資料６.「精神疾患の治療に機械学習の予後予測を使用することは倫理的か？」（AMA Journal of Ethics 誌・2018 年）

アブストラクト

機械学習は、早期介入の機会、潜在的治療反応、予後、健康への影響など、臨床上の変数を予測する一つの方法である。このコメンタリーでは、新たな精神疾患が発症した患者に関して（その予後を推測するために）機械学習を用いる際の、次のような倫理的課題を検証する。(1) 臨床におけるイノベーション（新しいが十分確立していない手法）はいつ倫理的に受け入れられるか？（2）臨床医は、機械学習予測モデルにより生じる倫理的問題について患者とどのように共有すべきか？

ケース

K 医師は、学術医療センターの入院患者の精神科病棟の常勤精 Q 神科医である。K 医師は、新たに精神疾患を発症したとして、救急部門から入院を認められた患者を診療する役割を担っている。これらの患者に関する大きな課題は、臨床医が個々の患者の臨床的な帰結（アウトカム）を予測できないということである。標準的なベースラインまで回復する患者、軽度の症状だけを経験する患者もおれば、一方では、悪化して、重度障害に至る可能性もある。

K 医師は、最近発表された、精神疾患に関する初期の兆候を示した患者を対象として用いる疾患予測モデルの試行に関心を持っている。このモデルは、ヨーロッパの精神疾患患者に関する大規模なデータベースの情報を用いた機械学習によって開発されたものであり、診療に役立つかもしれない次の 2 つの情報を臨床医に提示するものである。

第一に、そのモデルは患者の予後を予測する。性別、職業的地位、主要な抑うつ症状に関する病歴といった患者の基本的情報を用いて、このモデルは患者の一年後の転帰が良好か不良かを予測する。転帰が良好か不良かは、患者の全体的機能状態を定量化するために有効な方法である機能の全体的評定（GAF）によって定義されている。転帰良好は（GAF で 65 点かそれ以上と定義されている）典型的に、患者が最小の障害で機能できることを示している。転帰不良（GAF で 65 点未満）は、障害の重度が大規模であることを示す。転帰不良の上限の GAF では（50－60 点）患者は中程度の障害を社会的、職業的機能において経験する可能性がある。転帰不良の下限の GAF では（0－10 点）、患者は自分の衛生状態を処理できないか、恒常的に自殺願望を有する恐れがある。注目すべきは、このモデルが約 75%の精度で転帰が良好か不良の結果を予測するということである。

第二に、このモデルは一部の患者について、治療の選択肢の方針案を示す。精神科医は、精神疾患を治療するさまざまな抗精神疾患薬を入手できるが、転帰不良と予測された患者の場合、(他剤に比して)アミスルプリドまたはオランザピンの効き目が高い。

K 医師と同僚は、このモデルは精神疾患患者の治療を改善するため、臨床に取り入れたいと考えている。しかし、こうした予測した予後について患者側に情報提供すべきかどうか迷っている。この情報は患者や家族が将来についての計画を立て、また意思決定をするための助けとなるかもしれない。一方、この情報が益よりもむしろ害になることはないのか、またそのような害が起こりうるのはどのような場合なのか、はかりかねている。

コメンタリー

K 医師は、精神疾患の発症早期(初回エピソード)の患者を対象として、ヨーロッパの多施設および多数の患者から得られた大規模なデータセットをもとにした機械学習による疾患予測モデルの試行を考えている。

機械学習は、アルゴリズム構築のために用いられる技術であり、学習を重ねてさらに改善される性質のものである。アルゴリズムを用いて、パターン決定や未来の結果を予測する大量のデータセットを解析することができる。機械学習は、予測、診断、治療の改善を通じて、精神医学を大きく進歩させることが期待されている。

上述のシナリオは、機械学習技術の場が研究段階から臨床現場へ移行する際に生じる、倫理的ないくつかの検討課題を示している。このケースにおけるモデルは、統計的には立証されているが、患者の症状の改善につながるような臨床的な介入の手段としてはいまだに確立していない。

以下では、まず、K 医師がこの臨床的なイノベーション(新しいが十分確立していない手法)の実践が倫理的に正当化されるか否かを検討する。その後、疾患予測のアルゴリズムの対象集団(すなわち精神疾患の症状を有する患者)について、インフォームドコンセントに関する特別に考慮すべき倫理的問題が生じるかどうかを論じる。

臨床医は、いつ臨床上の新手法を導入すべきか?

