

医療に応用されるAI（人工知能）の現状と課題 —IBM「ワトソン」とディープラーニングを中心に

執筆者 KDDI 総合研究所 リサーチフェロー 小林雅一

▼記事のポイント

<サマリー>

近年、様々な分野に普及している AI（人工知能）が、ついに私達の健康や命を預かる「医療」の分野にまで進出しようとしている。既に IBM の AI スパコン「ワトソン」が難病患者の命を救ったり、先端 AI「ディープラーニング」が画像診断などで専門医に勝るとも劣らない能力を示しつつある。

こうした医療用 AI はインドやアフリカのように、膨大な人口の割には医師の数が足りない国や地域で「医療の新たな担い手」として期待されている。また少子高齢化による労働力人口の減少が進む日本でも、特に医師不足が深刻な地方部で、医療用 AI の必要性は今後高まっていくであろう。

ただ、こうした AI は開発段階における使用テスト等では素晴らしい性能を示すが、いざ病院やクリニックなど医療現場で使おうとすると、現場ならではの複雑な使用環境に足を取られて本領を発揮できない。ここに見られるように、実用化に際しての課題も幾つか浮かび上がってきている。

また医療用 AI は患者の「カルテ（診療記録）」や「ヘルスケア・データ」などを機械学習することで性能が磨かれるが、これら扱いに細心の注意を要する個人情報漏洩など、患者のプライバシーが侵害される恐れも指摘されている。

本レポートでは「人の命」という究極の領域に踏み出した AI の現状と課題、リスクなどを探っていく。

<主な登場人物>

IBM グーグル（アルファベット） ディープマインド

<キーワード>

AI 人工知能 医療 ワトソン エキスパート・システム ディープラーニング ニューラルネット

<地域>

米国 日本 アジア 英国 世界

Current Status and Issues Related to the Application of AI (Artificial Intelligence) to the Health Care Sector —The Cases of IBM "Watson" and Deep Learning

Masakazu Kobayashi

KDDI Research Institute Research Fellow

Abstract

In recent years AI (Artificial Intelligence) has been put to use in a wide variety of fields; and as the technology makes inroads into the field of "medical care" it will now start to impact upon our health and wellbeing. IBM's AI supercomputer "Watson" has already saved the lives of patients with intractable diseases, and advanced AI "Deep Learning" is showing an ability comparable to specialists in a number of areas, including in diagnostic imaging. Accordingly, this type of medical AI is expected to become the "new leading force in medical care" in countries and regions where the number of doctors is insufficient for their large populations, such as in India and Africa. Even in Japan, where labor force shrinkage is due to the declining birthrate and aging population, the need for medical AI will increase in the future, especially in rural areas where there is a serious shortage of doctors.

Although such AIs show excellent performance in tests at the development stage, this is challenging to replicate in practice because health care work sites, such as hospitals or clinics, are inherently complex use environments with unique characteristics at each site. Even in these early development stages, there are some emerging challenges in the commercialization phase. Medical AI has the potential to put its machine learning capacity to high performance use with regards to a patient's "medical chart" (medical records), "health care data" and associated data keeping tasks. However, widespread concerns about access to a patient's personal information, particularly the leakage of personal information, require careful attention to ensure that patient confidentiality is maintained. In this report, we will investigate the current situation, challenges, and risks of AI present as it enters this critical domain of "human life."

Key Players

IBM, Google (Alphabet), Deep Mind

Keywords

AI, artificial intelligence, medical care, Watson, expert system, deep learning, neural network

Region

USA, Japan, Asia, UK, World

1 はじめに

スマートフォンの音声操作、自動運転、対話型スピーカー、掃除ロボット、機械翻訳、宅配ドローン、ヘッジファンドの資産運用システム……。様々な領域でAI（人工知能）の導入が進んでいるが、そのフロンティアとして今、一際注目を浴びているのが、医療分野におけるAIの活用だ。

米国では2018年8月、連邦政府におけるヘルスケア分野の規制当局FDA（食品医薬品局）が、世界初と見られる医療用AI機器の販売を許可した¹。

この製品は眼科クリニックなどで、糖尿病患者のかかり易い眼疾患をAIが自動診断する装置だ。同様の技術は（囲碁AI「アルファ碁」の開発で有名な）英ディープマインド（米アルファベット傘下）をはじめ世界中のIT企業や大学などが、ここ数年、盛んに研究開発を続けている（詳細は後述）。

また、より包括的な医療用AIとしてはIBMの「ワトソン（Watson）」がよく知られている。元々、米国の人気クイズ番組に出演するAIコンピュータとして開発されたワトソンだが、その後、産業各界における商用化が図られ、中でも（製薬産業なども含めた広義の）医療分野への応用が最も大きな期待を集めている。そこでは「癌診断における医師の支援」や「新薬開発」など、多方面にワトソンの活用が図られている。

他にも様々な種類の医療用AIが存在するが、いずれも人の命や健康を左右するだけに誤作動が許されないなど、その技術開発に課せられた条件は極めて厳しい。しかし医療現場で実際に、これらのAIを使ってみた医師らの反応は、必ずしも肯定的とは限らない。

