

Aus der Klinik für Neurochirurgie
der Heinrich-Heine-Universität Düsseldorf

Direktorin: Univ.-Prof. Dr. med. Katharina Faust

Klinische Integration der Vorhersage von IDH-Mutationen in Gliomen mit
maschinellern Lernen basierend auf MRT-Bildgebung

Dissertation

Zur Erlangung des Grades eines Doktors der Medizin an der Medizinischen Fakultät
der Heinrich-Heine-Universität

vorgelegt von
Jan Philipp Lost
2025

Als Inauguraldissertation gedruckt mit Genehmigung der Medizinischen Fakultät der
Heinrich-Heine-Universität Düsseldorf

gez.:

Dekan: Prof. Dr. med. Nikolaj Klöcker

Erstgutachter: Prof. Dr. med. Michael Sabel

Zweitgutachter: PD Dr. med. Christian Rubbert

Man muss nichts im Leben fürchten, man muss nur alles verstehen.

– Marie Curie

Teile dieser Arbeit wurden veröffentlicht:

Lost, J. et al., Enhancing clinical decision-making: An externally validated machine learning model for predicting isocitrate dehydrogenase mutation in gliomas using radiomics from presurgical magnetic resonance imaging, *Neuro-Oncology Advances*, Volume 6, Issue 1, January-December 2024, vdae157

<https://doi.org/10.1093/noajnl/vdae157>

Zusammenfassung

Gliome, die häufigsten primären Hirntumoren des zentralen Nervensystems, stellen eine Herausforderung für Prognose und Behandlung dar. Sie treten in verschiedenen Entitäten auf, von denen das Glioblastom die häufigste und bösartigste mit schlechter Prognose und begrenzten Behandlungsmöglichkeiten darstellt. Die Diagnostik von Gliomen stützt sich zunehmend auf verschiedene molekulare Tumormarker, um ein individuelles Therapieschema und Prognose der Krankheit zu bestimmen.

Von zentraler Bedeutung für die Diagnostik von Gliomen ist die Mutation der Isocitratdehydrogenase (IDH), nach der Gliome in verschiedene Klassen eingeteilt werden. Wie bei anderen molekularen Subtypen muss IDH traditionell durch Untersuchung einer Tumorprobe nach einer Biopsie bestimmt werden. Mehrere Studien weisen aber auf Probleme hin, die mit der ausschließlichen Verwendung der molekularpathologischen Techniken verbunden sind. Darunter fallen das Risiko von Komplikationen, hohe Kosten, potenzielle technische Hindernisse und eine mögliche Fehlklassifizierung durch lokale Tumorheterogenität.[1] Viele Arbeitsgruppen haben daher die Vorhersage des IDH-Status mit Hilfe von maschinellem Lernen untersucht und dabei gute Ergebnisse erzielen können, jedoch mit relativ kleinen Stichprobengrößen, ohne externe Validierung und ohne ihre Ergebnisse in die klinische Praxis umsetzen zu können.[2]

Diese Forschungsarbeit soll nun den Nutzen der IDH-Vorhersage für die klinische Praxis bewerten und ein zuverlässiges Modell mit Daten aus einer einzelnen Einrichtung erstellen, ohne multizentrische Zusammenarbeit oder stark kurierte öffentliche Datenbank. In der retrospektiven Studie wurde ein Trainingsdatensatz mit 377 Patienten verwendet, die im Yale New Haven Health (YNHH) behandelt wurden, sowie 207 Fälle von US-Einrichtungen außerhalb des YNHH für die unabhängige Validierung. Es wurde eine 6-stufige Verarbeitungspipeline für ein maschinelles Lernmodell entwickelt, die die Bilderfassung, die halbautomatische Bildsegmentierung, die Extraktion quantitativer Bildmerkmale, die Modellerstellung einschließlich der Merkmalselektion sowie interne und externe Validierungstechniken für das Modelltraining und -bewertung umfasst.

Das Modell zeigte hohe Präzision mit Fläche unter der Kurve = 0,835, Genauigkeit = 0,851, Sensitivität = 0,850 und Spezifität = 0,847 in der externen Validierung. Dass das Modell auf einem heterogenen Datensatz basiert und robuste und hohe Ergebnisse liefert, unterstreicht seinen potenziellen klinischen Nutzen.[1]

Summary

Gliomas, the most prevalent primary brain tumor of the central nervous system poses challenges for prognosis and treatment. They occur in different types of classes, with glioblastoma being its most predominant subtype, with poor prognosis and limited treatment options. Glioma classification is more and more reliant on several molecular tumor markers to determine an individualized therapy scheme and prognosis of the disease.

Isocitrate dehydrogenase (IDH) mutation status is a disease defining molecular alteration and the predominant subtype in the most recent brain tumor classification of the world health organization. It is therefore pivotal for the diagnostic process in clinical care for patients with gliomas. Like other molecular subtypes, IDH traditionally needs to be determined with molecular testing of a tumor portion, received by tumor biopsy before it can be assessed and integrated into the diagnostic process. Several studies indicate issues using solely molecular pathology for IDH status determination, including risk of complications, high costs, potential technical limitations and a risk of misclassification through spacial tumor heterogeneity.[1] While many research groups therefore evaluated IDH status and molecular subtype, prediction using machine learning with high results often face high bias, use relatively small sample sizes, lack external validation and fail to convert their findings into clinical practice.[2]

To determine possible clinical utility, this study aims to evaluate the enhancement of IDH prediction for clinical practice and to build a reliable model with accessible data from a single institution, without the need of multicenter collaborations and usage of highly curated public databases. The retrospective study utilizes a large single center dataset with 377 patients treated at Yale New Haven Health (YNHH) for model training and 207 cases from US facilities outside of YNHH for independent testing. A 6-step processing pipeline for an extreme gradient boosting model was established, including image acquisition, semi-automated imaging segmentation, *radiomic* feature extraction, model building including feature selection, as well as internal and external validation techniques for model training and evaluation.[1]

As a result, our model demonstrated high performance metrics with area under the curve = 0.850, accuracy = 0.851, sensitivity = 0.850 and specificity = 0.847 in external validation. The model's strong performance on a heterogeneous dataset demonstrates its potential for broad clinical applicability.[1]

Abkürzungsverzeichnis

ZNS	Zentrales Nervensystem
WHO	Weltgesundheitsorganisation
CBTRUS	<i>Central Brain Tumor Registry of the United States</i>
EANO	<i>European Association of Neurooncology</i>
MRT	Magnetresonanztomographie
FLAIR	<i>Fluid Attenuated Inversion Recovery</i>
IDH	Isocitratdehydrogenase
Gy	Gray
MGMT	O6-Methylguanin-Methyltransferase
TTF	<i>Tumor Treating Field</i>
RANO	<i>Response Assessment in Neuro-Oncology</i>
NAD⁺/NADH⁺	Nicotinamidadeninukleotid (oxidiert/reduziert)
KI	Künstliche Intelligenz
SVM	<i>Support Vector Machine</i>
ANN	<i>Artificial Neural Network</i>
XGBoost	<i>Extreme Gradient Boosting Algorithm</i>
MICCAI	<i>Medical Image Computing And Computer Assisted Intervention</i>
YNHH	<i>Yale New Haven Health</i>
PACS	<i>Picture Archiving and Communication System</i>
AUC	<i>Area Under the Curve</i> / Fläche unter der Kurve
CEST	<i>Chemical exchange saturation transfer</i>

Inhaltsverzeichnis

1	Einleitung	1
1.1	Gliome	1
1.1.1	Zusammenfassung	1
1.1.2	Epidemiologie	2
1.1.3	Diagnostik	3
1.1.4	Therapie	5
1.1.5	Prädiktiver Marker: Isocitratdehydrogenase	7
1.2	Maschinelles Lernen	9
1.2.1	Grundprinzipien von maschinellem Lernen	9
1.2.2	Maschinelles Lernen mit XGBoost	10
1.2.3	Chancen von maschinellem Lernen in der Neuroradiologie	11
1.3	Ziele der Arbeit	12
2	Enhancing clinical decision-making an externally validated machine learning model for predicting isocitrate dehydrogenase mutation in gliomas using radiomics from presurgical magnetic resonance imaging, Lost J. et al., Neuro-Oncology Advances, Volume 6, Issue 1, January-December 2024, vdae157	13
3	Diskussion	14
3.1	Einordnung in den wissenschaftlichen Kontext	14
3.1.1	Einordnung der Datenbasis und Annotierung	14
3.1.2	Einordnung der Datenverarbeitung und des Klassifizierungsmodells ..	16
3.1.3	Einordnung der Resultate	17
3.2	Potenzieller Nutzen für die Patientenversorgung	18
3.3	Limitationen	20
3.4	Zukunftsperspektiven	21
3.5	Schlussfolgerungen	23
4	Literatur- und Quellenverzeichnis	24

1 Einleitung

1.1 Gliome

1.1.1 Zusammenfassung

Gliome sind mit 5 bis 6 Erkrankungsfällen auf 100.00 Einwohner pro Jahr die häufigsten Tumoren des zentralen Nervensystems (ZNS).[3]

Sie gelten als hirneigene Tumore oder auch primäre Hirntumore da sie aus den Gliazellen des Gehirnes entstehen.[4] Ihnen gegenüber sind unter anderem Neoplasien abzugrenzen, deren Ursprung außerhalb des ZNS liegt und die deshalb als sekundäre Hirntumore bezeichnet werden.[5] Gliome können anhand einer Vielzahl von Eigenschaften in Untergruppen unterteilt werden, welche sich unter anderem auf histomorphologischer oder molekularer Ebene und durch Unterschiede in bildgebenden Verfahren, wie der MRT, darstellen.

Die aktuelle Klassifizierung von Hirntumoren der Weltgesundheitsorganisation (WHO) von 2021 (im weiteren Verlauf gekürzt als WHO-Klassifikation bezeichnet) ordnet Gliome anhand der oben genannten Eigenschaften.[6] Dabei werden die Gliome in hoch und niedrig maligne Tumore und in Grade von 1 bis 4 unterteilt.

In der klinischen Praxis findet die Diagnostik von Gliomen anhand der WHO-Klassifikation statt, da sich aus ihr Konsequenzen für Prognoseeinschätzung und Therapieplanung ergeben (Abbildung 1).[7]

Die Prognose von hochgradig malignen Gliomen umfasst nur eine geringe Zeitspanne von 10 bis 17 Monaten.[8, 9]

Therapeutisch kommen bei Gliomen verschiedene Kombinationen aus Radiochemotherapien sowie operative Verfahren zum Einsatz. Das Ausmaß dieser Maßnahmen ist abhängig von WHO-Klassifikation und Patientendisposition.[7, 10]

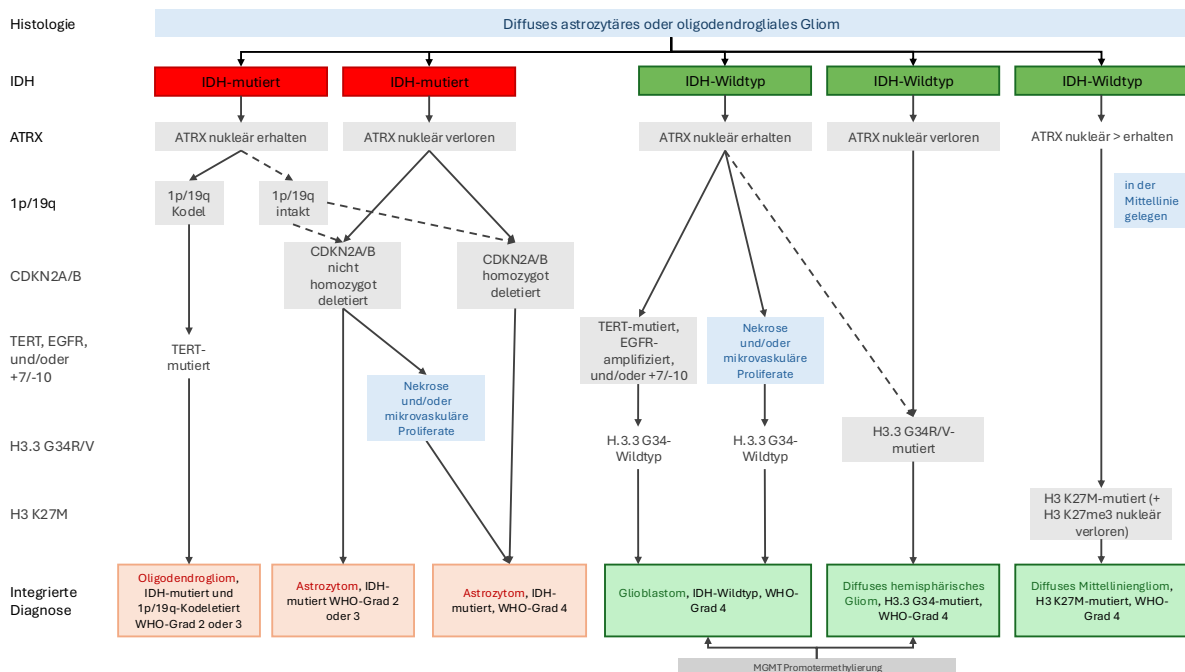


Abb. 1: Diagnostischer Algorithmus für die integrierte Diagnose der häufigsten diffusen Gliome im Erwachsenenalter. Modifiziert nach Weller et al.[7]

Gezeigt werden die verschiedenen molekularen Subtypen die nach radiologischer und histologischer Untersuchung (in Blau hervorgehoben) die genaue Einordnung des Tumors anhand der WHO-Klassifikation ermöglichen. Der besondere Fokus liegt hier auf dem Mutationsstatus der IDH. Um ein Gliom der Gruppe der Astrozytome oder Oligodendrogliome zuzuordnen zu können, ist es notwendig, dass in den Tumorzellen eine Mutation der IDH nachgewiesen werden kann (hier in Rot dargestellt). Liegt keine IDH-Mutation, also ein Wildtyp vor, so folgt eine Zuordnung zu den hochmalignen Glioblastomen oder hemisphärischen bzw. Mittelliniengliomen, mit dem WHO-Grad 4 (hier in Grün dargestellt). Für die vollständige integrierte Diagnose sind Informationen über eine Vielzahl von molekularen Subtypen (in Grau dargestellt), sowie der radiomorphologischen und histologischen Eigenschaften von Nöten.

1.1.2 Epidemiologie

Die aktuellste Ausgabe des *Central Brain Tumor Registry of the United States (CBTRUS)*, dem größten populationsbasierten Register für primäre Hirntumore und andere Tumore des ZNS in den Vereinigten Staaten, zeigt einige Trends in Bezug auf die Epidemiologie von primären Hirntumoren auf.[11]

Während die Inzidenz nicht-maligner Tumorentitäten in den Jahren 2004-2017 gestiegen ist, wurde ein signifikanter Rückgang der Inzidenz von malignen Tumoren verzeichnet. Die altersadjustierte Inzidenz für Glioblastome machte mit 51.2% den größten Anteil an allen malignen Tumorentitäten des gesamten ZNS, und mit 61% den größten Anteil aller Gliome aus und blieb weitgehend konstant. Glioblastome zeigten im neuesten Report des *CBTRUS* eine kurze mediane Überlebensrate von nur 9 Monaten. Andere maligne Gliome im Erwachsenenalter zeigten eine größere Spanne von 21-207 Monate medianen Überlebens.[11]

Einige Studien diskutieren, dass weibliche Geschlechtshormone einen protektiven Einfluss auf die Entstehung von Gliomen haben könnten. Diese These wurde durch

das systematische Review von Grochans et al. unterstützt.[12-14] Es zeigte, dass Männer in den untersuchten Kohorten häufiger von Glioblastomen betroffen sind als Frauen.[12] Außerdem konnte in diesem Review ein Anstieg der Inzidenz von Gliomen mit steigendem Alter korreliert werden. Die höchste Inzidenz wiesen hier Patient*innen in der Altersklasse von 75-84 Jahren auf. Das Review zeigte außerdem, dass mit steigendem Alter die Inzidenz von Glioblastomen ansteige, mit einem Maximum bei 75-84 Jahren.[12]

1.1.3 Diagnostik

Patient*innen mit Gliomen (und anderen soliden Hirntumoren) stellen sich häufig mit erstmaligen Krampfanfällen, fokalen neurologischen Defiziten oder Zeichen eines erhöhten Hirndruckes in behandelnden Kliniken vor. Anamnestisch werden außerdem oft unspezifische Symptome wie Abgeschlagenheit oder Kopfschmerz in der jüngsten Vergangenheit erhoben.[7, 15-17]

Bei klinischem Verdacht auf einen Hirntumor ist laut den aktuellen Leitlinien der *European Association of Neurooncology (EANO)* die Anwendung eines bildgebenden Verfahrens indiziert.

Dafür wird die Verwendung der Magnetresonanztomographie (MRT) empfohlen. Bei der MRT wird, anders als bei der weit verbreiteten Computertomographie, keine Röntgenstrahlung für die Erstellung des Bildes genutzt. Bei der MRT wird die Reaktion von Zellbestandteilen auf einen starken magnetischen Impuls gemessen und anhand von verschiedenen Modalitäten unterschiedlich dargestellt. Dabei können Rückschlüsse auf die Gewebebeschaffenheit geschlossen werden, die entweder besonders hell (hyperintens) oder besonders dunkel (hypointens) dargestellt werden.[18] Die MRT mit T2-gewichteten sowie T1-gewichteten Aufnahmen mit und ohne Gabe eines gadoliniumhaltigen Kontrastmittels gilt als Goldstandard in der Erstdiagnostik eines potenziellen Hirntumors.[7] Mit diesen Aufnahmen kann neben der Lage des Tumors und peritumoralem Ödem, durch Kontrastmittelgabe die Vaskularisierung und das Vorliegen von Tumornekrosen untersucht werden (Abbildung 2.)

