

臨床検査領域でのデータサイエンス技術の重要性と可能性

◎石田 秀和¹⁾
岐阜大学医学部附属病院¹⁾

【データサイエンスとは】

データサイエンスは大量のデータから有用な知見や洞察を得るための科学的な手法である。データサイエンスやそれを担うデータサイエンティストに明確な定義は存在しないが、情報工学や機械学習といった技術をビジネス（我々の業界では臨床）に活かすことが主たる目的となる。機械学習は「予測」のための手法であり、近年目覚ましい発展をしている AI にも利用されている。AI は機械が主体となり、プログラムされた学習パターンでデータ処理・分析・予測を行う。一方、データサイエンスは人が主体となり、AI を含む機械学習的手法などをツールとして使用する。最終的には必要なデータを取捨選択、可視化することで、利用できる形にまで昇華することが、データサイエンティストの役割となる。

【臨床検査とデータサイエンス】

臨床検査は、疾患診断や治療効果判定、予防医療や個別化医療における情報源として重要な役割を担っている。しかしながら、臨床検査データは多種多様で複雑な特性を持っており、その分析には高度な知識を必要とする。現在検査室に存在する臨床検査データの多くは、時系列の数値データだけでなく、定性データやカテゴリデータ、さらに画像やコメントなどの非構造化データも含まれる。これらを網羅的に理解し診療録情報なども包括的に活用する人材として臨床検査技師は適任であり、臨床検査のプロフェッショナルがデータサイエンスの知識を身につけ、実践することで医学・医療の発展に大きく寄与する可能性がある。

【臨床検査におけるデータサイエンスの活用】

データサイエンスを身近に感じられた例としては、SARS-CoV-2 感染症におけるダッシュボードツールが挙げられる。さまざまな組織から SARS-CoV-2 感染症に関する感染状況や関連死者数などの情報を可視化したダッシュボードツールが公表され、感染症の状況やそのインパクトが目に見える形で示された。その他にもデータサイエンスはマーケティングや物流、ファイナンス、スポーツなど幅広く活用されている。臨床検査に関連する分野では、骨髄検査の深層学習モデル開発や AI による脳波解釈に関する研究など、AI を含む機械学習を活用した研究が数多く行われている。

我々の研究グループでは、筋肉量の影響を受けにくい腎機能指標の開発として、一般的な臨床検査項目からシスタチン C による推定糸球体濾過量 (eGFR) に近い、機械学習モデルの eGFR の開発を行った。検査部に蓄積されている大量の検査データを活用し、機械学習的手法を用いることで、汎用的に測定される生化学検査・血液検査項目のみからシスタチン C の測定を行わずにクレアチニンよりも正確な腎機能評価が可能になると考えている。現在、臨床検査システムへの導入方法の検討とともにその有用性について詳細な解析を行っている。

また、近年急速に発展しているテキスト生成 AI の活用についても検証を行っている。OpenAI 社の ChatGPT に代表されるテキスト生成 AI はユーザーからの質問に対し自然な言語で回答する AI であるが、専門的な医療知識や臨床検査に関する知識には乏しい。我々の行った調査では、無料で使用できる ChatGPT (GPT-3.5) における臨床検査技師国家試験 (文章問題のみ) の正答率は 5 割程度、有料版 (GPT-4) でも 8 割程度の正答率であった。誤答の多くは、もっともらしい回答を生成していたため、生成系 AI の使用には注意が必要であることが考えられた。しかしながら、使い方によっては業務効率化に寄与する可能性があると考えている。現在、既存のテキスト生成 AI に臨床検査の情報をさらに学習させ精度を向上する取り組みを行うとともにその活用方法についても検討を行っている。

多くの臨床検査の現場は多種多様な検査業務に加え、ISO15189 維持活動やタスクシフト・シェアなどの病院組織への貢献など、膨大な業務を抱えている。データサイエンスは臨床検査の新しい価値を創出するとともに、業務効率化の一助となる可能性がある。臨床検査を含む医療データをそのように活用するかは、日々大量のデータに直接的に触れている臨床検査技師の重要な役割になることが考えられる。

連絡先 058-230-7251