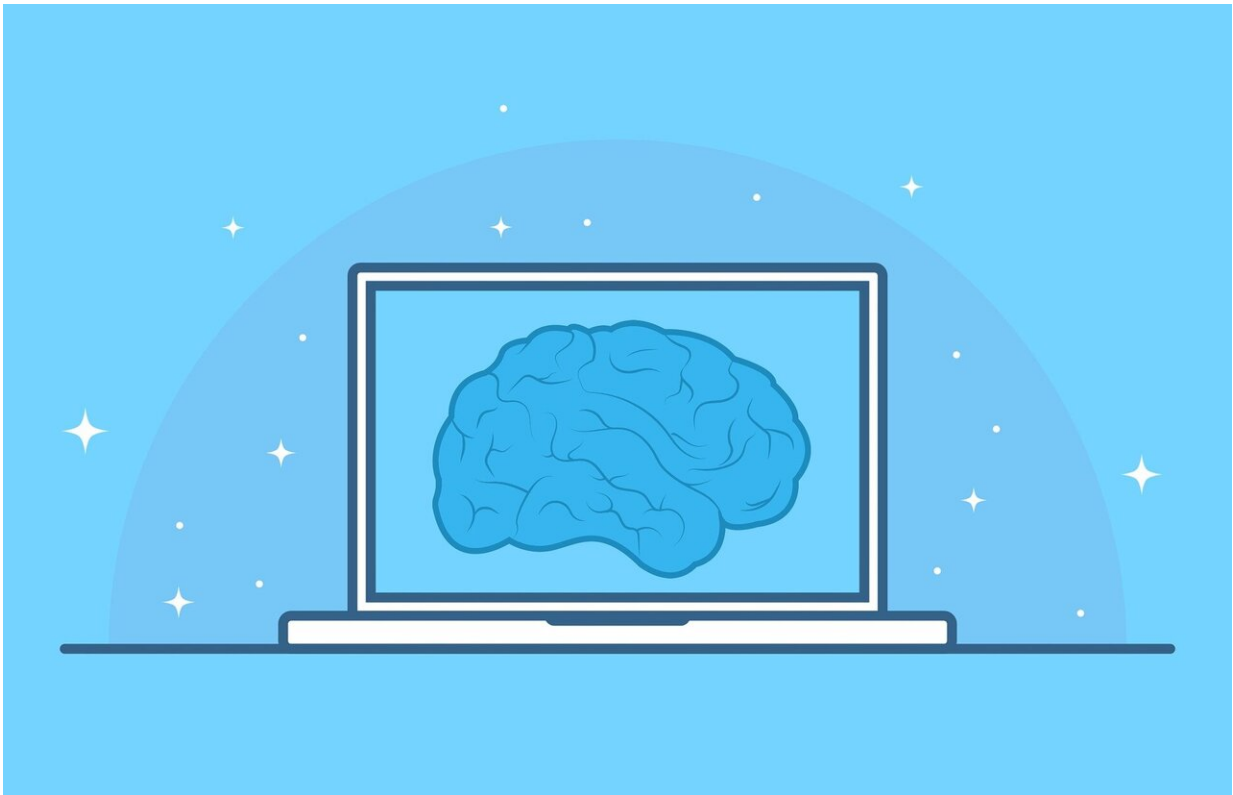# AI techniques in medical imaging may lead to incorrect diagnoses

May 12 2020

Credit: CC0 Public Domain

Machine learning and AI are highly unstable in medical image reconstruction, and may lead to false positives and false negatives, a new study suggests.

A team of researchers, led by the University of Cambridge and Simon Fraser University, designed a series of tests for medical image reconstruction algorithms based on AI and deep learning, and found that these techniques result in myriad artefacts, or unwanted alterations in the data, among other major errors in the final images. The effects were typically not present in non-AI based imaging techniques.

The phenomenon was widespread across different types of artificial neural networks, suggesting that the problem will not be easily remedied. The researchers caution that relying on AI-based image reconstruction techniques to make diagnoses and determine treatment could ultimately do harm to patients. Their results are reported in the *Proceedings of the National Academy of Sciences*.

"There's been a lot of enthusiasm about AI in medical imaging, and it may well have the potential to revolutionise modern medicine: however, there are potential pitfalls that must not be ignored," said Dr. Anders Hansen from Cambridge's Department of Applied Mathematics and Theoretical Physics, who led the research with Dr. Ben Adcock from Simon Fraser University. "We've found that AI techniques are highly unstable in medical imaging, so that small changes in the input may result in big changes in the output."

A typical MRI scan can take anywhere between 15 minutes and two hours, depending on the size of the area being scanned and the number of images being taken. The longer the patient spends inside the machine, the higher resolution the final image will be. However, limiting the amount of time patients spend inside the machine is desired, both to reduce the risk to individual patients and to increase the overall number of scans that can be performed.

Using AI techniques to improve the quality of images from MRI scans or other types of medical imaging is an attractive possibility for solving

the problem of getting the highest quality image in the smallest amount of time: in theory, AI could take a low-resolution image and make it into a high-resolution version. AI algorithms 'learn' to reconstruct images based on training from previous data, and through this training procedure aim to optimise the quality of the reconstruction. This represents a radical change compared to classical reconstruction techniques that are solely based on mathematical theory without dependency on previous data. In particular, classical techniques do not learn.

Any AI algorithm needs two things to be reliable: accuracy and stability. An AI will usually classify an image of a cat as a cat, but tiny, almost invisible changes in the image might cause the algorithm to instead classify the cat as a truck or a table, for instance. In this example of image classification, the one thing that can go wrong is that the image is incorrectly classified. However, when it comes to image reconstruction, such as that used in medical imaging, there are several things that can go wrong. For example, details like a tumour may get lost or may falsely be added. Details can be obscured and unwanted artefacts may occur in the image.

"When it comes to critical decisions around human health, we can't afford to have algorithms making mistakes," said Hansen. "We found that the tiniest corruption, such as may be caused by a patient moving, can give a very different result if you're using AI and deep learning to reconstruct medical images—meaning that these algorithms lack the stability they need."

Hansen and his colleagues from Norway, Portugal, Canada and the UK designed a series of tests to find the flaws in AI-based medical imaging systems, including MRI, CT and NMR. They considered three crucial issues: instabilities associated with tiny perturbations, or movements; instabilities with respect to small structural changes, such as a brain

image with or without a small tumour; and instabilities with respect to changes in the number of samples.

They found that certain tiny movements led to myriad artefacts in the final images, details were blurred or completely removed, and that the quality of image reconstruction would deteriorate with repeated subsampling. These errors were widespread across the different types of neural networks.

According to the researchers, the most worrying errors are the ones that radiologists might interpret as medical issues, as opposed to those that can easily be dismissed due to a technical error.

"We developed the test to verify our thesis that deep learning techniques would be universally unstable in medical imaging," said Hansen. "The reasoning for our prediction was that there is a limit to how good a reconstruction can be given restricted scan time. In some sense, modern AI techniques break this barrier, and as a result become unstable. We've shown mathematically that there is a price to pay for these instabilities, or to put it simply: there is still no such thing as a free lunch."

The researchers are now focusing on providing the fundamental limits to what can be done with AI techniques. Only when these limits are known will we be able to understand which problems can be solved. "Trial and error-based research would never discover that the alchemists could not make gold: we are in a similar situation with modern AI," said Hansen. "These techniques will never discover their own limitations. Such limitations can only be shown mathematically."