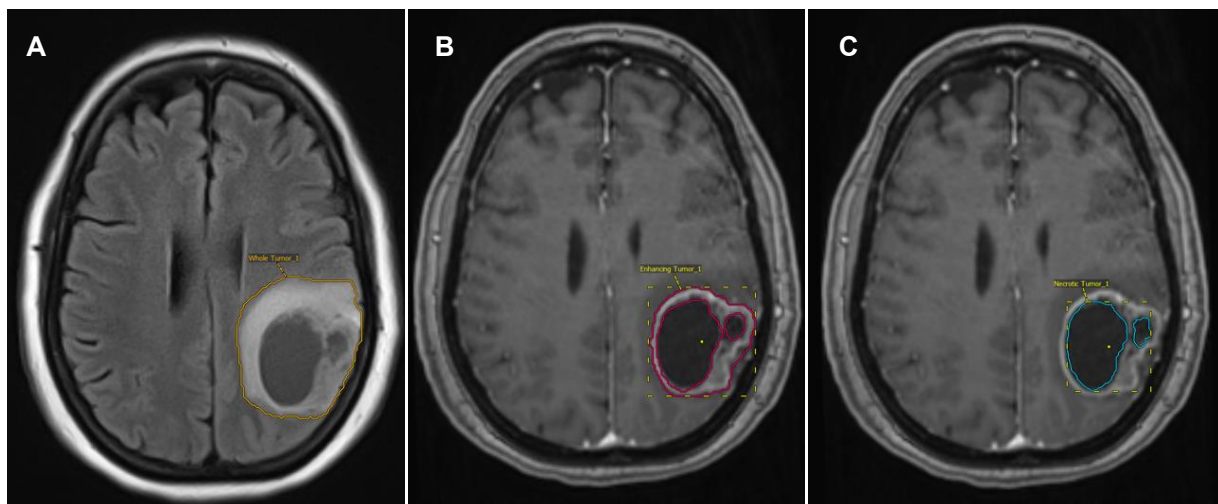


Abb. 2: MRT-Bilder im Transversalschnitt eines Patienten mit Glioblastom in verschiedenen Wichtungen (A-C) mit Segmentierung verschiedener Tumorareale.

A zeigt eine spezielle T2- Wichtung in der pathologische Flüssigkeitsansammlungen besonders hervorgehoben werden können (FLAIR-Wichtung). Hier soll die Ausdehnung des peritumoralen Ödems (hyperintenser Bereich) verdeutlicht werden. Die gelbe Segmentierung umrandet den gesamten Tumor, inklusive Ödem.

B und C zeigen eine T1-Wichtung mit zusätzlicher Gabe eines gadoliniumhaltigen Kontrastmittels. In B wird dadurch die kontrastmittelaufnehmende Portion des Tumors verdeutlicht (hyperintenser girlandenförmiger Bereich), welche Rückschlüsse auf die Vaskularisierung des Tumors zulässt. In C werden die dazu abzugrenzenden bereits nekrotisierten Tumorareale (hypointenser Kern des Tumors) in blau hervorgehoben.

Diese Art der Segmentierung wurde im Rahmen der Studie genutzt, um die durch künstliche Intelligenz gestützte Bildverarbeitung vorzunehmen.

Für die Erstellung eines spezifischen Therapiekonzepts und einer integrierten Diagnose nach der WHO Klassifikation, ist die Entnahme von Tumorgewebe im Rahmen eines neurochirurgischen Eingriffes zur histologischen und molekulargenetischen Untersuchung essentiell.[19]

Die Gewebegewinnung kann entweder mittels einer stereotaktischen Serienbiopsie oder einer offenen Biopsie erfolgen. Häufig werden auch direkt therapeutische Tumorextirpationen durchgeführt, wobei im Zuge dessen Gewebeproben asserviert werden können. In jedem Fall sollte intraoperativ eine Schnellschnittdiagnostik erfolgen, um das weitere operative Vorgehen dem histopathologischen Befund anzupassen.[19]

Unter Berücksichtigung der intraoperativen Risiken sollte zusätzliches Gewebe für weiterführende molekulare pathologische Diagnostik asserviert werden.[7, 19] Die Leitlinie empfiehlt, dass auf die Diagnosesicherung durch die Untersuchung von

Gewebeproben nur in Ausnahmefällen verzichtet werden solle.[7] Das gilt, wenn das Risiko der Operation, beispielsweise durch Nähe zu großen Blutgefäßen und dem damit verbundenen Blutungsrisiko, das durch unvollständige Diagnostik entstehende Informationsdefizit deutlich überwiegt.[19]

Doch auch das Gewinnen einer Gewebeprobe zur histologischen und molekulargenetischen Untersuchung ist kein fehlerfreies Verfahren. Neben dem Risiko und hohen Kosten eines neurochirurgischen Eingriffes, werden Fehlklassifikationen durch Tumorerogenität, gerade in hochgradigen Gliomen beschrieben.[20, 21] Auch technische Limitationen in Bezug auf Klassifikationen der Isocitratdehydrogenase (IDH) wurden in einigen Studien beschrieben.[22-24] Miller et al. zeigten außerdem, dass chirurgische Intervention den Zugang zu Studien mit neoadjuvanten Therapien vermindern könnten.[25]

1.1.4 Therapie

Nach dem Erstellen einer integrierten Diagnose, basierend auf MRT-Bildgebung, Histologie und molekulargenetischem Profil, ist die Behandlung von Gliomen meist multimodal, bestehend aus operativer und anschließender (Radio-)Chemotherapie.[7] Die Entnahme von Gewebeproben in Form einer Biopsie (stereotaktisch oder offen) oder einer gezielten vollständigen Tumorentfernung bildet die Basis der Diagnostik und stellt bei offener Exstirpation die erste therapeutische Maßnahme dar.

Tumor	Primärtherapie
Astrozytom, IDH-mutiert, WHO Grad 2	Resektion oder Biopsie und Verlaufsbeobachtung oder
Oligodendrogliom, IDH-mutiert, WHO-Grad 2	Resektion oder Biopsie und Chemoradiotherapie (mit Vincristin oder Temozolomid) oder Studie
Astrozytom, IDH-mutiert, WHO-Grad 3	Resektion oder Biopsie und Strahlentherapie und Temozolomiderhaltungstherapie bis 12 Zyklen
Oligodendrogliom, IDH-mutiert, WHO Grad 3	Resektion oder Biopsie und Chemotherapie mit Vincristin vor oder nach Strahlentherapie

Astrozytom, IDH-mutiert, WHO-Grad 4	Resektion oder Biopsie und Strahlentherapie (oder kombinierte Chemoradiotherapie mit Temozolomid) und Temozolomiderhaltungstherapie bis 12 Zyklen
Glioblastom, IDH-Wildtyp, WHO-Grad 4 Oder Diffuses hemisphärisches Gliom	Resektion oder Biopsie und Chemoradiotherapie mit Temozolomid und 6 Zyklen Temozolomiderhaltungstherapie (und ggf. TTF), bei älteren, vulnerableren Patienten ggf. Hypofraktionierte Chemoradiotherapie und 12 Zyklen Temozolomiderhaltungstherapie oder ggf. Radio- oder Chemotherapie nach MGMT-Status
Diffuses Mittelliniengliom	Resektion oder Biopsie und Radiotherapie (ggf. Chemotherapie) und Studie

Tabelle 1: Vereinfachte Darstellung der Optionen für die Primärtherapie von Gliomen nach S2K-Leitlinie aus dem Jahr 2021 der Deutschen Gesellschaft für Neurologie. Modifiziert nach Wick et al. 2021[19]

Die Therapiestruktur sollte als Risiko-Nutzen-Abwägung unter Berücksichtigung von klinischen Faktoren wie dem Alter, symptombezogene Beeinträchtigungen (Anhand des *Karnofsky Performance Status*) und Allgemeinzustand aufgestellt werden.[7, 26]

Die Basis der therapeutischen Interventionen stellt die Resektion des Tumorgewebes im Rahmen eines neurochirurgischen Eingriffes dar. Ziel für alle Gliomentitäten ist, vor allem von Grad 2 oder höher, die größtmögliche Masse an Tumorgewebe zu entfernen, ohne dabei neurologische Einschränkungen zu provozieren.[7]

Das Ausmaß der Resektion ist dabei von Lage und Art des Tumors, sowie der Patientensituation abhängig und sollte individuell mit den Patient*innen abgestimmt werden. Studien legen nahe, dass eine vollständige Resektion des in T2-Wichtung hyperintensiven Tumoranteile, in einigen Tumoren das progressionsfreie und allgemeine Überleben steigern kann.[27] So würde bei einem niedriggradigen Astrozytom mit IDH-Mutation, nach der Biopsie auch eine abwartendes Prozedere mit Verlaufskontrolle in Betracht gezogen werden, was bei IDH-Wildtyp Gliomen, wegen ihrer hohe Malignität, nicht mehr in Frage kommen würde.(Tabelle 1)[19]

Nach der operativen Therapie wird für die meisten Gliomentitäten die Radiotherapie empfohlen, welche mindestens drei bis fünf Wochen nach dem operativen Eingriff beginnen sollte. Zeitpunkt sowie Dosis der Therapie, sind abhängig von Alter,

Prognose, *Karnofsky Performance Status*, Klassifizierung des Glioms und Größe des Residualtumors. Typische Strahlendosen liegen bei 50-60 Gy, mit einer Tagesdosis von 1,8-2 Gy.[19] Folgend steht eine Pharmakochemotherapie, meist mit alkylierenden Substanzen wie Temozolomid. Auch hier ist die Verwendung von der klinischen Situation des Patienten und vom Tumortypen abhängig, wobei ältere Patient*innen, die bereits von neurologischen Einschränkungen betroffen sind, eher zurückhaltend chemotherapeutisch behandelt werden.[19] Für die Wahl der Chemotherapie ist bei IDH-Wildtyp Gliomen vor allem der Methylierungsstatus des Promotors der O6-Methylguanin-Methyltransferase (MGMT) innerhalb der Tumorzellen von Bedeutung. Ein methylierter Promotor inaktiviert das zugehörige Gen und suggeriert damit, dass eine Therapie mit Temozolomid effizienter sein kann, als bei Patient*innen mit intakter MGMT.[23] Aktuelle Untersuchungen, beispielsweise von Chai et al. zeigen außerdem, dass der MGMT Status ebenfalls prädiktiven Wert für das Überleben von Patient*innen mit IDH-mutierten Gliomen haben könnte, die mit Temozolomid behandelt werden.[28] Stupp et al. zeigten, dass ein längeres progressionsfreies Überleben erreicht werden kann, wenn bei der Chemotherapie von Glioblastomen mit Temozolomid eine Therapie mit *Tumor Treating Fields* (TTF) ergänzt wird.(Tabelle 1)[29] Diese werden als Haube auf den Kopf gesetzt und sollen durch Generierung eines elektrischen Feldes die Zellteilung des Tumors stören.[29, 30]

Nachsorgeuntersuchungen erfolgen bei Patient*innen typischerweise alle drei bis sechs Monate, wobei Patient*innen mit höhergradigen IDH-mutierten Gliomen typischerweise schon nach drei bis vier Monaten einbestellt werden. Glioblastompatient*innen werden schon nach drei Monaten kontrolliert, bei Anzeichen eines Tumorprogress entsprechend früher. Die Beurteilung erfolgt anhand der *Response Assessment in Neuro-Oncology* (RANO) Kriterien, wo MRT-Bildgebung genutzt wird um das Therapieansprechen zu überwachen.[19]

1.1.5 Prädiktiver Marker: Isocitratdehydrogenase

Der Mutationsstatus der IDH galt als einer der wichtigsten prognostischen Faktoren für das Überleben von Patient*innen mit Gliomen und ist somit für die Erstellung eines spezifischen Therapiekonzeptes für Patient*innen mit Gliomen mittlerweile unerlässlich.[31]

Die IDH ist ein Enzym, das an entscheidenden Schritten im Citratzyklus von Zellen beteiligt ist. Es liegt in verschiedenen Ausprägungen entweder im Zytosol (IDH1) oder Mitochondrium (IDH2/3) von Zellen vor.[25] Physiologisch soll es durch die

Umwandlung von Isocitrat zu alpha-Ketoglutarat die Reduktion von NAD⁺ zu NADH+H zur Energiegewinnung ermöglichen. Das entstandene alpha-Ketoglutarat wird im Citratzyklus weiterverwendet.[32] In manchen Gliomen kann eine somatische Mutation am Codon 132 des Gens der IDH dafür sorgen, dass fehlerhafte IDH-Monomere entstehen. Dimerisieren diese mit intakten IDH-Monomeren, katalysiert dieses Enzym die Umwandlung des entstanden alpha-Ketoglutarat zu 2-Hydroxyglutarat.[33, 34]

Dieses wirkt als Onkometabolit unter anderem für eine epigenetische und genomische Instabilität und eine DNA-Hypermethylierung. Außerdem wird beschrieben, dass das Tumormikromilieu durch 2-Hydroxyglutarat beeinflusst werde, indem es in den Extrazellularraum sezerniert wird und dort die tumorspezifische Immunantwort hemmen soll.[35-38]

Das Vorliegen einer IDH Mutation in Gliomen scheint damit einen prognosebegünstigenden Faktor darzustellen, was durch die Ergebnisse von Yan et al. unterstrichen wird.[33]

Yan et al. beschrieben das Auftreten der Mutation bei primär niedriggradigen Gliomen und sekundär hochgradigen Gliomen, die in der vorherigen WHO Klassifikation als IDH-mutierte Glioblastome bezeichnet wurden (heute IDH-mutierte Astrozytome Grad 4).[33] Zudem wird ein längeres medianes Überleben bei diesen sekundär hochgradigen Gliomen mit einer IDH-Mutation im Vergleich zu Glioblastomen ohne IDH-Mutation beschrieben (31 Monate vs. 15 Monate).[34] Dabei ist zu beachten, dass IDH-mutierte Gliome vor allem bei jungen Erwachsenen (medianes Alter 30-40 Jahre) und IDH-Wildtyp Gliome vor allem im höheren Alter (medianes Alter 50-60 Jahre) auftreten.[31]

Die aktuelle WHO-Klassifikation unterscheidet nun nicht mehr zwischen IDH-mutierten und IDH-Wildtyp Glioblastomen und legt den Nachweis einer IDH-Mutation seit 2021 als krankheitsdefinierender Faktor für Astrozytome und Oligodendrogliome fest (Abbildung 1). Damit ein Tumor als Glioblastom klassifiziert werden kann, muss nun ein IDH-Wildtyp nachgewiesen werden. Der IDH-Mutationsstatus ist damit von einem der wichtigsten prognostischen Markern zu einem unerlässlichen Teil der Hirntumordiagnostik geworden.[6, 7, 19]

In der MRT sind IDH-mutierte Gliome nicht eindeutig von IDH-Wildtyp Gliomen zu unterscheiden. Es wurden jedoch bereits bildmorphologische Unterschiede zwischen IDH-Wildtyp und IDH-mutierten Gliomen beschrieben. In der Literatur werden IDH-mutierte Gliome häufig im Frontallappen lokalisiert [39-42], als allgemein kleiner [43], weniger Gadolinium-kontrastmittelaufnehmend [39, 44], sowie mit erniedrigtem

Blutfluss [45] und besser differenzierbaren Tumorgrenzen [44] beschrieben. Wenn ein solider Tumor eine nahezu komplette Hyperintensität in der T2-Wichtung des MRTs und gleichzeitig eine hypointense Signalausprägung in der FLAIR-Wichtung zeigt (*T2-FLAIR-MISMATCH*), spricht das hochspezifisch für ein IDH-mutiertes Gliom.[46-48] Broen et al. zeigten bei Tumoren mit einem positiven *T2-FLAIR-Mismatch* eine Spezifität von 100%, jedoch war die Sensitivität mit 51% relativ gering.[47] Ein eindeutiges bildmorphologisches Profil, welches sowohl hoch spezifisch, als auch hoch sensitiv eine Unterscheidung von IDH-mutierten und IDH-Wildtyp Gliomen im MRT ermöglicht - beispielsweise durch MRT-Analysen mit maschinellem Lernen - ist folglich von wachsendem Interesse.

1.2 Maschinelles Lernen

1.2.1 Grundprinzipien von maschinellem Lernen

Maschinelles Lernen ist ein großes Teilgebiet der künstlichen Intelligenz (KI), das unter anderem für datenbasierte Vorhersagen genutzt wird. In unserem Alltag ist die Verwendung von maschinellem Lernen allgegenwärtig.[49] Auch in der Medizin wird in verschiedensten Bereichen bereits KI genutzt, um die Patientenversorgung weiter zu verbessern.[2, 50-54]

Grundlegend kann man für unterschiedliche Aufgaben verschiedene Herangehensweisen wählen. Beim überwachten Lernen (*supervised learning*) werden die Trainingsdaten mit vordefinierten Markierungen und Zielwerten versehen. Der Algorithmus lernt anhand dieser markierten Daten Zusammenhänge zu erkennen und eine gewünschte Schlussfolgerung zu ziehen. Dadurch eignet sich ein Modell mit überwachtem Lernen besonders gut für Klassifikations- und Vorhersageaufgaben. Die Daten werden dem Modell wiederholt zur Verfügung gestellt, damit aufgrund der Markierungen die Klassifikation gelernt werden kann. Der Lernprozess entsteht durch die richtige und falsche Klassifikation mit Abgleich zum gewünschten Zielwert. Das Modell soll nun die Markierung eines zuvor unbekanntes Datenpunktes vorhersagen können, ohne dass ihm dieses vorher bekannt war.

