

PRESS RELEASE

2026 年 1 月 9 日

理化学研究所

日本医科大学

東北大学

がん AI 予測の「二重の壁」を超える

ー施設差や検体差に左右されない医療の実現へー

概要

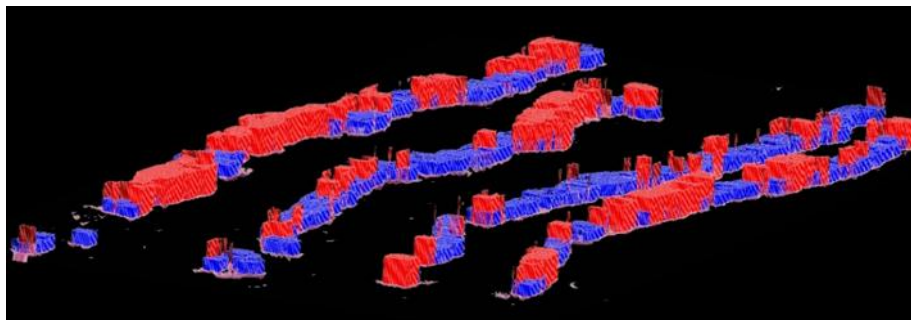
理化学研究所（理研）革新知能統合研究センター生命空間医科学チームの赤塚純客員研究員（日本医科大学泌尿器科准教授）、堤光太郎客員研究員、山本陽一朗チームディレクター（東北大学大学院医学系研究科人工知能医科学分野教授）らの共同研究グループは、AI の推論過程に医学的知識を組み込む新しいアプローチを提示し、がん AI 予測における「二重の壁」である施設差や検体差に左右されない、データ効率の高い技術を構築しました。

本研究成果は、「ドメインシフト^[1]問題」と呼ばれている AI の汎用性（はんようせい）に対する重要課題解決への一つの道筋を示すもので、地域差や施設規模に左右されず、誰もが公平に質の高い医療を受けられる未来の実現に貢献することが期待されます。

医療現場で用いられる病理画像は、病院ごとに色や質が異なり、手術前の生検検体^[2]と手術後の全摘検体^[3]では観察によって得られる情報も大きく変わります。AI にとってはこれらの“ばらつき”が予測精度を下げる「二重の壁（二つのドメインシフト）」となり、どこでも同じように使える AI の実現を阻んできました。

今回、共同研究グループは、前立腺がん患者の生検組織から、将来の病状進行を病理プロファイル^[4]を用いて AI 予測する際に、「人間の基準より少しだけ詳しい道しるべ」である中間推論スコア^[5]を用いるという新しいアプローチを採用することで、医学領域では長く困難とされてきた AI 予測の精度と汎用性を同時に向上することができました。

本研究は、科学雑誌『Nature』の関連誌『npj Digital Medicine』（1 月 7 日付）に掲載されました。



3次元病理生検組織上に可視化した病理プロファイル

背景

医療現場で用いられる病理画像は、診断や治療方針の検討を行う上で必要不可欠ですが、病院ごとに微妙に色や質が異なり、手術前の生検検体と手術後の全摘検体では観察によって得られる情報も大きく変わります。そのため、これまでのAIは「ある病院では高精度だが、別の病院では精度が下がる」（施設差）、また「手術前の小さな生検検体と、手術後の大きな検体で性能が変わる」（検体差）という二重の壁が立ちはだかっていました。少ないデータだけで「治療の結果（再発の有無など）」を直接学習させると、AIは偶然の揺らぎにも敏感に反応してしまい、学習が不安定になります。一方、人間が使うがんグレード分類などの医学基準は安心できる指標ですが、情報が粗く、AIの能力を十分に引き出すことができません。

医療AIが直面してきた最大の課題は、どの施設でもどの検体でも安定した性能を保つ「汎用性」です。医療分野では大規模言語モデルのように膨大なデータを常に集められるわけではありません。特に大量データの取得が難しい領域ではこの問題が顕著でした。

研究手法と成果

共同研究グループは、前立腺がん患者の手術前の生検検体から将来の経過を予測するAIを開発しました。このAIは、予測精度を向上するために大学病院から収集した手術後の全摘検体のデータを基に多様な特徴を学習し、さらに臨床知識を取り入れた中間推論ステップを導入しています（図1）。

