

2024 年度年次報告書

信頼される AI システムを支える基盤技術

2021 年度採択研究代表者

竹内 一郎

名古屋大学 大学院工学研究科
教授

AI 駆動仮説の静的・動的信頼性保証と医療への展開

主たる共同研究者:

川口 淳 (佐賀大学 大学院医学系研究科 教授)

佐久間 淳 (東京科学大学 情報理工学院 教授)

松井 茂之 (名古屋大学 大学院医学系研究科 教授)

三好 寛明 (久留米大学 医学部 主任教授)

研究成果の概要

本研究は、現代の AI 技術と機械学習が社会実装されつつある現状において、その判断結果などの信頼性をいかに評価するかという重要な課題に対し、理論と応用の両面からアプローチしている。特に、AI による判断が医療や科学研究といった誤った意思決定が高いリスクを持つ分野に応用される場面では、出力の統計的有意性を伴った信頼できる AI の構築が望まれる。本研究では、AI・機械学習に対する統計的有意性評価法の構築を第一の柱とし、第二の柱としてその評価法を医療応用、特に AI 病理診断に適用・実証する取り組みを行っている。

2024 年度は、これらの二本柱に基づき、複数の研究グループの協力のもとで重要な成果を挙げた。まず、統計的信頼性評価法に関する研究では、特に深層学習モデルに対する選択的推論 (Selective Inference: SI) の応用と一般化に焦点を当てた。SI は、データ駆動型で仮説を選択する現代の AI の特性に対応する新しい統計的推測手法であり、もともとは線形モデルの変数選択における有意性評価のために開発された技術である。しかし、近年の AI 技術の複雑化に伴い、より高度なモデルに対しても有意性を評価する必要性が高まっており、本研究ではこれを深層学習に拡張する試みを進めた。昨年度の段階では、選択的推論を深層学習モデルに適用するためには、モデル構造に応じた選択イベントの数理的特徴付けを行い、その都度、個別に計算手順を設計し、専用のソフトウェアを開発する必要があった。これは極めて専門性の高い作業であり、選択的推論の一般的な普及を阻む障壁となっていた。これに対し、2024 年度の研究により、区分線形構造を持つニューラルネットワークに限定することにより、あらかじめ各層の動作をモジュール化し、それぞれに選択イベントを実装することで、モデル全体に対する選択的推論が自動的に適用可能となる汎用フレームワークを構築した。例えば、ほとんどの畳み込みニューラルネットワーク (Convolutional Neural Networks) は区分線形関数への分解が可能であり、このフレームワークを適用できる。このフレームワークの重要な特徴は、学習済みモデルを ONNX (Open Neural Network Exchange) 形式で保存し、それを読み込むだけで、追加の実装作業なしに選択的 p-value の計算が可能となる点にある。ONNX は複数のディープラーニングフレームワーク間の互換性を確保する標準形式であり、これを活用することで、ユーザは自身の学習済みモデルをそのまま信頼性評価に利用することができる。

一方、医療応用に関する取り組みでは、AI 病理診断の実証実験に向けたデータ基盤と診断プラットフォームの整備が大きく進展した。久留米大学病院と連携し、悪性リンパ腫の症例を対象とした病理画像データベースの構築を進めるとともに、Q-path と呼ばれる高度な病理画像管理・解析プラットフォームを本プロジェクトに導入した。Q-path は、超高解像度の whole slide image (WSI) を効率的に管理・閲覧することが可能だけでなく、外部開発の解析アルゴリズムを柔軟に統合できる機能を持つため、研究成果の実装に適している。2024 年度は、これまでに各研究グループが開発してきた統計的信頼性評価の手法を Q-path に組み込み、それらを病理診断タスクに適用可能とする技術的な基盤が整えられた。これにより、病理医が AI を用いて得た診断結果に対し、その有意性や信頼性を定量的に確認できる新しい診断支援のあり方が現実のものとなりつつある。

2024 年度の各グループの取り組みは以下の通りである。竹内グループでは、前述の ONNX ベースの選択的推論フレームワークを開発し、多様なニューラルネットワークへの適用性と実装の容易さを両立させた。佐久間グループは、深層学習モデルの説明可能性指標に対する統計的信頼性評価法を構築し、さらに、細胞検出を起点としたグラフニューラルネットワーク(GNN)を用いることで、サブタイプ分類の精度向上に成功している。病理画像の構造的特性をうまく取り入れたこのアプローチは、実臨床での利用に向けた重要な一歩といえる。松井グループは、個別化医療に資するデータ解析法に焦点を当て、特に複雑な診断プロトコルにおいて生じる情報のリークや選択バイアスを統計的に補正する技術を開発し、これにより信頼性の高い分析を可能とした。この成果は、統計学分野において権威ある *Journal of the Royal Statistical Society* に掲載されるなど、国際的にも高く評価されている。川口グループは、脳画像解析における AI 応用の信頼性評価に取り組み、今年度はバイオマーカーに基づく症例のサブグループリングと、それに応じた脳画像解析フレームを開発した。これにより、疾患の進行度や個別差を考慮した解析が可能となり、診断精度や再現性の向上に資する成果が得られた。三好グループは、AI 病理診断システムの構築に向けた症例データの整理と Q-path の整備に引き続き取り組むとともに、実証実験に向けて、実際の臨床業務で活用可能な AI ツールやユーザーインターフェースについての具体的な検討を進めた。臨床医との対話を通じて、現場に即した機能性と操作性を両立させる設計が進められている。

総じて、2024 年度は、信頼される AI の実現に向けた理論と応用の融合が進展した一年であった。特に、選択的推論を深層学習に適用可能とする技術基盤の確立、それを医療分野において実装・実証するための体制整備は、国内外における AI 応用研究の先端的な成果である。今後は、これらの成果をさらに発展させ、より多くの疾患領域への展開や、医療現場での運用実証を進めることで、AI の信頼性と社会的受容性の両立を実現していく予定である。

【代表的な原著論文情報】

- 1) T. Shiraishi, D. Miwa, T. Katsuoka, VNL. Duy, K. Taji, I. Takeuchi. Statistical Test for Attention Maps in Vision Transformers. *Proceedings of The 41th International Conference on Machine Learning (ICML 2024)*
- 2) VNL. Duy, HT. Lin, Ichiro Takeuchi. CAD-DA: Controllable Anomaly Detection after Domain Adaptation by Statistical Inference. *Proceedings of The 27th International Conference on AI and Statistics (AISTats 2024)*
- 3) Y. Zhe, J. Sakuma. Zero-shot domain adaptation based on dual-level mix and contrast. *Proceedings of The IEEE Conference on Artificial Intelligence (CAI 2024)*.
- 4) R. Emoto, M. Igeta, K. Matsui, K. Ishii, T. Takamura, S. Matsui. Evaluating treatment effect modifiers using data from randomized two-sequence, two-period crossover clinical trials: application to a diabetes study. *Journal of the Royal Statistical Society Series C* (in press).
- 5) Y. Ishimaru, A. Kawaguchi. Subgroups identification using biomarkers from multi-modal brain imaging. In *Advances in health and disease*. Nova Science Publishers (2024).