

# 横隔膜検出タスクにおける AI 出力の正誤ラベリングと修正機能によるアルゴリズム忌避の抑制

Mitigating Algorithm Aversion in Medical Professionals: Investigating the Relationship between Psychological Factors and AI Output Usage Rates

三宅 圭音<sup>\*1\*2</sup>  
Keito Miyake

尾崎 公美<sup>\*3</sup>  
Kumi Ozaki

前東 晃礼<sup>\*4</sup>  
Akihiro Maehigashi

山田 誠二<sup>\*2\*1</sup>  
Seiji Yamada

<sup>\*1</sup>総合研究大学院大学

The Graduate University for Advanced Studies, SOKENDAI

<sup>\*2</sup>国立情報学研究所

National Institute of Informatics

<sup>\*3</sup>浜松医科大学

Hamamatsu University School of Medicine

<sup>\*4</sup>静岡大学

Shizuoka University

医療診断において医師の意思決定は、患者の健康に直結する重要な要素である。しかし、AI が人間よりも高性能とわかっていても人間の判断を優先する「アルゴリズム忌避」が課題となり AI 活用を妨げる可能性がある。本研究では、読影経験者を対象に横隔膜検出タスクを実施し、「AI 出力の正誤ラベリング」と「AI 出力修正」の 2 つのインタラクション機能を提案・検証した。結果として、インタラクション機能がある条件はない条件と比べて有意に AI 選択率が上昇した。

## 1. はじめに

AI は医療分野において大きな進歩を遂げており、特に画像診断の領域では高度な診断支援システムが開発されている。AI の画像解析能力が向上することで医療従事者の業務負担を軽減し、診断の精度向上が期待されている。しかし、AI が進化する一方で医療従事者が AI を信頼せず自身の判断を優先する傾向がある。これは「アルゴリズム忌避 (algorithm aversion)」と呼ばれ、AI が高性能だと分かっていても使用を避ける現象である [Dietvorst 15]。アルゴリズム忌避が発生すると AI の支援が十分に活かされず、結果として全体の診断パフォーマンスが低下することが懸念される。特に、医療現場では誤診や見落としが患者の安全に直結するため、この問題は極めて重要である [Filiz 23][Meyer 13][Bigman 18]。

先行研究では、一般の人よりも専門家の方がアルゴリズムを使用しないことが報告されており、専門的な知識や経験を有するほど AI を使用しなくなる傾向があるとされる [Jussupow 22][Gaube 21]。例えば、放射線科における読影は非常に専門的な技術が要求されており、このアルゴリズム忌避が起こりやすい状況にあると言える。特に経験豊富な医療従事者ほど自己効力感が高く、アルゴリズムの決定に頼らない傾向がある [Arkes 86]。

また、先行研究においてユーザがアルゴリズムの出力を修正できる機能を持つことで、そのシステムの使用率が向上され、アルゴリズム忌避が抑制されていることが示されている [Dietvorst 18]。しかし、専門家におけるアルゴリズム忌避抑制の試みはまだ研究の余地が残っている。特に、医療における意思決定という文脈において、どのようなインタラクション設計が効果的であるかは十分に検討されていない。

このような背景を踏まえ、本研究では読影経験者の AI に対するアルゴリズム忌避を正誤ラベリングと修正機能によって抑制することを目的とする。具体的には、AI の出力に対して読影経験者が「正解か不正解のラベルを付ける機能」と「ペンによる直接修正機能」という 2 つの機能を提供することで、AI の使用率 (リライアンスレート) を高めることを試みる。

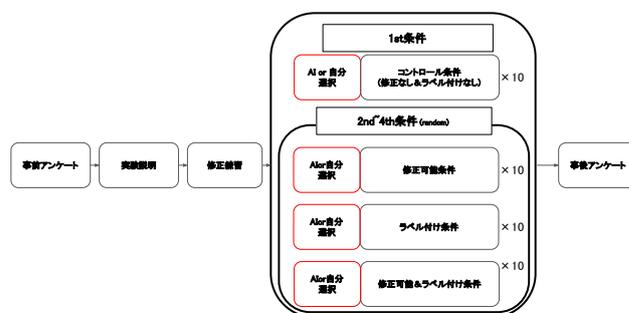


図 1: 実験全体の流れ

## 2. 実験方法

### 2.1 参加者

浜松医科大学付属病院の放射線科医、放射線技師、および読影経験者 25 名 (男性 19 名、女性 6 名; 年齢 24-54 歳、中央値 29 歳; 経験年数 0-30 年、中央値 3 年) が実験に参加した。実験は約 30 分で完了し、参加者には謝礼として 2,000 円の QUO カードを支払った。参加者数は事前に G\*Power を用いて算出した [Faul 07]。

