between performance in controlled multireader experiments

and performance in clinical practice. There is a range of likely causes for this failure, but whatever the reason, we now have decades of hard-won experience that we can use to inform AI design, testing, validation, policy, and regulation. It is heartening to see that these conversations are beginning to take place, with many groups in leadership roles, including RSNA, identifying safety and evidence-based AI use as key priorities (19,20).

It would be easy to point to the negative historical experience of CAD in radiology and dismiss the current claims around modern AI, but this isn't the CAD we have known. Outside of medicine, where the techniques that underpinned CAD failed, deep learning succeeds. Where CAD methods were narrow and brittle, deep learning can exploit patterns that are more meaningful and more broadly useful. We should consider this new CAD through a lens that makes the most of our historical experiences as well as our new knowledge. We should hold these systems to higher standards than we have in the past, avoiding the pitfalls we now know how to recognize, while at the same time recognizing the technological advances that finally might allow computers to fulfill their clinical potential. Only by doing so will we be able to plan for the coming changes in our profession, to predict where this technology will succeed and where it may still fail, and most importantly, to protect ourselves and our patients from harm.

**Disclosures of Conflicts of Interest:** L.O. disclosed no relevant relationships.

## References

- Rao VM, Levin DC, Parker L, Cavanaugh B, Frangos AJ, Sunshine JH. How widely is computer-aided detection used in screening and diagnostic mammography? *J Am Coll Radiol* 2010;7(10):802–805.
- Lehman CD, Wellman RD, Buist DS, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med* 2015;175(11):1828–1837.
- Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* 2007;356(14):1399–1409.
- Gilbert FJ, Astley SM, Gillan MG, et al. Single reading with computer-aided detection for screening mammography. *N Engl J Med* 2008;359(16):1675–1684.
- Tchou PM, Haygood TM, Atkinson EN, et al. Interpretation time of computer-aided detection at screening mammography. *Radiology* 2010;257(1):40–46.
- Philpotts LE. Can computer-aided detection be detrimental to mammographic interpretation? *Radiology* 2009;253(1):17–22.
- Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis* 2015;115(3):211–252.
- Li FF, Deng J. ImageNet: Where have we been? Where are we going? <https://www.youtube.com/watch?v=jYvBmJo7qjc>. Published 2017. Accessed April 2, 2019.
- Krafcik J. Where the next 10 million miles will take us. <https://medium.com/waymo/where-the-next-10-million-miles-will-take-us-de51bebb67d3>. Published 2018. Accessed April 2, 2019.
- Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;316(22):2402–2410.
- Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542(7639):115–118 [Published correction appears in *Nature* 2017;546(7660):686.].
- Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet* 2018;392(10162):2388–2396.
- Donahue J, Jia Y, Vinyals O, et al. Decaf: a deep convolutional activation feature for generic visual recognition. *International Conference on Machine Learning*, 2014; 647–655.
- U.S. Food and Drug Administration. Computer-assisted detection devices applied to radiology images and radiology device data—premarket notification

- [510 (k)] submissions—guidance for industry and Food and Drug Administration staff. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/computer-assisted-detection-devices-applied-radiology-images-and-radiology-device-data-premarket>. Published 2012. Accessed April 2019.
15. U.S. Food and Drug Administration. FDA permits marketing of clinical decision support software for alerting providers of a potential stroke in patients. <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-clinical-decision-support-software-alerting-providers-potential-stroke>. Published 2018. Accessed April 2019.
  16. U.S. Food and Drug Administration. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. <https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm>. Published 2018. Updated April 11, 2019. Accessed April 2019.
  17. Australian Therapeutic Goods Administration. Consultation: regulation of software, including Software as a Medical Device (SaMD). <https://www.tga.gov.au/node/873400>. Published 2019. Accessed April 2019.
  18. IMDRF Software as a Medical Device (SaMD) Working Group. “Software as a Medical Device”: Possible Framework for Risk Categorization and Corresponding Considerations. International Medical Device Regulators Forum. <http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf>. Published 2014. Accessed April 2019.
  19. Ethics of AI in Radiology: European and North American Multisociety Statement, 2019. [https://www.acrdsi.org/-/media/DSI/Files/PDFs/Ethics-of-AI-in-Radiology-Statement\\_RFC.pdf](https://www.acrdsi.org/-/media/DSI/Files/PDFs/Ethics-of-AI-in-Radiology-Statement_RFC.pdf). Accessed April 20, 2019.
  20. Chokshi FH, Flanders AE, Prevedello LM, Langlotz CP. Fostering a healthy AI ecosystem for radiology: conclusions of the 2018 RSNA Summit on AI in radiology. *Radiol Artif Intell* 2019;1(2):e190021.