

Direct Image to Subtype Prediction for Brain Tumors using Deep Learning

Katherine J. Hewitt, Chiara M. L. Löffler, Hannah Sophie Muti, Anna Sophie Berghoff, Christian Eisenlöffel, Marko van Treeck, Zunamys I. Carrero, Omar S. M. El Nahhas, Gregory P. Veldhuizen, Sophie Weil, Oliver L. Saldanha, Laura Bejan, Thomas O. Millner, Sebastian Brandner, Sascha Brückmann, Jakob Nikolas Kather

Department of Medicine III, University Hospital RWTH Aachen, Aachen, Germany (K.J.H, C.M.L.L, O.L.S, J.N.K); Else Kroener Fresenius Center for Digital Health, Technical University Dresden, Dresden, Germany (K.J.H, C.M.L.L, H.S.M, M.vT, Z.I.C, O.S.M.EN, G.P.V, J.N.K); Department of Visceral, Thoracic and Vascular Surgery, University Hospital Carl Gustav Carus Dresden, Dresden, Germany (H.S.M); Department of Medicine I, Division of Oncology, Medical University of Vienna, Vienna, Austria (A.S.B); Department of Pathology, St. Georg Teaching Hospital, University of Leipzig, Leipzig, Germany (C.E.); Department of Internal Medicine I, University Hospital Carl Gustav Carus, Dresden, Germany (C.M.L.L., J.N.K.); Neurology Clinic, National Center for Tumor Diseases (NCT), University Hospital Heidelberg, Heidelberg, Germany (S.W.); Clinical Cooperation Unit Neuro-oncology, German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany (S.W.); School of Medicine, University College London, London, United Kingdom (L.B.); Division of Neuropathology, Queen Square Institute of Neurology, University College London, London, WC1N 3BG, UK (T.O.M., S.B.); Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK (T.O.M.); Institut für Pathologie, University Hospital Carl Gustav Carus, Dresden, Germany (S.B.); Pathology & Data Analytics, Leeds Institute of

© The Author(s) 2023. Published by Oxford University Press, the Society for Neuro-Oncology and the European Association of Neuro-Oncology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Medical Research at St James's, University of Leeds, Leeds, United Kingdom (J.N.K.); Medical Oncology, National Center for Tumor Diseases (NCT), University Hospital Heidelberg, Heidelberg, Germany (J.N.K)

Corresponding Author: jakob-nikolas.kather@alumni.dkfz.de

Else Kröner-Fresenius Center for Digital Health

Technische Universität Dresden

DE – 01062 Dresden

+49 351 458-7558

Accepted Manuscript

Key Points

- Deep Learning can predict both molecular alteration status and subtype of diffuse adult-type gliomas directly from histopathology whole slide images.
- This is the first study to predict subtype and molecular status according to the 2021 WHO CNS 5th edition.

Importance of the Study

The 2021 update to the World Health Organisation (WHO) classification system for central nervous system (CNS) tumors places further emphasis on molecular characterisation in the diagnosis of adult-type diffuse gliomas. However, molecular assays are not necessarily available in all healthcare systems. Deep Learning (DL) offers an alternative approach for predicting molecular status directly from digitized histopathology slides. This study demonstrates that DL can accurately predict not only the subtype of glioma, but also the status of individual molecular alterations. To our knowledge, this is the first study predicting clinically-relevant markers according to the WHO 5th edition. This work provides further evidence supporting the use of digital workflows to support decision-making in clinical medicine.

Abstract

Background. Deep Learning (DL) can predict molecular alterations of solid tumors directly from routine histopathology slides. Since the 2021 update of the World Health Organization (WHO) diagnostic criteria, the classification of brain tumors integrates both histopathological and molecular information. We hypothesize that DL can predict molecular alterations as well as WHO subtyping of brain tumors from hematoxylin and eosin-stained histopathology slides.

Methods. We used weakly supervised DL and applied it to three large cohorts of brain tumor samples, comprising N=2,845 patients.

