

報道関係各位

脳腫瘍の遺伝子変異を画像データから AI で予測 —専門医等との比較により、AI による IDH 変異予測精度を検証—

2026 年 7 月 15 日

国立研究開発法人国立がん研究センター

国立研究開発法人理化学研究所

東海大学

国立大学法人旭川医科大学

株式会社ヒューマノーム研究所

発表のポイント

- 代表的な脳腫瘍の一つである神経膠腫において、IDH 変異の有無は、神経膠腫の分類、予後予測、治療反応性を判断するうえで重要です。通常、生検などの侵襲的検査が必要ですが、画像検査等の非侵襲的検査から予測できれば、患者さんの負担軽減につながります。
- 研究グループは、AI を用いて MRI 画像から IDH 変異の有無を予測する画像診断支援モデルを開発し、脳腫瘍診療の専門医等 18 名の読影結果と比較し、AI モデルの予測性能を科学的に評価しました。
- 本研究の結果、海外患者の脳腫瘍画像データセットにおいては、AI モデルが多くの医師を上回る予測性能を示しました。
- 一方で、日本人患者の脳腫瘍画像データセットを用いたところ、AI の予測性能の低下が認められ、施設や患者背景に応じた検証の重要性が示されました。
- 本研究により、実臨床での AI の応用可能性と課題が明らかになりました。今後、AI による予測結果に加え、不確実性も提示する AI システムを開発することにより、AI の強みを活かした医師の支援につながると考えられます。

概要

国立研究開発法人国立がん研究センター（所在地：東京都中央区、理事長：間野博行）は、理化学研究所（所在地：埼玉県和光市、理事長：五神真）、東海大学（所在地：神奈川県伊勢原市、学長：木村英樹）および全国の複数の医療機関と共同で、脳腫瘍の治療において重要な指標となる遺伝子変異（IDH^{注1} 変異）の有無を、生検などの侵襲的な検査で判定する従来法ではなく、より非侵襲的な方法で行えるかを検討するため、MRI 画像での人工知能（AI）による診断予測の有用性を、脳腫瘍の診療を専門とする医師等による予測と比較し評価しました。神経膠腫^{注2} は代表的な脳腫瘍の一つであり、その治療方針や予後は、腫瘍の遺伝子変異の特徴によって大きく異なります。なかでも IDH 変異の有無は、神経膠腫の分類、予後予測、治療反応性の評価において重要な指標です。本研究では、MRI 画像から IDH 変異の有無を予測する AI による画像診断支援モデルを、海外（日本人以外）患者と日本人患者の脳腫瘍

画像を用いて、従来の画像認識技術と最新技術の 2 種類で開発し、脳腫瘍の診療を行う専門医等 18 名による診断と比較し、予測性能を評価しました(図1)。

その結果、海外患者の脳腫瘍画像データセットを用いた検証において、AI モデルは高い予測精度を示しました。一方、日本人のデータを用いた検証では AI の予測性能の低下が認められ、予測結果の信頼性には課題があることも明らかになりました。特に、一部の経験豊富な医師は、AI に匹敵する予測精度を示していました。これらの結果により、神経膠腫での IDH 変異の有無を AI を用いて予測し医師を支援できる可能性と、一方では医師の専門性や患者背景の違いにより AI の性能が低下することが確認され、AI の実臨床への導入にあたっては十分に検証する必要性が示されました。

本研究は、国立がん研究センター研究所 医療 AI 研究開発分野の高橋慧外来研究員(理化学研究所革新知能統合研究センター上級研究員)、浜本隆二分野長(理化学研究所革新知能統合研究センターチームディレクター)、国立がん研究センター中央病院 脳脊髄腫瘍科の成田善孝科長、高橋雅道医長(現・東海大学医学部脳神経外科 教授)、旭川医科大学 脳神経外科学 教授 木下学らを中心に、全国の医療機関との研究コンソーシアムを構築して実施されました。