以上のように実用化が迫ってきた医療用AIの現状とそこに見られる問題、さらに利用に伴う懸念や危険性などを以下で詳しく見ていくこととする。

2 医療用AIとは何か

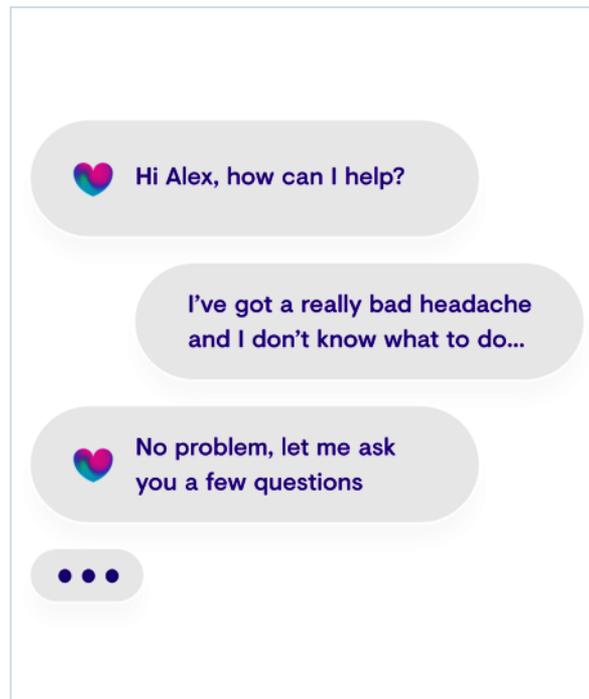
まず医療用AIとは具体的に何を指すのか？—この点を整理しておくことから始めよう。

（前述の）病院やクリニックなど医療機関で使われる、AI搭載の自動診断機器やAIコンピュータなどは医療用AIの代表である。しかし一般に医療用AIとは、それらに限定されるものではない。たとえば私達自身による日頃のヘルスケアやそれに対する医学的助言なども含め、広い意味での医療行為に使われるAI全般を指す。

¹ <https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357>

英国のITベンチャー「Babylon」が開発したスマホ・アプリでは、ユーザーがチャット画面から健康について気になる点を入力すると、それに仮想ドクター（会話型AI）が応えてアドバイスを返してくる（図1）。こうした手軽な製品（サービス）も医療用AIの一種と見られている。

【図1】ユーザーがスマホ画面から仮想ドクター（会話型AI）とチャットする様子



出典：Babylon¹

Babylonのスマホ・アプリは「AIによる自動診断」のような本格的な医療サービスではなく、あくまで日常的なヘルスケア・サービスという位置付けだ。従って仮にチャット画面から入力された情報によって何らかの病気の可能性が検知された場合、アプリ内の仮想ドクターはユーザーに（本物の）医師の診断を受けることを勧め、そのためにクリニックの受診予約などの手続きをオンラインで行ってくれる。

他にも米ベンチャー企業「Qventus」が提供する「病院の業務管理にAIを導入したシステム」なども、ある意味では「医療用のAI」と見ることができるだろう。このAIは電子カルテなど病院内のデータを効率的に管理し、医師や看護師らが従来と同じ時間で、より多くの患者に対応することを可能にする。

ただ、（上記の）「スマホ・アプリを介したヘルスケア」や「病院の業務管理」などは、広範囲に渡る医療行為の中であって、どちらかと言えば周辺的なサービスに位置付けられる。逆に言えば、医療行為の中心にあるのは、やはり医師による病気

¹ <https://www.babylonhealth.com/product/ask-babylon>

の診断や（看護師らも含めての）その治療と見ていだろう。

実際、患者の立場から見れば、AIによる自動化のインパクトが最も大きいのは「病気の診断とその治療行為」であることは間違いない。仮に、こうしたAIが（人間の）医師を凌ぐ正確さで病気の診断を行い、従来の医師には思い付かなかった画期的な治療法を提示してくれるのであれば、患者の延命・治癒率が高まるからだ。

しかし逆に、こうしたAIが誤診をしたり、誤った治療法を提示した場合、それは患者に文字通り致命的な結果をもたらす恐れがある。これらの点を考え合わせると、様々な医療用AIの中で最も重要性が高く、今後の医学界に与える影響が最も大きいのは「病気の診断や治療を自動化する本格的な人工知能」であろう。

以下、本稿では、こうした本格的な医療用AIに焦点を絞って、その実態を見ていく。

3 ワトソンとは何か

現在、患者の病気を診断したり、その治療法を提案する医療用AIの中で、世界的に有名なものは（冒頭で紹介した）IBMの「ワトソン」であろう。

ワトソンは元々、（米ニューヨーク州に位置する）「トーマス・J・ワトソン研究所」などIBMの研究部門を中心に、2007～2011年にかけて開発された。それは人間の言葉を理解して操ると共に、その能力を使って私達からの様々な質問に答えることが可能とされる。いわゆる質疑応答用のコンピュータとして開発されたのだ。

そもそもIBMは何故こうした風変りなコンピュータを作ろうとしたのだろうか？それは当時、米国で50年以上もテレビ放送されてきた人気クイズ番組「ジョパディ (Jeopardy!)」にワトソンを出演させ、そこで伝説的な二人の歴代チャンピオンと対戦させるためだ。

もしも番組でワトソンが彼ら（人間の）クイズ王に勝つことができれば、IBMは長年の研究で培ってきた「自然言語処理」をはじめ、AI関連の技術力を世間にアピールできるというわけだ。つまりワトソンは元々、商用目的ではなく、一種のデモンストレーション用に開発されたAIコンピュータだった。

ジョパディにおける歴代チャンピオン対ワトソンの戦いは、2011年2月14、15日の両日に渡って、（前述の）トーマス・J・ワトソン研究所内に設けられた特設会場で行われたが、終わってみればワトソンの圧勝だった。この対戦の様子は全米に生放送され、「自社の技術力を広くアピールする」というIBMの狙い通りの結果となった。

この勝利から間もなく、様々な企業から「ワトソンを自分たちのビジネスや業務に使えないか」という問い合わせや要望がIBMに多数寄せられた。ここから約3年の歳月をかけて、同社はワトソンをビジネスに応用するための検証作業を幾つもの企業や大学と共同で進めた。そして2014年1月、事業化の目途が立ったとして、IBMは

ワトソン事業部（正式にはCognitive Solutions事業部）を発足した。

以来、ワトソンがカバーするビジネス分野は、「金融」「ヘルスケア」「メディア」「製造」など多数の業種に及ぶ。中でもスタート時から大きな期待を集め、今でも同事業部の中心に位置するのが「(医療をはじめとする)ヘルスケア」だ。これは「ワトソン・ヘルス (Watson Health)」という商品名で提供されている。

