

2023 年 11 月 27 日

## 人工知能の高性能に潜む新たな危険性を解明 医用画像診断用途の信頼性向上に期待

### 【発表のポイント】

- 人工知能の高い性能に反する、不適切動作の新たな原因を特定しました。
- 新たな原因は、性能向上に寄与すると考えられてきたデータ特性に潜んでおり、とくに COVID-19 の診断用途など、開発の緊急性が高い場合に対策が不十分だったと考えられます。
- 今回の解明により、不適切動作を抑制するための対策が可能となり、医用画像診断などの高い信頼性が要求される用途への実装が期待されます。

### 【概要】

人工知能（AI）の高性能化の中心技術として注目されている深層学習では、大量の訓練データが必要とされ、医療分野ではデータ確保が課題となっています。COVID-19 診断など緊急性が高い状況では、慣例に従いデータ収集した場合でも、AI は一見高性能にも関わらず、実際には性能が悪い場合があります。今回新たに特定した性能低下の原因は、AI が医学的な特性とは関係のない、撮影装置や施設などの特性を学習してしまう、これまで注目されていなかった問題でした。

東北大学大学院医学系研究科医用画像工学分野の張彰研究員、張暁勇准教授（現所属：仙台高等専門学校総合工学科）、本間経康教授らの研究グループは、胸部 X 線画像を用いた COVID-19 診断を例に挙げ、新たな原因が深層学習性能に与える影響を詳細に調査しました。実験の結果、今回特定された原因を含むデータで訓練されたモデルは信頼性が低下することを実証しました。この知見は、今後の医療 AI 開発において、訓練データに要求される重要な性質を示唆しており、信頼性の高い AI に繋がる成果です。

本研究成果は、2023 年 11 月 3 日に Scientific Reports 誌に掲載されました。

## 【詳細な説明】

### 研究の背景

医療分野への応用においても、深層学習などに代表される人工知能（artificial intelligence, AI）は過去 10 年で大きく発展してきました。深層学習において欠かせない要素の一つはデータです。膨大なデータ量が必要とされるだけでなく、データの統計的性質が非常に重要で、AI の性能を左右します。たとえば、疾患群と正常群の分類では、訓練時に両群のデータ数が揃っていないと多い群の性質が過度に学習されてしまい、性能が低下する群間不均衡問題が知られています。このため、COVID-19 診断など緊急性が高くデータの確保が難しい状況では、疾患群と正常群を、それぞれ異なる医療施設等から別々に収集して、データ数を揃えた例も報告されました。この場合、群間不均衡は生じず、開発された AI も高性能に見えますが、実際には性能が悪い場合があります（図 1）。この新しい不均衡問題は、よく知られた群間不均衡を回避しているため、これまであまり注目されておらず、その原因は明らかになっていませんでした。したがって、対策も不十分であり、医用画像診断のような信頼性が要求される用途では、潜在的な脅威となっていたと考えられます。

モデルが実際のクラスに焦点を当てる代わりに、病院ごとの特徴を捉えるリスクがあります。これを「出自内不均衡」と呼び、各出自（医療機関）内でのクラスの不均衡を指します。

### 今回の取り組み

私たちの研究では、「疾患群と正常群の医学的な違いとは関係のない、撮影装置や施設などのデータ出自に起因する違いを学習してしまうことで性能低下が生じる」という仮説を立て、これを「出自内不均衡」問題と名付けました。具体的には、胸部 X 線画像を用いた COVID-19 の診断に焦点を当て、出自内不均衡問題が深層学習に与える影響を詳しく調査する実験を行いました。医学的な違いを学習できているか確認するため、診断上重要な肺の部分を隠したり、切り取ったりする画像処理を行い、処理毎にデータセットを分けました（図 2）。その上で、ある装置や施設等の出自内で疾患数と正常数の均衡の取れたデータセットと、それとは異なり出自内で疾患数と正常数が揃っていない不均衡なデータセットの 2 種類を用意し、訓練用と評価用に分けました。

均衡、不均衡それぞれのデータセットを使用して同じモデルを訓練し、2 種類の評価用データセットで相互に性能を評価した結果、出自内均衡の取れたデータで訓練されたモデルの信頼性が高いことが示されました。また、モデルが診断上注目した領域を可視化手法で比較したところ、出自内均衡による訓練モデルは、医学的により妥当な領域に注目していたことが示されました（図 3）。これは、COVID-19 診断などの緊急を要する AI 開発において、訓練データの出自内均衡がいかに重要かを意味しています。