本研究成果は、国際学術誌「npj Digital Medicine」オンライン版に 7 月 10 日付で掲載されました。

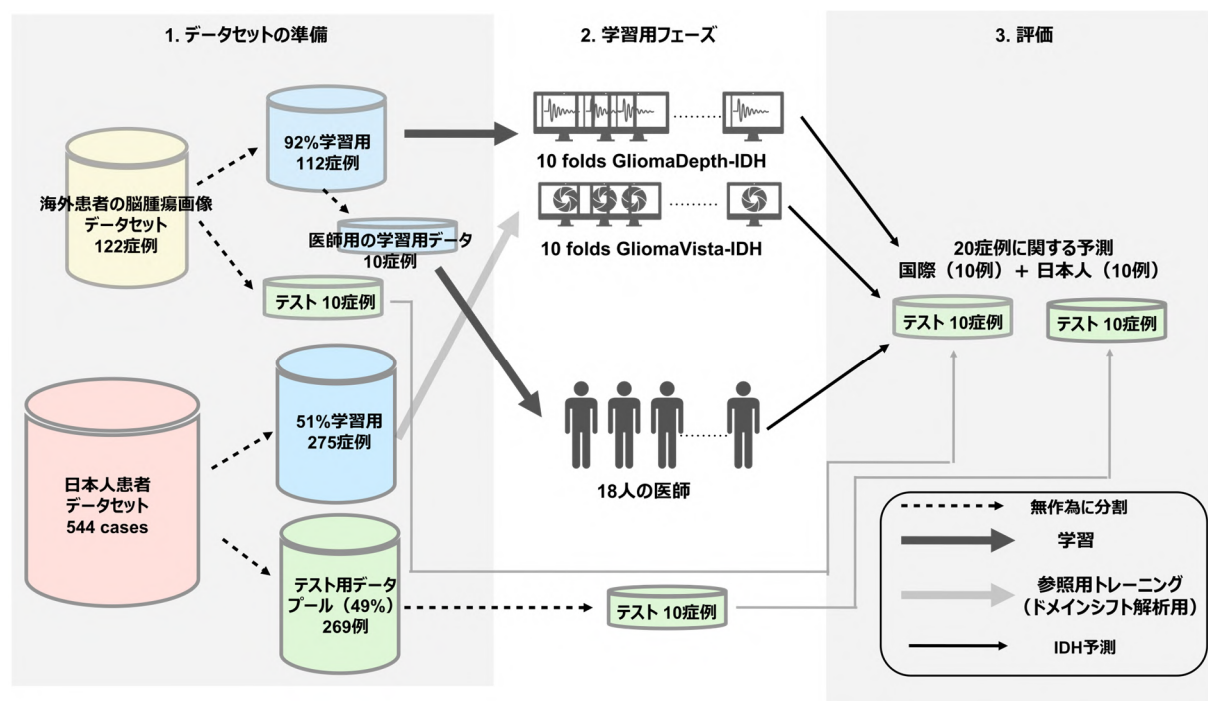
図 1: IDH 変異状態の予測における AI モデルと医師の比較のための実験計画法

本研究は、2 つの独立したデータセットを用いて 3 つのフェーズに分けて進められました。

フェーズ 1 - データセットの準備: 海外患者のデータセット脳腫瘍画像 (BraTS 注³)データセット(122 症例)は、学習用と評価用サブセットに分割し、学習用サブセットから医師のトレーニング用に 10 症例が選択されました。日本人 (JC)データセット(544 例)も、学習用と評価用サブセットに分割され、10 例の評価用ケースが選択されました。

フェーズ 2 - トレーニングフェーズ: 2 つの異なる AI モデル(畳み込みニューラルネットワークベースの GliomaDepth-IDH および Vision Transformer ベースの GliomaVista-IDH)が開発され、ドメイン固有のトレーニングがモデル性能に与える影響を評価するため、両モデルを JC トレーニングデータで再トレーニングしました。18 名の医師に対し、10 症例を用いてトレーニングを実施しました。

フェーズ 3 - 評価: トレーニング済みのすべての AI モデルおよび医師は、2 つのテストセット(BraTS: 10 症例、JC: 10 症例)について、IDH 変異状態を予測しました。