ワトソン・ヘルスはその実用化の過程で、多数の医師による、一種の教育プロセス（カスタマイズ）を経て商品化された。

その典型が「癌医療 (oncology)」への応用だ。IBMは2012年から（ニューヨークにある癌専門の医療・研究機関）メモリアル・スローン・ケタリング・キャンサー・センター (MSKCC) と提携。ここに勤務する研究医らがIBMの技術者と共同で、ワトソンを「各種癌の診断・治療」を行う医療用AIへとカスタマイズしていった。

やがて、この作業を完了したワトソンは、2015年頃から世界各国の病院へと試験的に投入され、少なくとも当初は周囲の期待にたがわぬ高い能力を示した。

米New York Times紙の報道によれば、最初期にワトソンを導入した米ノースカロライナ大学・医学大学院では、各種癌の症例1000ケースをワトソンに入力したところ、その99%でワトソンが提示した治療法は癌専門医によるものと一致した。そればかりか全体の30%では、ワトソンは医師が見落としていた別の治療法も発見したという¹。

日本では、さらに劇的な事例が報告された。難病で死の危険にさらされた女性をワトソンが救ったというのだ²。

命を救われたのは、「骨髄異形成症候群」で東京大学病院に入院していた66歳の女性。この病気は血液癌の一種で、症状が進行すると骨髄性白血病に至る。それまで彼女は複数の抗がん剤を投与されたが、いずれも効果がなく、一時は死を覚悟したとされる。

そうした中、2015年7月に東大病院が臨床研究の一環としてワトソンを導入。この女性患者の「ゲノム (DNA配列)」は既に検査で判明していたので、東大の研究チームはこのデータをワトソンに入力した。すると10分程度でデータを解析し、(医師らが想定していたのとは) 別の白血病を発症している可能性を提示した。

この情報を参考に医師が抗がん剤を変更したところ、てきめん効果が表れ、女性の病状は急速に回復。それから約2か月後に退院することができたという。

1

<https://www.nytimes.com/2016/10/17/technology/ibm-is-counting-on-its-bet-on-watson-and-paying-big-money-for-it.html>

2 <http://www.itmedia.co.jp/enterprise/articles/1610/28/news062.html>

<https://www.nikkei.com/article/DGXLZO05697850U6A800C1000000/>

4 多数の国で使われる医療用ワトソン

以上のように、2015年頃から試験的に導入が始まったワトソンは当初、物珍しさも手伝って高い評価を受けるケースが目立った。しかし、それから数年が経過した後、ワトソンは（それを本格的に使ってみた）医療関係者の間で、どのように評価されているのだろうか？

これについては、米国のバイオ・医療関連のメディアとしては最も信頼されている「STAT¹」や米ウォールストリート・ジャーナル紙²などが、かなり詳細な調査報道を実施している。

それによれば医療用ワトソン（以下、ワトソン）は米国、インド、韓国、台湾、スロバキアなど数多くの国・地域で使われている。これらの国で、ワトソンを導入した病院の数は総計200以上に上るといふ。

ワトソンはインターネットを経由したクラウド・サービスとして提供される。たとえば癌医療用のワトソンは「Watson for Oncology」、ゲノム医療（患者のDNA情報に基づく新型医療）用のワトソンは「Watson for Genomics」など別々のサービス（製品）として提供される（この2つ以外にも幾つかのサービスが用意されている）。

ワトソンの使用料金は、そのサービスの種類や数に応じて、患者1人当たり（米ドル換算で）200～1000ドルと開きがある。ワトソンを導入する病院では、主に（前述の）癌医療やゲノム医療用のワトソンを使うケースが多いという。

ワトソンは、病院側がIBMとワトソンの利用契約を結んだ時点で、すぐに使える状態にはなっていない。ワトソンを使用可能な状態にするためには、その準備段階として病院側のスタッフ、つまり医師やITシステム担当者らが共同で（日本では「カルテ」と呼ばれる）多数の患者の「診療記録（medical record）」をワトソンに入力しておく必要がある。

これは病院側が事前に予想していた以上に負担のかかる作業であったという。その主な原因は、（元々、医師という人間が作成した）診療記録の乱雑性や不規則性にある。つまり病院に蓄積されている膨大な診療記録には、しばしば略語が使われていたり、異なる医師によって異なる表記スタイルが使われていたり、ときには記載ミスなどもあるので、これらを全部修正して、ワトソンが処理できる形の標準フォーマットに統一するだけでも大変な作業になるからだ。このため病院側では、導入したワトソンを医師が使える状態にするまでに、長い時間を要する事が多いという。

¹ <https://www.statnews.com/2017/09/05/watson-ibm-cancer/>

²

<https://www.wsj.com/articles/ibm-bet-billions-that-watson-could-improve-cancer-treatment-it-hasnt-worked-1533961147>

5 ワトソンの原理的限界

このような作業を完了して、使用可能になったワトソンに対する医師の評価は複雑だ。

ワトソンの主な役割は医師のアシスタントだ。ワトソンは医師に、病気の診断や治療法に関する「アドバイス (recommendations)」を提供すると同時に、その根拠となる過去の医学論文や結論に至るまでの思考経路なども参考情報として提示する。

ここで医師がワトソンの性能を評価する際に、主要な評価基準となるのが、ワトソンによるアドバイスが医師自身による見立て（診断、治療法）と一致する程度だ。これについては、（ワトソンを試験的に導入した）デンマークの某病院では、その一致率が僅か33%に留まったため、協議の末、ワトソンの導入を見送ったという。つまり「ワトソンのアドバイスは信頼性に欠ける」と判断したのだ。

逆に、他の病院ではワトソンのアドバイスと医師の見立てが一致する割合が96パーセントに達したケースもあるという。しかし、このように一致率が高ければ、ワトソンに対する医師の評価も高いかということ、必ずしも、そうではないという。

