Report on Patient Privacy Volume 19, Number 11. November 07, 2019 Addressing 'Deep Fake' Scans Is Critical Amid Tech Advances

By Jane Anderson

"Deep fake" radiology scans - with altered results falsely showing either fake cancerous nodes or a clear scan where the patient actually has cancer - are poised to spread, Israeli researchers warn, and health care entities need to take steps to guard against attacks, along with teaching providers to recognize signs that a scan has been altered.

Rogue states or terrorist groups could use these attacks to instill fear or attempt to alter political dynamics in other countries, but they also could be used in a more pedestrian way: for insurance fraud, researchers say in a study posted online at Cornell University's publications website.

Author Yisroel Mirsky tells *RPP* that malicious tampering of radiology scans is both an immediate threat and one that will escalate over time. "This attack may be rather immediate if a state actor is involved. However, the tampering of medical images, in my opinion, will likely become a more evident issue in the next two-to-four years as the technology becomes more accessible."

Insurance fraud would be possible if the potential attacker receives a copy of his or her own scan, Mirsky says, noting that this type of threat could grow more prevalent, largely "because the attack is much simpler - there is no need to develop a malware or infect the hospital's network."

Hospitals and other health care entities can guard against this type of attack, Mirsky says. "However, doing so requires changes in settings, configurations and software. It is hard to say exactly how much or how long it would take, because every hospital is different in terms of their implemented security policies, physical security, bureaucracy, etc."

Worse, he says, "some hospitals out there have very outdated systems, and they cannot simply enable common security features. These hospitals will likely be exposed to the threat for some time."

Deep Learning Creates Fakes

In the study, "CT-GAN: Malicious Tampering of 3D Medical Imagery Using Deep Learning",[1] Mirsky and his colleagues showed how an attacker can use deep learning to add or remove evidence of medical conditions from volumetric (3D) medical scans.

Motivations for such an attack appear to be right out of a crime drama or political thriller. "An attacker may perform this act in order to stop a political candidate, sabotage research, commit insurance fraud, perform an act of terrorism, or even commit murder," the authors wrote.

In some ways, this is a more dangerous tactic than the more common phishing and ransomware hacks, the authors said: "An attacker with access to medical records can do much more than hold the data for ransom or sell it on the black market."

Although the idea of altering radiology scans is relatively new, Mirsky says he and his co-authors were not concerned about giving malicious actors ideas, even when they spelled out details of how they accomplished

their own hack in their paper. "I'm sure that there are malicious actors out there who have thought of the possibility, and there is a chance that some have even implemented it in one way or another," he says. "The concept of tampering media with deep learning has been around for a while and gained much publicity with regards to deep fakes in 2017."

Radiologists need to be educated about the possibility of such an attack, he says. "If a radiologist notices some artifact - noise, distortion, irregularity - it should be investigated. For example, recent scans taken from the same machine can be reviewed. It is also important to note that this attack affects AI [artificial intelligence] tools, which assist radiologists in detecting medical conditions in scans. Therefore, a radiologist should trust his/her own instincts and not rely solely on the tool itself."

Implement Defensive Tactics

Hospitals can take several approaches to prevent and defend against these types of attacks, Mirsky says. The research group showed in a video accompanying its study what it termed an "easy" attack vector involving physical intrusion and the installation of a man-in-the-middle device, but there are multiple avenues, Mirsky says.

"For example, the radiologist's terminal can be infected, especially if he or she uses a personal computer. Another example is propagation through the network to an attack point. This may occur due to vulnerabilities in internet-facing services such as web-based PACS [picture archiving and communication system] and others in the hospital," he says.

To address these problems, Mirsky recommends using digital signatures in the digital imaging and communications in medicine (DICOM) file format. "A digital signature can be used that [indicates] nobody has altered the content since its creation," he says. "Admins should enable this feature on all modality devices which support it so that they sign every DICOM scan with a proper digital signature."

This is already part of the DICOM standard, and many companies' devices support it, Mirsky says. However, the viewing devices also must verify the signatures. "Software providers should ensure that they can verify signatures—otherwise, a digital signature can't help," he says.

After a facility's information technology department has implemented the use of digital signatures, it should enable "proper end-to-end encryption" on all communications across the hospital's network, Mirsky says.

It's also possible to use digital signatures to catch an attack as it happens, he adds. "If a digital signature does not match that of the scanner's, then this is an immediate indication that the scan's pixel data or metadata has been altered."

If that occurs, then a facility's response would depend on how far the attack has gotten, and whether it targets one patient or all patients - "ransomware or terrorism," Mirsky says. "In the case of an individual, if there is proper logging enabled in the network, then one may be able to trace it back to the source after the attack has been detected. However, in the case of a wide attack, there is little that can be done after the fact if all scans in the PACS were altered."

The best solution in the case of a systemwide attack is to perform regular offline and offsite backups of recent scans so that data can be recovered in the event of such an attack, he says.

Contact Mirsky at yisroel@post.bgu.ac.il.

1 Yisroel Mirsky et al, "CT-GAN: Malicious Tampering of 3D Medical Imagery Using Deep Learning", USENIX Security Symposium, revised June 6, 2019, http://bit.ly/2WzJzNH.