Results. We found that the key molecular alterations for subtyping, *IDH* and *ATRX*, as well as 1p19q co-deletion, were predictable from histology with an area under the receiver operating characteristic curve (AUROC) of 0.95, 0.90 and 0.80 in the training cohort respectively. These findings were upheld in external validation cohorts with AUROCs of 0.90, 0.79 and 0.87 for prediction of *IDH*, *ATRX* and 1p19q co-deletion respectively.

Conclusions. In the future, such DL-based implementations could ease diagnostic workflows, particularly for situations in which advanced molecular testing is not readily available.

Keywords: Adult-type diffuse gliomas, Deep Learning, subtype, molecular signatures, IDH.

Introduction

Diffuse gliomas are the most frequent primary malignant brain tumors in adults and broad demographic shifts mean that their incidence is expected to increase^{1,2}. Due to the aggressive nature, infiltrative growth and central location, these tumors generally have a poor prognosis. Clinical outcome for these entities largely depends on subtype, the diagnosis of which has historically been based on histological assessment^{1,2}.

In 2021, the World Health Organization (WHO) released the 5th edition of diagnostic criteria for adult-type diffuse gliomas. These criteria are based on a combination of molecular alterations and histopathological assessment, which determines WHO CNS subtype and grade³. In addition, methylation profiles can be used to define the glioma diagnosis².

Based on their molecular profile, a significant proportion of adult-type gliomas can be subtyped by *IDH* (*IDH1* or *IDH2*) and *ATRX* mutation status, as well as 1p19q co-deletion. Glioblastomas CNS WHO 4 are currently defined by the absence of *IDH*- and *H3* mutations and occurrence of at least one of the following histopathological (microvascular proliferations, necrosis) or molecular (chromosome +7/-10 signature, *EGFR* amplification, *TERT* promoter mutation) alterations. *IDH*-mutant gliomas are subclassified into astrocytomas or oligodendrogliomas. The classification of *IDH*-mutant oligodendrogliomas CNS WHO 2/3 are defined by demonstration of a combined whole arm deletion of chromosomes 1p and 19q. Immunohistochemically retained nuclear *ATRX* expression as a surrogate of *ATRX*- *wildtype*- status is a very strong correlate for 1p19q- codeletion- status in *IDH*-mutant gliomas². *IDH*-mutant astrocytomas CNS WHO 2/3/4 are molecularly defined by retained 1p19q chromosomes and the presence of *ATRX*-mutation, or its immunohistochemical surrogate of nuclear *ATRX*- staining-loss. The grading of *IDH*-mutant astrocytomas also relies on a combination of histomorphological as

well as molecular traits: the presence of *CDKN2A/B* homozygous deletion and/or necrosis and/or microvascular proliferation defines a CNS WHO grade 4 lesion in *IDH-mutant* astrocytomas².

This evolution in the diagnostic approach represents a further shift away from the traditional histopathological process where pathologists typically assign subtypes based on specific tumor features, including morphology. The integration of genetic alterations into the diagnostic process reflects improved understanding of these tumors allowing for greater diagnostic precision. However, the necessity for next molecular assessment and in some cases, methylation array profiling, also makes the process more expensive⁴. These diagnostic requirements are further confounded by the fact that many advanced genetic tests are not available in all healthcare systems. For a fully integrated diagnosis, multiple different assays may be required to characterize different molecular features. This may further drive up costs associated with molecular testing, serving to widening the already pronounced inequities in precision medicine⁵. Furthermore, the requirement for multiple tests could delay treatment by up to several weeks, with adverse sequelae for the patient⁶.

An increasing body of evidence indicates that Deep Learning (DL) techniques are able to predict the phenotypes linked to individual molecular alterations directly from routine hematoxylin and eosin (H&E) stained histopathology slides⁷. This has the potential to accelerate diagnostic workflows and reduce the costs associated with molecular testing⁸. These studies have shown that clinically relevant features, such as biomarkers and subtype, can be predicted directly from histopathology slides for many tumor types^{7,9,10}, including in brain tumors^{11,12}. DL is a machine learning (ML) method from the field of artificial intelligence (AI) and is a common and powerful way to extract quantitative information from image data¹³. A typical DL workflow involves three main stages; pre-processing of the data, training and testing of an algorithm and assessment of model performance through interpretation of statistical results¹³.