Im Gegensatz dazu arbeitet das unüberwachte Lernen (*unsupervised learning*) mit unmarkierten Daten. Das Modell erhält keine vorgegebenen Zielwerte und muss stattdessen eigenständig Muster und Strukturen in den Daten identifizieren, um diese in Gruppen oder Clustern zusammenzuschließen. Unüberwachtes Lernen wird damit

gerne verwendet um Gemeinsamkeiten zwischen den Datenpunkten zu finden und durch Zuordnung in Clustern darzustellen.[55, 56]

Eine Mischform stellt das semi-überwachte Lernen dar, wo sowohl markierte als auch nicht markierte Datenpunkte in einem Datensatz vorliegen.

Das Modell für die Vorhersage von IDH-Mutationen in dieser Studie basiert auf einem überwachten Lernmodell, weshalb diese Methodik im Folgenden näher erläutert wird. Innerhalb des überwachten Lernens können verschiedene Methoden genutzt werden, um Vorhersageaufgaben durchzuführen. Gängige Methoden stellen dabei die *Support Vector Machines* (SVMs), *Artificial Neural Networks* (ANNs) und Entscheidungsbäume (*Decision Trees*) dar. SVMs sind Vektor basierte Methoden, in denen Datenpunkte aufgrund ihrer Eigenschaft in einem Bereich aufgetragen und durch eine Trennlinie (sogenannte *Hyperplane*) aufgeteilt werden, um sie in zwei Klassen zu einzuteilen. Dabei gilt je weiter die Punkte von der *Hyperplane* entfernt, desto generalisierbarer ist die Aussage, die das Modell über die Klassifizierung trifft.[55] ANNs stellen eine Methode dar, die mit Knotenpunkten und mehreren Schichten arbeitet, über die die Daten zu einem Endpunkt gelangen, wobei nur Eingangs und Endschicht dem Betrachter bekannt sind und die zwischengeschalteten Schichten unbekannt bleiben.[55]

Entscheidungsbäume sind als einer der ältesten maschinellen Lernalgorithmen die Basis für viele Modelle die heute für Klassifizierungen und Vorhersagen genutzt werden.[57] Dabei werden die Datenpunkte basierend auf einer vordefinierten Eigenschaft durch den Algorithmus aufgeteilt und dann einer Klasse zugeordnet. Die Technik kann nun von höher entwickelten Algorithmen genutzt werden, um größere und aufeinander aufbauende Entscheidungsbäume zu erstellen, um die Klassifizierung zu optimieren und die Generalisierbarkeit der Zuordnung zu erhöhen. Der *Extreme Gradient Boosting Algorithm* (XGBoost)[57], der für diese Arbeit verwendet wurde, nutzt diese Methodik ebenfalls.

1.2.2 Maschinelles Lernen mit XGBoost

Das von Chen et al. skalierbare Entscheidungsbaum-Modell *XGBoost* stellt ein etabliertes Verfahren für maschinelles Lernen in bereits vielen Facetten unseres Alltages wieder.[57] Es verbindet mehrere einzelne Entscheidungsbaum-Verfahren um eine schwache Klassifizierung Schritt für Schritt zu verbessern. Dies geschieht, indem die Klassifikation der Entscheidungsbäume mit dem Endpunkt verglichen und eine Abweichung vom vordefinierten echten Wert (Residuum) im nächsten

Entscheidungsbaum korrigiert werden soll. Diese Fehlerreduktion soll dafür sorgen, dass das Modell mit jedem Durchlauf zunehmend genauer Klassifizieren kann. In dem Modell werden dafür durch den Anwender steuerbare Hyperparameter genutzt, um diesen Prozess zu optimieren. Folgende Hyperparameter sind in dem Modell enthalten: *n_estimators* stellt die Anzahl der Bäume dar, die das Modell erstellen soll. Wird *n_estimators* größer, so steigt die Datenmenge, die der Algorithmus umsetzen muss und damit die Laufzeit, jedoch potenziell auch die Qualität der Klassifizierung. *n_learning*, die Lernrate, beschreibt wie stark eine Fehlerkorrektur für den nächsten Entscheidungsbaum vorgesehen ist. Hohe Lernraten lassen das Modell schneller präzise Klassifikationen durchführen, erhöhen aber das Risiko einer Überanpassung (oder *Overfitting*) an die Trainingsdaten. Das Modell würde Charakteristiken der Datenpunkte, die für die eigentliche Klassifizierung unerheblich sind, zu stark berücksichtigen. Dadurch lässt sich die Klassifizierung dieser bereits vom Modell gesehenen Daten nicht auf unbekannte Daten übertragen.[58]

Max_depth oder die maximale Tiefe, beschreibt die Anzahl an Zweigen, die die generierten Entscheidungsbäume erstellen. Eine zu hohe Anzahl an Zweigen erhöht das Risiko der Überanpassung, wobei eine kleinere Anzahl die Klassifikation weniger Präzise machen würde.

Um weiterhin Überanpassung zu vermeiden, werden mit *colsample_bytree* für einzelne Bäume nur ein zufällig ausgewählter Teil der Daten verwendet. Dadurch werden nicht immer alle Merkmale in jedem Baum verwendet und die Komplexität des Modells wird verringert, was Überanpassung verhindern und Rechenleistung einsparen soll. *Gamma* beschreibt den Schwellenwert, an dem eine weitere Aufteilung in einem Baum stattfinden soll. *Gamma* legt also fest, ab welchem Wert mit einer neuen Aufteilung keine ausreichende Fehlerreduktion mehr zu erwarten wäre und hindert das Modell daran überflüssig komplex zu werden, sowie Überanpassung zu vermeiden.

1.2.3 Chancen von maschinellem Lernen in der Neuroradiologie

Neben der Vorhersage von IDH-Mutationen wird der Nutzen von maschinellem Lernen in der Neuroradiologie für viele verschiedene Aufgaben untersucht. Verschiedenste Arbeitsgruppen konnten bereits Klassifizierungsaufgaben für Hirntumoren durch künstliche Intelligenz erstellen. So konnten bereits vielversprechende Resultate für Tumorsegmentierung[53], der Vorhersage vom Gesamtüberleben[59], Differenzierung von Gliomen zu verschiedenen Tumorentitäten des ZNS wie beispielsweise

Metastasen[50] oder Lymphomen,[51] Vorhersage des Grades von Gliomen[52] und von molekularen Subtypen erreicht werden.[2, 60, 61] Die Vorhersage molekularer Subtypen ist laut diesen Erkenntnissen zwar möglich, basiert jedoch oft auf stark bereinigten Datensätzen, während sich Daten aus der klinischen Praxis als deutlich heterogener erwiesen. Dies führt dazu, dass in Formaten, in denen die Algorithmen neue Daten auswerten, beispielsweise in den *MICCAI-Challenges*, die Ergebnisse nur schlecht reproduzierbar waren.[2]

1.3 Ziele der Arbeit

Im Rahmen dieser Arbeit wurden Patient*innen-Daten des Yale New Haven Health Krankenhauses (YNHH) verwendet. Ein positives Ethikvotum (IRB-Protokollidentifikationsnummer 2000029055) lag zu Beginn der Datenerfassung vor.

Ziel dieser Arbeit war es, mit Hilfe von maschinellem Lernen eine Vorhersage von IDH-Mutationen in Gliomen an einem ausreichend großen und realitätsnahen Datensatz in präoperativen MRT-Aufnahmen vorzunehmen. Dazu wurde das skalierbare Entscheidungsbaummodell *XGBoost* genutzt, welches schon in vorherigen Arbeiten der *Imagine Quant Research Group* etabliert war. Die verarbeiteten Daten wurden retrospektiv aus MRT-Bildern von Patient*innen mit Gliomen, die zuvor am YNHH behandelt wurden, gesammelt und verarbeitet. Das Hauptaugenmerk dieser Forschungsarbeit lag dabei auf folgenden Punkten:

- Entwicklung eines praktikablen maschinellen Lernmodells auf Basis eines klinisch relevanten und repräsentativen Datensatzes
- Erstellung eines Datensatzes für die externe Validierung des Modells, anhand von retrospektiven Daten eines einzelnen Krankenhauses
- Erörterung der klinischen Implementierung präoperativer Vorhersage von molekulargenetischen Faktoren, anhand der Prädiktion des IDH-Mutationsstatus

2 Enhancing clinical decision-making an externally validated machine learning model for predicting isocitrate dehydrogenase mutation in gliomas using radiomics from presurgical magnetic resonance imaging, Lost J. et al., Neuro-Oncology Advances, Volume 6, Issue 1, January-December 2024, vdae157

Enhancing clinical decision-making: An externally validated machine learning model for predicting isocitrate dehydrogenase mutation in gliomas using radiomics from presurgical magnetic resonance imaging

Jan Lost[✉], Nader Ashraf, Leon Jekel, Marc von Reppert, Niklas Tillmanns, Klara Willms, Sara Merkaj, Gabriel Cassinelli Petersen, Arman Avesta, Divya Ramakrishnan[✉], Antonio Omuro, Ali Nabavizadeh[✉], Spyridon Bakas, Khaled Bousabarah, MingDe Lin, Sanjay Aneja, Michael Sabel, and Mariam Aboian[✉]

All author affiliations are listed at the end of the article

Corresponding Author: Mariam S. Aboian, MD, PhD, 3401 Civic Center Blvd., Philadelphia, PA 19014, 215-590-7000, USA (mariam.aboian@gmail.com).

Abstract

Background. Glioma, the most prevalent primary brain tumor, poses challenges in prognosis, particularly in the high-grade subclass, despite advanced treatments. The recent shift in tumor classification underscores the crucial role of isocitrate dehydrogenase (IDH) mutation status in the clinical care of glioma patients. However, conventional methods for determining IDH status, including biopsy, have limitations. Exploring the use of machine learning (ML) on magnetic resonance imaging to predict IDH mutation status shows promise but encounters challenges in generalizability and translation into clinical practice because most studies either use single institution or homogeneous datasets for model training and validation. Our study aims to bridge this gap by using multi-institution data for model validation.

Methods. This retrospective study utilizes data from large, annotated datasets for internal (377 cases from Yale New Haven Hospitals) and external validation (207 cases from facilities outside Yale New Haven Health). The 6-step research process includes image acquisition, semi-automated tumor segmentation, feature extraction, model building with feature selection, internal validation, and external validation. An extreme gradient boosting ML model predicted the IDH mutation status, confirmed by immunohistochemistry.

Results: The ML model demonstrated high performance, with an Area under the Curve (AUC), Accuracy, Sensitivity, and Specificity in internal validation of 0.862, 0.865, 0.885, and 0.713, and external validation of 0.835, 0.851, 0.850, and 0.847.

Conclusions. The ML model, built on a heterogeneous dataset, provided robust results in external validation for the prediction task, emphasizing its potential clinical utility. Future research should explore expanding its applicability and validation in diverse global healthcare settings.

Key Points

- Machine learning model predicts IDH mutation in gliomas with a high AUC of 0.862
- Promising results in external validation with area under the curve of 0.835.
- Based on heterogeneous data, the model shows potential clinical utility.

Importance of the Study

In contrast to prior studies demonstrating promising results in IDH mutation prediction, this research addresses a critical gap in their translational applicability. Earlier investigations often relied on meticulously curated datasets, limiting generalizability to routine clinical scenarios. Furthermore, the absence of robust external validation undermines the confidence in the broader utility of these models. Our study stands out by

employing a large, heterogeneous dataset reflective of real-world clinical practice, ensuring a more inclusive representation. The success of the ML model validated both internally and externally, accentuates its potential for clinical implementation. By mitigating the limitations of past studies, our research lays a foundation for a more widely applicable and clinically relevant tool for predicting IDH mutation status in gliomas.

Glioma is the most common primary malignant central nervous system (CNS) tumor, with 6 cases of gliomas diagnosed per 100 000 people every year in the United States.^{1,2} Despite advanced treatments combining temozolomide with specialized radiotherapy, the prognosis of high-grade gliomas remains limited, with median survival spanning only 10 to 17 months.^{3,4} Over the past decade, the World Health Organization (WHO) Classification of Tumors of the CNS experienced a paradigm shift away from specific histological alterations towards molecular diagnosis of glioma, further reinforced by the recent 2021 classification.⁵ This recent WHO classification recognizes isocitrate dehydrogenase (IDH) mutation status as a pivotal marker for glioma classification, therapeutic decision-making, and prognosis⁶⁻⁸; hence, it categorizes gliomas into IDH mutant and IDH wild type, which serves as a foundation for determination of their grade.⁵ IDH wild-type gliomas have an aggressive and infiltrative appearance on MR imaging with evidence of blood-brain barrier breakdown and they demonstrate very low overall survival.^{9,10} Conversely, IDH mutant glioma patients typically have a more favorable prognosis characterized by significantly improved survival than those with wild-type gliomas (31 vs. 15 months)^{11,12} and demonstrate heterogeneous imaging characteristics on MRI, such as altering contrast enhancement intensity and tumor necrosis.

In the field of neuro-oncology, imaging plays a crucial role in determining the location of the tumor, defining tumor borders, and providing key information for establishing the diagnosis. Nonetheless, the gold standard for the definitive diagnosis of gliomas occurs via a stereotactic biopsy or tumor resection, followed by histopathological analysis and molecular profiling. This approach, however, has its drawbacks, including the risks of complications, high costs, and potential misclassification due to the spatial heterogeneity of gliomas leading to sampling bias.¹³⁻¹⁵ In addition, the necessity for a biopsy of the tumor limits the availability of neoadjuvant therapy trials to patients, which can potentially influence treatment outcomes. Recent studies suggest that targeted therapy of IDH mutated gliomas is gaining interest and might be an increasingly important staple of glioma treatment in the future.¹⁶ Some studies have also reported that the reliability of IDH mutation testing through pathology is challenged by technical limitations.¹⁷⁻¹⁹

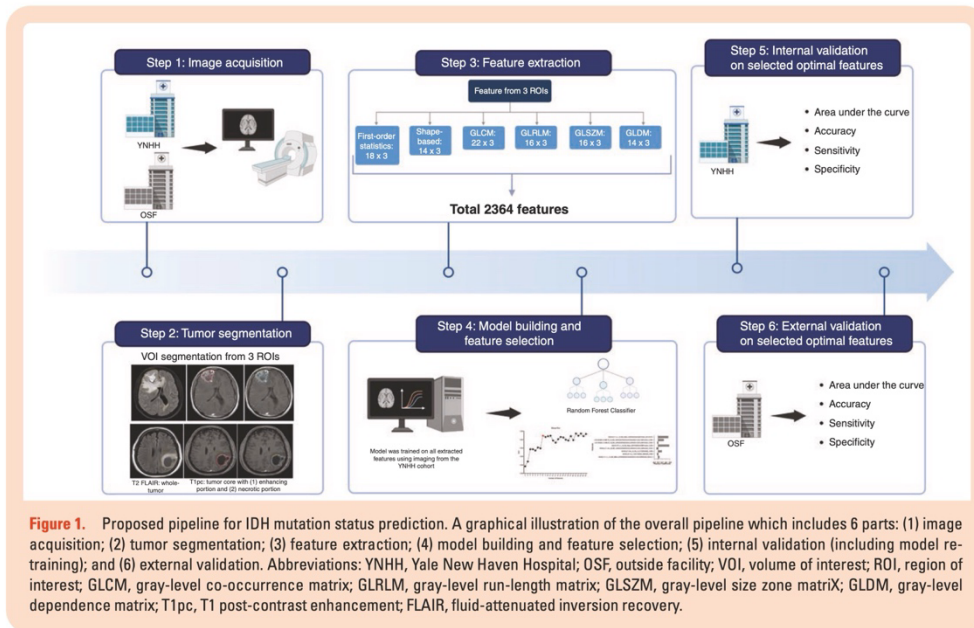
Considering the pitfalls of conventional methods for determining the IDH mutation status, magnetic resonance

imaging (MRI) emerges as a valuable tool in this context by offering a noninvasive technique for the preoperative evaluation of IDH mutation, aiding in directing targeted biopsies, opening consideration for neoadjuvant therapy, screening for potential genetic mutations, and tailoring treatment strategies.²⁰ However, IDH mutation status lacks a distinct radiological imaging profile.²¹ IDH mutant gliomas might exhibit characteristics like T2-FLAIR mismatch sign,²² reduced enhancement, lower blood flow on perfusion-weighted images, increased mean diffusion values, smaller sizes, and a tendency to occur in the frontal lobe.²³ Despite these trends, the sensitivity and specificity for identification of IDH mutation by MRI ranges from 56% to 100% and from 51% to 100%, respectively.²⁴ In cases with a positive T2-FLAIR mismatch sign, the specificity has been reported to be 100%. Hence, there is a growing need for sensitive and specific techniques to accurately predict IDH mutation status in gliomas through MRI.

