

## 医療情報の収集から活用、人材育成が必要—東京大学大学院医学系研究科准教授の今井健氏 講演レポート



東京大学未来ビジョン研究センター(旧・東京大学政策ビジョン研究センター)、慶應義塾大学メディカルAIセンター、エムスリー株式会社m3.com編集部が2019年1月から開催している「医療×AIセミナーシリーズ」のイベントレポートです。

2020年3月26日、医療×AIセミナーシリーズ第11回シンポジウム「医療現場で本当に価値あるAIを作るために」がオンライン配信形式で行われた。主催は東京大学未来ビジョン研究センター、慶應義塾大学メディカルAIセンター、エムスリー株式会社m3.com編集部。協力は日本ディープラーニング協会(JDLA)、世界経済フォーラム第四次産業革命日本センター(C4IRJ)。

臨床現場でのAI実装を進めるため、AI・ICTの現状の情報共有、課題整理を通じて、医師ら医療関係者、開発者、利用者、政策関与者らステークホルダーの交流をはかってきた医療×AIセミナー。第11回となる今回のシンポジウムでは、医療現場でのAIの開発・実装を行うエンジニアや医療情報関係者、医療現場でのAI活用人材育成に取り組むステークホルダーが参集して講演。話題提供とエコシステムの形成や臨床実装に向けてパネルディスカッションが行われた。

### 医療情報学と臨床でのAI活用

東京大学大学院医学系研究科准教授の今井健氏は「医療情報学と臨床でのAI活用について」と題して講演。医療AIの開発と臨床応用を進めている今井氏は、「医療情報学は広大な領域。医療における情報処理過程全てを扱う学問だ」と紹介した。いま、病院で運用されている電子カルテシステムの実運用を通じてデータが集まりつつあり、それを利活用しようという動きが起きている。医療情報を収集し活用するためにはデータの標準化が必要だ。また、収集されたビッグデータを使って現場や政策に還元させる必要もある。それら全体が、医療情報学が扱う領域だという。

病院情報システムは複雑だ。あらゆる場所に情報が散っている。様々なサーバ群に散っている診療データには、コード化／構造化されているデータ、波形・画像・ゲノム情報、自然言語情報などがある。構造化データは研究利用が容易だが、自然言語情報などの非構造化データは取り扱いが難しい。これらをどうやって標準化して様々な施設から収集するのか。わが国では「SS-MIX2」標準ストレージと拡張ストレージという規格が策定されている。標準ストレージには患者基本情報／処方・注射オーダー／検査オーダーなどの構造化情報が含まれている。いっぽう、波形・画像・ゲノム情報や自然言語情報は拡張ストレージの情報で、収集の有無や格納形式は病院によって異なる。

これまで、SS-MIX2標準ストレージの情報や診療報酬データの集積・利活用が進められている。「MID-NET (Medical Information Database Network)」は2018年度から本格運用されている。またデータベースの特性に違いはあるが同様に「レセプト情報・特定健診等情報データベース (NDB)」も構築されている。これらのデータベースによって、様々な臨床疫学的知見が得られている。だが、数が多いがアウトカムについてのデータは限定的で、入院前の経緯など詳細はわからない。そのため、AI活用推進のためには拡張ストレージがカバーする情報の活用が不可欠だと考えられている。

拡張ストレージ向け情報の収集例として、今井氏はがんゲノム医療の現状を紹介した。症例登録においては、がんのパネル検査データ、臨床データ、薬剤応答性、副作用、併存する疾患、生活習慣データなどを収集する。それらは標準ストレージにはないので、電子カルテに構造化テンプレートを埋め込んで入力してもらって収集している。つまり拡張ストレージ情報の活用は今後必須だが、非常に面倒だということだ。また、標準化のさらなる推進、保険医療共通の1人1番号の整備、電子カルテの特性の理解などが進まないとい保険医療データベースは有効な知識源とならないという。

たとえば、特定の疾患や状態をもつ患者集団を自動的に抽出することすら現実的には難しい。電子カルテはそもそも網羅的に書かれているわけではない。正確性の問題もある。自然言語の理解も大きな課題だ。記載のバイアスの問題もある。今井氏は、それらの特性・取り扱い方の理解と、その人材育成が重要だと指摘した。

2000年代以降、ビッグデータ時代が到来した。ディープラーニングなどの医療分野への応用も進んでいる。いっぽう、電子カルテデータからディープラーニングを使うとどんなことができるのか。2018年にGoogleはカリフォルニア大学サンフランシスコ校、スタンフォード大学医学部、シカゴ大学病院と共同で、ある発表を行った。通常は必要な前処理なしでディープラーニングを使って解析を行い、院内死亡、30日以内再入院、7日以上長期入院などを従来の確率モデルより精度高く予測することができたというものだ。これはすごい結果だったが、同時に限界を示すものであると今井氏は感じたと述べ、「どうしても医学知識自体の構造化と整備が必要になる」と指摘した。診療における知識はパターン認識やシンボル処理だけではなく様々な知識を駆使することで形成されている。現状の機械学習アルゴリズムがカバーできているのは一部でしかない。

東大病院では診療ガイドラインからの知識抽出とそれをういた高血圧の診療支援ツール、臨床医学知識の構造化とそれをういた診断支援、統合病態遷移マップなどの研究開発を進めている。機械学習などをういた医療データ解析と相補的な柱として、知識データベースは必ず役にたつ。一方で、知識データベースは更新が必須だ。症例報告など要約された診療情報から機械学習で自動的に学ぶことができるようになれば、知識更新も無理なくできるようになるのではないかと考えており、研究を行っている締めくくった。



森山和道  
サイエンスライター

サイエンスライター、科学書の書評屋。1970年生。広島大学理学部地質学科卒。NHKディレクターを経て現職。現在、科学技術分野全般を対象に取材執筆を行う。特に脳科学、ロボティクス、インターフェースデザイン分野。研究者インタビューを得意とする。