背景

神経膠腫は中枢神経系に発生する代表的な脳腫瘍の一つであり、腫瘍の種類や遺伝子変異の特徴によって、治療方針や予後が大きく異なります。特に *IDH* 変異の有無は、神経膠腫の分類、予後予測、治療反応性を判断するうえで重要な指標です。現在、*IDH* 変異の判定には病理検体を用いた免疫染色や DNA 解析が必要ですが、これらは手術や生検を伴う侵襲的な検査です。そのため、手術前に MRI 画像から *IDH* 変異を非侵襲的に予測できれば、治療方針の検討や手術計画の最適化に大きく貢献すると期待されます。

また、近年、医用画像と AI を組み合わせた診断支援技術の開発が進んでいますが、神経膠腫の *IDH* 変異予測に関しては、AI の性能が実際の医師と比べてどの程度有用か、また施設や患者背景の違いによって精度が低下しないかについて十分に検証されていませんでした。本研究は、AI と複数分野の医師を直接比較し、実臨床での応用可能性と課題を明らかにすることを目的に実施されました。

研究結果

本研究では、MRI 画像から *IDH* 変異の有無を予測する 2 種類の AI モデルを開発しました。一つは従来型の画像認識技術を用いた GliomaDepth-*IDH*、もう一つは画像の重要な部分に注目する最新技術 (attention mechanism) を用いた GliomaVista-*IDH* です。これら 2 つのモデルの予測性能を、18 名の医師 (神経放射線科医 8 名、脳神経外科医 5 名、脳神経外科レジデント 5 名) による読影結果と比較しました。

1. 海外患者データセットでは AI が多くの医師を上回る *IDH* 変異の予測性能を発揮

まず、海外患者の脳腫瘍画像データセット (BraTS) を用いて AI モデルを学習させました。その後、同じ海外患者データセットの別の症例で *IDH* 変異の予測性能を検証したところ、従来型と最新型の両 AI モデルですべての医師グループを上回る予測性能を示しました。最新型の GliomaVista-*IDH* は、予測の信頼性 (キャリブレーション^{注4)} の評価指標である Brier Score^{注5} でも 0.09 と優れた値を示し、予測結果の信頼性も高いことが確認されました。予測性能の評価指標には、AUC^{注6} (Area Under the Curve) を用い、「*IDH* 変異がある患者」と「*IDH* 変異がない患者」をどれだけ正確に区別できるかを確認しました。値は 0 から 1 の範囲で、1 に近いほど識別能力が高いことを意味します。

AI モデルの成績

GliomaDepth-*IDH* モデル (従来型): AUC 0.85、正解率^{注7} 0.79、F1 スコア^{注8} 0.84

GliomaVista-*IDH* モデル (最新型): AUC 0.97、正解率 0.90、F1 スコア 0.92

医師グループの成績

神経放射線科医: AUC 0.65

脳神経外科医: AUC 0.59

脳神経外科レジデント: AUC 0.54

2. 日本人患者データでは AI の性能低下が見られたが、一部の熟練医師は AI を上回る

次に、外部検証として海外患者データセットで学習した AI モデルを日本人患者データ (JC データセット) で検証したところ、AI の予測性能は低下 (GliomaDepth-*IDH*: AUC 0.85 から AUC 0.75、GliomaVista-

IDH: AUC 0.97 から AUC 0.82 へ減少)しました。これは、学習したデータの違いが AI の判定に影響する(ドメインシフト)と考えられ、今回のデータセットにおいては以下のような相違がありました。

一方で、3 名の医師(神経放射線科医 2 名、脳神経外科医 1 名)は、日本人患者データセットにおいて AUC 0.88、F1 スコア 0.78 を達成し、両方の AI モデルを上回る成績を示しました(図 2)。また、予測の信頼性を示す Brier Score でも 0.19 と優れた値を示し、AI モデル(0.21~0.32)と比較してより信頼性の高い予測を行っていました。これは、長年の経験で培われた専門医の診断能力を、現在の AI システムではいまだ完全には再現できない可能性があることを示しています。

AI モデルの成績

GliomaDepth-IDH(従来型): AUC 0.75、正解率 0.75(Youden 指数^{注9}による最適化後: 正解率 0.80、F1 スコア 0.74)

GliomaVista-IDH(最新型): AUC 0.82、正解率 0.64、再現率^{注10} 0.92(Youden 補正後: 正解率 0.83、適合率^{注11} 0.82、F1 スコア 0.79)