Recent advances in machine learning (ML) show promising results in several tasks such as tumor segmentation,²⁵ predicting overall survival,²⁶ differentiating gliomas from other CNS malignancies such as metastases²⁷ and lymphomas,²⁸ as well as classifying glioma grades²⁹ and molecular subtypes.^{15,30} While much of the literature in neuro-oncology that applies ML techniques demonstrates high area under the curve (AUC) values within internal datasets, the application of these models to external validation sets remains limited.¹⁵

A significant concern in current research is the reliance on small, meticulously selected ("cherry-picked") datasets without artifacts and well-curated imaging sequences, which poses challenges in applying these findings to the more diverse and complex data encountered in everyday clinical practice. A recent systematic review of ML algorithms for identification of IDH mutation by MRI demonstrated AUC and accuracy of 0.88 and 85% in internal validation and 0.86 and 87% in very limited external validation data sets, reporting studies with incomplete data and high bias.¹⁵ Recognizing these limitations, our study aims to bridge this gap.

By employing a large, annotated dataset encompassing external validation and heterogeneous data reflective of routine clinical scenarios, we aim to set a new standard for the clinical application of ML in neuro-oncology. Given the pivotal role of IDH in both clinical management and prognosis of glioma patients, our approach not only aligns with the current academic focus on noninvasive IDH prediction



but also contributes valuable insights into the real-world applicability of ML in clinical settings. Furthermore, our findings, grounded in a comprehensive and practical dataset, promise to provide a more relatable and transferable benchmark for comparison with ongoing academic efforts in this rapidly evolving field.

Methods

Study Design

This retrospective study was conducted in alignment with the Helsinki Declaration and approved in November 2021 by the Yale New Haven Hospital (YNHH) Institutional Review Board at our institute (IRB protocol ID 2000029055).

Figure 1 shows the overall workflow of our approach, which includes 6 parts: (1) image acquisition; (2) tumor segmentation; (3) feature extraction; (4) model building and feature selection; (5) internal validation (including model retraining); and (6) external validation. We used the Checklist for Evaluation of Radiomic Research (CLEAR), which can be found in [Supplementary Material Appendix 1](#).³¹

Subjects

Patients were categorized based on the location where their imaging scans were performed. Specifically, those who underwent imaging at YNHH were assigned as such and utilized for training and as an internal validation

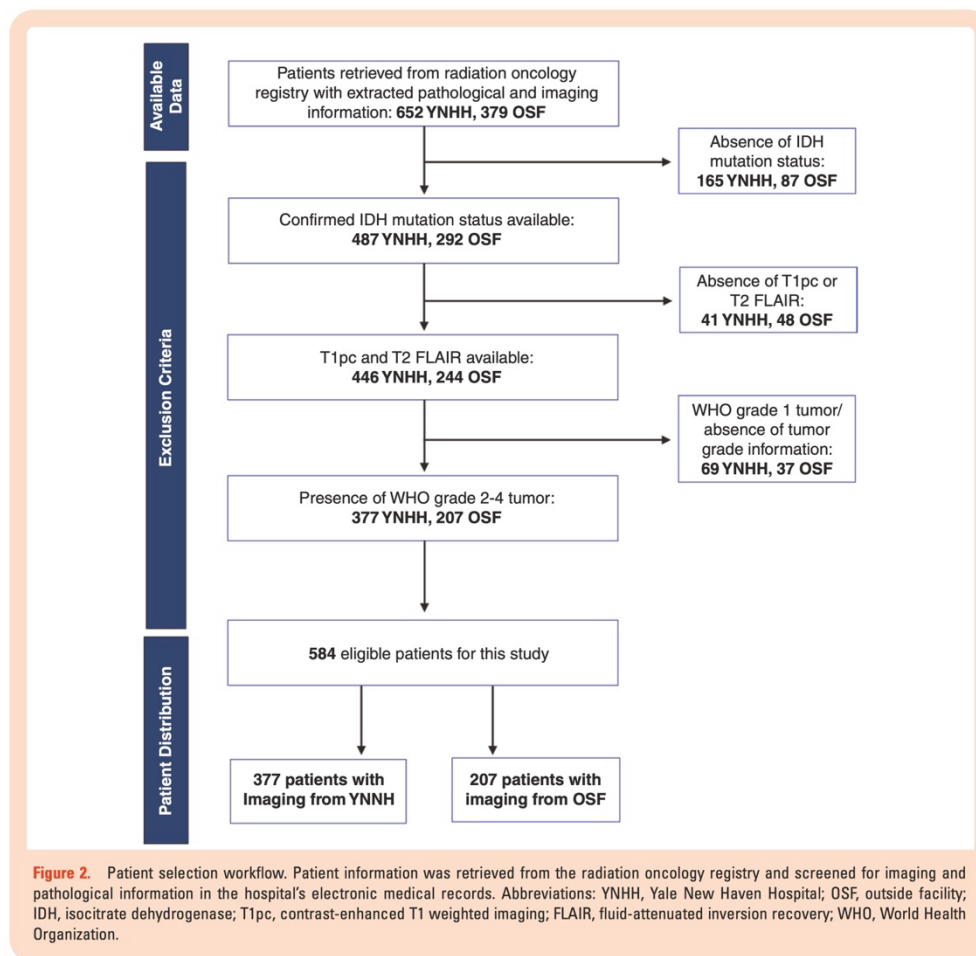
dataset. In contrast, patients whose imaging was conducted at national hospitals outside of YNHH were categorized under the “outside facility” (OSF) group and reserved for external validation, effectively serving as our testing cohort. This stratification strategy was critical in ensuring the robustness and generalizability of our predictive model, as it allowed for the evaluation of the model’s performance in diverse clinical settings.

A total of 1031 patients (652 YNHH and 379 OSF) with available imaging and pathological data from January 2012 to December 2019 were initially collected in our glioma database, with clinical information collected from our hospital’s radiation oncology registry. The necessary MRI and pathology data were extracted from the hospital’s electronic medical records.

To develop an ML pipeline to specifically predict IDH mutation status in glioma, patients were selected based on the following criteria: [I] grades 2–4 diffuse glioma according to the 2021 WHO classification criteria with histopathologically confirmed IDH mutation status, and [2] availability of conventional pretreatment MRI scans consisting at least of both contrasts enhanced T1 weighted imaging (T1pc) and T2 fluid-attenuated inversion recovery (FLAIR). Patients were excluded according to the following criteria: [I] absence of IDH mutation status ($n = 252$), [II] absence of T1pc or T2 FLAIR ($n = 89$), and [III] WHO grade 1 tumor ($n = 20$) or absence of information about tumor grade ($n = 93$).

Figure 2 illustrates the flow diagram of the study population.

Finally, data from 584 patients (219 females and 365 males; age range 5–95) met the inclusion criteria. The cohort was split into 377 cases from YNHH used for model training, and



207 cases under the OSF group for external validation of the prediction pipeline. The clinical characteristics of the selected patients from YNHH and OSF are summarized in Table 1.

Imaging Data Protocol

The incorporation of data from both YNHH-affiliated hospitals and external facilities introduces a diversity of imaging protocols within the ML pipeline. The predominant imaging protocol involves the utilization of 3T imaging systems from Siemens Vervio (Siemens Healthineers AG). Specifically, for FLAIR imaging, the parameters include TE (Echo Time) ranging from 82 to 112 milliseconds, TR (Repetition Time) set at 9000 milliseconds, Inversion Time of 2500 milliseconds, and a slice thickness of 4–5 mm. T1-weighted images were acquired using 2 different pulse sequences: Post Gadolinium Gradient Echo (GRE) with TE ranging from 2.48

to 3.09 milliseconds, TR between 1600 and 1900 milliseconds, Inversion Time set at 900 milliseconds, and a slice thickness of 0.9–1 mm. Post Gadolinium Spin Echo (PGSE) with TE varying from 2.48 to 9.3 milliseconds, TR spanning 277 to 5000 milliseconds, Inversion Time set at 900 milliseconds, and a slice thickness ranging from 1 to 5 mm. Notably, gadolinium-based contrast agents were administered for the T1-weighted images. DICOM metadata extraction was facilitated through the application programming interface of the PACS provider (Visage 7 by Visage Inc.).

Tumor Segmentation

For accurate tumor delineation, we employed a U-Net Deep Learning segmentation tool, integrated into our local PACS System.³² The UNETR was trained using volumetric patches of size 128 × 128 × 64 with a spacing of 1.5 × 1.5 × 2 cm³

Table 1. Clinical Characteristics of All Included Patients (N = 584)

Variable	YNHH (n = 377)	OSF (n = 207)	P-value
Age (years), mean ± SD	58.3 ± 17.5	58.8 ± 15.5	.722 ^a
Sex, n (%)			.982 ^b
Female	142 (37.7)	77 (37.2)	
Male	235 (62.3)	130 (62.8)	
IDH mutation status, n (%)			.653 ^b
Mutated	43 (11.4)	27 (13.0)	
Wild-type	334 (88.6)	180 (87.0)	
Tumor enhancing status, n (%)			.059 ^b
Enhancing	323 (85.7)	164 (79.2)	
Non-enhancing	54 (14.3)	43 (20.8)	
WHO tumor grade, n (%)			.589 ^b
Grade 2	29 (7.7)	21 (10.1)	
Grade 3	43 (11.4)	24 (11.6)	
Grade 4	305 (80.9)	162 (78.3)	
Imaging sequences, n (%)			
FLAIR, PGSE, GRE	279 (74.0)	91 (44.0)	
FLAIR, PGSE	30 (8.0)	94 (45.4)	
FLAIR, GRE	68 (18.0)	22 (10.6)	

a: Assessed using an independent samples t-test.

b: Analyzed using chi-squared tests.

Abbreviations: YNHH, Yale New Haven Hospital; OSF, outside facility; IDH, isocitrate dehydrogenase; WHO, World Health Organization; FLAIR, fluid-attenuated inversion recovery; PGSE, pulsed gradient spin echo contrast-enhanced T1 weighted imaging; GRE, gradient echo contrast-enhanced T1 weighted imaging.

using trilinear interpolation with sampling of the patches. The z-direction was resampled from 1 to 2, compared to a conventional Convolutional Neural Network, which resulted in a batch size of 128 × 128 × 64. The UNETR tool automatically segments glioma tumors on FLAIR imaging, using a combination of T1pc and FLAIR sequences as inputs. The segmentation process identified 3 key volumes of interest (VOI): whole tumor portion (including peritumoral edema) on FLAIR, tumor core (encompassing necrotic and enhancing portions), and exclusively necrotic tumor portion. These segmented VOIs were subsequently mapped onto the T1pc sequences.

Manual adjustments were made by a neuroradiologist to the segmented VOIs to ensure precision, particularly in challenging cases with complex tumor morphology. As part of our inclusion criteria, FLAIR segmentation with peritumoral edema was present in all the cases; however, core and necrotic segmentation on T1pc were present in 487 (83.4%) and 432 (74.0%) cases, respectively. This segmentation methodology aligns with the latest advancements in glioma MRI analysis, ensuring high fidelity in tumor delineation crucial for subsequent feature extraction.

Feature Extraction

Radiomic features were preprocessed and extracted using an open-source Python tool Pyradiomics (version 3.0.1).³⁹

These features encompassed 3 primary categories: (1) volume and shape, (2) intensity, and (3) texture. To further enhance the feature set, advanced image processing techniques were employed. Retrieval performance with the inclusion of 8 high- and low-pass wavelet filters for each of these feature classes was applied to the original images to generate transform-domain images. The following original Pyradiomics features were extracted from each of the 3 VOI, if present: 18 first-order statistics, 14 shape-based features, 22 gray-level co-occurrence matrices, 16 gray-level run length matrices, 16 gray-level size zone matrices, and 14 gray-level dependence matrices. Finally, 2364 features were extracted from the MR imaging of each patient.

ML Pipeline

Our ML approach aims to reliably predict IDH mutation from the extracted radiomic features of MR Images. Our binary prediction pipeline involves an extreme gradient boosting (XGBoost) algorithm for model training.³⁴ In the preprocessing phase, categorical features are converted, and numerical features are normalized using MinMaxScaler from the scikit library. Additionally, correlated features were removed using a custom preprocessing transformer. The resulting features are used in our XGBoost pipeline.

To address the heterogeneous nature of our data, we handle class imbalance using SMOTE (Synthetic Minority Over-sampling Technique), a widely used oversampling technique. The pipeline comprises a 5-fold cross-validation with 10 iterations, ensuring robust model evaluation. The reproducibility seed is set to 123 to ensure consistent results across runs. During the model training process, we employ Randomized SearchCV for hyperparameter tuning. The features used for prediction are then ranked by their median importance for binary classification. The best 25 features are visualized in an elbow plot, where diminishing returns are observed to find the optimal number of features for this model. The Pearson Coefficient is then analyzed for the optimal number of features. The resulting features are then used to re-train the XGBoost model.

The re-trained model is then used for external validation. The data used for external validation from outside hospitals is preprocessed in the same way as the training cohort. Lastly, the pipeline the performance metrics mean AUC, accuracy, sensitivity, and specificity with their respective 95% confidence interval (CI) for each of the 3 steps (Internal validation, internal validation after feature selection, and external validation).

IDH Mutation Assessment

The IDH mutation status was retrieved from our institution's internal electronic medical record pathology reports. Tumor samples were biopsied from a region in the brain that had been confirmed through imaging at YNHH. Specimens with a minimum of 50% of tumor cells in a microdissection target were accepted for analysis. The initial determination of IDH mutation status for all samples involved the utilization of immunohistochemistry (IHC), employing a specific clinically validated antibody (DIA-H09; Dianova GmbH, Hamburg, Germany) targeting the IDH1 R132H mutation.³⁵ For cases in which IDH1 R132H mutation was not detected via IHC, and patients were younger at the time of diagnosis, bidirectional Sanger sequencing was performed using the BigDye Terminator Kit on ABI3130/3730 (Applied Biosystems).

Statistical Analysis

Data for this study were organized and recorded in a Microsoft Excel spreadsheet (Version 16.81), utilizing a combination of descriptive statistics and tabulated presentations. Statistical analyses were performed to derive insights from the collected patient data, employing both the student's *t*-test for comparing means and the chi-square test for assessing associations between categorical variables. Statistical significance was determined by *P*-values < .05. Analysis was conducted using GraphPad Prism software (Version 10.0.03, GraphPad Software, LLC).

Results

Clinical Characteristics of Patients

The study evaluated the clinical characteristics of 584 patients (Table 1), divided into 2 cohorts: training and internal

validation YNNH group with 377 patients (37.7% females and 62.3% males) and external validation OSF group with 207 patients (37.2% females and 62.8% males). The mean age for the YNHH cohort was 58.3 ± 17.46 years, while for the OSF cohort, it was 58.78 ± 15.45 years. IDH mutation status revealed 43 (11.4%) and 334 (88.6%) patients of the YNHH group with IDH mutant (IDHmut) and IDH wild type (IDHwt) gliomas, respectively. Similarly, the OSF group comprised 27 (13.0%) and 180 (87.0%) patients with IDHmut and IDHwt gliomas, respectively. The majority of the patients in the YNHH group ($n = 323$, 85.7%) and OSF group ($n = 164$, 79.2%) had an enhancing tumor on T1pc. In terms of imaging sequences, most patients ($n = 279$, 74.0%) in the YNHH cohort had FLAIR, PGSE, and GRE sequences, whereas this combination was less common in the OSF group ($n = 91$, 44.0%). The FLAIR and PGSE combination were available for 30 (8.0% of YNHH group) 94 (45.4% of OSF group) patients, and the FLAIR and GRE combination for 68 (18.0% of YNHH group), and 22 (10.6% of OSF group) of patients.

There were no statistically significant differences in terms of clinical factors such as age ($P = .722$), sex ($P = .982$), IDH mutation status ($P = .653$), tumor enhancing status ($P = 0.059$), WHO tumor grade ($P = 0.589$) between the 2 groups.

Feature Selection

After extracting 2364 radiomic features from the imaging sequences, we employed an elbow plot analysis to determine the optimal number of features. This analysis revealed that the model's predictive performance, as measured by the AUC, reached an optimal point at 8 features, where the inclusion of additional features resulted in negative or diminishing returns (Figure 3).

Further examination of these 8 features using the Pearson correlation coefficient analysis demonstrated varying degrees of correlation with the outcome variable, with some features showing stronger associations than others (Figure 4). The 8 features selected through this process consist of 7 textures and 1 intensity feature. The detailed characteristics and implications of these 8 radiomic features are thoroughly documented in Supplementary Appendix 2.