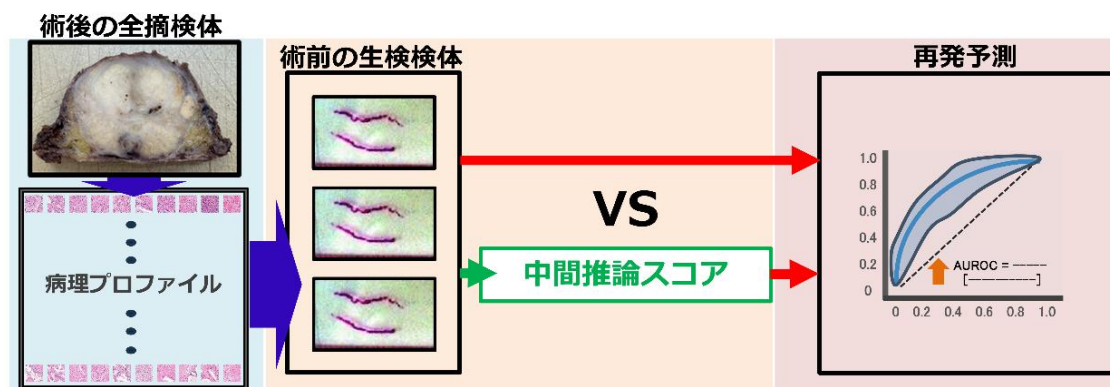


図1 研究手法のフローチャート

病理画像から治療結果を直接予測（再発予測）する従来手法（赤矢印）と、本研究で導入した中間推論スコアを用いる手法（緑矢印＋赤矢印）の違いを示している。本研究で提示するアプローチでは、大量のデータが得られる手術後の全摘検体から病理学的特徴を学習し、その特徴を手術前の生検検体へ適用して病理プロファイルを作成する。その後、臨床知識に基づき中間推論スコアを生成し、このスコアを経由して最終的な治療結果を予測している。

治療結果（Y）を病理画像（X）から直接予測すると、少数データでは学習が不安定になります（X→Y）。一方で、人間の医学基準（H）は情報が粗いため、

深層学習^[6]モデルの能力を十分に引き出すことができません ($X \rightarrow H \rightarrow Y$)。そこで、より詳しい指標 (H') を取り入れ、さらに医学的に考慮される最低限の補正を加えた中間推論スコア ($T(H')$) として用いることで、 $X \rightarrow T(H') \rightarrow Y$ と経路を安定させました。中間推論スコアには複数の設計が可能です。その一例として、本研究では、従来は予後予測のために五つにまとめられていたグリーンソン分類^[7]を、より詳細に序列化しました。臨床情報は、このスコアの方向付けを補助する「弱い重み付け (weakly supervised) ^[8]」として学習時のみに使用され、推論時には臨床情報を必要とすることなく中間推論スコアを生成できます。

本研究では、日本医科大学(NMSH)、愛知医科大学(AMUH)、順天堂大学(JUH)のそれぞれの大学病院から前立腺がん手術を受けた患者データを収集し、複数病院のデータを用いた正確な外部検証を含む予測性能評価に用いました。AI はまず、手術後の全摘標本から約 350 万枚の画像パッチ (断片) を抽出し、深層学習モデルである「Vision Transformer (Vit)」などを組み合わせて病理学的な特徴ベクトルに圧縮しました。次に、生検検体 (各施設合計約 5,200 万パッチ) に全摘検体から得た特徴を適用し、各症例に「どの特徴が何%含まれるか」を定量化した病理プロファイルを作成しました (図 2)。そして、(A) 病理プロファイルをそのまま予後予測に使う場合、(B) 病理プロファイルから中間推論スコアを経由して予後予測に使う場合、(C) これらに前立腺特異抗原 (PSA) といった血液データをさらに加えて予後予測する場合を比較しました。