なぜならワトソンが提示した診断結果や治療法が医師の見解と一致した場合、多くの医師は「そんなことは、お前から言われなくても分かっている」という感情をワトソンに対して抱いてしまうからだ。

逆にワトソンと医師の見立てが相反した場合、医師は「お前の言っていることは本当に正しいのか？信用して大丈夫なのか？」という不信感をワトソンに抱いてしまう。つまり、どっちに転んでもワトソンには分が悪いのだ。

もちろん、仮にワトソンのアドバイスが医師の見立てと食い違っても、ワトソンが自らの正しさを立証するために、目から鱗が落ちるほど斬新な発想や鮮やかな思考プロセスを補足情報として提示してくれば、医師も納得することができるかもしれない。そもそも病院がワトソンの導入を決めた主な理由は、このAIコンピュータが、医師という人間では到底叶わない程の独創的な診断や治療法を提案してくれるのを期待してのことだ。

ところがワトソンの原理上、それは不可能なのだという。なぜならワトソンが医師に提供する診断や治療法は元々、医師（人間）がワトソンに入力したものであるからだ。（前述のように）ワトソンを癌医療用にカスタマイズ（教育）する際には、癌専門の医療研究機関である「MSKCC」の医師団が長い時間をかけて、「患者の症状が、かくかくしかじかの場合には、このように診断し、このような治療法を提示せよ」という教えを、大量にワトソンに入力していた。

つまりワトソンの本質を比喩的に表現すれば、「MSKCCの医師団が丸ごと（ワトソンという）コンピュータに詰め込まれたもの」である。ワトソンが提示する診断や治療法は元々、彼ら医師（人間）によるものである以上、それが人間を超える独創的なアイデアや思考プロセスを提示することは有り得ない——（前掲の）米国メディアの報道は基本的に、こうした見方で統一されている。

実は、この種の限界は1970～80年代の「第二次AIブーム (AIバブル)」が弾けて、周囲の期待が失望へと転じた頃、既に指摘されていた。当時のAIは「ルールベースのAI」、中でも「エキスパート・システム」と呼ばれる方式が主流だった。この種のAIは、たとえば医師のような各界エキスパート (専門家) の知識やノウハウをプログラミング言語でルール化してコンピュータに移植すれば、このコンピュータ (AI) が専門家に代わって仕事をしてくれる、という基本思想に基づいて開発されていた。

これは現在のワトソンと共通するところがある。もちろんワトソンが今から40年以上も前のエキスパート・システムと全く同じものであるとは考え難いが、優に半世紀以上に及ぶAI研究の成果を蓄積してきたIBMが、一種古典的なAI方式に改良を施して現代に蘇らせたことは十分考えられる。実際、(後述するように) その節は随所に見られるのである。

6 米国での評価は今一つ

以上のような (癌医療用) ワトソンに対する評価は国や地域によって異なる。

前述の米メディア報道によれば、米国での評価は今一つのようなようだ。たとえばフロリダ州の先端医療病院「Jupiter Medical Center」で、73歳の肺癌患者に対しワトソンは「化学療法が妥当」とのアドバイスを提示したが、これは既に同病院の医師も考えていたことなので、例によって「まあ、いいか (It's fine)」という冷めた評価しか貰えなかった。

とは言え、ワトソンはそうした治療法を提示するに至った理由として、これを裏付ける医学専門誌の論文・記事なども提示してくれる。このため医師は自分の治療法に自信が持てると同時に、その治療法に関する最新情報などが得られる場合もある。この点でワトソンは、米国の医療関係者から相応の評価を受けている。

ここに見られるように、ワトソンにはMSKCCの医師達が入力した診断・治療法と共に、(ワトソンを癌医療用にカスタマイズした初期段階で) 42種類に渡る癌専門誌から約200万ページもの医学論文が入力されており、これら専門情報はその後も適宜更新されている。

ワトソンは医師にアドバイスを提供する際、これらの医学文献を検索し、実際の患者の症状に関係すると見られる情報を医師に示すのだ。医師という人間の記憶力には限界があるが、桁違いの記憶容量と高速プロセッサを併せ持つワトソンなら、ひとりの人間ではカバーし切れないほど大量の医学情報を蓄積し、必要に応じて、すぐに検索できるというわけだ。

しかし、ここには大きな問題、あるいは課題が存在する。それは「これらの医学情報は常時更新されなければならない」ということだ。世界中の大学や医療機関などでは激しい研究競争が昼夜繰り広げられおり、癌に関する医学論文だけでも年間20万本以上が発表されている。従ってワトソンの記憶装置に蓄積されている学術誌

の論文・記事など医学情報も、これら最新の研究成果を常に反映して更新しなければ、すぐに時代遅れになってしまう。

これは単に医学論文だけの話ではない。MSKCCの医師団が当初ワトソンに入力した診断・治療法についても同じことが言える。たとえば「転移性の肺癌」に関する治療ガイドラインは、癌に関する国際学会で何らかの研究成果が発表されてから僅か1週間後に変更されることもある。このように目まぐるしく変わる情報を、(ワトソンの運用担当者らが) 更新し続けるのは容易な事ではない。

この結果、特に米国のような医療先進国では、「ワトソンが提供する治療法や医学情報は最新とは言い難い」と失望する医師が少なくないという。

実は、この問題も1970～80年代のルールベースAI、つまりエキスパート・システムが廃れた大きな要因として指摘されていた。当時も、各界専門家のノウハウなどを蓄積した「データベース(AIが従うべきルール集)」を常時更新する必要に迫られ、そのために「知識エンジニア」と呼ばれるエキスパート・システム専門要員を(同システムを導入した) クライアント各社に常駐させる必要があった。

ところが、そのための人件費が膨大な金額に膨らんでしまったため、「こんなことなら、AIで仕事を自動化するより、いっそ本物の専門家(人間)を雇った方が安上がりだ」という認識がクライアント企業をはじめ産業界全体に広がった。これがエキスパート・システム失敗の一因となり、その後の「AIの冬」と呼ばれる停滞期を招くことにつながった。現在のワトソンも本質的には、当時のAIと同様の問題を抱えているようだ。