## 今後の展開

今回あらたな性能低下の原因を特定したことで、この出自内不均衡による影響を軽減する方法の開発が可能になりました。出自内不均衡の影響は異なる装置・施設間のデータ特性の隔たりに起因する可能性があるため、出自固有の特徴を最小限に抑える必要があります。このような訓練データ収集法の改善に加え、学習結果の医学的妥当性検証や、より根本的に医学的妥当性を保証する学習法などの解決策の開発を行っています。これにより、医療 AI の信頼性向上が可能になり、安心して適切な医療提供への貢献が期待されます。

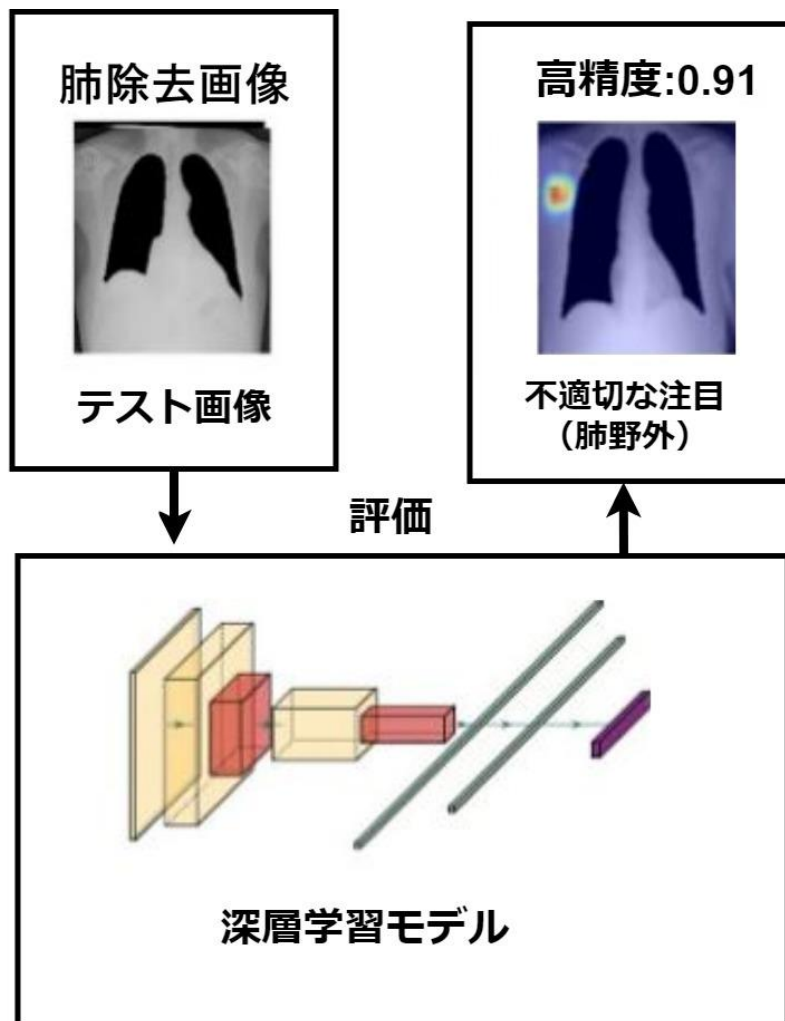


図 1. 深層学習モデルが、医学的に重要な領域である肺を除去（黒塗り）した画像でも高い精度で COVID-19 の感染を判断できた例です。しかし、注目した場所が肺領域外にあるなど、その根拠は医学的に不適切でした。このように、深層学習モデルが獲得した根拠は、医学的な所見とは異なり信頼性が低い場合がありますが、その原因はこれまで明らかになっていませんでした。