Performance of the Classification Model

Table 2 shows the classification results of the different combinations of sequences used in the model. We compared different runs using the available segmentation sequences. The combined approach of "Whole_Flair + Core + Necrotic" sequences showed notable performance metrics across various validation phases and achieved the highest overall performance in external validation (Table 2). Therefore, this combination will be the focus of our analysis. In the training phase, the model achieved an accuracy (ACC) of 0.885 and a sensitivity (SEN) of 0.926, indicating high efficacy in correctly identifying positive cases. The specificity (SPEC), however, was relatively lower at 0.562, suggesting a moderate rate of correctly identifying negative cases. The AUC was 0.846, with a 95% CI (0.827, 0.865),

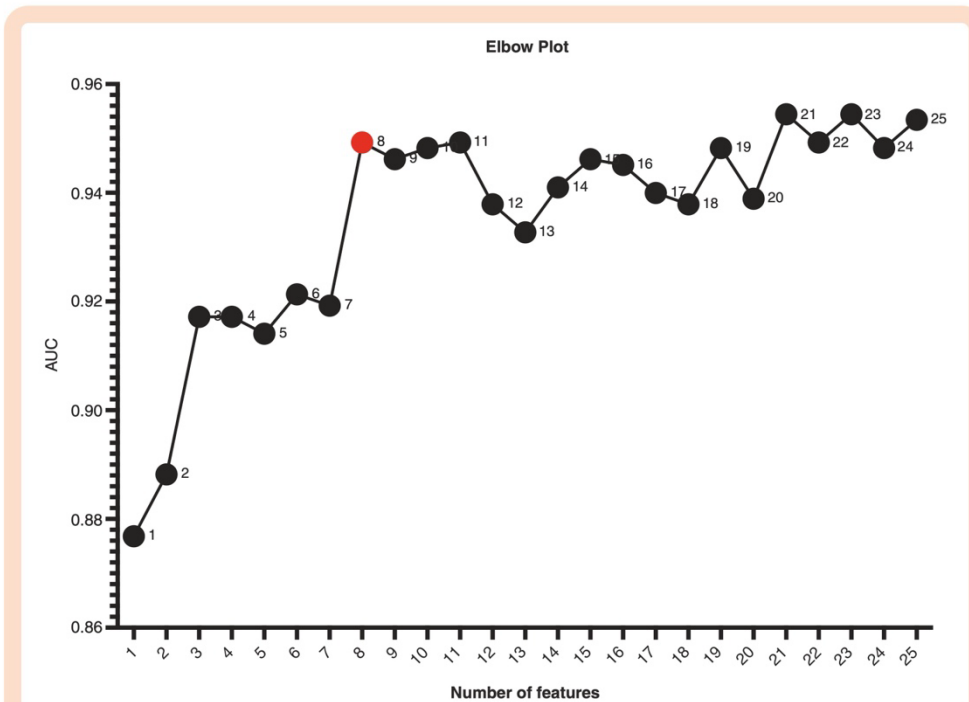


Figure 3. Elbow plot. Elbow plot visualizing the top 25 features ranked by median importance score. Eight is the point of optimal feature selection, as additional features stagnate performance and result in negative or diminishing returns.



Figure 4. Extracted optimal features and weights. Abbreviations: LLL, low low subband filter; GLDM, gray-level dependence matrix; GLCM, gray-level co-occurrence matrix.

reflecting strong model performance. During the internal validation phase, there was a slight decrease in ACC to 0.865 and in SEN to 0.885, but an increase in SPEC to 0.713, indicating improved identification of negative cases. The

AUC remained high at 0.862, with a 95% CI (0.842, 0.881). In the external validation phase, the model's performance metrics further evolved, with the ACC slightly decreasing to 0.851 and SEN to 0.850. However, SPEC saw a significant

Table 2. Evaluation Metrics in the Training and Validation Phases

Sequences	Phase	ACC	SEN	SPEC	AUC	95% CI
Whole FLAIR + Core + Necrotic	Training	0.885	0.926	0.562	0.846	(0.827, 0.865)
	Internal validation	0.865	0.885	0.713	0.862	(0.842, 0.881)
	External validation	0.851	0.850	0.847	0.835	(0.698, 0.978)
Core + Necrotic	Training	0.885	0.910	0.684	0.857	(0.836, 0.876)
	Internal validation	0.872	0.891	0.733	0.891	(0.877, 0.905)
	External validation	0.821	0.833	0.733	0.805	(0.706, 0.884)
Whole FLAIR + Necrotic	Training	0.871	0.908	0.591	0.842	(0.823, 0.860)
	Internal validation	0.862	0.872	0.786	0.872	(0.854, 0.889)
	External validation	0.739	0.744	0.707	0.749	(0.646, 0.906)
Whole FLAIR + Core	Training	0.891	0.928	0.594	0.859	(0.839, 0.878)
	Internal validation	0.850	0.864	0.739	0.864	(0.842, 0.883)
	External validation	0.792	0.789	0.813	0.814	(0.701, 0.973)

Each modality had a different number of optimal features since the optimal number of features suggested by the elbow plot varies from the data used for prediction. AUC, area under the curve.

improvement to 0.847, aligning more closely with the SEN. The AUC was 0.835, within a broader 95% CI (0.698, 0.978). Also, the F1-Score, generally seen as more biased resistant than conventional performance metrics, was noticeably high with 0.908 (Supplementary Appendix 3).

Discussion

The development of our ML model to predict IDH mutation status in glioma represents a crucial step in developing personalized medicine approaches in patients who are treated for these aggressive tumors with limited therapeutic options. The primary focus of our research is the development of an ML model on a clinically relevant dataset and the external validation of the ML model in a heterogeneous dataset. This approach is vital for ensuring the model's clinical applicability and generalizability.^{36,37} The high sensitivity and specificity achieved in our model, particularly in the external validation phase, highlight its potential for clinical deployment. Such performance is crucial for models intended for real-world clinical use, as they must reliably identify both IDH mutant and wild-type gliomas in diverse patient populations.³⁸ Especially in regard to clinical decision-making, aiding in targeted therapy of IDH mutated gliomas, robust predictive models might be of rising interest in the future.¹⁶

Our study's external validation in a diverse clinical setting aligns with recent trends in healthcare modeling, emphasizing the importance of model generalization beyond the initial development environment.^{39,40} The systematic allocation of patients into distinct cohorts for internal and external validation addresses the critical need for models that can adapt to varied clinical scenarios,^{41,42} especially in a clinical context where diagnostic accuracy is paramount.⁴³ This stratification strategy, as seen in our study, enhances the robustness of the predictive model, a key aspect that

has been highlighted in recent studies.^{44,45} The application of our model to an external dataset has provided insights into its performance variability, an aspect often overlooked in ML studies.

The model's high sensitivity in the training phase and its gradual improvement in specificity from the training to the external validation phase emphasize its potential in medical diagnostics. The slight decrease in accuracy and sensitivity from the training to the external validation phase might be attributed to the model's exposure to a more diverse set of data, a common occurrence in ML models.⁴⁶ Contrary to expectations based on prior studies, the specificity was higher in external validation compared to internal validation, which may stem from the relatively smaller size of our external validation set.¹⁵

Moreover, the broader CI in the external validation phase, particularly for the AUC, indicates variability in the model's performance across different external datasets. This variability could be due to differences in data quality, distribution, or other external factors.⁴⁷ Such findings highlight the need for further investigation into the factors influencing the model's performance in diverse clinical settings.

Our research demonstrates the model's high efficacy in identifying positive cases with notable accuracy and sensitivity. However, the variability observed in the external validation phase underscores the challenges faced in ML when models are applied to diverse datasets.³⁶ This finding is crucial for medical diagnostics, as it highlights the need for models that are not only accurate but also adaptable to different patient populations and imaging conditions.

The inclusion criteria, based on the latest WHO classification, along with rigorous imaging data acquisition and tumor segmentation processes within PACS, have contributed to the model's accuracy. The use of advanced radiomic feature extraction, including shape, intensity, and texture characteristics, parallels the methodologies employed in similar studies.^{48,49} The selection of these features using an elbow plot analysis has optimized the

model's performance, as evidenced by its AUC scores in both the internal and external validation phases.

The integration of multi-modal MRI data, as employed in our study, is in line with the current best practices in ML for healthcare applications.^{39,45} The use of advanced radiomics features and a U-Net Deep Learning segmentation tool reflects the state-of-the-art in medical image analysis, ensuring high fidelity in tumor delineation crucial for subsequent feature extraction.^{41,44}

Future research should focus on expanding the model's applicability to other glioma subtypes and incorporating additional clinical variables for more comprehensive predictions. The integration of genomic data, alongside radiomic features, could further enhance the model's predictive capabilities.⁵⁰ Additionally, ongoing validation with larger and more varied external datasets will be essential to continually assess and refine the model's performance.

Limitations

Our study, with its retrospective nature, encounters several limitations. The heterogeneity of the external dataset, particularly in terms of imaging protocols and equipment, may have introduced variability impacting model performance. The sample size, although substantial, may not capture the full spectrum of clinical scenarios, limiting the generalizability of our findings. Additionally, the complexity of the model might challenge clinical interpretability, and the absence of longitudinal data limits insights into long-term efficacy. The study's focus on radiomic features, excluding broader clinical and genomic data, potentially restricts its comprehensive applicability. Finally, further validation in diverse global healthcare settings is necessary to confirm the model's universal effectiveness.

Conclusion

Our study contributes significantly to the field of neuro-oncology by providing a robust, externally validated ML model for predicting IDH mutation status in gliomas. Its successful validation across heterogeneous patient cohorts and clinical settings lays the groundwork for its potential clinical implementation, offering a promising tool for personalized patient management in glioma treatment. The implications of this research extend to potentially improving patient outcomes and informing treatment strategies, marking a pivotal step towards more tailored and effective neuro-oncological care.

Supplementary material

Supplementary material is available online at *Neuro-Oncology Advances* (<https://academic.oup.com/noa>).

Keywords:

gliomas | machine learning | MRI | neuro-oncology

Funding

JL received funding from the doctoral scholarship of Heinrich-Heine-University, American Society of Neuroradiology Fellow Award 2018 (MA). This publication was made possible by KL2 TR001862 (MA) from the National Center for Advancing Translational Science (NCATS), components of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NIH. Arman Avesta is a PhD Student in the Investigative Medicine Program at Yale which is supported by CTSA Grant Number UL1 TR001863 from the National Center for Advancing Translational Science, a component of the National Institutes of Health (NIH). Arman holds securities in Hyperfine. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. MingDe Lin is an employee and stockholder of Visage Imaging, Inc., and unrelated to this work, receives funding from NIH/NCI R01 CA206180 and is a board member of Tau Beta Pi Engineering Honor Society.

Conflict of interest statement

MingDe Lin is an employee and stockholder of Visage Imaging, Inc., and unrelated to this work, receives funding from NIH/NCI R01 CA206180 and is a board member of Tau Beta Pi Engineering Honor Society. Khaled Bousabarrah is an employee of Visage Imaging, GmbH. Michael Sabel is a consultant for Novocure and Codman.

Authorship statement

J.L. conceptualized the study framework, collected/ analyzed data, and prepared the primary manuscript. N.A. contributed to data analysis, and the conception of the study, and prepared the primary manuscript. L.J., M.vR., and K.W. contributed to data analysis, data collection, and manuscript revisions. N.T., S.M., and G.C.P. contributed to data collection and manuscript revisions. A.A. and S.A. contributed to data analysis and manuscript revisions. D.R., A.O., A.N., S.B., K.B., and M.L. contributed to manuscript revisions. M.S. contributed to the study framework and manuscript revisions. M.A. was the primary supervisor of the project and contributed to the study framework, data collection, and manuscript revision.

Affiliations

Department of Neurosurgery, Heinrich-Heine University, Dusseldorf, Germany (J.L.); College of Medicine, Alfaisal University, Riyadh, Saudi Arabia (N.A.); DKFZ Division of Translational Neurooncology at the WTZ, German Cancer Consortium, DKTK Partner Site, University Hospital Essen, Essen, Germany (L.J.); University of Leipzig, Leipzig, Germany

(M.R.); Department of Diagnostic and Interventional Radiology, Medical Faculty, University Dusseldorf, Dusseldorf, Germany (N.T.); University of Leipzig, Leipzig, Germany (K.W.); University of Ulm, Ulm, Germany (S.M.); University of Göttingen, Göttingen, Germany (G.C.P.); Department of Radiology, Massachusetts General Hospital, Boston, Massachusetts, USA (A.A.); Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, Connecticut, USA (D.R.); Department of Neurology and Yale Cancer Center, Yale School of Medicine, New Haven, Connecticut, USA (A.O.); Department of Radiology, Perelman School of Medicine, Hospital of University of Pennsylvania, University of Pennsylvania, Philadelphia, Pennsylvania, USA (A.N.); Division of Computational Pathology, Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, Indiana, USA (S.B.); Visage Imaging, Inc., Berlin, Germany (K.B.); Visage Imaging, Inc., San Diego, California, USA (M.D.L.); Department of Therapeutic Radiology, Yale School of Medicine, New Haven, Connecticut, USA (S.A.); Department of Neurosurgery, Heinrich-Heine-University, Duesseldorf, Germany (M.S.); Department of Radiology, Children's Hospital of Philadelphia (CHOP), Philadelphia, Pennsylvania, USA (M.A.)

References

- Hanif F, Muzaffar K, Perveen K, Malhi SM, Simjee Sh U. Glioblastoma multiforme: A review of its epidemiology and pathogenesis through clinical presentation and treatment. *Asian Pac J Cancer Prev*. 2017;18(1):3–9.
- Mesfin FB, Karsonovich T, Al-Dhahir MA. *Gliomas*. In: *StatPearls [Internet]*. Treasure Island, FL: StatPearls Publishing; 2024.
- Mohammed S, Dinesan M, Ajayakumar T. Survival and quality of life analysis in glioblastoma multiforme with adjuvant chemoradiotherapy: A retrospective study. *Rep Pract Oncol Radiother*. 2022;27(6):1026–1036.
- Molinaro AM, Taylor JW, Wiencke JK, Wrensch MR. Genetic and molecular epidemiology of adult diffuse glioma. *Nat Rev Neurol*. 2019;15(7):405–417.
- Louis DN, Perry A, Wesseling P, et al. The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro Oncol*. 2021;23(8):1231–1251.
- Peng H, Huo J, Li B, et al. Predicting isocitrate dehydrogenase (IDH) mutation status in gliomas using multiparameter MRI radiomics features. *J Magn Reson Imaging*. 2021;53(5):1399–1407.
- Wen PY, Kesari S. Malignant gliomas in adults. *N Engl J Med*. 2008;359(5):492–507.
- Giantini-Larsen AM, Abou-Mrad Z, Yu KK, et al. Treatment and outcomes of IDH1-mutant gliomas in elderly patients. *J Neurosurg*. 2023;140(2):367–376.
- Dang L, White DW, Gross S, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature*. 2009;462(7274):739–744.
- Zhang CB, Bao ZS, Wang HJ, et al. Correlation of IDH1/2 mutation with clinicopathologic factors and prognosis in anaplastic gliomas: A report of 203 patients from China. *J Cancer Res Clin Oncol*. 2014;140(1):45–51.
- Han S, Liu Y, Cai SJ, et al. IDH mutation in glioma: Molecular mechanisms and potential therapeutic targets. *Br J Cancer*. 2020;122(11):1580–1589.
- Hartmann C, Hentschel B, Wick W, et al. Patients with IDH1 wild type anaplastic astrocytomas exhibit worse prognosis than IDH1-mutated glioblastomas, and IDH1 mutation status accounts for the unfavorable prognostic effect of higher age: Implications for classification of gliomas. *Acta Neuropathol*. 2010;120(6):707–718.
- Parker NR, Khong P, Parkinson JF, Howell VM, Wheeler HR. Molecular heterogeneity in glioblastoma: Potential clinical implications. *Front Oncol*. 2015;5:55.
- Friedmann-Morvinski. Glioblastoma heterogeneity and cancer cell plasticity. 2014.
- Lost J, Verma T, Jekel L, et al. Systematic literature review of machine learning algorithms using pretherapy radiologic imaging for glioma molecular subtype prediction. *AJNR Am J Neuroradiol*. 2023;44(10):1126–1134.
- Miller JJ, Gonzalez Castro LN, McBrayer S, et al. Isocitrate dehydrogenase (IDH) mutant gliomas: A Society for Neuro-Oncology (SNO) consensus review on diagnosis, management, and future directions. *Neuro Oncol*. 2023;25(1):4–25.
- Cryan JB, Haidar S, Ramkissoon LA, et al. Clinical multiplexed exome sequencing distinguishes adult oligodendroglial neoplasms from astrocytic and mixed lineage gliomas. *Oncotarget*. 2014;5(18):8083–8092.
- Hegi ME, Diserens A-C, Gorlia T, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med*. 2005;352(10):997–1003.
- Gutman DA, Dunn WD, Jr, Grossmann P, et al. Somatic mutations associated with MRI-derived volumetric features in glioblastoma. *Neuroradiology*. 2015;57(12):1227–1237.
- Fathi Kazerooni A, Bakas S, Saligheh Rad H, Davatzikos C. Imaging signatures of glioblastoma molecular characteristics: A radiogenomics review. *J Magn Reson Imaging*. 2020;52(1):54–69.
- Johnson DR, Guerin JB, Giannini C, et al. 2016 updates to the WHO brain tumor classification system: What the radiologist needs to know. *Radiographics*. 2017;37(7):2164–2180.
- Jain R, Johnson DR, Patel SH, et al. "Real world" use of a highly reliable imaging sign: "T2-FLAIR mismatch" for identification of IDH mutant astrocytomas. *Neuro Oncol*. 2020;22(7):936–943.
- Chow D, Chang P, Weinberg BD, et al. Imaging genetic heterogeneity in glioblastoma and other glial tumors: Review of current methods and future directions. *AJR Am J Roentgenol*. 2018;210(1):30–38.
- Suh CH, Kim HS, Jung SC, Choi CG, Kim SJ. Imaging prediction of isocitrate dehydrogenase (IDH) mutation in patients with glioma: A systematic review and meta-analysis. *Eur Radiol*. 2019;29(2):745–758.
- Niklas Tillmanns AEL, Cassinelli G, Merkaj S, et al. Identifying clinically applicable machine learning algorithms for glioma segmentation: Recent advances and discoveries. *Neurooncol Adv*. 2022;4(1):vdc093.
- Bakas S, Shukla G, Akbari H, et al. Overall survival prediction in glioblastoma patients using structural magnetic resonance imaging (MRI): Advanced radiomic features may compensate for lack of advanced MRI modalities. *J Med Imaging (Bellingham)*. 2020;7(3):031505.
- Jekel L, Brim WR, von Reppert M, et al. Machine Learning applications for differentiation of glioma from brain metastasis—A systematic review. *Cancers*. 2022;14(6):1369.
- Cassinelli Petersen GI, Shatalov J, Verma T, et al. Machine learning in differentiating gliomas from primary CNS lymphomas: A systematic review, reporting quality, and risk of bias assessment. *AJNR Am J Neuroradiol*. 2022;43(4):526–533.
- Bahar RC, Merkaj S, Cassinelli Petersen GI, et al. Machine learning models for classifying high- and low-grade gliomas: A systematic review and quality of reporting analysis. *Front Oncol*. 2022;12:856231.
- Jian A, Jang K, Manuguerra M, et al. Machine learning for the prediction of molecular markers in glioma on magnetic resonance imaging: A systematic review and meta-analysis. *Neurosurgery*. 2021;89(1):31–44.
- Kocak B, Baessler B, Bakas S, et al. CheckList for Evaluation of Radiomics research (CLEAR): A step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMI. *Insights Imaging*. 2023;14(1):75.
- Aboian M, Bousabarah K, Kazarian E, et al. Clinical implementation of artificial intelligence in neuroradiology with development of a novel

