Artificial Intelligence can identify illnesses "ranging from cancer to eye diseases" as accurately as trained doctors, new research has revealed. Using a technique known as deep learning, computers can examine thousands of medical images to identify patterns of disease. This has "enormous potential" for improving the precision and speed of diagnosis, according to scientists from University Hospitals Birmingham NHS Foundation Trust.

By pooling data from 14 trials, the team showed that deep learning correctly detected disease in 87 per cent of cases - compared to 86 per cent achieved by doctors. And the ability to accurately rule out patients who did not have disease was similar - 93% for the machine algorithms compared to 91 per cent for doctors.

"We found deep learning could indeed detect diseases ranging from cancers to eye diseases as accurately as health professionals," said lead author Professor Alastair Denniston.
"But it is important to note AI did not substantially outperform human diagnosis." But this is expected when the judge is Human for him to not see past his own idea.

More than 30 AI algorithms for healthcare have already been approved by the US Food and Drug Administration, and there have been strong market forces driving their development. Reports of AI outperforming humans in diagnostic testing have also generated much excitement and debate.

"Diagnosis of disease using deep learning algorithms holds enormous potential," Prof Denniston said.

However, the study, published in The Lancet Digital Health, notes that only a few studies were of sufficient quality to be included in the analysis. It means the true power of AI which combines algorithms, big data and state of the art computers to emulate human intelligence remains uncertain.

There was a lack of direct comparisons between the performance of humans and machines, and validation in real clinical environments.

"We reviewed over 20,500 articles, but less than 1 per cent of these were sufficiently robust in their design and reporting that independent reviewers had high confidence in their claims," said Prof Denniston.

"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."

Prof Denniston and colleagues called for higher standards of research and reporting to improve future evaluations.
"Evidence on how AI algorithms will change patient outcomes needs to come from comparisons with alternative diagnostic tests in randomised controlled trials," said co-author Dr Livia Faes, of 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."

Dr Tessa Cook, of the University of Pennsylvania, USA, who was not involved in the study, said comparing AI to doctors in the real world where data is "messy, elusive, and imperfect" is challenging.

Writing in the journal, she said: "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."