To assess the current status of DL research in diagnostic neuropathology, we performed a systematic literature review (**Suppl. Fig 1 and 2**). DL applications in brain tumors have mostly focused on classification tasks. The most common classification tasks consisted of molecular status prediction (n=13) followed by grade (n=10) and subtype prediction (n=4). This demonstrates that brain tumors remain understudied in the field of computational pathology. This is particularly evident when compared to breast or colorectal cancer, which have been investigated in hundreds of computational pathology studies⁷. The brain-cancer-related studies identified via literature review were generally performed on small datasets^{11,14-16}, without external validation¹⁷⁻¹⁹, or used networks that were not pre-trained on histopathology images^{18,20-22}. This demonstrates a pressing need to address brain tumor research using larger, more heterogeneous cohorts employing up-to-date methods. To our knowledge, there are no previously published studies attempting to predict subtype in adult-type diffuse gliomas on the basis of the WHO 2021 5th edition.

To this end, we collected and analyzed three independent datasets of digitized routine histological whole slide images (WSI) with the primary aims of predicting genetic alterations and differentiating brain tumor subtypes by using DL.

Materials and Methods

Ethics Statement

This study was performed in accordance with the Declaration of Helsinki. The research involved analysis of anonymized archival digital images of human tissue. Ethical approval was obtained at collaborating centers prior to collection and pseudonymization. Data was obtained from University College London as part of the UK Brain Archive Information Network (BRAIN UK) which is funded by the Medical Research Council and Brain Tumour Research. BRAIN UK reference number: 22-011

– Artificial intelligence-based reconstruction of the WHO 2021 diagnostic algorithm for adult-type diffuse gliomas. The overall analysis was approved by the Ethics commission of the Medical Faculty of Technical University of Dresden (BO-EK-444102022). The STARD 2015 checklist can be found in **Suppl. Tab. 2.**

Cohort Description

The first cohort was obtained from University College London (UCL) via Brain UK (REF:22/011). This cohort consisted of n=1882 digitized H&E histopathology WSIs from n=1877 patients with corresponding clinicopathological and molecular data. These are non-consecutive cases collected from routine diagnostic work between 2011 and 2019. The dataset, curated from 40 sites across seven countries, encompasses both typical clinical cases and a small number of rare entities. All cases in which a diagnosis of an adult-type diffuse glioma was made were included in our study. These cases had an integrated molecular diagnosis according to WHO 2021, updated from a diagnosis according to WHO 2016, i.e. all cases were retrospectively assessed to comply with the CNS WHO 2021 classification.

The second cohort was derived from The Cancer Genome Atlas (TCGA, n=864) network www.portal.gdc.gov for the tumor entities of Low Grade Glioma (LGG, n=493) and Glioblastoma (GBM, n=371). The dataset is composed of cases contributed from 38 sites across seven countries. All cases in which a diagnosis of an adult-type diffuse glioma or a relevant genetic alteration was made, were included in our study. Digitized histopathological WSIs with matching clinical-pathological and molecular data were obtained from www.cbioportal.org. Classification for this cohort was made according to WHO 2016. Molecular alteration data available for this cohort were used to update the subtype diagnosis to comply with WHO 2021, with assistance from an expert neuropathologist (SB). The third cohort was the Clinical

Proteomic Tumor Analysis Consortium (CPTAC, n=99) from <https://www.cancerimagingarchive.net/collections/>. All cases that were tested for *CDKN2A/B* homozygous deletion were included in our study. Matching clinicopathological and molecular data were downloaded from www.cbioportal.org. Data was obtained from these online sources as of 13th March 2022.