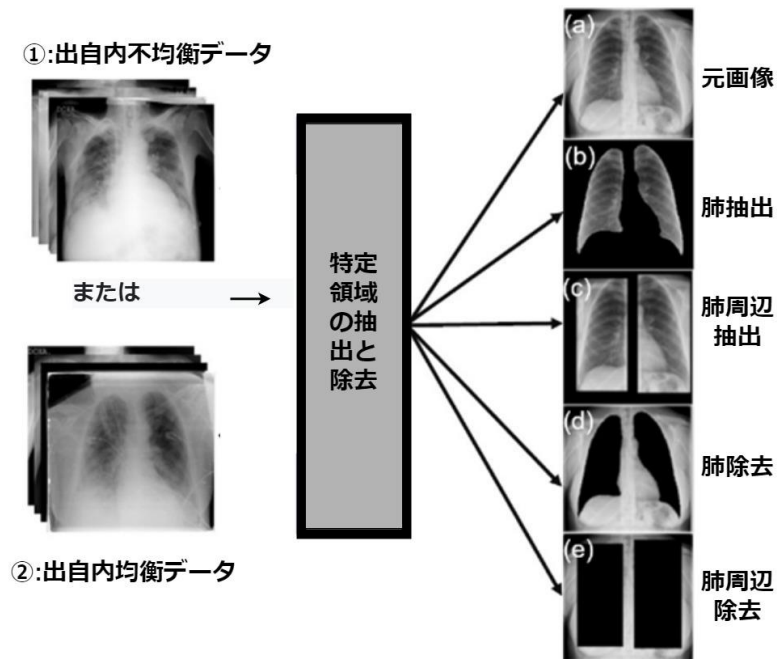


図 2. 深層学習モデルが、胸部 X 線画像中のどの領域の特徴に基づいて COVID-19 の感染判断を行っているかを調べるため、元画像から肺などの特定領域を抽出したり、除去したりする画像処理を用いてデータセットを用意しました。今回新たに、出自内均衡と出自内不均衡な 2 つのデータセットを、深層学習モデルの訓練ならびに性能評価に使用しました。これら出自内均衡と不均衡の結果を比較することで、その影響を検証しました。

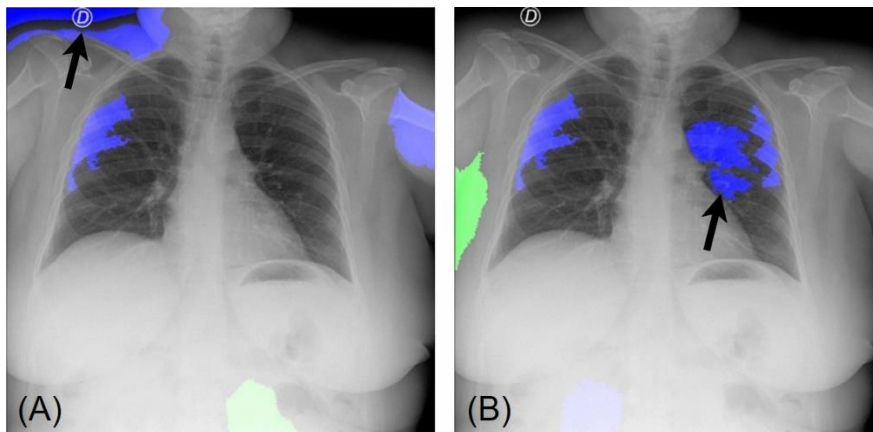


図 3. モデルが COVID-19 感染判断の根拠として注目した領域：(A) 出自内不均衡データセットで訓練されたモデルの根拠領域、(B) 出自内均衡データセットで訓練されたモデルの根拠領域。画像内の矢印は、感染判断に最も寄与した領域です。(A)の出自内不均衡を用いたモデルは、医学的に不適切な肺領域外の文字 (D) 付近に注目してしまっていますが、(B)の出自内均衡を用いたモデルは医学的に妥当な肺領域に注目しており、判断根拠の信頼性が高いことが示されました。

### 【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, and JP19H04479 の一部の支援を受けています。

### 【論文情報】

タイトル : How intra-source imbalanced datasets impact the performance of deep learning for COVID-19 diagnosis using chest X-ray images

著者 : Zhang Zhang\*, Xiaoyong Zhang, Kei Ichiji, Ivo Bukovský and Noriyasu Homma

\*責任著者 : 東北大学大学院医学系研究科 医用画像工学分野

ZHANG ZHANG

掲載誌 : Scientific Reports

DOI : 10.1038/s41598-023-45368-w

URL: <https://www.rii.med.tohoku.ac.jp/>

### 【問い合わせ先】

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、学術研究員 張彰

TEL: 022-717-8190

Email: [web\\*rii.med.tohoku.ac.jp](mailto:web*rii.med.tohoku.ac.jp) (\*を@に置き換えてください)

### 【取材に関すること】

東北大学大学院医学系研究科・医学部広報室

東北大学病院広報室

TEL: 022-717-7149

E-mail: [press\\*pr.med.tohoku.ac.jp](mailto:press*pr.med.tohoku.ac.jp) (\*を@に置き換えてください)