医師グループの成績

神経放射線科医: AUC 0.75、Brier Score 0.25

脳神経外科医: AUC 0.6、Brier Score 0.35

脳神経外科レジデント: AUC 0.55、Brier Score 0.34

本結果から、2 つの AI モデルともに、学習したデータセットと異なるデータでの予測性能は、低下はするものの高い性能を維持していましたが、予測性能は GliomaVista-IDH(最新型)では Brier Score が 0.32 と課題があり、GliomaDepth-IDH(従来型)の Brier Score は 0.21 で、より信頼性の高い予測を行っていたことがわかりました。

海外患者と日本人患者でのデータセットの相違点

- 撮影装置や撮影条件

海外患者データセットは国際的な医療機関で標準化された条件の下で撮影した画像であるのに対し、日本人データセットは日本の 10 施設で 20 年以上にわたり、それぞれ異なる装置・設定で撮影された画像です。撮影時のスライス厚(画像の間隔)も施設ごとに異なっていました。

- 腫瘍の特徴

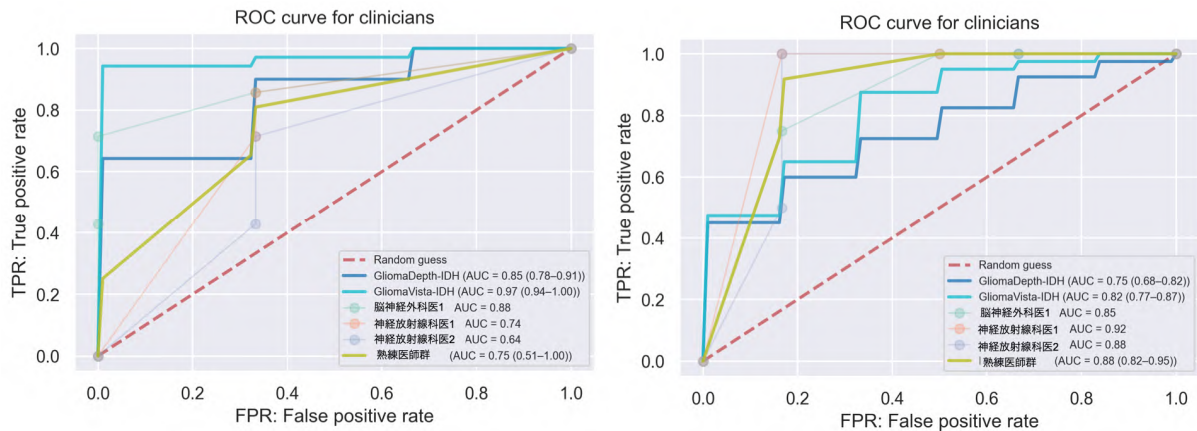
日本人データセットの腫瘍は平均的に小さく(中央値 70 cm³ vs 海外患者 107 cm³)、画像上の特徴も微妙に異なっていました。

- 患者背景の違い

日本人データセットは日本人患者のみで構成され、年齢や腫瘍の種類分布も海外患者データセットと異なっていました。

これは、英語の文章で学習した AI が日本語の文章を読もうとするようなもので、基本的な原理は同じでも、学習したデータの細かな特徴の違いが判定に影響し、結果的に性能が低下したものです。AI は学習時に見たデータの特徴(パターン)を記憶しており、それと似た特徴を持つ新しいデータに対しては高い性能を発揮しますが、異なる特徴を持つデータでは性能が低下するのです。

図 2: AI モデルと医師の予測性能の比較(ROC 曲線)



左図(a)は海外患者データセット(BraTS)、右図(b)は日本人患者データ(JC)での結果を示しています。グラフの曲線が左上に近いほど予測性能が高いことを意味します。日本人データでは、熟練医師グループが AI モデル(青色と水色の線)を上回る性能を示していることがわかります。