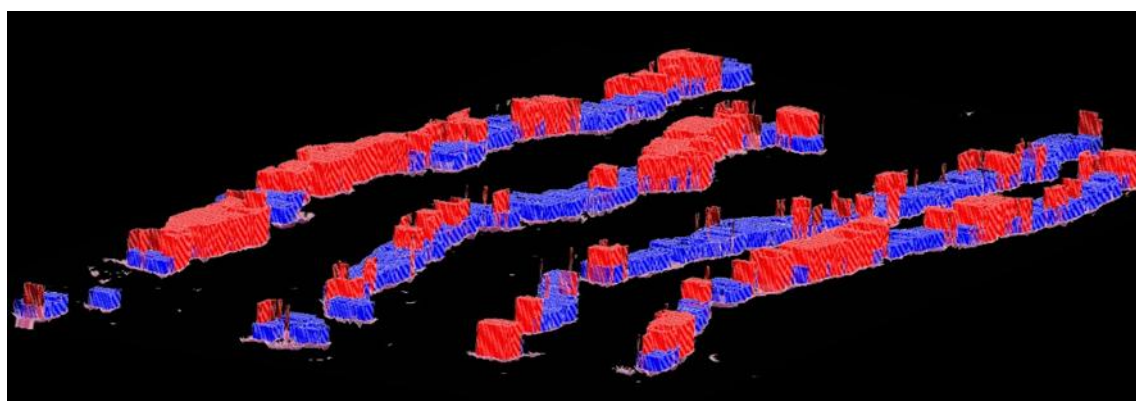


図 2 3次元病理生検組織上に可視化した病理プロファイル

生検検体の 3 次元再構成データ上に、病理プロファイルの基となる AI が抽出した 100 種類の病理学的特徴の分布を可視化したものである。各パッチに対して全摘検体から学習した特徴表現を適用し、「どの特徴が何%含まれるか」が赤と青の色分布 (赤は高リスク、青は低リスク) とその高低 (リスクの値) として示されている。これにより、生検検体という小さな標本内でも、腫瘍の異質性や微細な病理的パターンが空間的にどのように存在しているかを直観的に把握でき、中間推論スコアの基盤となる特徴の空間的広がりを視覚的に理解できる。

AI モデルの性能は、生化学的再発予測の受信者動作特性 (ROC) 曲線下の面積 (AUROC) ^[9]などで評価しました。主要な結果は以下の通りです (() 内の数値は 95%信頼区間)。

AUROC による予測精度 (0~1 で表され 1 に近いほど予測精度が高い) :

(A) 病理プロフィールをそのまま予後予測に使用

NMSH=0.604 (0.517-0.691)

AMUH=0.659 (0.497-0.807)

J U H=0.699 (0.500-0.872)

(B) 病理プロフィールから中間推論スコアを経由して予後予測

NMSH=0.741 (0.664-0.812)

AMUH=0.755 (0.623-0.871)

J U H=0.779 (0.602-0.927)

(C) 上記中間推論スコアに PSA を併用して予後予測

NMSH=0.796 (0.727-0.858)

AMUH=0.783 (0.653-0.895)

J U H=0.805 (0.648-0.931)

中間推論スコアの導入により、全施設で AUROC が向上し、PSA 併用では最高 0.805 に達しました。一方で、グリーンソン分類の AUROC は NMSH=0.600、AMHU=0.640、JUN=0.677 でした。また NMSH では確率予測の総合的な品質を評価するブライア (Brier) スコア (数値が低いほど確率予測の精度が高い) が、病理プロフィールに中間推論スコアを導入することにより 0.360 から 0.219 へ、AMUH・JUH でもそれぞれ 0.295→0.173、0.480→0.239 と改善し、AI の予測と実際の再発率の一致度が向上しました。さらに多変量解析では、この中間推論スコアが年齢や PSA を超える独立した再発予測因子であることが複数の施設で示され (p (有意水準) <0.05)、臨床的意義の高さが裏付けられました。

今後の期待

本研究の最大の成果は、「二重の壁 (二つのドメインシフト)」を超える道筋を示したことです。一つ目の壁は「病院ごとの違い」であり、施設や機器、染色法の違いによるばらつきです。そして二つ目の壁は「検査の種類の違い」であり、生検と全摘で画像の特徴が異なることです。

今回の AI は、これら二つの壁を同時に克服し、複数病院のデータを用いた外部検証でも安定した性能を発揮しました。さらに、予測の正確さだけでなく、世界基準であるグリーンソン分類と比べても高い汎用性を実現したことも重要です。

