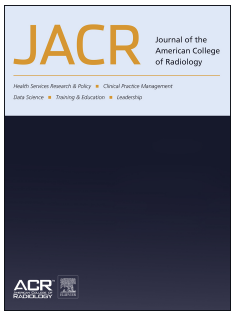
Journal Pre-proof



The Potential Dangers of AI for Radiology and Radiologists

Linda C. Chu, MD, Anima Anandkumar, PhD, Hoo Chang Shin, PhD, Elliot K. Fishman, MD

PII: S1546-1440(20)30403-8

DOI: https://doi.org/10.1016/j.jacr.2020.04.010

Reference: JACR 5172

To appear in: *Journal of the American College of Radiology*

Received Date: 31 March 2020

Revised Date: 9 April 2020

Accepted Date: 10 April 2020

Please cite this article as: Chu LC, Anandkumar A, Chang Shin H, Fishman EK, The Potential Dangers of AI for Radiology and Radiologists, *Journal of the American College of Radiology* (2020), doi: https://doi.org/10.1016/j.jacr.2020.04.010.

Title: The Potential Dangers of AI for Radiology and Radiologists

Authors: Linda C. Chu MD[1], Anima Anandkumar PhD[2,3], Hoo Chang Shin PhD[3], and Elliot K. Fishman MD[1]

Affiliations:

[1]The Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, 600 North Wolfe Street, Baltimore, MD, 21287

[2]Department of Computing and Mathematical Science, California Institute of Technology, 1200 E California Blvd, Pasadena, CA, 91125

[3]NVIDIA Corporation, 2788 San Tomas Expy, Santa Clara, CA, 95051

Corresponding Author: Linda C. Chu, MD, Hal B168, Johns Hopkins Hospital, 600 North Wolfe Street, Baltimore, MD, 21287; phone: 443-831-4342; email: lindachu@jhmi.edu

Each author substantially contributed to the conception of the work, writing and revision of the manuscript. Each author has approved the final version of the manuscript.

Conflict of interest: The authors declare no conflict of interest related to materials discussed in this article.

Statement of data access and integrity: The authors declare that they had full access to all of the data in this study and the authors take complete responsibility for the integrity of the data and accuracy of the data analysis.

The Potential Dangers of AI for Radiology and Radiologists

With the advent of artificial intelligence (AI) across many fields and subspecialties, there are considerable expectations for transformative impact. However, there are also concerns regarding the potential abuse of AI. Many scientists have been worried about the dangers of AI leading to "biased" conclusions in part due to the enthusiasm of the inventor or over-enthusiasm by the general public. Here, though, we are considering some scenarios in which people may intend to cause potential errors within data sets of analyzed information, resulting in incorrect conclusions and leading to potential problems with patient care and outcomes.

A generative adversarial network (GAN) is a recently developed deep-learning model aimed at creating new images. It simultaneously trains a generator and a discriminator network, which serves to generate artificial images and to discriminate real vs. artificial images, respectively. We have recently described how GANs can produce artificial images of people and audio content that fool the recipient into believing that they are authentic. As applied to medical imaging, GANs can generate synthetic images that can alter lesion size, location, and transpose abnormalities onto normal exams (Fig. 1) [1]. GANs have the potential to improve image quality, reduce radiation dose, augment data for training algorithms, and perform automated image segmentation [2]. However, there is also the potential for harm if these artificial images infiltrate our healthcare system by hackers with malicious intent. As proof of principle, Mirksy et al. showed that they were able to tamper with CT scans and artificially inject or remove lung cancers on the images. When the radiologists were blinded to the attack, this hack had a 99.2% success rate for cancer injection and a 95.8% success rate for cancer removal. Even when the

radiologists were warned about the attack, the success of cancer injection decreased to 70%, but the cancer removal success rate remained high at 90% [3]. This illustrates the sophistication and realistic appearance of such artificial images. These hacks can be targeted against specific patients or can be used as a more general attack on our radiology data. It is already challenging enough to keep up with the daily clinical volume when the radiology system is running smoothly. Our clinical workflow would be paralyzed if we cannot trust the authenticity of the images and must spend extra effort to search for evidence of image tampering on every case.

There are multiple access points within the chain of image acquisition and delivery that can be corrupted by attackers, including the scanner, picture archiving and communication system (PACS), server, and workstations [3]. Unfortunately, data security is poorly developed and poorly standardized in radiology. In 2016, Stites et al. performed a scan through the World Wide Web of networked computers and devices and showed that there were 2774 unprotected radiology or digital imaging and communications in medicine (DICOM) servers worldwide, most of them located in the United States [4]. To date, there has been no known hack into the radiology system, aside from the research study demonstrating its feasibility [3]. However, the vulnerability is clearly present, and may be exploited by hackers.