3. AI の性能向上には適切な学習データと総合的な評価が重要

学習したデータの違いによるドメインシフトの影響を評価するため、日本人患者データで学習し直した AI モデルを両方のデータセットで検証しました。その結果、日本人データで学習したモデルは日本人患者の予測でより良い性能を示しましたが、逆に海外患者データでの性能は低下しました。また、医師グループによる正診率も、普段見慣れない海外患者データセットではやや低下していました。このことから、ドメインシフトは AI と医師の両方に影響を与えることがわかりました。重要な発見として、GliomaVista-IDH(最新型)では、AUC(識別能力)が同じ0.82であっても、学習データによって正解率(0.64から0.73)、適合率(0.53から0.63)、Brier Score(0.32から0.24)が大きく改善しました。これは、AI の臨床応用を評価する際には、AUC だけでなく、予測の信頼性(キャリブレーション)や正解率などを総合的に評価する必要があることを示しています。

展望

本研究成果により、高度な AI アーキテクチャは、神経膠腫における IDH 変異予測において、特に優れた診断能力を有する一部の専門医を除き、多くの医師と同等、またはそれ以上の予測性能を発揮することが示されました。また、施設やデータ環境が異なる場合には、AI の性能や信頼性が低下する可能性があるため、実装にあたっては各臨床現場に応じた適応と検証が重要です。今後、予測結果に加えて不確実性も提示できる AI システムを開発することで、医療従事者の支援や診断の標準化に貢献することが期待され、神経膠腫における新たな診療体制の構築につながると考えられます。

研究支援(研究コンソーシアム)

本研究では、日本人の脳腫瘍画像データの収集にあたり多数の施設よりご支援をいただきました。ご協力いただきました医療関係者の皆様ならびに患者さんに深く感謝申し上げます。

コンソーシアム参画施設:

国立研究開発法人国立がん研究センター、国立研究開発法人理化学研究所、東海大学、国立大学法人旭川医科大学、国立大学法人東京大学、国立大学法人熊本大学、杏林大学、獨協医科大学、国立大学法人岡山大学、国立大学法人九州大学、国立大学法人東京科学大学、横浜市立大学、埼玉医科大学、国立大学法人大阪大学、関西医科大学、大阪公立大学、国立病院機構大阪医療センター、和歌山医科大学、大阪国際がんセンター、国立大学法人弘前大学、順天堂大学、国立大学法人東海国立大学機構名古屋大学、国立大学法人滋賀医科大学、京都府立医科大学、国立大学法人宮崎大学、国立大学法人鹿児島大学、株式会社ヒューマノーム研究所

論文情報

雑誌名: *npj Digital Medicine*

タイトル: Comparing artificial intelligence and physician performance in predicting *IDH* mutation status in glioma

著者: Satoshi Takahashi, Masamichi Takahashi (* Corresponding Author), Manabu Kinoshita, Mototaka Miyake, Risa Kawaguchi, Naoki Shinojima, Akitake Mukasa, Kuniaki Saito, Motoo Nagane, Ryohei Otani, Fumi Higuchi, Shota Tanaka, Nobuhiro Hata, Kaoru Tamura, Kensuke Tateishi, Ryo Nishikawa, Hideyuki Arita, Masahiro Nonaka, Takehiro Uda, Junya Fukai, Yoshiko Okita, Naohiro Tsuyuguchi, Yonehiro Kanemura, Fumiyasu Tsushima, Shingo Kakeda, Toshiaki Akashi, Toshiaki Taoka, Yoshiyuki Watanabe, Kei Yamada, Toshinori Hirai, Minako Azuma, Takashi Yoshiura, Jun Sese, Koichi Ichimura, Yoshitaka Narita, Ryuji Hamamoto (* Corresponding Author)

DOI: [10.1038/s41746-026-02695-2](https://doi.org/10.1038/s41746-026-02695-2)

掲載日: 2026 年 7 月 10 日

URL: <https://www.nature.com/articles/s41746-026-02695-2>

研究費

- 内閣府科学技術・イノベーション推進事務局 研究開発と Society5.0 との橋渡しプログラム (BRIDGE)【研究総括:国立がん研究センター研究所 医療 AI 研究開発分野 分野長 浜本隆二】
- 文部科学省 次世代人工知能技術等研究開発拠点形成事業補助金

用語解説

注 1: IDH(イソクエン酸脱水素酵素)

