

臨床検査領域におけるデータサイエンス教育の重要性と課題

片岡 浩 巳*

要 旨 現代の医療分野では、データサイエンスの知識と技術の修得が不可欠であり、医療従事者がデータサイエンス技術を駆使することで医療の質の向上が期待される。データサイエンス教育には統計学、機械学習、データベース管理、プログラミングが含まれ、実際の医療データを用いた演習やプロジェクトへの参加が重要である。臨床検査領域では、バイオマーカーの探索やAI技術の活用が期待され、診断精度の向上や検査の効率化を進めていく必要がある。これには、匿名化された大規模なリアルワールドデータを収録したデータベースの整備、そして、標準化とデータの精度管理が重要であり、これにより異なる医療機関間での情報共有が容易になる。教育プログラムの見直しと、データサイエンスを基盤とした教育と技術開発の推進により、医療の質の向上が期待される。本稿は、臨床検査領域におけるデータサイエンス教育の重要性と課題を紹介し、医療現場での実践に役立つことを目指している。

キーワード データサイエンス教育、臨床検査技師育成、リアルワールドデータ

はじめに

現代の医療分野において、データサイエンス領域の知識の修得は不可欠であり、それぞれの対象領域の専門知識を持つ人が、直接的にデータサイエンス技術を駆使して問題を解決できるスキルを持つことで、大きな飛躍を遂げることが予想される。また、日常業務を担う、すべての医療従事者がデータを適切に収集・解析・活用できるスキルを持つことが望まれる。

データサイエンス教育は、統計学、機械学習、データベース管理、プログラミングなどのトピックが含まれているのが一般的であるが、実際の医療データを用いた演習やプロジェクトに参加して

経験を積むことも重要と考える。これにより、医療従事者はより高度なデータ解析を行うことができるようになり、臨床現場での意思決定のサポートや医療の質を向上させることができる^{1)~3)}。

特に、臨床検査領域では、多くの革新的な技術や手法の利用が期待されている。バイオマーカーの探索は、ゲノム、プロテオーム、代謝物などのバイオマーカーを用いて、疾患の早期診断や治療法の個別化を進めることが可能となり、さらなる医療の質が向上できる。また、AI技術の活用も臨床検査において重要な役割を果たす。画像解析、パターン認識、自然言語処理などのAI技術を活用することで、診断精度の向上や、臨床検査の効率化、そして、精度向上が期待できる。

* 川崎医療福祉大学医療技術学部臨床検査学科 kataokah@mw.kawasaki-m.ac.jp

データサイエンス技術を用いた研究を支えるには、匿名化された大規模なリアルワールドデータ (RWD) を収録したデータベースの整備が必要であり、さらに、データの精度は、診断や治療、そして研究の質に直結するため、高精度な検査技術の開発と普及、そして、徹底したコード管理が求められる。このコード管理は、各種コードの標準化により、検査結果の一貫性が確保され、異なる医療機関間での比較や情報共有が容易となる。臨床検査の役割も、大きな変革が必要であり、それらを支える教育の大幅な見直しも必要となっている。データサイエンスを基盤とした教育と技術開発の推進により、医療の質がさらに向上することが期待できる。

本稿では、臨床検査領域におけるデータサイエンス教育の重要性と臨床検査領域の課題と取り組み事例について紹介する。

I. データサイエンス教育の国内外の取り組み

1. 日本の取り組み

日本政府は、AI リテラシー教育の推進と医療DX (デジタルトランスフォーメーション) の推進を強力に進めている。特に内閣府では、2016年に第5期科学技術基本計画として Society 5.0⁴⁾が提唱されている。これにより、AIをはじめとする、データサイエンス教育の枠組みが構築され、大きな変革が訪れようとしている。教育プログラムの開発と普及も進められており、以下のような具体的な取り組みが行われている。

