

4. 医療AI研究開発のためのデータ収集・活用の現状と将来展望

1) イヨウガゾウラボにおけるAI研究開発のための医用画像データ収集・加工・流通

國枝 良 (株)イヨウガゾウラボ代表取締役兼CEO

医療分野におけるAI活用は、診断支援、治療計画支援、予後予測、業務効率化など、多方面で期待が高まっている。しかし、現場に近い立場から見れば、医療AI開発の律速段階は、モデルの設計そのものよりも、学習に堪えるデータをいかに集め、いかに使える形へ整えるかにある。特に医用画像分野では、データの取得、匿名化、選別、標準化、教師データ化に多大な時間を要し、AI研究者や企業が、本来注力すべきモデル開発や評価に十分なリソースを振り向けにくい構造が続いてきた。当社は、このボトルネックを解くことこそ、日本における医療AIの社会実装を前に進めるカギであると考えている。

医療AI開発のボトルネックがデータ整備にあるという認識から、当社は医用画像データの収集・加工・流通の実証に多く取り組んできた。その背景には、国内ではなお事業領域そのものが、医用画像データの収集・加工・流通をAI研究開発の観点から再設計するという、市場形

成の途上にある分野だという事情がある。そのため当社は、事業会社として具体的な提供価値を形にしながらか、AI開発に有効なデータの条件は何か、どの工程を標準化・自動化すべきかを見極めるための検証を重ねてきた。当社にとってこうした取り組みは、単なる研究的検証ではなく、得られた知見を事業へ継続的にフィードバックしながら、まだ市場が十分に形成されていないこの新たな事業領域において、利用者に選ばれる事業者をめざすための実践でもある。当社は、医用画像通信技術研究組合(CIP)を起点として事業会社化し、東京大学との産学連携である「Beyond AI連携事業」の流れの中で生まれた経緯を持つ。こうした背景もあり、研究的な検証と事業化の検討とを往復しながら、医療AI開発の現場で役立つデータ供給のあり方を具体化してきた。本稿では、現在の事業として重視している内容と、こうした取り組みを通じて得られた知見とを分けて記すこととしたい。

医療AI開発において不足しているのは「データ量」ではなく「使えるデータ」

日本では、医療AIをめぐる議論において、しばしば「大量データの不足」が課題として語られる。もちろん量は重要である。ただし、量だけでは不十分である。電子カルテ上で対象疾患を検索して画像を集めても、確定診断前の症例や疑い例が含まれるため、AI開発にそのまま用いるには適さない。また、治療前後が混在した症例や、不完全なメタデータ(撮像条件、診断名、治療前後の別など、画像に付随する属性情報が欠落または不統一な状態)を伴う症例も混入しやすい。このようなデータは、一見すると件数を稼げるが、精度を要求される医療AIの学習には必ずしも適さない。

当社が重視しているのは、疾患の有無や治療フェーズが明確であり、匿名化と利用条件が整理され、さらに、AI学習上意味のある属性情報が付与された「使えるデータ」である。医療データの二次利用基盤においては、収集、匿名化、標準化、加工、提供、フィードバックまでを一気通貫で管理する仕組みが不可欠である。そして、これら一連の流れを循環させることで、単なる収集量の拡大ではなく、質的改善を前提とした流通モ