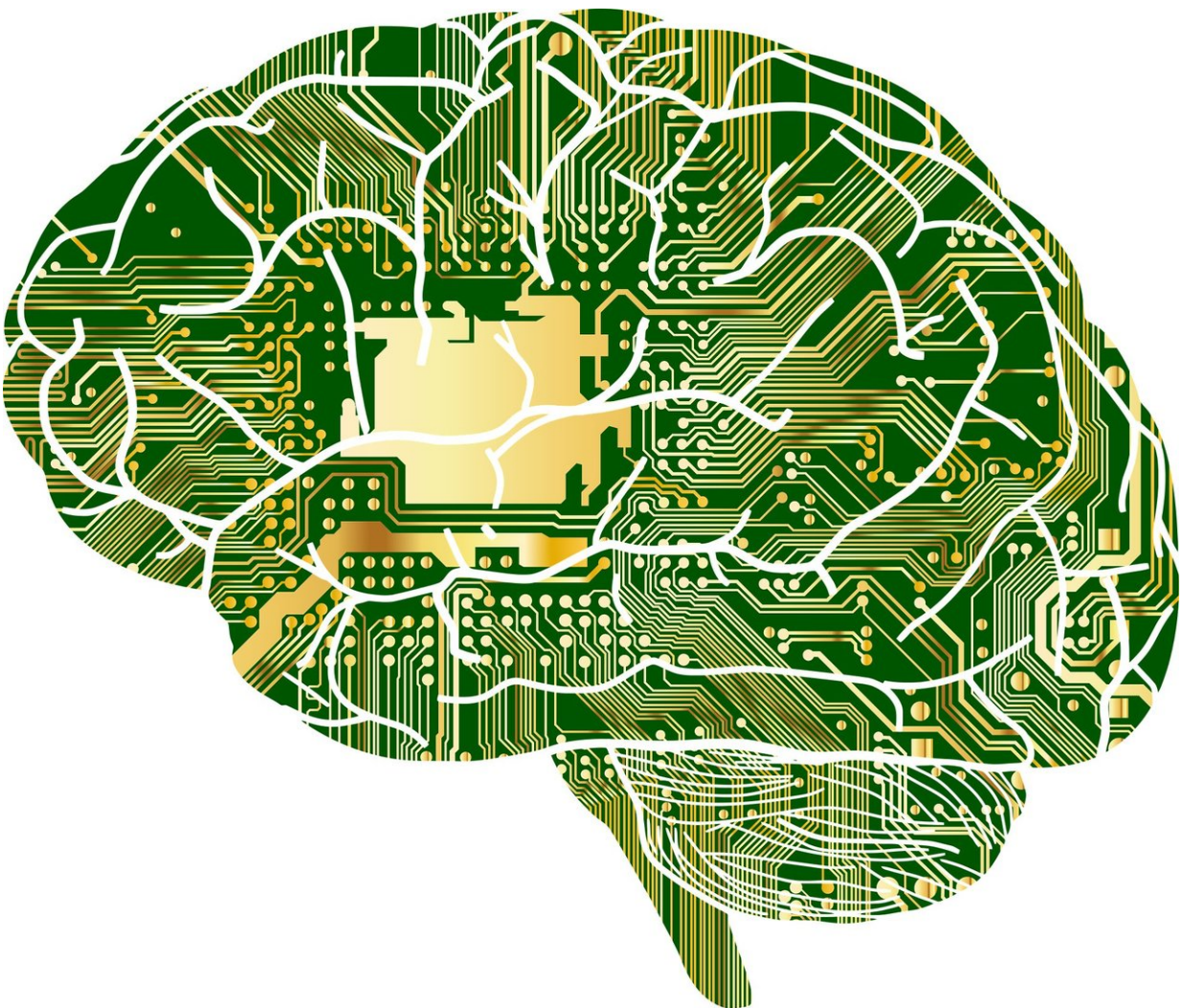


# First systematic review and meta-analysis suggests AI may be as effective as health professionals at diagnosing disease

September 25 2019

---



Credit: CC0 Public Domain

Artificial intelligence (AI) appears to detect diseases from medical imaging with similar levels of accuracy as health-care professionals, according to the first systematic review and meta-analysis, synthesising all the available evidence from the scientific literature published in *The Lancet Digital Health* journal.