A. 小中高等学校におけるAI教育の導入

AI リテラシーを育むため、小中高等学校のカリキュラムにプログラミング教育が組み込まれ、早期からAIやデータサイエンスの基礎知識を習得することが期待できる。

B. 大学および専門学校での高度な教育

大学や専門学校においては、データサイエンスやAIの専門教育が強化されようとしている。例えば、医療系大学では、医療データサイエンス専攻やデジタルヘルスケアに関するコースが設置され、学生が実践的なスキルを習得できるよう支援されはじめている。

C. オンライン教育プログラム

社会人や現役医療従事者を対象にしたオンライン教育プログラムが拡充されつつある。これにより、現場で働きながら新しいスキルを習得することが容易になっている。例えば、厚生労働省や経済産業省は、データサイエンスとAIに関するオンライン講座を提供し、広範な層にリテラシー向上の機会を提供している。

2. 日本と米国のデータサイエンス教育の比較⁵⁾⁶⁾

A. カリキュラムの柔軟性と幅広さ

多くの米国の大学や教育機関では、データサイエンスや関連分野に特化した専門的な学位プログラムが提供されている。カリキュラムは幅広く、コンピュータサイエンス、統計学、機械学習、ビジネスアナリティクスなどの要素を組み合わせているのが特徴である。一方、日本の教育システムでは、従来の分野との垣根がまだ厚く、データサイエンスを包括的にカバーするプログラムは限られているのが現状である。また、カリキュラムの柔軟性や実践的な要素の欠如が課題となっている。

B. 実践的な学習機会

米国では、実践的な学習を重視したプロジェクトや産業界との連携を通じて、実際のデータ解析や問題解決に取り組む機会が提供されている。日本の教育機関では、実践的な学習機会が不足しており、理論的な知識に偏る傾向が見受けられる。このため、実務で必要とされるスキルが大幅に不足しているとの指摘がある。

C. 産業界との連携

産業界との密接な連携を通じて、実務に必要なスキルや知識を教育プログラムに反映させる取り組みが盛んである米国に比べて、日本の教育機関と産業界との連携は十分ではなく、産業界のニーズに合ったカリキュラムや実務経験を積むための機会の提供が不足している。

II. 臨床検査領域に求められるデータサイエンス教育とは

臨床検査領域では、病歴、薬歴、検査歴等の膨大な医療データを取り扱う点と臨床側が要求するデータ解析スキルを持つことが重要である。これ

には、従来型の統計学、機械学習、データベース、プログラミング技術に加えて、実践的かつ統合的なデータ解析能力とチームとしてプロジェクトを推進できるプロジェクトベースの学習が必要と考えている。

1. 統計学

統計学はデータ解析の基礎であり、医療分野におけるデータサイエンスにおいても不可欠な知識である。データのサンプル数を考慮した実験計画法の立案、データの分布による解析手法の選択、データのまとめ方など、基本的な統計指標の理解は医療データの分析において重要である。さらに、仮説検定や回帰分析などの統計手法を学ぶことで、医療データから正確な結果を導き出す能力を養うことができる。

2. 機械学習

機械学習は、大量のデータからルールやパターンを自動的に見つけ出し、予測や分類を行う技術である。医療データに基づく診断支援や治療予測のために、医療従事者は機械学習の基本概念(教師あり学習、教師なし学習、強化学習など)を理解し、実際のデータに適用する方法や結果の解釈方法を学ぶ必要がある。

3. データベース

データベース管理は、医療データの効果的な保存、検索、更新を行うための技術である。データベースを取り扱うためには、関係データベースの基本概念、データの正規化、SQL (Structured Query Language) を用いた問い合わせ文(クエリ)の書き方などの知識を持つことで、様々な課題を容易に解くことが可能となる。多くの臨床研究の問い(Research Question)や診療の質指標(Quality Indicator)の課題は、SQLを用いてデータの切り出しや集計を行うことで解決できる場合が多い。

