AI出力の調整機能による読影課題でのアルゴリズム忌避の抑制

Suppression of algorithm aversion through modifying AI's decision in an interpretation of radiogram task

三宅 圭音 [1, 2*]　山田 誠二 [2, 1]
Keito MIYAKE[1, 2]　Seiji YAMADA[2, 1]

[1] 総合研究大学院大学
[1] The Graduate University for Advanced Studies, SOKENDAI
[2] 国立情報学研究所
[2] National Institute of Informatics

Abstract: 現代社会では，人間と AI の協力がますます重要性を増している．しかし，読影医のような専門性の高い人はアルゴリズム忌避が高い傾向にあり，アルゴリズムの出した結果を参考にしないことがある．先行研究では AI の出力をわずかでも調整することができればアルゴリズム忌避を軽減できる可能性が示唆されている．本実験の目的は，X 線画像の読影タスクにおいて AI 出力の調整によるアルゴリズム忌避の抑制を行い調査することである．

1　はじめに

現在の社会で人工知能 (AI) の使用は一般的になり，様々な場面で AI との協力が必要となってきている．しかし，AI の性能が人間よりも高いにも関わらず，AI が出した結果よりも人間が出した結果を優先してしまう現象がある．これをアルゴリズム忌避 (algorithm aversion) と言う [2].

アルゴリズム忌避が起こることで，様々な問題が発生する可能性がある．例えば，医療分野において，アルゴリズムが出力した正しい診断結果を受け入れずに，患者の健康や治療に影響を及ぼす恐れがある [4][9].

このアルゴリズム忌避は多様な要因によって引き起こされるが，その 1 つとして self-efficacy(自己効力感) がある [8]. 特定の分野での経験が一般の人よりも多い専門家において，自己効力感の高い傾向が見られる．読影医は非常に専門性の高い職業であり，アルゴリズムの使用を避ける傾向がある [5][7].

その一方で，アルゴリズム忌避を軽減させる方法も提案されている．その 1 つとして，不完全な AI による出力結果を調整できるようにすることで，アルゴリズム忌避を軽減できる可能性が示されている [3].

この背景から，本研究の目的は，胸部 X 線画像のセグメンテーションタスクにおいて，AI の出力調整を行うことで読影医のアルゴリズム忌避を抑制することができるかを実験的に解明することである．これにより，AI との協力が促進されることが期待される．

*連絡先： 総合研究大学院大学/国立情報学研究所
〒 101-8430 東京都千代田区一ツ橋 2 丁目 1 番地 2 号
E-mail: m-keito@nii.ac.jp

2　アルゴリズム忌避

人間よりも優れたアルゴリズムやモデルが提供されているにも関わらず，人々はそれらが出した結果を避ける傾向にあること，または信頼しない現象をアルゴリズム忌避と呼ぶ．例えば，天気予報 AI が出力した天気予報より，天気予報士が出した天気予報の方が信頼できて採用されやすくなる．Dietvorst らは，アルゴリズムのエラーに対する人々の反応が，人間のエラーに対する反応よりも厳しいことを示した [2]．このアルゴリズムによるエラーは，否定的な情報の影響が強く，アルゴリズムへの信頼が著しく損なわれることがある [6].

3　関連研究

3.1　アルゴリズム忌避が起こる要因

Mahmud らは，アルゴリズム忌避に関するシステマティックな文献レビューを行い，アルゴリズムによる意思決定に影響を与える要因に焦点を当て，アルゴリズム忌避がどのように形成され，どのような要因がその影響を説明するのかについて、包括的な概観を提供している [8].

3.2 医療分野での AI 利用

医療分野での AI 利用は顕著に増加しており，その応用範囲は診断支援や医療教育，疾病予防など多岐にわたる [7][10][11]．AI 技術は精度の高い診断情報の提供に不可欠な役割を担っている．2024 年時点では ChatGPT の急速な成長もあり，医療教育にパーソナライズされた学習や協働学習など，多様な応用を提供している [1]．

3.3 アルゴリズム忌避軽減の試み

Dietvorst らは，学生の数学テストの点数予測を行うタスクで，AI の出力結果をわずかに修正できる場合にアルゴリズム忌避を軽減することを示唆している [3]．

4 実験方法

4.1 実験目的と実験計画

本研究の目的は，胸部 X 線画像のセグメンテーションタスクにおいて，AI が出力した結果をわずかに調整できる状況での，AI へのアルゴリズム忌避が抑制されるかを調査することである．

本研究での仮説として以下の二つを考えた．

H1: AI の結果を調整できると，AI を選択する傾向が高くなる．

H2: AI の結果を調整できると，AI への印象が良くなる．

この 2 つの仮設を検証するために実験を行う．実験計画は 1 要因で行い，AI 出力調整要因が 2 水準 (出力を調整できる，調整できない) である．参加者は実験前に AI 出力結果を調整するかどうかの選択をする．