Information on the diagnostic process was sourced from WHO via their website <https://tumourclassification.iarc.who.int/login>. An overview of the neuropathological approach to diffuse adult-type gliomas can be found in **Figure 1A-B**. The diagnostic criteria informed which molecular data we collected for this study. Full details of the data collected and data preprocessing can be found in **Supplementary Methods**. An overview of the subtype data available in the UCL and TCGA cohorts can be found in **Figure 1C-D**. Further cohort data and consort charts can be found in **Table 1** and **Suppl. Fig. 3** respectively.

Deep Learning Methods

We have previously established^{10,23} and validated^{24,25} a Deep Learning pipeline to predict molecular alterations directly from histopathology images, including neuropathology²⁶. Here, we used an attention-based multiple-instance learning approach (attMIL)^{26,27}. Preprocessing included normalization²⁸ and tessellation of WSIs before features were extracted from each tile²⁹. The attMIL model makes predictions for a patient, based on a collection of tiles extracted from the patient's slides. We call the aggregate of a patient's features a *bag*, with the features itself being the bag's *instances*. Since it is probable that not all of the instances have the same amount of information on the patient-level label, our model considers the entire bag at once. This enables it to consider information which may only be present in some of the instances while ignoring instances which contain little to no valuable information³⁰. **Fig 1E** gives an overview of our approach. Full details of these methods can be found in **Supplementary Methods**. Our preprocessing and pipeline scripts are freely available on github (<https://github.com/KatherLab/marugoto>).

Statistical Analysis

Model performance was assessed with the Area Under the Receiver Operating Curve (AUROC), which is the primary statistical endpoint in most studies on AI-based medical image analysis ⁷. A bootstrap 95% confidence interval (CI) is also given to provide a measure on AUROC accuracy. Furthermore, we provide additional statistics including sensitivity, specificity, accuracy and precision.

Data and Code Availability

Access to the UCL data can be requested via the Brain UK study proposal platform at <https://www.southampton.ac.uk/brainuk/index.page>. Data from the TCGA and CPTAC cohorts can be accessed via the cBioPortal website (www.cbioportal.org, or literature: ³¹).

Experimental Design

Our aim was to use DL for subtype and molecular status prediction in adult-type diffuse gliomas according to the WHO 2021 5th edition. We investigated two approaches for subtype prediction:

1. Direct prediction: The tumor subtype is predicted from the WSI directly, according to both 2016 and 2021 WHO approaches. The 2016 results were used as the baseline from which to compare the 2021 results.
2. Sequential prediction: Each alteration in the diagnostic pathway is predicted separately. As a result of this output, the diagnostic pathway can be reproduced in a stepwise fashion, where the combination of alteration present and/or absent would then determine the subtype.

Both experimental approaches were run with internal cross-validation and external validation settings. The UCL cohort was chosen as the primary cohort for internal cross-validation and as a training set for external validation, due to having a larger number of patients and more balanced classes, with the exception of Ch+7/-10 and *CDKN2A/B*, as described in **Suppl. Fig. 3**. Details of the cohorts that were used for each experiment are listed in **Table 2**.

Results

Direct Prediction: 2016 and 2021 WHO Subtypes

The 2016 WHO subtypes of brain tumors are based on molecular alteration status plus histomorphology. Deep Learning utilizes the characteristics of an image to inform classification decisions. Hence, a Deep Learning model should be able to predict these subtypes from images more readily than the other targets. Once established, these results can be used as a baseline from which to assess our further experiments.

In the cross-validation setting, our approach yielded AUROCs of 0.89 (CI±0.02), 0.94 (CI±0.01) and 0.93 (CI±0.01) for detection of astrocytoma, glioblastoma and oligodendroglioma, respectively (experiments 1-3 in **Table 2** and **Figure 2**). We also predicted the 2016 subtype in an external validation setting. These experiments yielded AUROCs of 0.86 (CI±0.03), 0.91 (CI±0.03) and 0.86 (CI±0.04) for astrocytoma, glioblastoma and oligodendroglioma respectively (experiments 14-16 in **Table 2** and **Figure 2**).