4. プログラミング

プログラミングは、データ解析を自動化し、効率的に行うためのスキルである。医療従事者は、RやPythonなどのデータ解析に適したプログラミング言語を学び、データの前処理、解析、可視化を行う能力を身に付ける必要がある。これにより、大規模なデータ処理や繰り返し計算が必要な解析

課題でも、自分自身で実行できるようになる。

5. 実践的かつ統合的なデータ解析

実際の医療データを用いた演習は、理論だけでなく実践的なスキルを身に付けるために重要と考える。医療従事者は、電子カルテデータや臨床試験データを使って、データのクレンジング、統計解析、機械学習モデルの構築を実践することによりデータ解析の全プロセスを理解し、臨床現場でのデータ活用能力を高めることができる。

6. プロジェクトベースの学習

プロジェクトベースの学習は、チームで協力して実際の課題に取り組む形式の学習方法である。バイオマーカーの探索や診断支援システムの利用などの具体的なプロジェクトを通じて、データサイエンスの技術を実践的に学ぶべきである。この方法は、実際の業務や研究の場に即したスキルを習得するのに有効と考える。

III. 臨床検査領域のデータサイエンス教育の現状と取り組み

臨床検査領域に求められるデータサイエンス教育の理想像と現実的なカリキュラム枠の制限を加味しながら工夫した、川崎医療福祉大学での取り組みについて紹介する。

1. 旧カリキュラムと新カリキュラムの比較(表1)

旧カリキュラムでは、基礎教育科目(選択)として数学、統計学、コンピュータ基礎演習、コンピュータ活用演習が含まれ、専門教育科目(必修)としては、情報科学概論(1年秋[後期]、8回)、情報科学演習(1年秋[後期]、15回)、医療統計学(1年秋[後期]、7回)、検査総合管理学実習(3年病院実習期間内2週間)で行われていた。一方、新カリキュラムでは、基礎教育科目の選択科目にデータサイエンス入門が追加された。さらに、医療統計学演習(3年春[前期]、15回)も新たに加えられ、検査総合管理学実習は期間が2週間から3週間に拡大することにした。検査総合管理学実習では、6人の小グループ班で自ら設定した臨床研究の問いに対して、実践的かつ統合的なデータ解析を行う。1年生の医学知識のない時期にリテラシーレベルの学習を行う旧カリキュラムから、

表1 データサイエンス関連教科の旧・新カリキュラムの比較

		旧カリキュラム			新カリキュラム		
基礎 教育 科目	数学	1年15回	選択	数学	1年15回	選択	
	統計学	1年15回		統計学	1年15回		
	情報リテラシー	1年15回	選択	情報リテラシー	1年15回	選択	
	コンピュータ基礎演習	1年15回		コンピュータ基礎演習	1年15回		
	コンピュータ活用演習	1年15回		コンピュータ活用演習	1年15回		
			データサイエンス入門	1年15回			
専門 科目	情報科学概論	1年8回	必修	情報科学概論 (医療統計一部吸収)	1年15回	必修	
	医療統計学	1年7回	必修				
	情報科学演習	1年15回	必修	情報科学演習	1年15回	必修	
				医療統計学演習	3年15回	必修	
	検査総合管理学実習	3年2週	必修	検査総合管理学実習	3年3週	必修	

ある程度の医学知識が身についた3年生の段階で、実践的なデータサイエンス関連の授業内容を盛り込む改正を行った。このカリキュラム改正により3年次の病院実習時点で、以下のような効果が認められた。

- EXCELのみでなく、R言語やSQLを駆使して短時間に課題解決ができる。
- SQLを用いて病歴、薬歴、検査歴を自由に結合して解析ができる。
- R言語を用いた適切な作図や統計的な検定が可能。
- プログラム作成により網羅的解析を可能とし、特徴の探索が可能。
- 解析結果の考察について学生同士で議論し、深め合うことができる。
- 臨床データを読む力と疑問に対する適切なデータ処理が行える能力の向上。