- workflow-efficient picture archiving and communication system-based automated brain tumor segmentation and radiomic feature extraction. Original Research. *Front Neurosci.* 2022;16:16:860208.
33. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 2017;77(21):e104–e107.
 34. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016; San Francisco, California, USA. doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)
 35. Capper D, Weisert S, Balsas J, et al. Characterization of R132H mutation-specific IDH1 antibody binding in brain tumors. *Brain Pathol.* 2010;20(1):245–254.
 36. Qin Y, Alaa A, Floto A, Schaar MV. External validity of machine learning-based prognostic scores for cystic fibrosis: A retrospective study using the UK and Canadian registries. *PLOS Digit Health.* 2023;2(1):e0000179.
 37. Youssef A, Pencina M, Thakur A, et al. External validation of AI models in health should be replaced with recurring local validation. *Nat Med.* 2023;29(11):2686–2687.
 38. Du P, Liu X, Wu X, et al. Predicting histopathological grading of adult gliomas based on preoperative conventional multimodal MRI Radiomics: A Machine learning model. *Brain Sci.* 2023;13(6):912.
 39. de Hond AAH, Kant IMJ, Fornasa M, et al. Predicting readmission or death after discharge from the ICU: External validation and retraining of a machine learning model. *Crit Care Med.* 2023;51(2):291–300.
 40. Chen TL-W, Buddhiraju A, Seo HH, et al. Internal and external validation of the generalizability of machine learning algorithms in predicting non-home discharge disposition following primary total knee joint arthroplasty. *J Arthroplasty.* 2023;38(10):1973–1981.
 41. Kunze KN, Kaidi A, Madjarova S, et al. External validation of a machine learning algorithm for predicting clinically meaningful functional improvement after arthroscopic hip preservation surgery. *Am J Sports Med.* 2022;50(13):3593–3599.
 42. Mari T, Asgard O, Henderson J, et al. External validation of binary machine learning models for pain intensity perception classification from EEG in healthy individuals. *Sci Rep.* 2023;13(1):242.
 43. Li G, Li L, Li Y, et al. An MRI radiomics approach to predict survival and tumour-infiltrating macrophages in gliomas. *Brain.* 2022;145(3):1151–1161.
 44. Luo AL, Ravi A, Arvisais-Anhalt S, et al. Development and internal validation of an interpretable machine learning model to predict readmissions in a United States healthcare system. *Informatics.* 2023;10(2):33.
 45. Huang C-Y, Güiza F, Wouters P, et al. Development and validation of the creatinine clearance predictor machine learning models in critically ill adults. *Crit Care.* 2023;27(1):272.
 46. Liu Y, Zheng Z, Wang Z, et al. Using radiomics based on multicenter magnetic resonance images to predict isocitrate dehydrogenase mutation status of gliomas. *Quant Imaging Med Surg.* 2023;13(4):2143–2155.
 47. Dal Bo M, Polano M, Lus T, et al. Machine learning to improve interpretability of clinical, radiological and panel-based genomic data of glioma grade 4 patients undergoing surgical resection. *J Transl Med.* 2023;21(1):450.
 48. Kumar A, Jha AK, Agarwal JP, et al. Machine-learning-based radiomics for classifying glioma grade from magnetic resonance images of the brain. *J Person Med.* 2023;13(6):920.
 49. Kim B-H, Lee H, Choi KS, et al. Validation of MRI-based models to predict MGMT promoter methylation in gliomas: BraTS 2021 radiogenomics challenge. *Cancers.* 2022;14(19):4827.
 50. Xu J, Ren Y, Zhao X, et al. Incorporating multiple magnetic resonance diffusion models to differentiate low- and high-grade adult gliomas: A machine learning approach. *Quant Imaging Med Surg.* 2022;12(11):5171–5183.

3 Diskussion

3.1 Einordnung in den wissenschaftlichen Kontext

3.1.1 Einordnung der Datenbasis und Annotierung

Die nach Stand der aktuellen Leitlinien der WHO-Klassifikation orientierten Einschlusskriterien, sowie die retrospektive Datenerfassung und semiautomatisierte Bilddatenverarbeitung, innerhalb des *Picture Archiving and Communication System (PACS)*, konnten erfolgreich zu der hohen Genauigkeit des hier verwendeten maschinellen Lernmodelles beitragen. Die Prozessierung der Bilddaten erfolgte zunächst durch eine semiautomatisierte Segmentierung in drei Bestandteile. *Whole FLAIR* für den gesamten Tumor, inklusive hyperintens darstellendes peritumorales Ödem in der FLAIR Wichtung, sowie *Core T1* mit dem kontrastmittelaufnehmenden hyperintens Anteil der T1-Wichtung und der *Necrotic T1* mit den hypointensen nekrotischen Gewebeanteilen.

Die Segmentierung der Schnittbilder ist eine verbreitete Annotierungstechnik, da eine Verwendung der vollständigen Schnittbilder eine zu große Datenmenge für Algorithmen darstellt, wobei ein Großteil der Informationen nicht mit der eigentlichen Pathologie zusammenhängen. Auch Wu et al. nutzen eine Segmentierung in Form einer Habitatanalyse, für die hyperintense von hypointensen Arealen in der T1 Wichtung unterschieden wurden.[62] Das könnte ein vergleichbarer Ansatz mit der in der eigenen Forschungsarbeit verwendeten Segmentierungstechnik sein, wobei anders als Wu et al., neben hypointensen und hyperintensen Tumorarealen (*T1 Core* und *T1 Necrotic*) noch eine FLAIR basierte peritumorale Ödemsegmentierung etabliert wurde. Van der Voort et al. nutzten ebenfalls eine manuelle Segmentierung für die Datenreduktion.[63] Beide Studien führten die Segmentierungen manuell durch und erzielten hohe Klassifizierungsergebnisse.

Aboian et al. zeigten jedoch in ihrer Studie, dass manuelle Segmentierungsansätze für die Annotierung von Datensätzen in Echtzeit im klinischen Alltag zu ineffizient und zeitaufwendig für Neuroradiolog*innen seien und dadurch nicht realistisch in die Praxis zu überführen wären. Die semiautomatisierte Segmentierung hingegen erreichte in kurzer Zeit ähnliche DICE-Scores verglichen mit Segmentierungen die manuell durchgeführt wurden.[54] In dieser Arbeit wurden deswegen ein im klinischen PACS integrierter Algorithmus für die Segmentierung verwendet. Im Anschluss wurden die Daten von erfahrenen Neuroradiolog*innen überprüft und korrigiert. Dieser zeit- und dateneffiziente Aufbau der Datenbasis machte die große Ansammlung an verfügbaren

Bilddaten erst möglich.[53, 54, 60] Aufgrund dessen sollte, um eine Transition der Vorhersage von molekularen Subtypen in Gliomen in klinische prospektive Untersuchungen zu ermöglichen, die Effizienz der Datenaufarbeitung unbedingt mitbedacht werden.[54] Die Segmentierungsgenauigkeit und Schnelligkeit von manuellen und automatisierten Konzepten aus dieser Forschungsarbeit sollten weiterführend einander gegenübergestellt werden, um die genannten Vor- und Nachteile zu quantifizieren.

Die Verwendung dieses semiautomatisierten Ansatzes stellt eine erfolgreiche Weiterentwicklung für diese Art von retrospektiven Studien dar, die von weiterführenden Arbeiten genutzt werden kann, um generalisierbarere maschinelle Lernmodelle zu erstellen, ohne große multizentrische Kollaborationen zu benötigen. Die hohe Anzahl von 584 Patient*innen innerhalb der Studie ist damit auch für kleinere Zentren ein realistischer Rahmen, die ähnliche Studien durchführen wollen. Van der Voort et al. haben mit ihrer großen multizentrischen Studie auch ohne semiautomatisierte Annotierung einen sehr großen Datensatz für ihr Modell entwickelt. Durch ein multizentrisches Studiendesign mit 16 Instituten wurde ein Trainingsdatensatz von 1.508 Gliompatient*innen für das Modelltraining rekrutiert und an 240 Patient*innen von 13 verschiedenen Instituten unabhängig getestet. Im Gegensatz dazu sind die Trainingsdaten für diese Studie aus einem Krankenhaus und lediglich die Testdaten aus externen Institutionen. Dadurch kann die Studie von Van der Voort et al. eine deutlich größere Datenbasis vorweisen.[63] Da jedoch mehrere Klassifizierungsaufgaben erfüllt werden sollten, wurde ein Großteil der Patient*innen eingeschlossen, obwohl keine Information über den IDH-Mutationsstatus vorlagen. Die Anzahl an Patienten mit IDH-Mutationsstatus war dadurch mit 666 im Trainings und 217 im Test Datensatz schon eher vergleichbar mit der hier verwendeten Patientenkohorte.[63] Außerdem bestand die Datenbasis zu einem großen Teil aus Bilddaten aus öffentlichen Datensätzen die bereits kuratiert wurden.[63] Während durch diese große multizentrische Datenbasis bestehend aus Datenbanken und kollaborierenden Krankenhäusern sehr gute Ergebnisse erzielt werden konnten, ist die Entwicklung solcher Modelle anhand von prospektiven Untersuchungen für einzelne Arbeitsgruppen nur schwer reproduzierbar.[2, 54]

Eine limitierende Hürde für die Entwicklung verlässlicher maschineller Lernalgorithmen scheint die Verwendung solcher großen multizentrischen Datenmengen zu sein, weswegen sich diese Forschungsarbeit darauf fokussiert, ein Modell zu entwickeln, welches nicht auf solche angewiesen ist.[50, 54] Um diese Limitation zu adressieren,

aber das Modell trotzdem anhand eines externen Validierungsdatensatzes testen zu können, wurden Bilddaten als externer Testdatensatz verwendet, die vor der Aufnahme in der behandelnden Klinik entstanden sind. Die Rekrutierung eines externen Datensatzes ist entscheidend, um mit geringerem Risiko der Datenverzerrung (engl. *Bias*), also einer zu hohen Ähnlichkeit der Validierungsdaten zu den Trainingsdaten, die Klassifizierungsstärke des Modells testen zu können.[64-66]

Dieses Vorgehen scheint bisher in keiner vergleichbaren Studie genutzt worden zu sein und stellt ein Alleinstellungsmerkmal dieser wissenschaftlichen Arbeit dar.[2, 53, 54] Eine so große Datenaufbereitung innerhalb des lokalen PACS zu gewährleisten hebt den Mehrwert dieser Studie für die klinische Integration von maschinellen Lerntechniken hervor.[2]

3.1.2 Einordnung der Datenverarbeitung und des Klassifizierungsmodells

Die segmentierten Bilddaten aus der Datenbasis wurden im Anschluss für das maschinelle Lernmodell in die numerischen Bildmerkmale von *PyRadiomics*[67] umgewandelt, um die Daten zu bündeln. Diese bestehen aus Form-, Intensität- und Textur-Charakteristika und sind bereits erfolgreich von einer Vielzahl vergleichbarer Studien genutzt worden.[68, 69] Alternativ speisen manche die Pixel- oder Voxeldaten aus den MRT-Schnittbildern direkt in ihr Modell ein, wobei dafür ein *Deep Learning* Algorithmus von Nöten ist, der so hoch dimensionale Datenmengen verarbeiten kann. Van der Voort et al. nutzten erfolgreich ein *Convolutional Neural Network* um neben einer molekularen Typisierung auch Graduierungsvorhersagen durchzuführen.[63] In dieser Arbeit wurde der skalierbare maschinelle Lernalgorithmus *XGBoost* genutzt, der in Klassifizierungsaufgaben in vergleichbaren Arbeiten vielversprechende Ergebnisse erzielen konnte. Die Verwendung von *Deep Learning* Algorithmen zeigt sich jedoch auch als äußerst vielversprechend, konnte in einigen Studien sogar bessere Ergebnisse erzielen. Dies zeigt die Vielseitigkeit von maschinellen Lernmodellen und dass sehr viele Faktoren auf die letztendliche Klassifizierungsstärke des Modells Einfluss haben.[2]

Da die verwendeten Daten in Bezug auf die Verteilung von IDH-mutierten und IDH-Wildtyp Gliomen sehr ungleichmäßig verteilt war, wurde eine synthetische Generierung von zusätzlichen Datenpunkten in der Minderheitsklasse (der IDH-mutierten Gliomen) vorgenommen, um eine einseitige Klassifizierung zugunsten der IDH-Wildtyp Gliome zu vermeiden. Dafür wurde der bereits vielfach genutzte *SMOTE*

Oversampler in das Modell integriert.[70] Die hohe *Area Under the Curve* (AUC) in der externen Validierung zeigt, dass der *Oversampler* die Klassifizierung der Minderheitsklasse erfolgreich in der Trainingsphase verbessert hat und das Modell nun auch diese robust in einem neuen Datensatz klassifizieren kann.

Die Selektion von besonders entscheidenden Merkmalen für die Vorhersage durch die Verwendung einer *Elbow-Plot* Analyse hat die Klassifizierung sichtbar erhöht.[1]

Die Verwendung aller drei verfügbaren Segmentierungen der Tumoranteile in der FLAIR und T1-Wichtung hat sichtbar zu den besten Ergebnissen geführt und war der Verwendung einzelner Modalitäten überlegen. Dies wird vor allem an der hohen AUC von 0.835 in der externen Validierung deutlich.[1] Auch Jian et al. kamen in ihrer Metaanalyse zu dem Ergebnis, dass ein multimodaler Ansatz bei vergleichbaren Studien der Erfolgreichste ist.[60]

Die Datenbasis für dieses Projekt besteht aus Routine MRT-Modalitäten die im regelhaften diagnostischen Prozess bei allen Patient*innen mit Gliomen erhoben werden. Dies stellte sicher, dass unser Modell eine hohe Generalisierbarkeit aufweist, da die verwendeten Modalitäten in der derzeitigen Leitlinie die empfohlenen präoperativen Bildgebungen darstellen.[7] Die Arbeit von Park et al. zeigt jedoch auf, dass die Nutzung von diffusions- und perfusionsbasierter Bildgebung in eine weitere Verbesserung bei *Radiomics* basierten Vorhersagen resultieren kann, obwohl diese Techniken in der Initialdiagnostik aktuell nicht vorgesehen sind.[19, 71] Zusätzlich konnten Wu et al. mit ihrem auf *Chemical exchange saturation transfer* (CEST) basierendem maschinellem Lernmodell erfolgreich IDH-mutierte von IDH-Wildtyp Gliomen unterscheiden.[72] Der Nutzen von spezialisierter Bildgebung für die Vorhersage des IDH Mutationsstatus sollte deswegen in weiteren Studien der ausschließlichen Verwendung von Routinebildgebung gegenübergestellt werden.