Such threats could affect not only radiology departments but also entire health systems. We have all read articles about security breaches of medical records. There have been almost 3000 breaches (involving more than 500 medical records) in the United States within the past 10 years. This includes high-profile cases such as the 2015 breach of the Anthem medical insurance company that potentially exposed the medical records of 78 million Americans and led to a $115

million settlement [5]. Hospitals and clinics have been held hostage as their data were corrupted by a third party who demanded payment (ransom) to release the data [5]. In 2017, ransomware WannaCry and NotPetya spread through thousands of institutions worldwide, including many hospitals, and caused $18 billion dollars in damages [5]. Hospitals and clinics have not been the only targets. The city of Baltimore was essentially out of business for a month this past year due to such a ransomware attack. At first glance, all of these situations seem more likely in a movie made for Netflix or HBO. However, the truth is that we must be prepared to deal with such scenarios in the near future. As electronic health records and hospital data become more centralized and more computerized the dangers only multiply.

However, there are several ways to mitigate potential AI-based hacks and attacks. These include clear security guidelines and protocols that are uniform across the globe. As deep-fake technology gets more sophisticated, there is emerging research on AI-driven defense strategies. One example features the training of an AI to detect artificial images by image artifacts induced by GAN [6]. However, AI-driven defense mechanisms have a long way to catch up, as seen in the related problem of defense against adversarial attacks. Recognizing these challenges, the Defense Advanced Research Projects Agency (DARPA) has launched the Media Forensics (MediFor) program to research against deep fakes [7]. Hence, for now, the best defense against deep fakes is based on traditional cybersecurity best practices: secure all stages in the pipeline, and enable strong encryption and monitoring tools.

In the current Coronavirus Disease-19 (COVID-19) pandemic, many clinicians and radiologists have turned to working remotely in attempts to “flatten the curve” and slow the spread of

disease. In the body imaging division at our institution, currently approximately half of the radiologists are working remotely. Many of our clinicians are transitioning to telemedicine visits, which adds tremendous stress on our networks. Our Informational Technology (IT) department has been proactive in setting up a dedicated Virtual Private Network (VPN) for radiology to ensure that there is sufficient bandwidth for our clinical work. On our few onsite rotations, we practice "social distancing" and we have suspended our all side-by-side readouts and in-person lectures. We have turned to Zoom and other mobile platforms for managing our rapidly changing clinical operations, educating trainees, or simply staying in touch during these uncertain times. The daily meeting participants rose from 10 million daily users in December 2019 to 200 million daily users in March 2020 [8]. Our reliance on Zoom and other mobile platforms has exposed a new vulnerability. There has been proliferation of "Zoombombing", in which intruders hijack video calls and past hate speech and offensive images. Furthermore, additional vulnerabilities in Zoom can allow hackers to gain control of the users' microphone, webcam, and steal login credentials. The Zoom video meetings did not provide end-to-end encryption as promised, and a large number of Zoom video meeting recordings, many of which contain private information, are left unprotected and viewable on the web. The Federal Bureau of Investigation (FBI) has issued security warnings about Zoom, and a number of organizations including SpaceX, Google, New York's Department of Education, and the US Senate have banned or discouraged the use of Zoom [8]. The meteoric rise and fall of Zoom is a cautionary tale about the importance of data security.

With the development of AI and all its potential wonders in terms of increasing the accuracy of our diagnostic capabilities and potentially improving patient care, we must also be concerned

about the potential dark side by bad actors. The sooner organized radiology and organized medicine address these issues with clarity the more stable and protected the healthcare system and our patients will be from those intent on creating harm and havoc by abusing AI. The acceleration of data sharing during the current pandemic exposes critical vulnerabilities in data security. It reminds us of the pervasive threat that bad actors can and will exploit any technology for their selfish gains. Doing nothing is not a viable strategy but acting in a concerted effort will lead us to the protection we need and is important as we push AI development over the next several years.

Acknowledgments

The authors thank senior science editor Edmund Weisberg, MS, MBE, for his editorial assistance.

References

1. Shin, H.C., et al., *Medical Image Synthesis for Data Augmentation and Anonymization using Generative Adversarial Networks*, in *Simulation and Synthesis in Medical Imaging*, A. Gooya, et al., Editors. 2018: Grenada, Spain.
2. Sorin, V., et al., *Creating Artificial Images for Radiology Applications Using Generative Adversarial Networks (GANs) - A Systematic Review*. Acad Radiol, 2020.
3. Mirsky, Y., et al., *CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning*, in *28th USENIX Security Symposium*. 2019: Santa Clara, CA. p. 461-478.

4. Stites, M. and O.S. Pianykh, *How Secure Is Your Radiology Department? Mapping Digital Radiology Adoption and Security Worldwide.* AJR Am J Roentgenol, 2016. **206**(4): p. 797-804.