We next investigated prediction of the 2021 subtype in an internal cross-validation and our approach was able to yield AUROCs of 0.92 (CI±0.03), 0.95 (CI±0.02) and 0.93 (CI±0.02) for astrocytoma, glioblastoma and oligodendroglioma, respectively (experiments 4-6 in **Table 2** and **Figure 2**).

We then performed external validation, which gave AUROCs of 0.84 (CI±0.03), 0.90 (CI±0.02) and 0.91 (CI±0.03) for prediction of astrocytoma, glioblastoma and oligodendroglioma, respectively (experiments 17-19 in **Table 2** and **Figure 2**).

When comparing these results to our baseline 2016 results, the performances were very similar. The external validation AUROCs for astrocytoma and glioblastoma (2016 vs 2021) were within 0.02 and the internal validation AUROCs were within 0.03 and 0.01 for astrocytoma and glioblastoma respectively (**Figure 2 A&B, E&G**). For oligodendroglioma, the AUROCs remained congruous across the internal validation experiments (2016 vs 2021), however, on external validation, the AUROC for the 2021 experiment was 0.05 greater than in the 2016 experiment (**Figure 2 C&G**).

Sequential Prediction: Molecular Alterations

For our sequential prediction experiments, we began by predicting the core diagnostic molecular alterations (*IDH*, 1p19q and *ATRX*) in a cross-validation setting. Prediction of *IDH* within this cohort was highly successful giving an AUROC of 0.95 (CI±0.02), with *ATRX* and 1p19q giving AUROCs of 0.91 (CI±0.03) and 0.80 (CI±0.03) respectively. For external validation of the core molecular alterations, results were similar to external validation for subtype prediction, with AUROCs of 0.90, (CI±0.02) 0.79 (CI±0.04 and 0.87 (CI±0.03) for *IDH*, *ATRX* and 1p19q prediction respectively (experiments 7-9 and 20-22 in **Table 2** and **Figure 3**).

Furthermore, we performed subgroup analyses to ascertain how well the model was able to predict molecular alterations within a subgroup composed of each tumor subtype. However, unfortunately our results did not find a strong link between any alteration and tumor subtype. ROC curves for these experiments can be found in **Suppl. Fig. 3**.

We also aimed to investigate prediction of four additional molecular alterations; *TERT* mutation, amplification of *EGFR*, *CDKN2A/B* homozygous deletion and chromosome trisomy 7 with monosomy 10 (ch+7-10). For *EGFR* amplification, internal validation gave an AUROC 0.83 (CI±0.04) and external validation yielded an AUROCs of 0.85 (CI±0.03) (experiments 11 & 24 in **Table 2** and **G** in **Fig 3**). Unfortunately, we were unable to fully assess predictability of *TERT* mutation, *CDKN2A/B* homozygous deletion and ch+7/-10 due to low prevalence of cases in our datasets. Preliminary results using the available data for these alterations are available in **Suppl. Fig. 4**.

Comparison of Approaches: Direct versus Sequential

In order to compare our two approaches, we stacked the external validation predictions for the three core alterations *IDH*, *ATRX* and 1p19q to determine a final subtype prediction according to WHO CNS 2021. We then performed statistical analysis of these stacked predictions and compared them to statistical analysis of the external validation results for the 2021 direct prediction experiment (**Fig. 4**). Overall, the results for the sequential prediction were superior to the direct prediction in all metrics (**Fig. 4**), except for precision (0.92 in sequential, 0.95 in direct) and specificity (0.97 in sequential, 0.99 in direct) of glioblastoma prediction, and sensitivity for oligodendroglioma prediction (0.84 for direct and 0.83 for sequential). Excellent performance for specificity in both the direct and sequential approaches was noted.

Interpretability: Plausible Features can be Linked to Predictions

To provide insight into the morphological features driving our network's predictions, we produced heatmaps for a random sample of external validation WSIs for each target. These were then reviewed by an expert neuropathologist (SB, **Supp. Fig. 7**).