Nevertheless, only a few studies were of sufficient quality to be included in the analysis, and the authors caution that the true diagnostic power of the AI technique known as deep learning—the use of algorithms, [big data](#), and computing power to emulate human learning and intelligence—remains uncertain because of the lack of studies that directly compare the performance of humans and machines, or that validate AI's performance in real clinical environments.

"We reviewed over 20,500 articles, but less than 1% of these were sufficiently robust in their design and reporting that independent reviewers had high confidence in their claims. What's more, only 25 studies validated the AI models externally (using medical images from a different population), and just 14 studies actually compared the performance of AI and [health professionals](#) using the same test sample," explains Professor Alastair Denniston from University Hospitals Birmingham NHS Foundation Trust, UK, who led the research.

"Within those handful of high-quality studies, we found that deep learning could indeed detect diseases ranging from cancers to eye diseases as accurately as health professionals. But it's important to note that AI did not substantially out-perform human diagnosis."

With deep learning, computers can examine thousands of medical images to identify patterns of disease. This offers enormous potential for improving the accuracy and speed of diagnosis. Reports of deep learning models outperforming humans in diagnostic testing has generated much excitement and debate, and more than 30 AI algorithms for healthcare

have already been approved by the US Food and Drug Administration.

Despite strong public interest and market forces driving the rapid development of these technologies, concerns have been raised about whether study designs are biased in favour of machine learning, and the degree to which the findings are applicable to real-world clinical practice.

To provide more evidence, researchers conducted a systematic review and meta-analysis of all studies comparing the performance of deep learning models and health professionals in detecting diseases from medical imaging published between January 2012 and June 2019. They also evaluated study design, reporting, and clinical value.

In total, 82 articles were included in the systematic review. Data were analysed for 69 articles which contained enough data to calculate test performance accurately. Pooled estimates from 25 articles that validated the results in an independent subset of images were included in the meta-analysis.

Analysis of data from 14 studies comparing the performance of deep learning with humans in the same sample found that at best, deep learning algorithms can correctly detect disease in 87% of cases, compared to 86% achieved by health-care professionals.

The ability to accurately exclude patients who don't have disease was also similar for [deep learning](#) algorithms (93% specificity) compared to health-care professionals (91%).

Importantly, the authors note several limitations in the methodology and reporting of AI-diagnostic studies included in the analysis. Deep learning was frequently assessed in isolation in a way that does not reflect clinical practice. For example, only four studies provided health professionals

with additional clinical information that they would normally use to make a diagnosis in clinical practice. Additionally, few prospective studies were done in real clinical environments, and the authors say that to determine diagnostic accuracy requires high-quality comparisons in patients, not just datasets. Poor reporting was also common, with most studies not reporting missing data, which limits the conclusions that can be drawn.

"There is an inherent tension between the desire to use new, potentially life-saving diagnostics and the imperative to develop high-quality evidence in a way that can benefit patients and health systems in clinical practice," says Dr. Xiaoxuan Liu from the University of Birmingham, UK. "A key lesson from our work is that in AI—as with any other part of healthcare—good study design matters. Without it, you can easily introduce bias which skews your results. These biases can lead to exaggerated claims of good performance for AI tools which do not translate into the real world. Good design and reporting of these studies is a key part of ensuring that the AI interventions that come through to patients are safe and effective."

"Evidence on how AI algorithms will change patient outcomes needs to come from comparisons with alternative diagnostic tests in randomised controlled trials," adds Dr. Livia Faes from Moorfields Eye Hospital, London. "So far, there are hardly any such trials where diagnostic decisions made by an AI algorithm are acted upon to see what then happens to outcomes which really matter to patients, like timely treatment, time to discharge from hospital, or even survival rates."

Writing in a linked Comment, Dr. Tessa Cook from the University of Pennsylvania, USA, discusses whether AI can be effectively compared to the human physician working in the real world, where data are "messy, elusive, and imperfect". She writes: "Perhaps the better conclusion is that, the narrow public body of work comparing AI to

human physicians, AI is no worse than humans, but the data are sparse and it may be too soon to tell."

**More information:** *The Lancet Digital Health*,  
[www.thelancet.com/journals/lan ... \(19\)30123-2/fulltext](http://www.thelancet.com/journals/lan... (19)30123-2/fulltext)

Provided by Lancet

Citation: First systematic review and meta-analysis suggests AI may be as effective as health professionals at diagnosing disease (2019, September 25) retrieved 23 April 2024 from <https://medicalxpress.com/news/2019-09-systematic-meta-analysis-ai-effective-health.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.