3.1.3 Einordnung der Resultate

Entsprechend der primären Endpunkte konnten ein maschinelles Lernmodell auf einer realistischen und heterogenen Datenbasis entwickelt werden und in einer externen Validierung verlässliche Ergebnisse erzielen. Dieser Ansatz ist insbesondere relevant für die Entwicklung eines Modells, das in die klinische Praxis überführt werden soll.[73, 74]

In der externen Validierung erreichte das Modell eine AUC von 0.835, eine Genauigkeit von 0.851, eine Sensitivität von 0.850 und eine Spezifität von 0.847.[1] Diese Werte sind vergleichbar mit den Ergebnissen anderer Studien zur Vorhersage von IDH-

Mutationen und molekularen Subtypen in Gliomen.[2, 60] Die systematische Aufteilung der Patientenkohorten aus verschiedenen klinischen Kontexten in interne und externe Validierungsgruppen entspricht geltenden Trends der Forschung und zeigt die notwendige Anpassungsfähigkeit von maschinellen Lernmodellen in Bezug auf vielfältige Datenansammlungen, die besonders relevant für die klinische Integration dieser Modelle ist.[64-66, 75]

Die Sensitivität und Genauigkeit des Modells zeigte sich vor allem in der Trainingsphase als außerordentlich hoch und konnte im weiteren Verlauf der Validierung ein relativ hohes Niveau halten. Das graduelle Abfallen der Sensitivität und Genauigkeit bei externer Validierung ist ein typisches Phänomen bei maschinellen Lernmodellen und auf die Verwendung von neuen, dem Modell gänzlich unbekannt, Datenpunkten zurückzuführen. Dadurch wird natürlicherweise das Maß an falsch-negativen Klassifizierungen erhöht.[66, 76]

Anders als in vorherigen Studien, zeigte sich die Spezifität initial niedrig und verzeichnet auch in der externen Validierung ein Ansteigen im Vergleich zur Trainingsphase, was möglicherweise darauf zurückzuführen sein könnte, dass die externe Validierungskohorte kleiner als die Trainingskohorte ist.[2]

In der externen Validierung zeigen sich die Konfidenzintervalle mit einer breiteren Streuung als im Modelltraining und der internen Validierung. Dies könnte ein Indikator für die Variabilität der Datenqualität und Verteilung der Patientenkohorten, oder anderer externen Faktoren sein.[77] Es sollten weitere Untersuchungen unternommen werden, um diese Einflüsse aus vielfältigen Patientendaten auf Lernmodelle zu erörtern.

Zusammenfassend ist das Modell mit hoher Sensitivität und Genauigkeit erfolgreich in der sicheren Klassifizierung von richtig-positiven Daten, jedoch mit erhöhter Variabilität in der externen Validierung. Dies macht die Herausforderungen von maschinellem Lernen in vielfältigen und heterogenen klinischen Patientengruppen deutlich.[73] Weiterhin unterstreicht es, dass Modelle, die für die unmittelbare Patientenversorgung entworfen werden, nicht nur genau, sondern auch Anpassungsfähig in Bezug auf hochvariable Daten, in Form verschiedener Patientenkohorten und Bildgebungsprotokollen sein müssen.

3.2 Potenzieller Nutzen für die Patientenversorgung

Während maschinelles Lernen in vielen Facetten des alltäglichen Lebens bereits fest integriert und auch in der Medizin angekommen ist, besteht die wissenschaftliche

Landschaft für die Vorhersage molekularer Subtypen in Gliomen noch weitgehend aus retrospektiven Arbeiten.[2, 60, 78] Bisher fehlen weitreichende prospektive Studien im klinischen Rahmen, um den Nutzen dieser Techniken in der Praxis zu überprüfen.

Ein gänzlich Verzicht auf molekularpathologische Untersuchungen, beziehungsweise ein Ersetzen dieses Vorgehens in naher Zukunft, ist aufgrund der niedrigen Morbidität und Mortalität einer stereotaktischen Serienbiopsie vorerst nicht realistisch.[19] Damit nicht-invasiv, also basierend auf Vorhersage durch Bildgebungsanalyse, der molekulare Status berücksichtigt wird, sollten einige Faktoren erfüllt werden. Zunächst sollte das genutzte Modell ausreichend verlässlich sein, im Sinne einer hohen Sensitivität, Spezifität und AUC.[2] Diese Eigenschaften sollten, wie bereits in Abschnitt 3.1.1 und 3.1.3 ausführlich erörtert, an externen und heterogenen Datensätzen nachgewiesen worden sein, um eine erfolgreiche Anwendung an dem Modell unbekanntem Daten nachweisen zu können.[64-66] Diese Modelle sollten dann in longitudinalen prospektiven Studien validiert werden. Durch die bereits genannten Probleme der fehlgeschlagenen Transition der Ergebnisse solcher Studien in prospektive Settings ist dieser Schritt als nächstes besonders wichtig. Die nicht invasive Vorhersage könnte vor allem bei Patient*innen mit besonders hohem Interventionsrisiko eine Unterstützung in der Tumorkonferenz darstellen. Weiterhin bietet sich gerade bei IDH-mutierten Astrozytomen vom Grad 2 ein zunächst konservatives Therapieschema an, wodurch bei einer sicheren Vorhersage keine invasive Biopsie für weitere Diagnostik benötigt wäre. Das Interventionsrisiko und das Narkoserisiko müssten dann nicht eingegangen werden, was gerade bei multimorbiden Patient*innen im hohen Alter von Vorteil sein könnte.[79] Dabei ist jedoch zu bedenken, dass solche niedriggradige Gliome vor allem bei jüngeren Patient*innen auftreten.[1, 2]

Vor allem die Indikation für offene Biopsien oder Tumorextirpation mit Kraniotomien sollte bei multimorbiden Patient*innen kritisch gestellt werden.[80] Auch bei IDH-mutierten Astrozytomen Grad 4 und Glioblastomen könnte durch eine solche besonders frühe Unterscheidung das therapeutische Vorgehen weiter individualisiert werden. Bei Patient*innen mit besonders hohem Risiko für postoperative neurologische Defizite könnte eine weniger aggressive und Parenchym schonendere Operationstechnik gewählt werden.

Der therapeutische Nutzen des IDH-Inhibitors Vorasidenib wurde bereits im Jahr 2023 durch Mellinghoff et al. nachgewiesen.[81] Derzeit wird die bereits präoperative Applikation von IDH-Inhibitoren untersucht, wobei ein Rückgang des Onkometaboliten

2-Hydroxyglutarat bei niedriggradigen Gliomen bereits nachgewiesen werden konnte.[82] Eine präoperative Vorhersage der IDH könnte in Zukunft Therapieregime frühzeitig in diese Richtung lenken.

Abgesehen von der IDH könnte eine Vorhersage des MGMT-Methylierungsstatus die Therapieschemata beeinflussen, da durch diesen eine Aussage auf den Therapieerfolg der Chemotherapie geschlossen werden kann. MGMT zeigt sich in mehreren Studien jedoch als schwer vorherzusagender Marker, wodurch noch weitere Untersuchungen von Nöten sind.[2, 60, 63]

3.3 Limitationen

Diese retrospektive Studie weist einige Limitationen auf. Während die Verwendung heterogener Daten, insbesondere des externen Validierungsdatensatzes, das maschinelle Lernmodell mit realitätsnahen Herausforderungen konfrontiert, erschwert die große Variabilität der Bilderfassungsprotokolle und des radiologischen Equipments eine genaue Einschätzung ihres Einflusses auf die Ergebnisse.

Obwohl die Studie mit einem Patientenkollektiv von 584 Patient*innen, darunter 207 aus einem externen Validierungsdatensatz, als groß einzustufen ist, könnten bestimmte Facetten des Patientenspektrums nicht ausreichend repräsentiert sein. Dies könnte die Generalisierbarkeit der Ergebnisse einschränken. Für verlässlichere Rückschlüsse auf die klinische Praxis wären aufgrund der Komplexität des Modells weiterführende Studien mit einem noch breiteren Datensatz und einer noch größerer Patientenkohorte erforderlich.

Da IDH-mutierte Gliome im Vergleich zu Glioblastomen mit IDH-Wildtyp verhältnismäßig selten sind, sind Patient*innen mit IDH-mutierten Gliomen in den Kohorten weniger häufig vertreten als die IDH-Wildtyp Glioblastome. Durch die Verwendung des *SMOTE Oversamplers* wurde diesem Verzerrungseffekt entgegenwirkt, um beide Klassen künstlich auszugleichen. Obwohl dies eine wirksame und weit verbreitete Möglichkeit darstellt, um maschinellen Lernmodelle an unbalancierten Datensätzen zu trainieren, kann das Risiko der systematischen Verzerrung nicht gänzlich behoben werden. Außerdem birgt dies die Gefahr einer Überanpassung anhand der künstlichen Merkmale. Dieser Effekt konnte jedoch durch externe Validierung gering gehalten werden.[70]

Darüber hinaus sind Aussagen über den Langzeitnutzen dieser Arbeit begrenzt, da longitudinale Daten aus einer Verlaufsstudie fehlen, die eine retrospektive Analyse nicht vollständig bieten kann. Zudem könnte die umfassende Anwendbarkeit dieser

Ergebnisse durch die fehlende Einbeziehung klinischer und genomischer Daten eingeschränkt sein, da die Vorhersagen ausschließlich auf quantitativen Bildmerkmalen, in Form von *Radiomics*, basieren.

Obwohl die Auffassung vertreten wird, mit der vorliegenden Patientengruppe ein möglichst repräsentatives Abbild des gängigen Patientenkollektivs geschaffen zu haben, sind für eine universelle Aussage über die Effektivität des Algorithmus weitere Validierungen in diverseren und globalen klinischen Kontexten erforderlich.

3.4 Zukunftsperspektiven

Im Rahmen dieser Arbeit wurde ein praktikabler Arbeitsfluss für die Vorhersage anhand von retrospektiven Bilddaten entwickelt. Der Fokus lag dabei vorrangig auf der Entwicklung eines reproduzierbaren sowie effizienten Modells und weniger auf einem Methodikvergleich im Rahmen einer *Benchmark*-Studie. Die dafür genutzten Techniken wurden vorher innerhalb der Arbeitsgruppe entwickelt und der Nutzen für die klinische Praxis wurde nun in verschiedenen Konzepten überprüft.[54, 83] Weiterführende Studien sollten verschiedene maschinellen Lernalgorithmen und Datenprozessierungstechniken vergleichen, um die bestmögliche Klassifizierung zu eruieren.

Um den klinischen Nutzen der Methodik und des Modelles weiterführend zu prüfen, sollten zukünftige Arbeiten diese Techniken in weiteren Bereichen anwenden. Zukünftige Studien sollten dieses Modell auf weitere molekulare Subtypen in Gliomen oder anderen Tumoren ausweiten. Für Gliome wäre besonders der MGMT-Methylierungsstatus von Interesse, wobei in vergleichbaren Studien weniger starke Ergebnisse als die IDH-Mutationsstatus Vorhersage erzielt werden konnten.[2, 63] Obwohl der MGMT-Methylierungsstatus derzeit vor allem für IDH-Wildtyp Gliome Relevanz für das therapeutische Vorgehen hat, suggerieren einige Studien, dass auch Rückschlüsse auf das Gesamtüberleben unter Chemotherapie bei IDH-mutierten Gliomen geschlossen werden können.[7, 28] Dies würde den Nutzen einer sicheren präoperativen Vorhersage des MGMT-Methylierungsstatus für die Patientenversorgung unterstreichen.

Andere Studien haben zusätzlich zu Bilddaten klinische Charakteristika wie das Alter und Geschlecht genutzt und damit vielversprechende Ergebnisse erzielt.[60, 84, 85] Während dieser Umstand den thematischen Rahmen dieser Arbeit überschritt, wäre eine weitere Untersuchung in fortlaufenden Studien über den Nutzen von klinischen

Daten für die Vorhersage von molekularen Subtypen sinnvoll um eine dadurch stattfindende Verbesserung der IDH-Vorhersage zu Untersuchen.

Auf Basis dieser Arbeit wäre es empfehlenswert, dass andere Arbeitsgruppen, die den Aufbau eines maschinellen Lernmodells anstreben, die hier verwendete Methodik für den Aufbau ihrer Datenbasis nutzen.

Durch die Rekrutierung von Patient*innen aus einem einzelnen Krankenhaus, jedoch mit Bilddaten aus verschiedenen Zentren, sowie der PACS integrierten semiautomatischen Bildverarbeitung, wurde ein effizientes und ressourcenarmes Vorgehen etabliert, welches auch von kleineren Arbeitsgruppen problemlos übernommen werden könnte.[54]

Da sich die Ergebnisse als zufriedenstellend herausgestellt haben, sollte der Nutzen dieser Klassifizierungen nun weiterführend in longitudinalen Studien innerhalb der klinischen Praxis evaluiert werden.

3.5 Schlussfolgerungen

Durch eine AUC von 0.835, Genauigkeit von 0.851, Sensitivität von 0.850 und Spezifität von 0.847 konnte das zuvor aufgestellte Ziel, ein extern validiertes maschinelles Lernmodell, welches den IDH-Mutationsstatus in Gliomen anhand von MRT-Bildgebung verlässlich vorhersagt, zu erstellen erfolgreich erreicht werden. Die weiteren Endpunkte konnten ebenfalls erreicht werden, in dem ein retrospektiver Datensatz verwendet wurde, der im Vergleich zu ähnlichen Studien eine größere und vielfältigere Datenmenge aufweist. Zudem wurde das angewendete Modell extern validiert und der klinische Nutzen dieser Technik erörtert.

Das maschinelle Lernmodell stellte sich, durch die externe Validierung an einer vielfältigen Patientenkohorte, als robust und konform zu dem aktuellem Forschungsstandard dar. Damit wird ein vielversprechendes Werkzeug für die Implementierung von maschinellem Lernen in der Behandlung von Patient*innen mit Gliomen geboten, welches nun in prospektiven Studien in der Patientenversorgung weiter getestet werden kann. Ich bin überzeugt, dass zukünftige Studien von diesen Erkenntnissen profitieren werden, um eigene maschinellen Lernmodelle zu entwickeln. Die Ergebnisse dieser Forschung könnten dazu beitragen, Behandlungsergebnisse von Patient*innen zu verbessern und Therapieansätze gezielter zu gestalten und stellen damit einen bedeutenden Schritt hin zu einer individuelleren und wirkungsvolleren Versorgung von Patient*innen mit neuro-onkologischen Erkrankungen dar.

4 Literatur- und Quellenverzeichnis

1. Lost, J., et al., *Enhancing clinical decision-making: An externally validated machine learning model for predicting isocitrate dehydrogenase mutation in gliomas using radiomics from presurgical magnetic resonance imaging*. *Neuro-Oncology Advances*, 2024. **6**(1).
2. Lost, J., et al., *Systematic Literature Review of Machine Learning Algorithms Using Pretherapy Radiologic Imaging for Glioma Molecular Subtype Prediction*. *AJNR Am J Neuroradiol*, 2023. **44**(10): p. 1126-1134.
3. Ostrom, Q.T., et al., *CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012-2016*. *Neuro Oncol*, 2019. **21**(Suppl 5): p. v1-v100.
4. Jiang, T., et al., *Clinical practice guidelines for the management of adult diffuse gliomas*. *Cancer Letters*, 2021. **499**: p. 60-72.
5. Franceschi, E., et al., *EANO–EURACAN clinical practice guideline for diagnosis, treatment, and follow-up of post-pubertal and adult patients with medulloblastoma*. *The Lancet Oncology*, 2019. **20**(12): p. e715-e728.
6. Louis, D.N., et al., *The 2021 WHO Classification of Tumors of the Central Nervous System: a summary*. *Neuro Oncol*, 2021. **23**(8): p. 1231-1251.
7. Weller, M., et al., *EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood*. *Nature Reviews Clinical Oncology*, 2020. **18**(3): p. 170-186.
8. Mohammed, S., M. Dinesan, and T. Ajayakumar, *Survival and quality of life analysis in glioblastoma multiforme with adjuvant chemoradiotherapy: a retrospective study*. *Rep Pract Oncol Radiother*, 2022. **27**(6): p. 1026-1036.
9. Molinaro, A.M., et al., *Genetic and molecular epidemiology of adult diffuse glioma*. *Nat Rev Neurol*, 2019. **15**(7): p. 405-417.
10. Ohka, F., A. Natsume, and T. Wakabayashi, *Current Trends in Targeted Therapies for Glioblastoma Multiforme*. *Neurology Research International*, 2012. **2012**(1): p. 878425.
11. Price, M., et al., *CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2017–2021*. *Neuro-Oncology*, 2024. **26**(Supplement_6): p. vi1-vi85.
12. Grochans, S., et al., *Epidemiology of Glioblastoma Multiforme–Literature Review*. *Cancers*, 2022. **14**(10).
13. Michaud, D.S., et al., *Reproductive Factors and Exogenous Hormone Use in Relation to Risk of Glioma and Meningioma in a Large European Cohort Study*. *Cancer Epidemiology, Biomarkers & Prevention*, 2010. **19**(10): p. 2562-2569.
14. Cowppli-Bony, A., et al., *Brain tumors and hormonal factors: review of the epidemiological literature*. *Cancer Causes & Control*, 2011. **22**(5): p. 697-714.
15. Posti, J.P., et al., *Presenting symptoms of glioma in adults*. *Acta Neurologica Scandinavica*, 2015. **131**(2): p. 88-93.
16. Rasmussen, B.K., et al., *Epidemiology of glioma: clinical characteristics, symptoms, and predictors of glioma patients grade I–IV in the the Danish Neuro-Oncology Registry*. *Journal of Neuro-Oncology*, 2017. **135**(3): p. 571-579.
17. Peeters, M.C.M., et al., *Prediagnostic symptoms and signs of adult glioma: the patients' view*. *Journal of Neuro-Oncology*, 2020. **146**(2): p. 293-301.
18. Azhar, S. and L.R. Chong, *Clinician's guide to the basic principles of MRI*. *Postgraduate Medical Journal*, 2023. **99**(1174): p. 894-903.
19. Wick, W., et al., *Gliome, S2k-Leitlinie, 2021, in: Deutsche Gesellschaft*