これにより、学生はより実践的で高度なデータサイエンスの技術と知識を身につけ、問題解決ができるようになった。

2. 医療データを用いた実践的な演習

実践的かつ統合的なデータ解析とプロジェクトベースの学習の双方を兼ね備えた検査総合管理学実習の詳細の取り組みの詳細を紹介する(図1)。実習班は、学生6人の小グループで構成され、自ら設定したテーマに向けて、RWDを用いたデー

タ解析を実施する。データ解析の補助的な情報検索が可能な、検査診断特性ネットワーク図を瞬時に作成できるシステム⁷⁾⁸⁾を利用して各自の課題に則した結果を探索する。次に、探索結果と教科書や論文、あるいは、ガイドラインの情報とを比較して考察することで、疑問点を明確にすることができる。検査診断特性ネットワーク図作成システムは、病名を指定すると、その病名に関連した検査診断特性に優れた検査項目セットの関係を作図できる。また、検査項目を指定すると、それぞれの検査項目に関連する検査診断特性(AUC)の高い検査項目のネットワーク図を作成することができる。さらに、各症例の検査値の問い合わせに対しては、尤度比やオッズ比による評価を短時間で行うことができる。これらの探索機能を用いて生じた疑問点を新たなテーマとして、明確な目的を設定し、目的に応じたデータ切り出しを行った後、各自のテーマに対する解析を実施する。

3週間の実習期間の最後には研究発表としてまとめ、今後の課題として残った疑問は、4年生の卒業研究テーマとして、さらに掘り下げることにした。

IV. データベース整備の課題と対策

1. データベースの整備

大規模なリアルワールドデータ(RWD)を収録

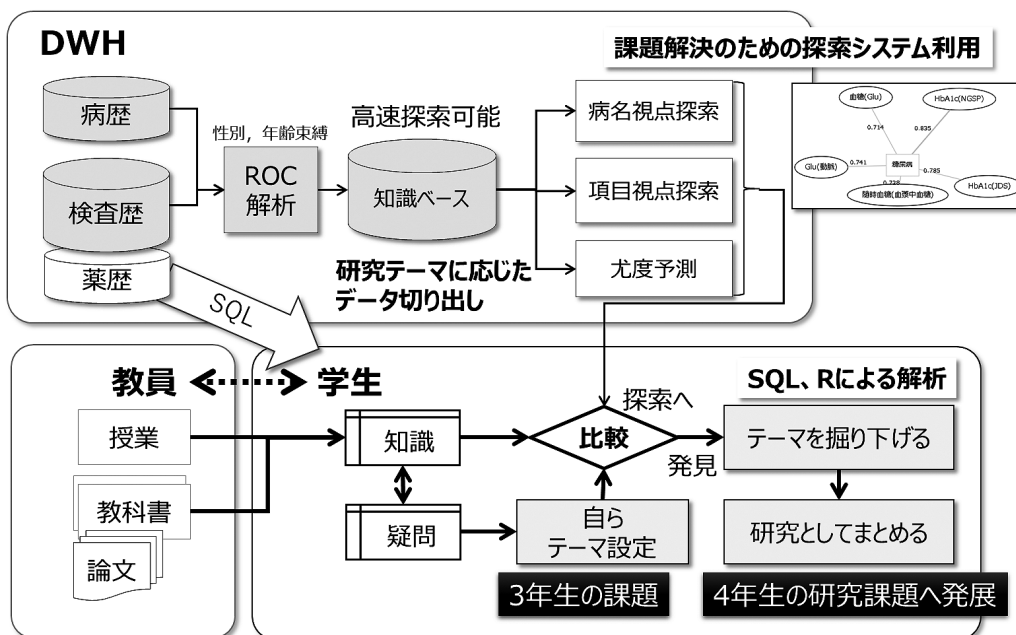


図1 病院実習におけるデータ解析の取り組み

したデータベースの整備は、データサイエンス研究を支える基盤となる。病歴、薬歴、検査歴を基にデータ解析に適した構造のデータウェアハウス (DWH) を作成した。このDWHは、高度なクエリ処理能力と柔軟なデータ集約機能を備えており、研究者がSQLを用いて迅速かつ効率的にデータ解析できるように設計している。

