

3. 日本語に特化した 医療用視覚言語モデルの開発

安道健一郎 理化学研究所革新的知能統合研究センター

画像と言語を統合してユーザーと対話を行う大規模視覚言語モデル (vision language model : VLM) が急速に普及している。GPTシリーズやGeminiなどを筆頭に、多数のVLMが開発されており、国内でもAsagi¹⁾やLLM-jp VL²⁾など、多くのVLMが公開されている。

一方、医療分野での活用に目を向けると、医療機関における生成AI活用に関しては各種ガイドラインが整備され、クラウド利用を含む使用環境が整いつつある。しかし、依然として多くの医療機関では、医療情報の保護の都合上、組織外への医療情報の持ち出しには強い制約が存在する。したがって、医療分野では、患者データをまったく外部へ送らずに運用できる、自組織内ネットワークで運用できるモデルが強く求められる。

その流れを受けて、医療分野で高い性能を有するような医療領域特化の基盤モデルが開発され始めており、すでに大規模言語モデル (large language model : LLM) では、多様な医療文書から学習された日本語医療用LLMが開発されている³⁾。しかし、VLMでは、いまだ日本語で医療に特化したオープンなモデルは存在していなかった。そのため、われわれは、2026年3月に日本語医療用のVLMを構築した⁴⁾。本稿では、そのモデル開発について詳解する。

本モデルが開発目的とした 基盤モデルの種類

基盤モデルは目的別に大まかに分類できる。まず、使用用途により分けると、汎用モデルは雑談、翻訳、要約、プログラミング支援など、幅広い用途に対応することを目的として開発される。一方、ドメイン特化モデルは医療、法律、金融など、特定領域のデータを学習させることで専門知識に適応させたモデルである。特に医療分野ではドメイン特化モデルの開発が活発であり、LLaVA-Med⁵⁾や、2026年に入ってからGoogleからMedGemma 1.5⁶⁾が公開されている。

次に、対象とする言語による分類がある。複数言語に対応するように開発されたモデルに対して、日本語、中国語、韓国語など、特定の言語に適応したモデルは言語特化モデルと呼ぶことができる。日本語に特化した基盤モデルは、日本語の文法、語彙、敬語表現、文化的背景などに適応することを目的としている。特定言語に集中できるため、同規模の多言語モデルより対象言語で高い性能を出しやすいことが知られており、使用できる計算資源に制限がある場合は手をつけやすい。なお、最近では、英語の知識を転用する目的でバイリンガルモデルを作ることがよくある。

最後に、モデルの公開形態による分類がある。モデル全体が公開され、手元の環境で実行を行えるものは、一般にオープンウェイトモデルと呼ばれる。一

方、GPTシリーズやGeminiなどのように、APIやWebサービスを通じてのみ利用可能なものはクローズドモデルと呼ばれる。

われわれは、医療ドメイン特化×日本語特化×オープンウェイトモデルを開発することで、医療知識を備えた、日本語に強い、オープンなモデルをめざした。また、学習に用いるデータはすべて商用可能ライセンスのものを採用することで、使用ライセンスが緩いデータやモデルの作成をめざした。

学習用データの構築

日本語医療用VLMが今まで存在していなかった理由は学習データの不足にある。特に、ライセンスが自由であり、かつ大規模なデータを用意するのは大変困難である。既存の医用画像と日本語テキストの対データは多くても数万件規模であり、VLMを訓練するには規模が不足している。そのため、本モデル開発では、さまざまな基盤モデルを用いて英語データを日本語データに整形することで、約1200万件の画像テキスト対を用意し、学習に用いた。データはすべて商用利用可のライセンスで、一般公開している⁷⁾。

1. データ構築の過程

PubMed Central (PMC) から医療論文の図とキャプションを集めたOpenPMC-18M⁸⁾というデータセットをデータ元として用いた。このデータセットは、PMCから複合図を分離し、さらにグラ