7 ワトソンに見られる偏向とは

以上のような米国とは対照的に、ワトソンが非常に高い評価を受けているのは、アジア圏を中心に医療に関して比較的、開発途上にある国だ。具体的にはインド、あるいは(後述するタイなど一部例外を除く) 東南アジア諸国の病院に勤務する医師らは、「ワトソンのお蔭で時間を節約し、より多くの患者に質の高い医療サービスを提供できるようになった」と口を揃えて言う。

中でも中央アジアに位置するモンゴルのケースは際立っている。その首都、ウランバートルにある韓国系の総合病院「UB Sondo Hospital」に導入された(癌治療用)ワトソンは、医師たちの間で引っ張りだこだ。

なぜなら同病院には癌専門医が一人もいないからだ。ワトソンが導入されるまで、一般医療に従事する医師らが癌患者に対応していた。彼らは癌に関する専門知識を全く持たないか、あるいは極めて限られた知識しか有していない。

このためワトソンを導入して以降、同病院の医師らはワトソンのアドバイスにほぼ100パーセント従っているという。前述のようにワトソンには、「MSKCC」という世界でも屈指の癌研究機関による診断・治療法や専門知識が詰め込まれている。

従って、同病院で癌患者を診断・治療する医師らにしてみれば、（癌については、ほぼ素人とも言える）自分の見立てに従うよりはワトソンのアドバイスに従う方が、余程安全で確かな診療ができるというわけだ。

他方、同じアジア圏でも、韓国や台湾、あるいはタイなど医療水準の比較的高い国・地域では、ワトソンに対して複雑な受け止め方をしている。なぜならワトソンはMSKCCなど米国の研究機関によって医療用にカスタマイズされたため、それが提示する診断・治療法などは、どうしても米国の患者向けにバイアス（偏向）がかかってしまうからだ。

ごく単純なケースでは、患者の体格の違いが挙げられるだろう。白人やアフリカ系の人達が多く住む米国では、患者の体格は平均してアジア圏の患者よりも大きくなる。このため同じ薬を投与する場合でも、（ワトソンのアドバイスに従って）米国人に対する投与量をそのままアジア圏の患者にも適用すると、過剰投与に陥る危険性がある。

こうした問題を未然に回避するには、アジア圏の病院の医師らが「薬の投与量」をはじめワトソンの提示する治療法を、アジア人の患者に向けて「ローカライズ（現地化）」する必要に迫られる。

実際、タイの首都バンコクにある総合病院「Bumrungrad International Hospital」では、一旦導入したワトソンを（同じアジア圏に属する）日本の医療ガイドラインなどを参考にローカライズしたという。

ところが、折角このように再調整したワトソンを同病院の医師らは、それほど使っていないという。その主な理由は、医師らがワトソンをローカライズする過程で、日頃自分たちが地元患者を診断・治療する際のノウハウや医学知識をワトソンに入力していったことにある。

この結果、生まれ変わったワトソンには、彼らタイ人医師による診断・治療情報が詰め込まれている。彼らが患者を診療する際にワトソンを使っても、彼ら自身が予め入力した情報が返ってくるだけだ。これでは確かに使う意味があまり感じられない。

8 ワトソンのメリットとは

以上のように、本格導入から数年を経たワトソンに対する、世界の医療関係者の見方は総じて辛口だ。しかし今後の可能性まで含めると、これを高く評価する専門家も少なくない。彼らはワトソンに代表される医療用AIが「科学的エビデンス（証拠）に基づく新たな医療を実現し、これまでのヒエラルキー（階級）的な医療現場を改革できる」と期待を寄せる。

韓国の大学病院「Gachon University Gil Medical Center」でワトソンを利用する医師は次のような事例をあげて、その可能性を説明する：

これまでの病院に見られる階級的な医療現場では、常に年長で地位の高い医師の意見が、若くて地位の低い医師の意見よりも優先される傾向があった。しかし今後は若手の医師でも、ワトソンを使って最新の医療情報を入手し、それを会議等で示すことで「こちらの治療法の方が患者を助けるためには効果的です」と説得できる。

そのように科学的な根拠を示されれば、上司に当たる医師も皆の前で自分の意見をゴリ押しできないだろうと言う。

またインドの総合病院チェーン「Manipal Hospitals」の医師によれば、同病院に入院している大腸癌と乳癌患者に対し、ワトソンは僅か数秒で最適な治療法を提示することができた。これらの情報は通常、同病院に勤務する20人の癌専門医が、約1週間に及ぶ会議を経て導き出す結論であるという。

現在のワトソンには「医学情報を常に更新する必要性」や「情報のバイアス」など課題が少なくないが、これらが解決された暁には、AIによる医療現場の民主化や効率化など大きなメリットが期待される。

9 ニューラルネットの医療応用

医療用AIの開発動向において、(ワトソンと並ぶ) もう一つの大きな焦点は「ニューラルネット (neural network)」の導入だ。

ニューラルネットとは1950年代に米国を中心に研究開発が始まったAIの一種で、私達人間や動物の脳を構成する無数の「ニューロン (神経細胞)」と「シナプス (接合部)」など基本的な仕組みを参考に考案された。その後、実用的には近年まで大した成果は見られなかったが、2006年頃から「ディープ・ニューラルネット」あるいは「ディープラーニング (深層学習)」などと呼ばれるようになって以降、画像認識や音声認識をはじめ多方面で長足の進歩を遂げ、一大ブームを巻き起こした¹。

この分野をリードする企業の一つが、2010年に英国で起業し、その後グーグル(アルファベット)に買収されて、その傘下に入ったディープマインドだ。ディープマインドは囲碁用のAI「アルファ碁」を開発し、これが2016~2017年にかけて韓国や中国のトップ棋士を次々と打破したことで世界的に知られるようになった。

