

2024年2月28日

報道機関 各位

国立大学法人東北大学

医用画像診断 AI に落とし穴 答えは正しくても考え方が正しいとは限らない 臨床応用に向けた課題を明確化

【発表のポイント】

- 人工知能（AI）による医用画像診断の性能は高いとの報告が多数あります。しかし今回、その診断に至った根拠と専門医の所見との一致度は高いとは限らないことを発見しました。
- 医学的に妥当でない不適切な根拠は、思わぬ結果を招く危険があります。
- このような危険性を認識し、対策することで、より安全性の高い AI の臨床応用が期待されます。

【概要】

深層学習^(注1)などの人工知能（AI）は進歩が著しく、医用画像診断への応用が進められています。しかし、AI が訓練データから何を学んだかなどの詳細はこれまで十分に解明されていませんでした。

東北大学大学院医学系研究科医用画像工学分野の曾昱雯助教らの研究グループは、深層学習が医用画像中のどこに注目して診断したのかを可視化する技術を用いて、その注目領域の医学的な妥当性を詳しく解析しました。先行研究で高性能を達成した深層学習モデルの注目領域と、医師の診断に基づく重要領域を比較した結果、深層学習モデルの高い分類性能に反して、その注目領域の 30%~80%は医学的な重要領域と無関係であることがわかり、両者に大きな齟齬があることが明らかになりました。本研究は AI による医用画像診断の医学的な妥当性に懸念があることを示しており、今後、新たな訓練法の開発など、さらなる検証と対策を進めることで、より安全性の高い AI の臨床応用が期待されます。

本研究成果は、2024年2月9日付で医学に関する画像分析の専門誌 Journal of Imaging Informatics in Medicine に掲載されました。

【詳細な説明】

研究の背景

近年、人工知能（AI）の分野では、深層学習などの新たな技術による顕著な進展があり、医療画像診断への応用においては専門医に匹敵する高い性能を示すという報告が多くあります。しかし、これまでの評価方法で高い性能を示していても、実際の臨床現場で同じ高性能を発揮できるのか、その信頼性について懸念が高まっています。特に、深層学習モデルが注目した医用画像の特徴が医学的所見とどの程度一致しているかなどの妥当性の検証は不十分であり、臨床において、医師の診断結果との乖離を引き起こす可能性が危惧されています。

今回の取り組み

東北大学大学院医学系研究科医用画像工学分野の曾昱雯（ソウ イブン）助教、張曉勇（チョウ ギョウユウ）非常勤講師（現所属：仙台高等専門学校総合工学科）、本間経康（ほんま のりやす）教授らの研究グループは、法医学における死後画像を用いた溺水診断を例に取り上げ、先行研究で高性能を達成した深層学習モデルの医学的妥当性を詳しく検証しました。具体的には、深層学習モデルが注目した画像特徴を可視化技術で特定し「注目領域」としました。また、放射線診断医の画像所見に基づいて注釈した画像領域を医学的な「重要領域」と定義し、モデルの注目領域と比較しました（図 1）。その結果、モデルの注目領域は、少ない場合だと 30%しか医学的な重要領域と一致しませんでした。また、80%程度一致する場合でも、領域中で重要視する位置が異なっていました（図 2）。今回検証した深層学習モデルが先行研究において 90%以上の高い正答率で溺死を分類可能であったと報告されていることから考えると、モデルと臨床上の医学的所見の間に予想外に大きな齟齬があったといえます。

今後の展開

これまでの研究で、深層学習モデルが不適切な根拠により導き出した答えは思わぬ誤診の原因になり、異なる性質を持つ症例画像に対して予想外の深層学習モデルの性能劣化に繋がる危険性が報告されています。今回明らかになった齟齬は、これまでとは異なる検証方法の重要性を示すとともに、人工知能の性能評価の難しさの一端を実証したものであり、臨床応用に向けた新たな課題を提起するものです。

今回の結果を踏まえ、医学的に妥当な根拠を獲得させるための、新たな訓練法の必要性が明らかになりました。これは、深層学習モデルの内部数理を人間の専門知識に基づく複雑な意思決定過程に近づける必要性を意味します。現在の訓練法は、数理的な簡易性と実現性のため、医学的な側面では単一の評価基準を採用することが多いのに対し、より多角的な医学的基準に沿った訓練法へ