5. Desjardins, B., et al., *DICOM Images Have Been Hacked! Now What?* AJR Am J Roentgenol, 2019: p. 1-9.

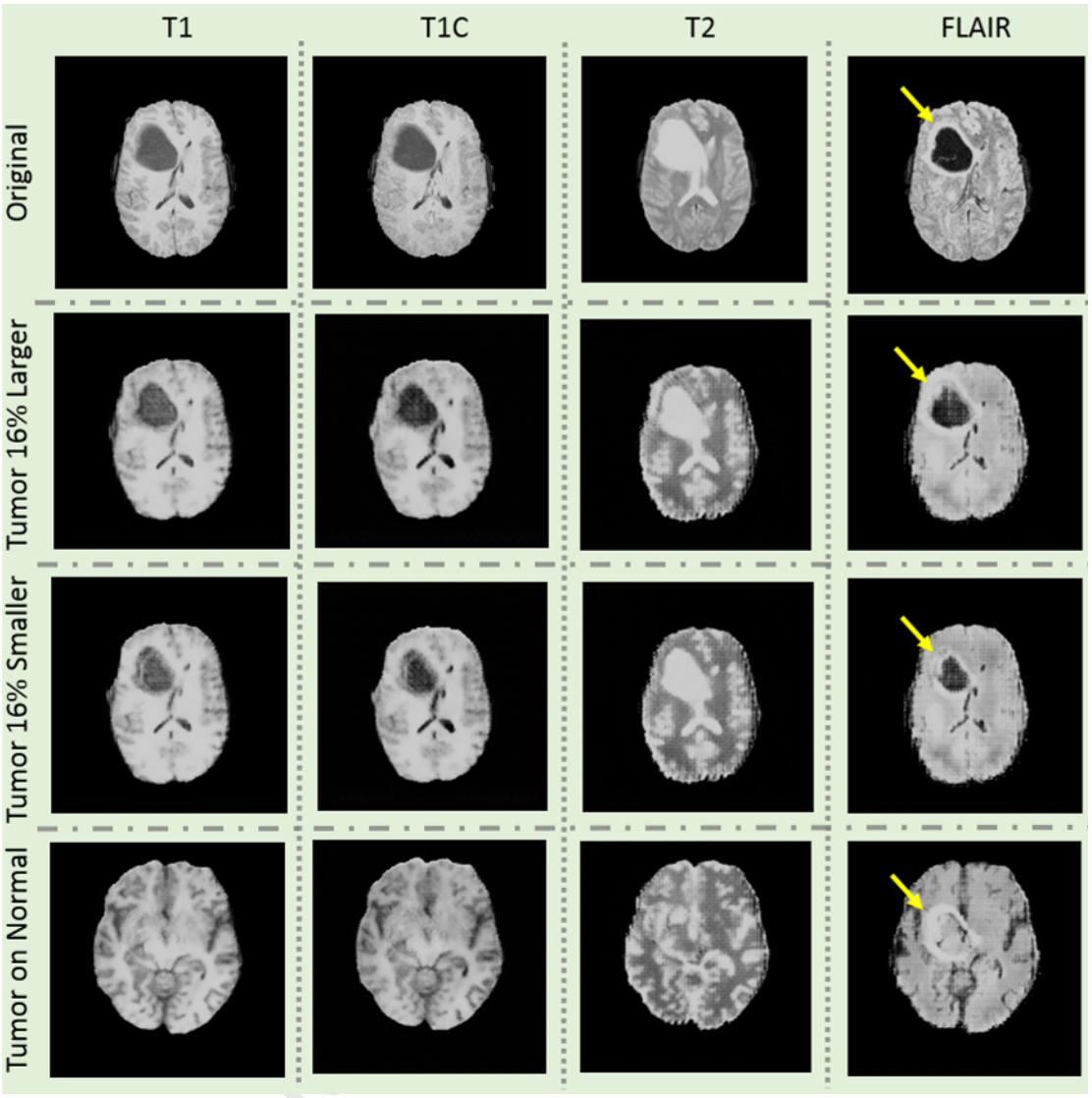
6. Zhang, X., S. Karaman, and S.F. Chang, *Detecting and Simulating Artifacts in GAN Fake Images*, in *The IEEE International Workshop on Information Forensics and Security*. 2019: Delft, The Netherlands.

7. Turek, M. *Media Forensics (MediFor)*. 2020 4/9/2020]; Available from: https://www.darpa.mil/program/media-forensics.

8. Hodge, R. *Zoom: Every security issue uncovered in the video chat app*. 2020 [cited 2020 4/9/2020]; Available from: https://www.cnet.com/news/zoom-every-security-issue-uncovered-in-the-video-chat-app/.

Figure Legend

Fig 1. Examples of images artificially generated using generative adversarial network (GAN) of brain tumor MRI images. First column: T1-weighted images; second column: T1-weighted images with contrast; third column: T2-weighted images; fourth column: FLAIR images. First row: Original images with tumor in the right frontal lobe (arrows). Second row: Tumor is made 16% larger. Third row: Tumor is made 16% smaller. Fourth row: Tumor is artificially placed on an otherwise tumor-free brain.


T1
T1C
T2
FLAIR
Original
Tumor 16% Larger
Tumor 16% Smaller
Tumor on Normal