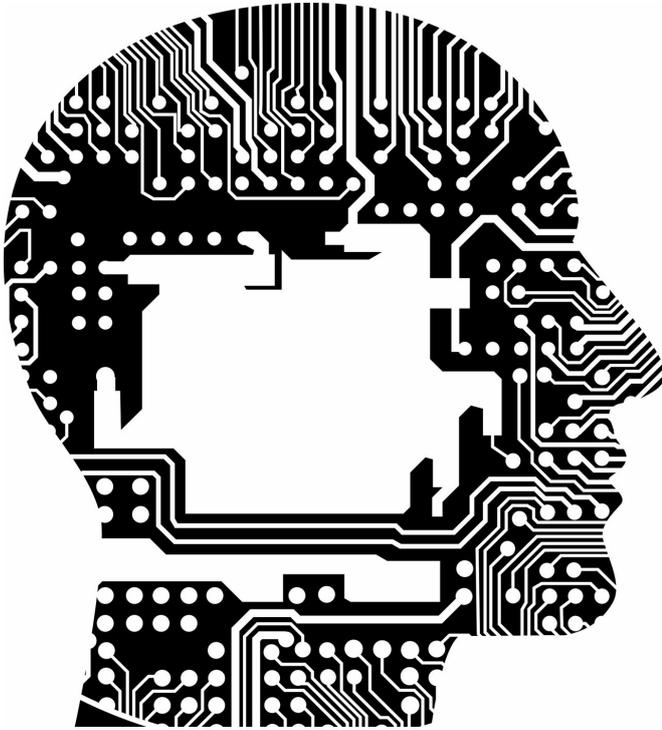# Concerns over 'exaggerated' study claims of AI outperforming doctors

25 March 2020

Credit: CC0 Public Domain

Many studies claiming that artificial intelligence is as good as (or better than) human experts at interpreting medical images are of poor quality and are arguably exaggerated, posing a risk for the safety of 'millions of patients' warn researchers in *The BMJ* today.

Their findings raise concerns about the quality of evidence underpinning many of these studies, and highlight the need to improve their design and reporting standards.

Artificial intelligence (AI) is an innovative and fast moving field with the potential to improve patient care and relieve overburdened health services. Deep learning is a branch of AI that has shown particular promise in medical imaging.

The volume of published research on deep learning is growing, and some media headlines that claim superior performance to doctors have fuelled hype for rapid implementation. But the methods and risk of bias of studies behind these headlines have not been examined in detail.

To address this, a team of researchers reviewed the results of published studies over the past 10 years, comparing the performance of a deep learning algorithm in medical imaging with expert clinicians.

They found just two eligible randomised clinical trials and 81 non-randomised studies.

Of the non-randomised studies, only nine were prospective (tracking and collecting information about individuals over time) and just six were tested in a 'real world' clinical setting.

The average number of human experts in the comparator group was just four, while access to raw data and code (to allow independent scrutiny of results) was severely limited.

More than two thirds (58 of 81) studies were judged to be at high risk of bias (problems in study design that can influence results), and adherence to recognised reporting standards was often poor.

Three quarters (61 studies) stated that performance of AI was at least comparable to (or better than) that of clinicians, and only 31 (38%) stated that further prospective studies or trials were needed.

The researchers point to some limitations, such as the possibility of missed studies and the focus on deep learning medical imaging studies so results may not apply to other types of AI.

Nevertheless, they say that at present, "many arguably exaggerated claims exist about equivalence with (or superiority over) clinicians,

which presents a potential risk for patient safety and population health at the societal level."

Overpromising language "leaves studies susceptible to being misinterpreted by the media and the public, and as a result the possible provision of inappropriate care that does not necessarily align with patients' best interests," they warn.

"Maximising patient safety will be best served by ensuring that we develop a high quality and transparently reported evidence base moving forward," they conclude.

More information: Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies in medical imaging, *The BMJ*, DOI: 10.1136/bmj.m689

Provided by British Medical Journal