In the oligodendroglioma and 1p19q cases, so-called “fried egg” cells (densely packed cells with round nuclei and perinuclear clearing that resemble fried eggs) were consistently highlighted in our heatmaps, indicating that they are important for the prediction of these targets. This was the case for oligodendroglioma subtype prediction in both 2016 and 2021 experiments. Similarly, the *IDH*-wildtype and glioblastoma heatmaps (both 2016 and 2021 classifications) focused on areas of microvascular proliferation and pseudopalisading necrosis. Interestingly, *ATRX* mutation heatmaps were mostly ‘cold’. In some cases, gemistocytes appeared to be relevant to classification, but these cells were not present in all cases (Suppl. Fig. 5G).

Discussion

The aim of this work was to recreate the diagnostic neuropathological workflow for diffuse adult-type gliomas with Deep Learning. Hence, we utilized a large dataset acquired through the Brain UK consortium - the UCL dataset - and demonstrated the efficacy of an attention-based DL pipeline on this task. We subsequently illustrated the general applicability of our pipeline by deploying on independent data using the publicly available TCGA and CPTAC datasets.

Through these methods, we were able to demonstrate that Deep Learning can infer both genetic alteration status and WHO subtype directly from WSI. Our results for subtype prediction and the three core diagnostic alterations were consistently above 0.79 in both internal and external validation. While an AUROC of 0.5 would indicate a random classifier, AUROCs of 0.7 - 0.8 indicate that a phenotype associated with the target is detectable to some degree by DL³². Thus, we postulate that our model successfully identified and leveraged relevant phenotypes for these targets.

A variance in AUROC between internal- and external- validation experiments was notable for some targets. This is a common issue termed domain shift³³. Domain shift can occur due to multiple factors,

particularly due to variations in tissue processing methods, staining and scanner properties between cohorts³³. While we employed multiple approaches to alleviate this problem, including training with a diverse cohort and normalizing tiles, model generalization remains an open problem in the field³³.

Visual interpretability performed as part of this study highlighted morphologies that are recognized as being associated with specific subtypes as important to model decision-making. Fried egg cells and chicken-wire vasculature were found to correlate with prediction of 1p19q co-deletion and oligodendroglioma subtype. Microvascular proliferation and necrosis correlated with *IDH*-wildtype and glioblastoma predictions, while gemistocytes and Rosenthal fibers correlated with astrocytoma and *IDHmut* predictions (**Suppl. Fig. 7**). These features are used to make morphological diagnoses by neuropathologists². This indicates that our network used morphology in decision-making similarly to that of a neuropathologist and supports the correctness of our model.

Our study assessed two experimental approaches. Our direct approach predicted tumor subtype directly from WSI, whereas the sequential approach predicted mutational alteration statuses. Statistical analysis of these approaches indicated that overall, the sequential approach performed better. The sequential approach provides the added benefit of interpretability, as it allows pathologists to understand which alterations are present in a WSI, and thus support the final predicted subtype diagnosis. However, the excellent specificity of the direct approach for glioblastoma prediction should be noted. Glioblastoma is important to rule out in the clinical setting due to the poor prognosis³⁴.

The limitations of our study primarily relate to our data. We were unable to evaluate all targets in the 2021 WHO CNS classification due to lack of data. Classes were particularly unbalanced for alterations such as *CDKN2A/B* in the UCL cohort and *TERT* in the TCGA and CPTAC cohorts, which can create biased predictions. Similarly, we were not able to externally validate ch+7-10 as we only had data in the TCGA cohort. Furthermore, our study only included adult-type diffuse gliomas without considering other differential diagnoses, as would take place in clinical practice.

Further work would include acquiring more data from different sites. This would allow us to improve balance in our classes and provide more varied training data, including common differentials. This should help address the issue of domain shift and make our network more applicable to additional cohorts.