予測モデルの K 医師による試験的運用は、「臨床イノベーション」と考えることができる、すなわち、標準的な手法に比して新しいが、臨床的に明確に優れているとは証明

されていない介入あるいはモデルの実践である。これらの活動は、「一般的な臨床行為」と「研究段階の行為」との中間段階に位置付けられるものであり、（ベルモント・レポートにあるように）倫理的に注意すべき状況である。こうした臨床イノベーションの実践は、どのような条件が満たされれば、倫理的に正当化されるのだろうか？

まず、この段階の行為をおこなう必要性が明確なものでなければならないだろう。精神疾患は、その症状の影響を被る人にとって、個人的、社会的、経済的な負担を伴う。精神疾患の症状が最初に現れてから回復に至る割合は、それが通常の臨床的対応を受けたとしても、数十年の間、変化はない（10%-15%）。

最初の発症以降の適切な治療や抗精神疾患薬の投与は症状を改善できるかもしれないが、医薬品への耐性、治療方針へのアドヒアランス、回復効果は患者によって大きな差がある。発症した疾患が重症化しうること、そして適切な治療が欠如している現状を併せて考えるならば、上記のようなイノベーションを用いることの必要性は高いといえるだろう。

次に、イノベーションによりもたらされるリスクが基本的条件のリスクと関連して倫理的に受け入れられるか我々は考察しなければならない。まず、K 医師は、提案されたイノベーションが約束された利益を提供できるという十分な証拠を必要とする。提案された精神疾患予測モデルの正確性が、ヨーロッパで行われた研究によって支持されるものである一方で、それらが前提とした「地域的な変数」（精神科診療や社会的支援の相違など）が、K 医師の臨床現場にも当てはまるかどうか、患者集団の結果を改善するモデルの妥当性や性能に適用可能かどうか分かっていない。モデルは、関連する地域的変数を説明するために調整される必要があるだろう。機械学習アルゴリズムの「ブラックボックス」的性質のため、ソフトウェア開発者は、システムが決定へと至るにおいてどのように入力データが寄与したか、把握しているとは限らず、あるいはそもそも理解もしていないかもしれない。したがって、システム設計者は、更新されたモデルを検証するためにどの変数を確認する必要があるか、確たることは分からないだろう。調整・補正（「キャリブレーション」）のために、各臨床現場からの追加の患者データが必要になる。

キャリブレーションには、臨床現場ごとの違いだけではなく、誤差や偏りも考慮に入れなければならない。機械学習は、しばしば人間の判断よりも客観的であるとして提示されるが、操作上のエラーに弱い。例えば、誤ったデータが入力されると、間違った分析結果がもたらされる。機械学習のアルゴリズムは、データに含まれるそれまでの固定観念をさらに際立たせてしまう可能性もある。例えば、アルゴリズムが社会経済的背

景や人種を考慮する構成になっている場合、こうしたアルゴリズムの判定結果には、弱い立場にある患者の「社会的な欠陥」が意図せずとも強調されることになるかもしれない。一方、適切なキャリブレーションによって、医療の偏りを少なくするためにアルゴリズムを用いることもできるかもしれない。

最後に、疾患予測モデルの使用そのものが、患者の受けるケアに影響し、精神科医の治療判断や資源の配分に影響を与えるだろう。最初のうちは、予測モデルが治療に及ぼす効果は、十分に考慮されないかもしれない。疾患予測モデルの実践が、倫理的に許されるようなリスクと恩恵のバランスのうえに立って展開されるためにも、臨床医は、それぞれの臨床現場で、その予測アルゴリズムが有用であることの検証に積極的に関与する必要がある（アルゴリズムの方式が特定の患者集団に適合することを確認し、予測から治療へ移行するための手順を検討することを通じて）。

一方、アルゴリズムを使用する医師にとって、アルゴリズムが生成する予測の背後の変数や論拠は難解で、選択された治療方針を自身で評価したり正当化したりすることが難しいこともあるかもしれない。疾患予測モデルの使用が正当化されるためは、機械学習につきものの、透明性の向上や結果の偏りに関する課題に対応することが必要であるが、それには、機械学習システムの作動方法について医師を訓練するなどの戦略を実施することや、モデルが特定の予測をした理由がわかるような機械学習システムの導入を推進することが考えられるだろう。機械学習システムを使用する臨床医は、その構造、これらのアルゴリズムの判定の土台となったデータセットとその限界についてより深く知っている必要がある。

対象集団へのインフォームドコンセント

疾患予測モデルの使用に関する信頼と透明性を確保するため、インフォームドコンセントに関する倫理的課題に注意を払う必要がある。今のところ、患者のデータを予測アルゴリズムにあてはめたり、その改善のために使用したりすることについて、インフォームドコンセントは特に必要とはされていない。さらに、患者は通常の診療において、医師がコンピュータに基づく意思決定支援を使用するのに気づかないだろうし、医師の判断に影響した情報源を聞かされることもほとんどない。これらの事実から次の質問が生じる。精神疾患予測モデルなどの機械学習予測アルゴリズムには、これら従来とは異なるアプローチを必要とするような、新しい倫理的課題があるのだろうか？

機械アルゴリズムが、心臓発作や脳卒中のリスクを評価するのに使われるような既存のリスク評価ツールとどう異なるか。この問いは、それが治療に及ぼす影響と関係するだろう。医師が診断及び治療決定の判断をするために、こうした機械学習アルゴリズ