今回の成果は、AI が単なるデータ駆動型の解析を行うのではなく、専門家の総合知ともいえる医学知識を織り込みながら段階的に推論するという新しい発想が、実際の医療現場の“ばらつき”を乗り越える力になることを示しました。これは、どの病院でも同じ質の医療を実現するための基盤として大きな意義を持つとともに、AI の説明可能性の向上にも寄与しています。

本研究は、AI が臨床知識を内在化したスコアの段階的推論を行うことで、大量データの入手が困難な少数データ環境でも安定した予測が可能になることを示しました。山本チームディレクターらはこれまでも、教師なし深層学習によるがん再発に関わる未知の特徴の自動探索^{注)}など、がんの理解を深め医療に貢献する AI 研究を進めてきました。今回の研究は、これまでの研究で課題となって

いた、少数データが原因となる汎用性の低下という問題に対して、一つの解決アプローチを提示するものです。一方で、もう一つの課題である AI が見つけた所見を生物学的に深く理解するため、AI for Science^[10]も現在進行中です。これらを両輪として、汎用的な予後予測、高精度な疾患解析、そして創薬開発の加速などを通して、患者一人一人に役立つ研究を今後も継続していきます。

注) 2019 年 12 月 18 日プレスリリース「がんの未知なる特徴を AI が発見」

https://www.riken.jp/press/2019/20191218_2/

論文情報

<タイトル>

Clinically informed intermediate reasoning enables generalizable prostate cancer prognostication through machine learning in limited settings

<著者名>

Jun Akatsuka*, Kotaro Tsutsumi*, Mami Takadate, Yasushi Numata, Hiromu Morikawa, Atsushi Marugame, Hayato Takeda, Yuki Endo, Yuka Toyama, Takayuki Takahashi, Kaori Ono, Junya Iwazaki, Ryuji Ohashi, Akira Shimizu, Tomoharu Kiyuna, Maki Ogura, Masao Ueki, Takuma Kato, Toshiyuki China, Mikio Sugimoto, Hisamitsu Ide, Naoto Sassa, Naonori Ueda, Shigeo Horie, Toyonori Tsuzuki, Go Kimura, Yukihiro Kondo, and Yoichiro Yamamoto

**equally contributed to this work*

<雑誌>

npj Digital Medicine

<DOI>

[10.1038/s41746-025-02193-x](https://doi.org/10.1038/s41746-025-02193-x)

補足説明

[1] ドメインシフト

学習データと、実際に使うデータの分布が異なることで、AI の性能が落ちる現象。

[2] 生検検体

手術前に診断のために体内から部分的に取り出した組織標本で、前立腺がんでは針を使って採取する小さな組織サンプルのこと。

[3] 全摘検体

手術で臓器を切除して得られる大型の組織標本で、前立腺全摘術では臓器全体が観察可能であるため生検検体に比べて情報量が非常に多い。

[4] 病理プロファイル

病理的な所見の集まりとして、本研究では深層学習（[6]参照）モデルにより病理画像から抽出した 100 種類の特徴について各症例にどの特徴がどれだけ含まれるかを定量化した特徴構成表。

[5] 中間推論スコア

最終結果を予測する前に AI に適した「中間的な尺度」を一度推定することで、予測を安定させるための手掛かりとなる指標。

[6] 深層学習

大規模な機械学習モデルを構築し、データに潜む関係性を学習することで、高度な予測を行う情報技術。AI の中核技術として、画像認識や翻訳などさまざまな分野で著しい成功を収めている。

[7] グリーソン分類

前立腺組織を顕微鏡で検査し、がんの悪性度を評価するために用いられる指標。世界中の病院で使用され、その後の治療方針に影響する重要な要素の一つとなる。原案は 1966 年に米国のドナルド・グリーソンにより提唱された。

[8] 弱い重み付け (weakly supervised)

ここでいう「弱い」とは、厳密な臨床情報ラベルを直接教え込むのではなく、スコアの方向付けをわずかに補正する限定的な重み付けであるという意味。

[9] 受信者動作特性 (ROC) 曲線下の面積 (AUROC)

検査などの性能を表す際に使用されるグラフの一つである ROC (Receiver Operating Characteristic) 曲線を作成したときの、グラフ下部の面積のこと。0 から 1 までの値を取り、値が 1 に近いほど判別能が高い。AUROC は Area Under the Receiver Operating Characteristic Curve の略。