### 2.2 仮説

以下の 2 つの仮説を設定した。

**H1:** AI の出力に対する修正機能は AI 使用率を向上させる。

**H2:** AI の出力に対するラベル付け機能は AI 使用率を向上させる。

H1 は一般人を対象とした先行研究に基づいていて、この機能が医療従事者にも適用できるかの効果を検証する。

H2 は修正機能と同様にコントロール感を得られ、AI 使用率を向上させ、アルゴリズム忌避を抑制できることを検証する。

### 2.3 実験デザイン

実験全体の流れを図 1 に示す。本研究では、リライアンスレートを従属変数として、独立変数を 2 つ (修正機能とラベル

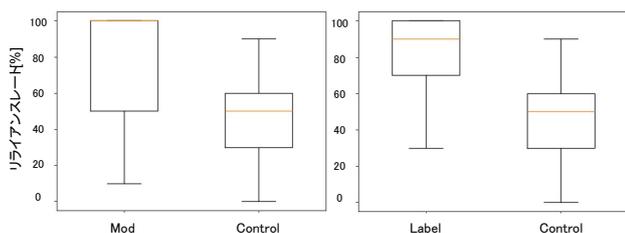


図 2: 各条件のリライアンスレートと条件間の分析結果

付け機能) 設定した。各要因は 2 水準 (機能あり・機能なし) あり, 合計 4 条件の参加者内実験計画を採用した。

これらの条件下で, 参加者は横隔膜検出タスクを行った。各試行において, 参加者はまず AI を使用するか, 参加者自身で行うかの選択をし, その後の操作は選択と条件に応じて変化する設計とした。条件間の順序効果を制御するため, 初回 10 回のコントロール条件終了後, 残りの 30 試行は 3 つの実験条件 (修正条件, ラベル付け条件, 両方あり条件) においてランダムに呈示された。

### 3. 結果

リライアンスレートを従属変数として, 2 (修正機能: あり・なし)  $\times$  2 (ラベル付け機能: あり・なし) の参加者内 2 要因分散分析を実施した。

分析の結果, 交互作用が有意であることが示された ( $F(1, 24)=19.39, p < .001, \text{partial } \eta^2=0.45$ )。この効果をより詳細に検討するため, 単純主効果の分析を行った。その結果, 修正機能のない水準下でラベル付け条件とコントロール条件で有意差が見られた ( $F(1, 24)=27.24, p < .000, \text{partial } \eta^2=0.00$ )。また, ラベル付け機能のない水準下で修正条件とコントロール条件で有意差が見られた ( $F(1, 24)=21.19, p < .000, \text{partial } \eta^2=0.08$ )。

### 4. 考察

統計分析の結果は, 本研究で設定した 2 つの仮説を支持するものであった。修正機能とラベル付け機能の両方が, 医療従事者の AI に対するリライアンス率を有意に向上させることが確認された。

修正機能の効果は, 先行研究 [Dietvorst 18] の知見と一致している。医療という専門性の高い文脈においても, ユーザーが出力を修正できる機能を持つことで, システムの使用意欲が向上することが示された。これは, AI の出力に対する制御可能性が, 医療従事者の懸念を軽減し, システムへの信頼を醸成する効果があることを示唆している。

ラベル付け機能については, AI の出力を正誤判定するという行為が, 医療従事者のシステムに対する理解を深め, より主体的な関与を促進する効果があったと考えられる。この機能は, 医療従事者の専門知識を活かしながら AI と協調できる仕組みとして機能した可能性がある。