4.2 実験内容

本実験では，対面実験による胸部 X 線画像のセグメンテーション調整タスクを実施する．実験の流れを図 1 に示す．まず実験を始める前に，実験内容の説明と診断への自己効力感や AI に対してどのような印象を持っているかの事前アンケートを行う．次に，参加者はセグメンテーションの調整をする際に，AI の結果を調整できるようにするか，使用しないのかを選択してもらう．その後，タスク画面に遷移し，セグメンテーションタスクを 40 回行う．終了したら事前アンケートと同様の内容の事後アンケートを行う．

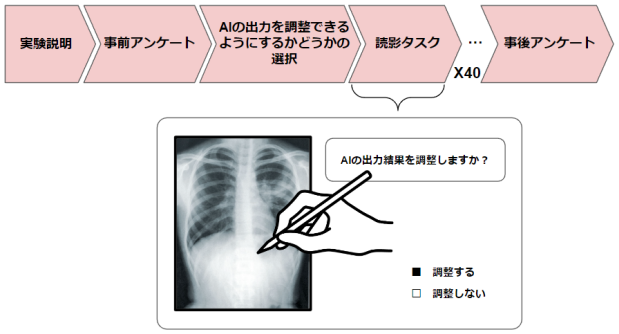

図 1: 実験の流れとタスクの例

5 まとめ

本研究では，胸部 X 線画像のセグメンテーションタスクで，AI の出力結果をわずかに調整できる場合にアルゴリズム忌避が抑制されるかどうかを調査する．実験結果の分析により，AI の出力結果を調整できる状況がアルゴリズム忌避の抑制にどのような影響を与えるかを明らかにすることが期待される．今後の研究では，より多くの参加者や異なる条件での実験を行い，結果の信頼性を高め，アルゴリズム忌避を抑制する実用的な手法の検討を進める．

参考文献

[1] Tessa Breeding, Brian Martinez, Heli Patel, Hazem Nasef, Hasan Arif, Don Nakayama, and Adel Elkbuli. The utilization of ChatGPT in reshaping future medical education and learning perspectives: A curse or a blessing? *Am. Surg.*, p. 31348231180950, June 2023.

[2] Berkeley J Dietvorst, Joseph P Simmons, and Cade Massey. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *J. Exp. Psychol. Gen.*, Vol. 144, No. 1, pp. 114–126, February 2015.

[3] Berkeley J Dietvorst, Joseph P Simmons, and Cade Massey. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Manage. Sci.*, Vol. 64, No. 3, pp. 1155–1170, March 2018.

[4] Ibrahim Filiz, Jan René Judek, Marco Lorenz, and Markus Spiwoks. The extent of algorithm aversion in decision-making situations with varying gravity. *PLoS One*, Vol. 18, No. 2, p. e0278751, February 2023.

[5] Susanne Gaube, Harini Suresh, Martina Raue, Alexander Merritt, Seth J Berkowitz, Eva Lermer, Joseph F Coughlin, John V Guttag, Errol Colak, and Marzyeh Ghassemi. Do as AI say: susceptibility in deployment of clinical decision-aids. *NPJ Digit Med*, Vol. 4, No. 1, p. 31, February 2021.

[6] Leanna Ireland. Who errs? algorithm aversion, the source of judicial error, and public support for self-help behaviors. *Journal of Crime and Justice*, Vol. 43, No. 2, pp. 174–192, March 2020.

[7] Ekaterina Jussupow, Kai Spohrer, and Armin Heinzl. Radiologists' usage of diagnostic AI systems. *Business & Information Systems Engineering*, Vol. 64, No. 3, pp. 293–309, June 2022.

[8] Hasan Mahmud, A K M Najmul Islam, Syed Ishtiaque Ahmed, and Kari Smolander. What influences algorithmic decision-making? a systematic literature review on algorithm aversion. *Technol. Forecast. Soc. Change*, Vol. 175, p. 121390, February 2022.

[9] Ashley N D Meyer, Velma L Payne, Derek W Meeks, Radha Rao, and Hardeep Singh. Physicians' diagnostic accuracy, confidence, and resource requests: A vignette study. *JAMA Intern. Med.*, Vol. 173, No. 21, pp. 1952–1958, November 2013.

[10] Richard Pak, Nicole Fink, Margaux Price, Brock Bass, and Lindsay Sturre. Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics*, Vol. 55, No. 9, pp. 1059–1072, July 2012.

[11] Shintaro Sukegawa, Sawako Ono, Futa Tanaka, Yuta Inoue, Takeshi Hara, Kazumasa Yoshii, Keisuke Nakano, Kiyofumi Takabatake, Hotaka Kawai, Shimada Katsumitsu, Fumi Nakai, Yasuhiro Nakai, Ryo Miyazaki, Satoshi Murakami, Hitoshi Nagatsuka, and Minoru Miyake. Effectiveness of deep learning classifiers in histopathological diagnosis of oral squamous cell carcinoma by pathologists. *Sci. Rep.*, Vol. 13, No. 1, pp. 1–9, July 2023.