ムを重用するほど、こうしたアルゴリズムは単なる支援ツール以上のものになってくるかもしれない。

さらに、機械学習システムが医療の中で恒常的に使用されるにつれ、予測モデルにもとづく治療や資源配分に関する判断は、治療を行っている医師ではなく、病院運営者が設定した規則や診療プロトコルによって左右されるかもしれない。従って、機械学習ツールは、患者との関係において、医師の役割を大きく変えてしまうかもしれない。機械学習システムが医療とさらに一体化すれば、医療施設における患者と機械学習による意思決定システムとの関係における「信頼」について注意深い検証が必要となるだろう。

患者の治療のための決定や資源の配分に影響を与えることは、医師患者関係に変更を迫るものになるかもしれないし、電子記録と機械学習が一体となることは同じく医師患者関係の間での守秘に影響を及ぼすことになるかもしれない。この場合、患者には自分の医療施設における疾患予測アルゴリズムの使用について知らされているべきだろう。患者には、機械学習が自分の治療に影響を与え得る方法、自身の情報の秘密確保、データのプライバシーについて考慮する十分な情報が必要となる。我々としては、患者のデータが予測アルゴリズムを作成あるいは改善するのに使用され得ること、さらに疾患予測ツールが患者の治療において役割を果たすかもしれないことについて、該当する患者に一定の注意情報を提供しておくべきであると我々は提言する。

初期精神疾患のケースでは、患者やその家族が、精神疾患のため患者が入院するときに、深刻な病気に関連した、そして心理社会的なストレスに常に向き合わなければならないことも考えられるだろう。その際、予測アルゴリズムの使用などのような、複雑な情報を消化し保持することがさらに困難になっている可能性があるため、いつ患者に情報提供するかについては検討のうえで判断がなされる必要がある。コミュニティの利害関係者からは、患者と家族と有意義な形で連携できるよう、こうした情報提供の内容及び手順を整備する方法について、アイデアを得ることができるかもしれない。

疾患予測アルゴリズムが示した「予後」は、治療に関するインフォームドコンセントの中に含めて患者に示されるべきだろうか？インフォームドコンセントは、治療に関する提案について、その詳細の「すべて」の説明を必須とする訳ではないが、患者に実施されうる治療オプションの性格、そのリスクや利点についての情報は、その患者に伝えられる必要がある。したがって、アルゴリズムが示した治療の選択肢についての情報に関するコミュニケーションを患者側と行うために、臨床医には、機械学習システムに関する十分な教育を受けることが必要となるだろう。予測モデルによって生成された「予後」について情報提供する前に、患者の苦痛及び転帰に関する「予後」を共有する

ことの効果がどのようなものか、少なくとも機械学習のアルゴリズムをもとに一定のデータを得ておくことは助けになるだろう。

K 医師が受け持つ患者は新たに発症した精神疾患に直面していることから、アルゴリズムによる予測情報を提供することが、かえって患者や家族にとっての心理的な苦痛にもつながり得る懸念がある。概して、「重度の精神症状を伴う患者は自分で主体的に情報を探してリサーチしたり治療上の決定を行ったりすることが十分にできない」という仮定は十分に精査されたものではない。このような懸念は、そのまま受け取るのではなく、実証的に検証される必要がある。教育レベル、自らの意思能力をどう評価しているか、治療にどの程度満足しているかといった要因によって、患者が治療に関する詳細を希望する度合いは異なるかもしれない。自立主義の能力（すなわち、強制からではなく、自分の価値観と生き方に調和する選択を自分でする能力）は、インフォームドコンセントのもう一つの極めて重要な構成要素であり、それぞれの判断について患者が選択を行うために、その患者が向き合うべきものである。その患者にとってのニーズや患者のインフォームドコンセントに関する対応能力に関心を向けることは（患者が医療上の決定に有意義に関与できるよう支えること、意思決定能力の評価を支援するツールを見出すことなど）、重要な作業であり続ける。

結論

疾患予測ツールを倫理的な方法で導入するためには、K 医師は予測モデルの使用に関する適切な情報をどのように（理解できる方法で）患者と家族に提供するか、慎重に検討する必要があるだろう。予測モデルの利点を最大化し、リスクを最小化するために、K 医師そして医療機関は一体となって、ソフトウェアを取り巻く倫理的に適切な手順およびプロトコルを策定する必要があるだろう。その中では、例えば、疾患予測ツールの実践に際して、倫理的あるいは臨床的な根拠をもとにそうすることが支持される場合には、医師の判断が疾患予測モデルの示した結果に優先されることが盛り込まれるべきである。このような措置を講じることが、患者の生き方を改善する形で、（精神疾患の疾患予測モデルのような）機械学習のツールが臨床での実用を推し進めることに寄与するだろう。

（仮訳：井上悠輔）

著者：Martinez-Martin N, Dunn LB, Roberts LW
原題：Is It Ethical to Use Prognostic Estimates from Machine Learning to Treat Psychosis?
出典：AMA J Ethics. 2018 Sep 1;20(9):E804-811. doi: 10.1001/amajethics.2018.804.