In conclusion, our study demonstrates that DL can predict the WHO CNS 2021 subtypes with high accuracy in a single and external cohort. Although a small number of studies have previously predicted WHO 2016 subtype and some molecular alterations, to our knowledge, no external validation experiments for 2016 subtype prediction have been previously performed. Furthermore, this is the first study to predict WHO 2021 subtype and the three core diagnostic alterations in one study.

Accepted Manuscript

Funding

JNK is supported by the German Federal Ministry of Health (DEEP LIVER, ZMV11-2520DAT111) and the Max-Eder-Programme of the German Cancer Aid (grant #70113864), the German Federal Ministry of Education and Research (PEARL, 01KD2104C), the German Academic Exchange Service (SECAI, 57616814) and the European Union (ODELIA). The UCLH Biomedical Research Centre is funded by the National Institute for Health Research (BRC399/NS/RB/101410). SB is also supported by the Department of Health's NIHR Biomedical Research Centre's funding scheme. TOM was supported by The Brain Tumour Charity (GN-000389 clinical research training fellowship) and by the National Institute of health research (NIHR) with clinical lecturer fellowship (CL-2019-19-001).

Acknowledgements

Digitized images of tissue samples were obtained from University College London (UCL) Queens Square institute of Neurology, through BRAIN UK, which is funded by the Medical Research Council and Brain Tumour Research. The authors would like to thank Brain UK and UCL for enabling this part of the study. We would also like to thank the Institute of Pathology at University Hospital Carl Gustav Carus Dresden (UKD) / Technische Universität Dresden (TUD) for their assistance interpreting heatmap results.

Author contributions

Concept for the paper and experimental design: K.J.H., C.M.L., H.S.M and J.N.K.. UCL dataset preparation: T.O.M. and L.B.. Software development: by M.v.T.. Data preprocessing: C.M.L., O.S.M.EN, G.P.V., O.L.S., H.S.M and K.J.H.. K.J.H. ran all experiments and wrote the original manuscript draft. All authors contributed to interpretation of results, writing of the final manuscript and agreed to submission of the manuscript.

Conflicts of Interest

JNK declares consulting services for Owkin, France; DoMore Diagnostics, Norway and Panakeia, UK; furthermore he holds shares in StratifAI GmbH and has received honoraria for lectures by Bayer, Eisai, MSD, BMS, Roche, Pfizer and Fresenius. No other potential conflicts of interest are noted by any of the authors.

References

1. Whitfield BT, Huse JT. Classification of adult-type diffuse gliomas: Impact of the World Health Organization 2021 update. *Brain Pathol.* e13062 (2022).
2. WHO Classification of Tumours Editorial Board. Lyon (France): International Agency for Research on Cancer. WHO Classification of Tumours Editorial Board. Central nervous system tumours. [cited 2022 Aug 15]. (WHO classification of tumours series, 5th ed.; vol. 6). (2021).
3. Weller M, van den Bent M, Preusser M, et al. EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. *Nat. Rev. Clin. Oncol.* **18**, 170–186 (2021).
4. Kang YJ, Killen J, Caruana M, et al. The predicted impact and cost-effectiveness of systematic testing of people with incident colorectal cancer for Lynch syndrome. *Med. J. Aust.* **212**, 72–81 (2020).
5. Drake TM, Knight SR, Harrison EM, Søreide K. Global Inequities in Precision Medicine and Molecular Cancer Research. *Front. Oncol.* **8**, 346 (2018).
6. Kawauchi D, Ohno M, Miyakita Y, et al. Early Diagnosis and Surgical Intervention Within 3 Weeks From Symptom Onset Are Associated With Prolonged Survival of Patients With Glioblastoma. *Neurosurgery* **91**, 741 (2022).
7. Cifci D, Foersch S, Kather JN. Artificial intelligence to identify genetic alterations in conventional histopathology. *J Pathol.* 257(4):430-444 (2022).
8. Kacew AJ, Strohhahn GW, Saulsberry L, et al. Artificial Intelligence Can Cut Costs While Maintaining Accuracy in Colorectal Cancer Genotyping. *Front Oncol.* 2021 Jun 8;11:630953.