の展開が期待され、臨床応用に向けて現在開発を進めています。

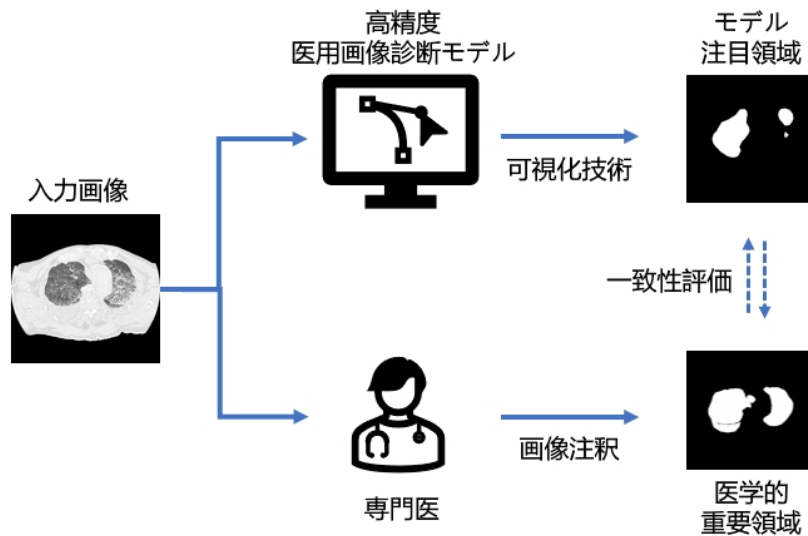


図 1. 人工知能による医用画像診断の信頼性を検証するため、可視化技術でモデルの「注目領域」を抽出し、同じ画像に対して放射線診断医の画像所見に基づいて注釈した医学的な「重要領域」とどの程度一致するかを評価した。

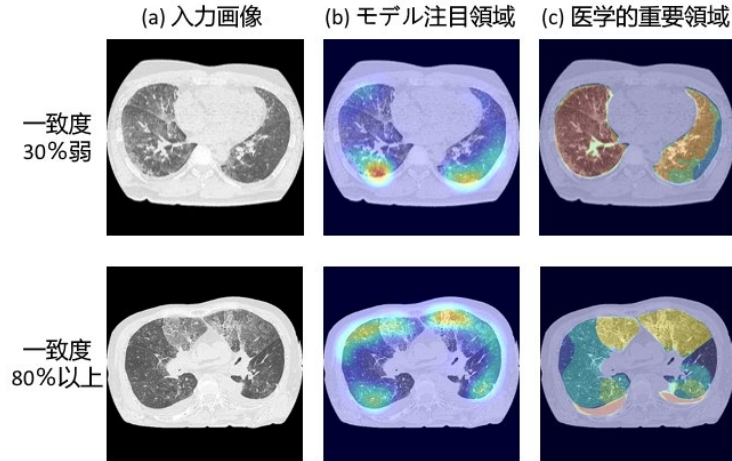


図 2. 元画像(a)に対し、(b)モデルの「注目領域」と(c)医学的な「重要領域」を投影した例。投影の色は赤いほど診断結果への貢献が高くなることを示す。正しく分類された二例のうち、上段の症例は 30%弱の一致度しかなく、下段の症例は 80%以上の一致度を有するが、重要視すべき部分（赤色）に大差が見られた。モデルと医学的所見の間に予想外に大きな齟齬があったといえる。

【謝辞】

この研究は、東北大学大学院医学系研究科附属 創生応用医学研究センター、東北大学 スマート・エイジング学際重点研究センターと JSPS 科研費 JP18K19892, JP20K08012, JP19H04479 の一部の支援を受けています。

【用語説明】

注 1. 深層学習：人工知能（AI）の一種であり、大量のデータを用いて多層のニューラルネットワークモデルを訓練し、自動的にデータから特徴を学習することができる。

【論文情報】

タイトル：Inconsistency between Human Observation and Deep Learning Models: Assessing Validity of Postmortem Computed Tomography Diagnosis of Drowning

著者：Yuwen Zeng*, Xiaoyong Zhang, Jiaoyang Wang, Akihito Usui, Kei Ichiji, Ivo Bukovsky, Shuoyan Chou, Masato Funayama, Noriyasu Homma

*責任著者：東北大学医学系研究科医用画像工学分野 助教 曾昱雯 Yuwen Zeng

掲載誌：Journal of Imaging Informatics in Medicine

DOI：10.1007/s10278-024-00974-6

URL: <https://link.springer.com/content/pdf/10.1007/s10278-024-00974-6.pdf>

【問い合わせ先】

（研究に関すること）

東北大学大学院医学系研究科医用画像工学分野

教授 本間 経康（ほんま のりやす）

助教 曾 昱雯（ソウ イブン）

TEL: 022-717-8190

Email: web@rii.med.tohoku.ac.jp

（報道に関すること）

東北大学大学院医学系研究科・医学部広報室

TEL: 022-717-8032

Email: press@pr.med.tohoku.ac.jp