この技術開発と半ば並行する形で、同社はアルファ碁の基盤技術であるディープラーニングを医療分野へと応用する研究開発を進めていた。一般にディープラーニングが得意とするのは、「パターン認識」と呼ばれる作業である。パターン認識とは、現代社会に溢れる大量データ、いわゆる「ビッグデータ」を解析し、ここから何等かのパターン(模様、傾向、様式など)を見出す作業だ。

¹ <https://gendai.ismedia.jp/articles/-/35512>

パターン認識の一例は画像解析 (画像認識) である。ごく身近な例で説明すれば、フェイスブックのようなソーシャル・メディアに時々刻々とアップロードされる写真のような画像をコンピュータ、つまりAIが自動解析し、その画像に何が写っているかを識別する (=ある種のパターンを認識する) 作業が画像解析だ。そのためにフェイスブックもディープラーニングを採用している。

ディープラーニングが持つ、この画像解析 (パターン認識) の能力を医療に使うとすれば、最も単刀直入な応用事例は「画像診断」であろう。画像診断とは、患者の患部を撮影した X 線写真、あるいはMRIやCTスキャンなどによる断層画像を (少なくとも、これまでは) 専門医が目視して、様々な病気を診断する作業だ。ここにディープラーニングを導入すれば、こうした画像診断が自動化できると考えられている。

実際、医療分野への進出を図るディープマインドが真っ先に手掛けたのも、画像診断にディープラーニングを応用することだった。2016年、同社は英国やインドにある幾つかの病院と提携し、「糖尿病網膜症」など眼疾患を、ディープラーニングで画像診断する医療システムの開発に着手した¹。

糖尿病網膜症とは糖尿病患者の約3分の1が発症する病気で、早期に発見すれば治療で進行を抑えることができるが、発見が遅れると失明に至る深刻な病だ。特にインドでは近年の急速な経済成長に伴い、栄養過多による糖尿病の患者が急増しており、その数は現時点で少なくとも6000万人に上る。しかし眼科クリニックやそこで働く医師の数が不足しているため、糖尿病網膜症の発見が遅れて失明する患者が多いと言われる。

そこでディープマインドはこうした進行性の眼疾患を (医師、つまり人間に代わって) ディープラーニングのようなAIに診断させることで、インドの医師不足に対処しようというわけだ。同社は2016年、提携する英国の病院から全部で約12万8000枚もの網膜画像を取り寄せた。

ここで網膜画像とは、眼科医が「光干渉断層計 (OCT)」など専用装置を使って、(眼疾患の) 患者の眼底を撮影することで得られる「網膜の断層画像」のことだ。(前述の) 英国の病院には、長年の眼底診断から大量の網膜画像が蓄積されている。

1

<https://www.technologyreview.com/s/602958/an-ai-ophthalmologist-shows-how-machine-learning-may-transform-medicine/>

10 教師有り学習のプロセス

これらの病院からディープマインドが取り寄せた12万8000枚の網膜画像は、ディープラーニングの「トレーニング・セット」として使われる。

ディープラーニングとは、ディープ（多層）・ニューラルネットを使った機械学習の一種だ。そこではニューラルネットに大量のデータ（ビッグデータ）を入力し、これを消化（学習）させることで、ニューラルネットがある種のパターンを認識できるようになる。この学習用データがトレーニング・セットである。

一般にニューラルネットによる機械学習には「教師有り学習 (supervised learning)」と「教師無し学習 (unsupervised learning)」の2種類があるが、教師有り学習に使われるトレーニング・セットでは、予め「ラベル付け」と呼ばれる作業が必要になる。

これは、ある分野における専門家（つまり人間）がAIに対する一種の教師役となって、（機械学習の準備段階として）トレーニング・セットを構成する大量のデータに「これはA」「これはB」「これはC」・・・といった形でラベルを付けていく作業だ。

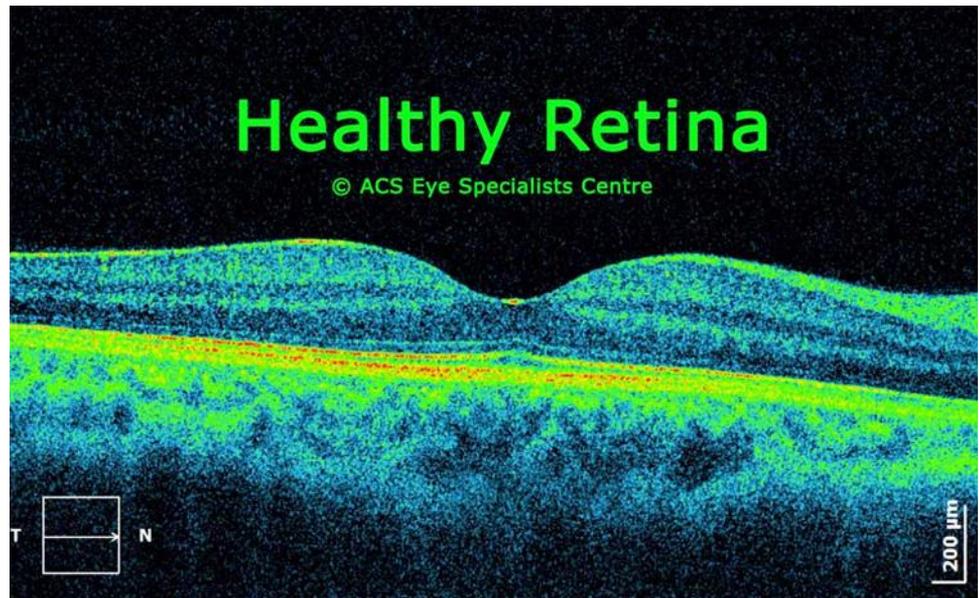
ディープマインドが開発する医療用ニューラルネットも「教師有り学習」を採用しているため、トレーニング・セットのラベル付け作業が必要となる。このための教師役に選ばれたのが3人の眼科医だった。彼らが12万8000枚の網膜画像を一々、目視で診断して、「これは加齢黄斑性変性」「これは糖尿病網膜症」「これは健康な目」・・・といった形で、個々の網膜画像（データ）にラベル付けしていった。