データベース整備の課題としては、日常診療から得られたデータを常に最新の状態に保ち、解析可能な形式に変換する必要がある。また、個人情報保護のための匿名化作業も定期的に行う必要があるため、これらの作業に精通した人材が必要である。

2. データの精度と標準化

データの精度は、診断や治療の質に直結するため、データ管理においては測定法や単位の違いを正確に管理できる体制が望まれる。これらの管理には、検査情報システム (LIS) や病院情報システム (HIS) に加え、政府が推進する全国の医療機関が診療情報を共有できる電子健康記録 (EHR) に必要不可欠な標準コード (JLAC11) へのマッピングを行う一連の仕組みを理解する必要がある。

演習用のDWHでは、これらのマスタ管理の重要性を理解させるために、測定法や単位の違いによるデータのクレンジング作業を前処理として学生に実施させるようしている。これにより、学生は日常業務での部内コード管理や、データの精度に関する重要性を理解することができる。

3. 技術と倫理のバランス

患者データのプライバシー保護、意思決定支援ツールの適切な使用など、倫理的課題を考慮した教育基盤の整備が重要である。

RWDをデータサイエンス教育に利用するためには、技術的な知識だけでなく倫理的な理解も必要である。この問題に対し、個人が特定できる患者番号等は、ハッシュ化を行い仮名化後、番号の再割り振りを行うことで匿名化を行った。さらに、日付は、各個人について誕生日を起点とした相対日付に置き換えることで、検査日等の情報から個人の特特定ができないように工夫している。これらの匿名化加工されたDWHは、データのプライバシーを確保しつつ、正確なデータ解析が可能となっている。

おわりに

データサイエンスの基礎知識と実践的なスキルを持つことは、現代の医療現場において不可欠である。特に臨床検査の領域では、バイオマーカーの探索やAI技術の活用による診断精度の向上が期待されている。データの標準化と共有、教育プログラムの構築を通じて、医療従事者がデータサイエンス技術を駆使して医療の質を向上させることが可能である。本稿を通じて、データサイエンス教育の重要性と臨床検査の未来についての理解が深まり、医療現場での実践に役立つことを期待する。

文 献

- 1) Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC Med Educ* 2023; 23 (1): 689.
- 2) Enticott J, Johnson A, Teede H. Learning health systems using data to drive healthcare improvement and impact: A systematic review. *BMC Health Serv Res* 2021; 21 (1): 200.
- 3) Bote-Curiel L, Muñoz-Romero S, Gerrero-Curieses A, Rojo-Álvarez JL. Deep learning and big data in healthcare: A double review for critical beginners. *Applied Sciences* 2019; 9 (11): 2331.
- 4) Society 5.0 とは, 内閣府, https://www8.cao.go.jp/cstp/society5_0/
- 5) 佐藤 貴海. 日米データサイエンティスト教育の違い. *日本ソーシャルデータサイエンス論文誌* 2019; 3 (1): 23-6.
- 6) Takayama K. A comparativist's predicaments of writing about 'other' education: A self-reflective critical review of studies of Japanese education. *Comparative Education* 2011; 47 (4): 449-70.
- 7) 片岡浩巳, 畠山 豊, 奥原義保. 医療意思決定支援データベースおよび医療意思決定支援方法. 特開2013-93019 (P2013-93019A).
- 8) 片岡浩巳. EBLM (evidence based laboratory medicine) の新展開: 医療ビッグデータの活用—医療ビッグデータ解析法. *臨床検査* 2020; 64 (5): 592-9.