細胞のエネルギー代謝に関わる酵素です。神経膠腫では *IDH* 遺伝子に変異があるかどうか、腫瘍の分類、予後予測、治療方針の検討に重要な指標(分子マーカー)となります。

注 2: 神経膠腫(グリオーマ)

脳や脊髄の神経細胞を支えるグリア細胞から発生する腫瘍の総称です。悪性度や分子学的特徴により治療方針や予後が異なり、代表的な原発性脳腫瘍の一つです。

注 3: BraTS (Brain Tumor Segmentation 2019) データセット

神経膠腫を対象とした海外患者の脳腫瘍 MRI 画像データセットです。複数施設の術前 MRI 画像と専門医による腫瘍領域ラベルを含み、AI による腫瘍分割や予後予測の評価に用いられます。

注 4: 較正 (キャリブレーション) 解析

AI や予測モデルが出力する確率と、実際にその事象が起こる割合との一致度を示します。較正が良いほど、予測確率を信頼しやすいことを意味します。

注 5: Brier Score

予測モデルが出力した確率と実際の結果との差を評価する指標です。値が小さいほど予測確率と実際の結果がよく一致しており、モデルの予測の信頼性や較正の良さを示します。0.2 以下が良好とされています。

注 6: AUC (ROC-AUC)

診断・予測モデルの識別性能を示す指標です。値は 0~1 で表され、1 に近いほど陽性例と陰性例を正しく区別できる性能が高いことを意味します。例えば、AUC 0.97 というのは、ランダムに選んだ「変異あり」と「変異なし」の患者ペアを見た時に、97%の確率で正しく区別できる能力があることを示しています。これは単純な「正解率」とは異なり、モデルの識別能力そのものを評価する指標です。

注 7: 正解率 (Accuracy)

全症例のうち、AI や医師による予測が実際の結果と一致した割合を示す指標です。値が高いほど全体として正しく判定できていることを意味します。一般的に 0.7 以上が良好、0.8 以上が非常に良好、0.9 以上が優秀とされています。

注 8: F1 スコア

陽性と判定した結果の正確さを示す適合率と、実際の陽性例をどれだけ見逃さず検出できたかを示す再現率のバランスを表す指標です。値が高いほど総合的な判定性能が高いことを意味します。

注 9: Youden 指数

診断法や予測モデルの性能を評価する指標で、感度と特異度の和から 1 を引いて算出します。値が大きいほど、陽性例と陰性例をバランスよく判別できることを意味します。

注 10: 再現率 (Recall)

実際に陽性である症例のうち、AI や医師が正しく陽性と判定できた割合を示す指標です。値が高いほど、陽性例の見逃しが少ないことを意味します。

注 11: 適合率 (Precision)

AI や医師が陽性と判定した症例のうち、実際に陽性であった割合を示す指標です。値が高いほど、陽性と判断した結果の信頼性が高いことを意味します。

お問い合わせ先

- 研究に関するお問い合わせ

国立がん研究センター研究所 医療 AI 研究開発分野

分野長 浜本隆二

E メール: rhamamot@ncc.go.jp

理化学研究所 革新知能統合研究センター AI 医用工学チーム

上級研究員 高橋慧

E メール: satoshi.takahashi.fy@riken.jp

東海大学 医学部 脳神経外科

教授 高橋雅道

E メール: masataka@tokai.ac.jp

旭川医科大学 脳神経外科学講座

教授 木下学

E メール: mail@manabukinoshita.com

- 広報窓口

国立研究開発法人国立がん研究センター

企画戦略局広報企画室

電話番号: 03-3542-2511(代表)

E メール: ncc-admin@ncc.go.jp

理化学研究所 広報部 報道担当

電話番号: 050-3495-0247

E メール: ex-press@ml.riken.jp

東海大学医学部附属病院

事務部事務課(広報)

電話番号: 0463-90-2001(直通)

E メール: prtokai@tokai.ac.jp

旭川医科大学

総務課広報・社会連携係

電話番号: 0166-68-2118(直通)

E メール: kouhou@asahikawa-med.ac.jp

株式会社ヒューマノーム研究所 広報担当

E メール: info@humanome.jp