このようにして用意された大量の（ラベル付き）網膜画像が、この医療用ニューラルネットのトレーニング・セット（学習用データ）となる。これを入力されたニューラルネットは、大量の網膜画像を解析することによって、各々の病気を示す画像上の「特徴点 (features)」を自動的に識別できるようになる。

たとえば「健康な目」の網膜（図2）は滑らかなカーブを描いているのに対し、「糖尿病網膜症」（図3）では網膜表面の「不自然な段差」や網膜内部の「黒い影」などが見て取れる。これらが断層画像上の特徴点となる。

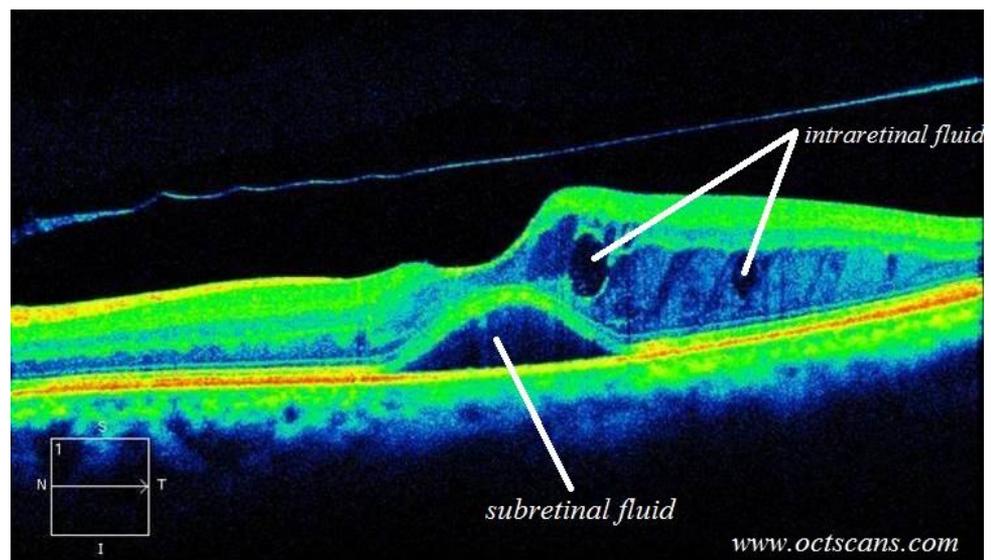
トレーニング・セットとして入力された大量の網膜画像（断層画像）を次から次へと解析することによって、このニューラルネットは糖尿病網膜症ならではの特徴点を徐々に学習していく。この学習を完了したニューラルネットは、次からはラベル付けされていない網膜画像を入力されても、「これは糖尿病網膜症」などと判定できるようになる。

【図2】 OCT (光干渉断層計) で撮影した、健康な目の網膜画像 (断層画像)



出典 : http://eyespecialist.name/english_OCT.html

【図3】 糖尿病網膜症の断層画像



出典 : <https://www.octscans.com/diabetic-retinopathy.html>

以上のプロセスがディープ・ニューラルネットによる機械学習、つまりディープラーニングである。このような機械学習を経て、ディープマインドが開発する医療用AIは、いよいよ実際の患者 (の網膜画像) を診断できるようになる。

それに先立ち、同社が約1万2000枚の (ラベル付けされていない) 網膜画像から成るテスト・セット (システムの性能を評価するためのテスト用データ) を入力した

ところ、全体の90%以上で経験豊富な眼科医と同等、あるいは彼らを凌ぐ高い精度で糖尿病網膜症を診断することに成功した。

このAIが実用化された暁には、それをインドの主要都市にある病院等のサーバーにインストールしておけば、クラウド・コンピューティングによって（眼科クリニックや病院の存在しない貧困地域でも）多数の糖尿病患者が網膜症の診断を受けることができるようになる。

11 実用化に際しての課題

ディープマインド（グーグル）は2019年1月、開発が一段落した（上記）眼底診断用ニューラルネットを「ARDA:Automated Retinal Disease Assesment（網膜症の自動診断）」と命名して公開。これを米ウォールストリート・ジャーナル紙が報じた¹。

それによれば、ARDAは（前述のテスト・セットを使った）研究開発段階における性能評価では、本物の眼科医に勝るとも劣らない診断能力を示す。ところが（利用地域として想定されている）インド国内の貧困地域にある医療現場では全く使い物にならなかったという。

その原因は画質の違いにある。ディープマインドが当該ニューラルネットのトレーニング・セット、つまり学習用データとして使った12万8000枚の網膜画像は、比較的新しいOCTによって撮影された高画質の断層映像だった。これに対しインドの貧困地域にあるクリニックや病院等で撮影された網膜画像は、旧式の検査装置による低画質の映像だった。ディープマインドの眼底診断システムはこれら低画質のデータを全く受け付けなかったため、こうした貧困地域ではAIシステムの性能評価以前に、そもそも、それによる画像診断ができなかったのだ。

今後の実用化に際しては、このように思わぬ障害が立ち塞がる事が予想される。理想的な開発環境と、理想とは乖離した医療現場との間にある、複雑な溝を埋めていくことが大きな課題となる。

12 プライバシー漏洩の懸念

以上のような眼底診断に加え、医療用ニューラルネットはMRIやCTスキャンなど、より適用範囲の広い画像診断にも応用可能で、既にディープマインドをはじめ世界中の企業や大学などがその研究開発を進めている。

¹

<https://www.wsj.com/articles/googles-effort-to-prevent-blindness-hits-roadblock-11548504004>