一方で, 本研究にはいくつかの限界点が存在する。実験は統制された環境で特定の横隔膜検出タスクを対象としており, より多様な臨床状況での検証が必要である。また, AI システムへの

過度な依存 (over-reliance) のリスク [Cecil 24][Klingbeil 24] や, 責任の所在に関する課題も考慮する必要があると考えられる。さらに, タスクの性質上, 線を引くという作業には明確な正解が定めにくく, 異なる条件間でのパフォーマンスの比較が困難であった点も本研究の限界として挙げられる。

これらの知見と課題は, 医療現場における AI システムの効果的な活用方法を検討する上で重要な示唆を提供するものである。今後は, より多様な診断タスクでの検証や, 長期的な効果の評価を通じて, 適切なリライアンスの実現方法を探求していく必要がある。

## 5. まとめ

本研究では, 横隔膜検出タスクにおいて, AI の出力に対する修正機能とラベル付け機能が医療従事者のリライアンス率に与える影響を検討し, 両機能ともにリライアンス率を有意に向上させることを確認した。

これらの機能は, 医療従事者に AI システムに対するコントロール感を与え, 出力の評価に積極的に関与させることで, システムへの信頼感を高める効果があることが示唆された。今後は, より多様な診断タスクでの検証や長期的な効果の評価を通じて, 医療現場における AI システムの最適な活用方法を探求していく必要がある。

## 参考文献

- [Arkes 86] Arkes, H. R., Dawes, R. M., and Christensen, C.: Factors influencing the use of a decision rule in a probabilistic task, *Organizational Behavior and Human Decision Processes*, Vol. 37, No. 1, pp. 93–110 (1986)
- [Bigman 18] Bigman, Y. E. and Gray, K.: People are averse to machines making moral decisions, *Cognition*, Vol. 181, pp. 21–34 (2018)
- [Cecil 24] Cecil, J., Lermer, E., Hudecek, M. F. C., Sauer, J., and Gaube, S.: Explainability does not mitigate the negative impact of incorrect AI advice in a personnel selection task, *Scientific Reports*, Vol. 14, No. 1, pp. 1–15 (2024)
- [Dietvorst 15] Dietvorst, B. J., Simmons, J. P., and Massey, C.: Algorithm aversion: People erroneously avoid algorithms after seeing them err, *J. Exp. Psychol. Gen.*, Vol. 144, No. 1, pp. 114–126 (2015)
- [Dietvorst 18] Dietvorst, B. J., Simmons, J. P., and Massey, C.: Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them, *Manage. Sci.*, Vol. 64, No. 3, pp. 1155–1170 (2018)
- [Faul 07] Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A.: G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences, *Behavior Research Methods*, Vol. 39, No. 2, pp. 175–191 (2007)
- [Filiz 23] Filiz, I., Judek, J. R., Lorenz, M., and Spiwoks, M.: The extent of algorithm aversion in decision-making situations with varying gravity, *PLoS One*, Vol. 18, No. 2, p. e0278751 (2023)

- 
- [Gaubé 21] Gaubé, S., Suresh, H., Raue, M., Merritt, A., Berkowitz, S. J., Lerner, E., Coughlin, J. F., Guttag, J. V., Colak, E., and Ghassemi, M.: Do as AI say: susceptibility in deployment of clinical decision-aids, *NPJ Digit Med*, Vol. 4, No. 1, p. 31 (2021)
- [Jussupow 22] Jussupow, E., Spohrer, K., and Heinzl, A.: Radiologists' Usage of Diagnostic AI Systems, *Business & Information Systems Engineering*, Vol. 64, No. 3, pp. 293–309 (2022)
- [Klingbeil 24] Klingbeil, A., Grützner, C., and Schreck, P.: Trust and reliance on AI — An experimental study on the extent and costs of overreliance on AI, *Computers in Human Behavior*, Vol. 160, p. 108352 (2024)
- [Meyer 13] Meyer, A. N. D., Payne, V. L., Meeks, D. W., Rao, R., and Singh, H.: Physicians' Diagnostic Accuracy, Confidence, and Resource Requests: A Vignette Study, *JAMA Intern. Med.*, Vol. 173, No. 21, pp. 1952–1958 (2013)