[10] AI for Science

AI を用いて科学研究の発見プロセス全体を加速・高度化する取り組みで、データ解析からシミュレーションまで複雑で大規模な科学データの理解と探究を AI が支援する。

共同研究グループ

理化学研究所

革新知能統合研究センター 目的指向基盤技術研究グループ

生命空間医科学チーム (旧病理情報学チーム)

チームディレクター 山本陽一郎 (ヤマモト・ヨウイチロウ)

(東北大学 大学院医学系研究科 人工知能医科学分野 教授)

研究員 高橋孝幸 (タカハシ・タカユキ)

上級技師 丸亀 敦 (マルガメ・アツシ)

技師 沼田康志 (ヌマタ・ヤスシ)

技師 森川 啓 (モリカワ・ヒロム)

技師 岩崎淳也 (イワザキ・ジュンヤ)

客員研究員 赤塚 純 (アカツカ・ジュン)

(日本医科大学 泌尿器科 准教授)

客員研究員 堤 光太郎 (ツツミ・コウタロウ)

(カリフォルニア大学 神経内科 レジデント)

客員研究員 高舘真美 (タカダテ・マミ)

(東北大学 大学院医学系研究科 大学院生)

客員研究員 尾野香織 (オノ・カオリ)

(東北大学 大学院医学系研究科 大学院生)
革新知能統合研究センター
副センター長 上田修功 (ウエダ・ナオノリ)

日本医科大学

泌尿器科

大学院教授	近藤幸尋	(コンドウ・ユキヒロ)
特命教授	木村 剛	(キムラ・ゴウ)
教授	戸山友香	(トヤマ・ユカ)
准教授	遠藤勇氣	(エンドウ・ユウキ)
助教	武田隼人	(タケダ・ハヤト)

解析人体病理学

大学院教授	清水 章	(シミズ・アキラ)
-------	------	-----------

統御機構診断病理学

大学院教授	大橋隆治	(オオハシ・リュウジ)
-------	------	-------------

愛知医科大学病院

病理診断科

教授	都築豊徳	(ツツキ・トヨノリ)
----	------	------------

泌尿器科

教授	佐々直人	(サッサ・ナオト)
----	------	-----------

順天堂大学 泌尿器科

主任教授	堀江重郎	(ホリエ・シゲオ)
特任教授	井手久満	(イデ・ヒサミツ)
准教授(研究当時)	知名俊幸	(チナ・トシユキ)

香川大学 泌尿器科

教授	杉元幹史	(スギモト・ミキオ)
講師	加藤琢磨	(カトウ・タクマ)

長崎大学 情報データ科学部

教授	植木優夫	(ウエキ・マサオ)
----	------	-----------

日本電気株式会社 医療ソリューション統括部

喜友名朝春	(キユナ・トモハル)
小掠真貴	(オグラ・マキ)

研究支援

本研究は、理研 TRIP イニシアティブ(空間病態モデル開発 AGIS*)により実施し、科学技術振興機構(JST)ムーンショット型研究開発事業「生体内ネットワークの理解による難治性がん克服に向けた挑戦(JPMJMS2022-9)」、日本学術振興会(JSPS)科学研究費助成事業若手研究(JP23K17235)、同基盤研究(C)(JP24K10131)による一部助成および日本電気株式会社による支援を受けました。*AGIS: Advanced General Intelligence for Science Program(科学研究基盤モデル開発プログラム)

発表者・機関窓口

<発表者>

理化学研究所 革新知能統合研究センター 生命空間医科学チーム
客員研究員 赤塚 純 (アカツカ・ジュン)
(日本医科大学 泌尿器科 准教授)
客員研究員 堤 光太郎 (ツツミ・コウタロウ)
チームディレクター 山本陽一郎 (ヤマモト・ヨウイチロウ)
(東北大学 大学院医学系研究科 人工知能医科学分野 教授)



赤塚 純



堤 光太郎



山本陽一郎

<機関窓口>

理化学研究所 広報部 報道担当
Tel: 050-3495-0247
Email: ex-press@ml.riken.jp

日本医科大学事務局学事部 庶務課
Tel: 03-3822-2131
Email: nms-shomuka@nms.ac.jp

東北大学大学院医学系研究科・医学部広報室
Tel: 022-717-8032
Email: press.med@grp.tohoku.ac.jp