さらにディープマインドは、ある人が発症しそうな病気を予知・予防するための医療用ニューラルネットの開発にも力を注いでいる¹。

このシステムでは、電子カルテ等に記載されている「身長」、「体重」、「血液・尿検査の結果」など多彩なヘルスケア・データをディープラーニングで分析し、様々な病気の発症確率を弾き出す。

ただし、そこには深刻な問題も指摘されている。ディープラーニングが画像診断や病気の予知・予防を行えるようにするためには、前述のように「トレーニング・セット」と呼ばれる大量のデータをシステムに入力して、これを機械学習させる必要がある。

このためにディープマインドは2016年に英国の「国民健康保険サービス (NHS)」と提携し、ロンドン市内にある3つの国立総合病院から、約160万人の患者に関する広範囲の医療ヘルスケア・データを取得・利用する契約を交わした。これを医療用AI、つまりディープラーニングの学習用データ (トレーニング・セット) として使おうというわけだ。

ところが、これら大量の医療データの中には、病院を訪れた受診者らの「HIV陽性」「薬物過剰摂取」、そして「中絶」に関する情報など、極めて扱いが微妙な個人情報が含まれていたため、ディープマインドと (親会社の) グーグルは英米の主要メディアから手厳しい非難を浴びた。

これに対しディープマインド (グーグル) は「我々が病院から取得した医療データは匿名化されており、個人の身元が判明することはない」と反論したが、これまでプライバシー関係の摩擦を引き起こすことが多かったグーグルだけに (たとえば Google Street View など)、この程度の弁明では焼石に水だった。

また近年のフェイスブックによる大量の個人情報流出に見られるように、彼らIT企業が個人データを適切に管理する保証はない。特に上記ディープマインドのケースでは、一際慎重な扱いを要するヘルスケア・データであるだけに、医療関係者らが懸念するのも止むを得ないと見る向きが多い。

13 医療用AIが暴走する恐れ

医療用ニューラルネット (ディープラーニング) には、別の側面からも問題が指摘されている。それはディープラーニングの「ブラックボックス化」と「暴走」の懸念だ。

これが顕在化した象徴的なケースは、2016年に韓国ソウルで開催された囲碁AI「ア

¹ <https://www.technologyreview.com/s/602963/is-deepminds-health-care-app-a-solution-or-a-problem/>

ルファ碁」とトップ・プロ「李世ドル九段」の対戦だ。その第4局の最中、李九段が打った奇手が功を奏し、そこからアルファ碁は意味不明の悪手を連発して敗れた。

（前述のように）このアルファ碁を開発したのがディープマインドだ。第4局対戦後の記者会見で、「もしもアルファ碁のAIが医療に応用され、それが今回の対戦のように暴走したら、どうなるのか？」という質問が記者から投げかけられたが、ディープマインドの関係者は質問を事実上受け流して、本質的な答えを返そうとはしなかった。

その理由は現在のディープラーニングがあまりにも複雑化して、内部の情報伝達ルートを外から推し量ることが出来なくなってしまったからだ。つまり一種のブラックボックス化してしまったため、暴走の原因を究明できないのだ。

この問題に対処するため、米DARPA（国防高等研究計画局）をはじめ世界各国で「Explainable AI（説明可能な人工知能）」の研究開発が進んでいるが、それは未だ十分な成果を上げるまでには至っていない。

14 まとめ

「ワトソン」や「ディープラーニング」など、先端的なAI技術が医療に応用されつつある。それはモンゴルなど、高度医療に携わる専門医の数が不足している国で既に重用されている。今後、日本でも労働力の高齢化などから医師不足が深刻化していくに連れ、医療用AIは必須の技術となっていく。それはまた高騰する医療コストの抑制・削減にも寄与するだろう。

このように大きな期待がかかる医療用AIだが、既に研究開発段階における性能評価では、一部で人間の医師に勝るとも劣らないレベルに達している。しかし実際の医療現場では、たとえば「医学情報を常時更新することが難しい」あるいは「断層映像の画質が低い」など思いがけぬ問題・障害に妨げられることもある。

さらに「医学情報の偏向」「ブラックボックス化による暴走」あるいは「プライバシーの侵害」など、安全性や倫理面を中心に数々の問題や懸念も抱えている。純粋に技術的な面だけに集中できる開発環境と、経済的要因などにも左右される複雑な医療現場との間にある溝を埋めていくことが、今後の実用化に際して大きな課題となる。

また医療関係者の理解を得ることも必要だろう。彼らの多くは医学や生物学に関する専門家だが、AIに関しては概ね素人なので、その機能や効用を理解してもらうことが医療用AIの普及に向けた第一段階となる。

現時点で様々な問題や短所が指摘される医療AIだが、これから加速する人手不足、中でも医師の超過勤務や過重労働などが懸念される中、その解決策として、これに勝る技術は当面見当たらない。普及期に見られる問題点の解決に時間を要する覚悟の上、医療用AIの育成に努めていく必要があるだろう。

【執筆者プロフィール】

氏名：小林 雅一（こばやし まさかず）

所属：KDDI総研

専門：ITやバイオなど先端技術の調査研究

経歴：東京大学大学院理学系研究科を終了後、雑誌記者などを経てアメリカに留学。ボストン大学でマスコミ論を専攻し、ニューヨークで新聞社勤務。慶應義塾大学メディア・コミュニケーション研究所などで教鞭をとった後、現職。

主な著書：

『ゲノム編集からはじまる新世界 超先端バイオ技術がヒトとビジネスを変える』（朝日新聞出版）

『AIが人間を殺す日 車、医療、兵器に組み込まれる人工知能』（集英社新書）

『ゲノム編集とは何か 「DNAのメス」クリスパーの衝撃』（講談社現代新書）

『AIの衝撃 人工知能は人類の敵か』（講談社現代新書）

『クラウドからAIへ アップル、グーグル、フェイスブックの次なる主戦場』（朝日新書）、ほか多数。