- für Neurologie (Hrsg.), Leitlinien für Diagnostik und Therapie in der Neurologie. Online: www.dgn.org/leitlinien (abgerufen am 07.08.2025). Deutsche Gesellschaft für Neurologie (Hrsg.): Online: www.dgn.org/leitlinien.
20. Parker, N.R., et al., *Molecular heterogeneity in glioblastoma: potential clinical implications*. *Front Oncol*, 2015. **5**: p. 55.
 21. Friedmann-Morvinski, *Glioblastoma Heterogeneity and Cancer Cell Plasticity*. 2014.
 22. Cryan, J.B., et al., *Clinical multiplexed exome sequencing distinguishes adult oligodendroglial neoplasms from astrocytic and mixed lineage gliomas*. *Oncotarget*, 2014. **5**(18): p. 8083-92.
 23. Hegi, M.E., et al., *MGMT gene silencing and benefit from temozolomide in glioblastoma*. *N Engl J Med*, 2005. **352**(10): p. 997-1003.
 24. Gutman, D.A., et al., *Somatic mutations associated with MRI-derived volumetric features in glioblastoma*. *Neuroradiology*, 2015. **57**(12): p. 1227-37.
 25. Miller, J.J., et al., *Isocitrate dehydrogenase (IDH) mutant gliomas: A Society for Neuro-Oncology (SNO) consensus review on diagnosis, management, and future directions*. *Neuro Oncol*, 2023. **25**(1): p. 4-25.
 26. Karnofsky, D.A., et al., *The use of the nitrogen mustards in the palliative treatment of carcinoma. With particular reference to bronchogenic carcinoma*. *Cancer*, 1948. **1**(4): p. 634-656.
 27. Li, L., et al., *Role of molecular biomarkers in glioma resection: a systematic review*. *Chinese Neurosurgical Journal*, 2020. **6**(1).
 28. Chai, R., et al., *Predictive value of MGMT promoter methylation on the survival of TMZ treated IDH-mutant glioblastoma*. *Cancer Biol Med*, 2021. **18**(1): p. 272-282.
 29. Stupp, R., et al., *Effect of Tumor-Treating Fields Plus Maintenance Temozolomide vs Maintenance Temozolomide Alone on Survival in Patients With Glioblastoma: A Randomized Clinical Trial*. *Jama*, 2017. **318**(23): p. 2306-2316.
 30. Kirson, E.D., et al., *Alternating electric fields arrest cell proliferation in animal tumor models and human brain tumors*. *Proc Natl Acad Sci U S A*, 2007. **104**(24): p. 10152-7.
 31. Fortin Ensign, S.P., et al., *Translational significance of CDKN2A/B homozygous deletion in isocitrate dehydrogenase-mutant astrocytoma*. *Neuro-Oncology*, 2023. **25**(1): p. 28-36.
 32. Rassow, J., *Biochemie: 50 Tabellen*. 2008: Thieme.
 33. Yan, H., et al., *IDH1 and IDH2 Mutations in Gliomas*. *New England Journal of Medicine*, 2009. **360**(8): p. 765-773.
 34. Han, S., et al., *IDH mutation in glioma: molecular mechanisms and potential therapeutic targets*. *Br J Cancer*, 2020. **122**(11): p. 1580-1589.
 35. Bunse, L., et al., *Suppression of antitumor T cell immunity by the oncometabolite (R)-2-hydroxyglutarate*. *Nature Medicine*, 2018. **24**(8): p. 1192-1203.
 36. Friedrich, M., et al., *Perspectives of immunotherapy in isocitrate dehydrogenase-mutant gliomas*. *Current Opinion in Oncology*, 2018. **30**(6): p. 368-374.
 37. Schumacher, T., et al., *A vaccine targeting mutant IDH1 induces antitumour immunity*. *Nature*, 2014. **512**(7514): p. 324-327.
 38. Waitkus, M.S., B.H. Diplas, and H. Yan, *Biological Role and Therapeutic Potential of IDH Mutations in Cancer*. *Cancer Cell*, 2018. **34**(2): p. 186-195.
 39. Carrillo, J.A., et al., *Relationship between Tumor Enhancement, Edema, IDH1 Mutational Status, MGMT Promoter Methylation, and Survival in Glioblastoma*. *American Journal of Neuroradiology*, 2012. **33**(7): p. 1349-1355.

40. Paldor, I., et al., *Frontal glioblastoma multiforme may be biologically distinct from non-frontal and multilobar tumors*. Journal of Clinical Neuroscience, 2016. **34**: p. 128-132.
41. Sonoda, Y., et al., *Association between molecular alterations and tumor location and MRI characteristics in anaplastic gliomas*. Brain Tumor Pathol, 2015. **32**(2): p. 99-104.
42. Xiong, J., et al., *Combination of diffusion tensor imaging and conventional MRI correlates with isocitrate dehydrogenase 1/2 mutations but not 1p/19q genotyping in oligodendroglial tumours*. Eur Radiol, 2016. **26**(6): p. 1705-15.
43. Metellus, P., et al., *Absence of IDH mutation identifies a novel radiologic and molecular subtype of WHO grade II gliomas with dismal prognosis*. Acta Neuropathol, 2010. **120**(6): p. 719-29.
44. Qi, S., et al., *Isocitrate dehydrogenase mutation is associated with tumor location and magnetic resonance imaging characteristics in astrocytic neoplasms*. Oncol Lett, 2014. **7**(6): p. 1895-1902.
45. Kickingereder, P., et al., *IDH mutation status is associated with a distinct hypoxia/angiogenesis transcriptome signature which is non-invasively predictable with rCBV imaging in human glioma*. Sci Rep, 2015. **5**: p. 16238.
46. Deguchi, S., et al., *Clinicopathological analysis of T2-FLAIR mismatch sign in lower-grade gliomas*. Scientific Reports, 2020. **10**(1).
47. Broen, M.P.G., et al., *The T2-FLAIR mismatch sign as an imaging marker for non-enhancing IDH-mutant, 1p/19q-intact lower-grade glioma: a validation study*. Neuro-Oncology, 2018. **20**(10): p. 1393-1399.
48. Jain, R., et al., *"Real world" use of a highly reliable imaging sign: "T2-FLAIR mismatch" for identification of IDH mutant astrocytomas*. Neuro Oncol, 2020. **22**(7): p. 936-943.
49. Sarker, I.H., *Machine Learning: Algorithms, Real-World Applications and Research Directions*. SN Comput Sci, 2021. **2**(3): p. 160.
50. Jekel, L., et al., *Machine Learning Applications for Differentiation of Glioma from Brain Metastasis—A Systematic Review*. Cancers, 2022. **14**(6).
51. Cassinelli Petersen, G.I., et al., *Machine Learning in Differentiating Gliomas from Primary CNS Lymphomas: A Systematic Review, Reporting Quality, and Risk of Bias Assessment*. AJNR Am J Neuroradiol, 2022. **43**(4): p. 526-533.
52. Bahar, R.C., et al., *Machine Learning Models for Classifying High- and Low-Grade Gliomas: A Systematic Review and Quality of Reporting Analysis*. Frontiers in Oncology, 2022. **12**.
53. Niklas Tillmanns, A.E.L., Gabriel Cassinelli, Sara Merkaj, Tej Verma, Tal Zeevi, M.Sc, Lawrence Staib, PhD, Harry Subramanian, MD, Ryan C Bahar, Waverly Brim, Jan Lost, Leon Jekel, Alexandria Brackett, MA, Sam Payabvash, MD, Ichiro Ikuta, MD, MingDe Lin, PhD, Khaled Bousabarah, M.Sc, Michele H Johnson, MD, Jin Cui, Ajay Malhotra, MD, Antonio Omuro, MD, Bernd Turowski, Mariam S Aboian, MD/PhD, *Identifying clinically applicable machine learning algorithms for glioma segmentation: recent advances and discoveries*. Neuro-Oncology Advances, 2022.
54. Aboian, M., et al., *Clinical implementation of artificial intelligence in neuroradiology with development of a novel workflow-efficient picture archiving and communication system-based automated brain tumor segmentation and radiomic feature extraction*. Frontiers in Neuroscience, 2022. **16**.

55. Kourou, K., et al., *Machine learning applications in cancer prognosis and prediction*. Computational and Structural Biotechnology Journal, 2015. **13**: p. 8-17.
56. Morimoto, J. and F. Ponton, *Virtual reality in biology: could we become virtual naturalists?* Evolution: Education and Outreach, 2021. **14**.
57. Chen, T. and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, Association for Computing Machinery: San Francisco, California, USA. p. 785–794.
58. Ying, X., *An Overview of Overfitting and its Solutions*. Journal of Physics: Conference Series, 2019. **1168**: p. 022022.
59. Bakas, S., et al., *Overall survival prediction in glioblastoma patients using structural magnetic resonance imaging (MRI): advanced radiomic features may compensate for lack of advanced MRI modalities*. J Med Imaging (Bellingham), 2020. **7**(3): p. 031505.
60. Jian, A., et al., *Machine Learning for the Prediction of Molecular Markers in Glioma on Magnetic Resonance Imaging: A Systematic Review and Meta-Analysis*. Neurosurgery, 2021. **89**(1): p. 31-44.
61. Chung, C.Y.C. and L.E. Pigott, *Predicting IDH and ATRX mutations in gliomas from radiomic features with machine learning: a systematic review and meta-analysis*. Frontiers in Radiology, 2024. **4**.
62. Zhu, Y., et al., *Deep Learning and Habitat Radiomics for the Prediction of Glioma Pathology Using Multiparametric MRI: A Multicenter Study*. Academic Radiology, 2025. **32**(2): p. 963-975.
63. van der Voort, S.R., et al., *Combined molecular subtyping, grading, and segmentation of glioma using multi-task deep learning*. Neuro-Oncology, 2023. **25**(2): p. 279-289.
64. de Hond, A.A.H., et al., *Predicting Readmission or Death After Discharge From the ICU: External Validation and Retraining of a Machine Learning Model*. Critical Care Medicine, 2023. **51**(2): p. 291-300.
65. Chen, T.L.-W., et al., *Internal and External Validation of the Generalizability of Machine Learning Algorithms in Predicting Non-home Discharge Disposition Following Primary Total Knee Joint Arthroplasty*. The Journal of Arthroplasty, 2023. **38**(10): p. 1973-1981.
66. Mari, T., et al., *External validation of binary machine learning models for pain intensity perception classification from EEG in healthy individuals*. Scientific Reports, 2023. **13**(1): p. 242.
67. van Griethuysen, J.J.M., et al., *Computational Radiomics System to Decode the Radiographic Phenotype*. Cancer Res, 2017. **77**(21): p. e104-e107.
68. Kumar, A., et al., *Machine-Learning-Based Radiomics for Classifying Glioma Grade from Magnetic Resonance Images of the Brain*. Journal of Personalized Medicine, 2023. **13**(6): p. 920.
69. Kim, B.-H., et al., *Validation of MRI-Based Models to Predict MGMT Promoter Methylation in Gliomas: BraTS 2021 Radiogenomics Challenge*. Cancers, 2022. **14**(19): p. 4827.
70. WP, C.N.B.K.H.L.K., *SMOTE: Synthetic Minority Over-sampling Technique*. Journal of Artificial Intelligence Research, 2002. **16**: p. 321-357.
71. Park, J.E., et al., *Radiomics prognostication model in glioblastoma using diffusion- and perfusion-weighted MRI*. Scientific Reports, 2020. **10**(1).

72. Wu, M., et al., *Amide proton transfer-weighted imaging and derived radiomics in the classification of adult-type diffuse gliomas*. *European Radiology*, 2023. **34**(5): p. 2986-2996.
73. Qin, Y., et al., *External validity of machine learning-based prognostic scores for cystic fibrosis: A retrospective study using the UK and Canadian registries*. *PLOS Digit Health*, 2023. **2**(1): p. e0000179.
74. Youssef, A., et al., *External validation of AI models in health should be replaced with recurring local validation*. *Nature Medicine*, 2023. **29**(11): p. 2686-2687.
75. Kunze, K.N., et al., *External Validation of a Machine Learning Algorithm for Predicting Clinically Meaningful Functional Improvement After Arthroscopic Hip Preservation Surgery*. *The American Journal of Sports Medicine*, 2022. **50**(13): p. 3593-3599.
76. Siontis, G.C.M., et al., *External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination*. *Journal of Clinical Epidemiology*, 2015. **68**(1): p. 25-34.
77. Dal Bo, M., et al., *Machine learning to improve interpretability of clinical, radiological and panel-based genomic data of glioma grade 4 patients undergoing surgical resection*. *J Transl Med*, 2023. **21**(1): p. 450.
78. Zhang, C.B., et al., *Correlation of IDH1/2 mutation with clinicopathologic factors and prognosis in anaplastic gliomas: a report of 203 patients from China*. *J Cancer Res Clin Oncol*, 2014. **140**(1): p. 45-51.
79. Howell, S.J. and J.K. Dhesi, *Frailty in the perioperative setting: lessons from SNAP-3*. *British Journal of Anaesthesia*, 2025. **135**(2): p. 296-299.
80. Licina, A., A. Silvers, and C. Thien, *Association between frailty and clinical outcomes in patients undergoing craniotomy—systematic review and meta-analysis of observational studies*. *Systematic Reviews*, 2024. **13**(1).
81. Mellingshoff, I.K., et al., *Vorasidenib in IDH1- or IDH2-Mutant Low-Grade Glioma*. *New England Journal of Medicine*, 2023. **389**(7): p. 589-601.
82. Mellingshoff, I.K., et al., *Vorasidenib and ivosidenib in IDH1-mutant low-grade glioma: a randomized, perioperative phase 1 trial*. *Nat Med*, 2023. **29**(3): p. 615-622.
83. von Reppert, M., et al., *Image-Based Search in Radiology: Identification of Brain Tumor Subtypes within Databases Using MRI-Based Radiomic Features*. *AJNR Am J Neuroradiol*, 2025. **46**(7): p. 1421-1428.
84. Li, Z.C., et al., *Multiregional radiomics profiling from multiparametric MRI: Identifying an imaging predictor of IDH1 mutation status in glioblastoma*. *Cancer Med*, 2018. **7**(12): p. 5999-6009.
85. Xu, W., et al., *Predicting the molecular subtypes of 2021 WHO grade 4 glioma by a multiparametric MRI-based machine learning model*. *BMC Cancer*, 2025. **25**(1).

Danksagung

Zuerst möchte ich mich bei meinem Doktorvater Prof. Michael Sabel bedanken. Ihre Begeisterung und Leidenschaft für die Neuroonkologie war der Funke, der diese Arbeit überhaupt erst initiiert hat. Ich danke Ihnen außerdem für das Vertrauen, diese Arbeit mit mir durchzuführen, und für den Rückhalt sowohl während meiner Zeit als studentische Hilfskraft als auch während der Erstellung dieser Promotionsarbeit.

Mein ganz besonderer Dank gilt außerdem meiner Co-Betreuerin Mariam Aboian, MD/PhD. Die Zeit in deinem Labor werde ich auf ewig wertschätzen und ist ein zentraler Teil meines akademischen Werdegangs. Dein Vertrauen und unermüdlicher Einsatz für mich und die anderen Mitglieder des Labors ist inspirierend und überhaupt erst der Grund, warum diese Arbeit zustande kommen konnte. Dank deines Einsatzes konnte ich unsere Arbeiten auf mehreren Kongressen vorstellen, wofür ich dir ebenfalls sehr dankbar bin. Von niemandem sonst habe ich mehr über wissenschaftliches Arbeiten lernen können.

Als Nächstes möchte ich die *PACS-based Boys*, Niklas Tillmanns, Leon Jekel und Marc von Reppert, nicht unerwähnt lassen. Dank euch waren selbst die stürmischsten Stunden in CB30 eine Freude. Eure fachliche Expertise und Freundschaft waren entscheidend für den Erfolg dieser Arbeit. Auch dem Rest der *Imagine Quant Research Group* sowie des *CORE Lab* bin ich zu tiefem Dank verpflichtet. Der enge Zusammenhalt in und außerhalb des Labors hat mir nicht nur großen Spaß bereitet, sondern war auch unerlässlich für den erfolgreichen Abschluss dieser Forschungsarbeit.

Diese Arbeit und die damit verbundene Forschungsreise wäre ohne die Unterstützung des HHU-Promotionstipendiums der Forschungskommission nicht möglich gewesen, der ich an dieser Stelle auch ganz herzlich für ihr Vertrauen danken möchte.

Danken möchte ich auch meiner Familie und meinen Freunden. Von Beginn meiner Bewerbung für den Forschungsaufenthalt an der Yale-Universität bis hin zum schlussendlichen Verfassen der Promotionsschrift wart ihr immer für mich da und habt mir die nötige Kraft gegeben, alle Höhen und Tiefen dieser Zeit durchzustehen. Ganz besonders möchte ich meinen Eltern Judit und Erwin danken. Dank euch war mir das Medizinstudium an sich überhaupt erst möglich, und ohne euren Beistand hätte ich die Forschungsreise und diese Arbeit nicht antreten

können. Auch Stefan, Lara und Manuel möchte ich danken, dass ihr in allen Lebenslagen immer für mich da seid und mich fortwährend unterstützt in allem, was ich tue. Erwin und Manuel möchte ich zusätzlich für die Zeit, das ehrliche Feedback und die Durchsicht dieser Arbeit danken.

Zuletzt danke ich ganz besonders meiner Partnerin Kristina. Du warst vom Anfang bis zum Abschluss des Projekts immer für mich da. Dank deiner anhaltenden Unterstützung habe ich auch die schwierigsten Phasen dieser Arbeit durchstehen können.

Danke!