9. Fu Y, Jung AW, Torne RV, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat Cancer* **1**, 800–810 (2020).
10. Kather, JN, Heij LR, Grabsch HI, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nature Cancer* **1**, 789–799 (2020).
11. Liu S, Shah Z, Sav A, et al. Isocitrate dehydrogenase (IDH) status prediction in histopathology images of gliomas using deep learning. *Sci. Rep.* **10**, 7733 (2020).
12. Pei L, Jones KA, Shboul ZA, Chen JY, Iftekharuddin KM. Deep Neural Network Analysis of Pathology Images With Integrated Molecular Data for Enhanced Glioma Classification and Grading. *Front. Oncol.* **11**, 668694 (2021).
13. Shmatko A, Laleh NG, Gerstung M, Kather JN. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nat Cancer* **3**, 1026–1038 (2022).
14. Kurc T, Bakas S, Ren X, et al. Segmentation and Classification in Digital Pathology for Glioma Research: Challenges and Deep Learning Approaches. *Front. Neurosci.* **14**, 27 (2020).
15. Prokop G, Örtl M, Fotteler M, et al. Quantifying Heterogeneity in Tumors: Proposing a New Method Utilizing Convolutional Neuronal Networks. *Stud. Health Technol. Inform.* **289**, 397–400 (2022).
16. Ker J, Bai Y, Lee HY, Rao J, Wang L. Automated brain histology classification using machine learning. *J. Clin. Neurosci.* **66**, 239–245 (2019).
17. Jin W, Fatehi M, Abhishek K, Mallya M, Toyota B, Hamarneh G. Artificial intelligence in glioma imaging: challenges and advances. *J. Neural Eng.* **17**, 021002 (2020).

18. Im S, Hyeon J, Rha E, et al. Classification of Diffuse Glioma Subtype from Clinical-Grade Pathological Images Using Deep Transfer Learning. *Sensors* **21**, (2021).
19. Nalisnik M, Amgad M, Lee S et al. Interactive phenotyping of large-scale histology imaging data with HistomicsML. *Sci. Rep.* **7**, 14588 (2017).
20. Jin L, Shi F, Chun Q, et al. Artificial intelligence neuropathologist for glioma classification using deep learning on hematoxylin and eosin stained slide images and molecular markers. *Neuro. Oncol.* **23**, 44–52 (2021).
21. Faust K, Xie Q, Han D, et al. Visualizing histopathologic deep learning classification and anomaly detection using nonlinear feature space dimensionality reduction. *BMC Bioinformatics* **19**, 173 (2018).
22. Xu Y, Jia Z, Wang LB, et al. Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. *BMC Bioinformatics* **18**, 281 (2017).
23. Kather JN, Pearson AT, Halama N, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* **25**, 1054–1056 (2019).
24. Echle A, Grabsch HI, Quirke P, et al. Clinical-Grade Detection of Microsatellite Instability in Colorectal Tumors by Deep Learning. *Gastroenterology* **159**, 1406–1416.e11 (2020).
25. Muti HS, Heij LR, Keller G, et al. Development and validation of deep learning classifiers to detect Epstein-Barr virus and microsatellite instability status in gastric cancer: a retrospective multicentre cohort study. *Lancet Digit Health* **3**, e654–e664 (2021).
26. Saldanha OL, Loeffler CML, Niehues JM, et al. Self-supervised deep learning for pan-cancer mutation prediction from histopathology. *npj Precis. Onc.* **7**, 35 (2023).

27. Seraphin TP, Luedde M, Roderburg C, et al. Prediction of heart transplant rejection from routine pathology slides with self-supervised Deep Learning. *Eur Heart J Digit Health*. Mar 2;4(3):265-274 (2023).
28. Macenko M, Niethammer M, Marron JS, et al. A method for normalizing histology slides for quantitative analysis. in *Proceedings of the Sixth IEEE international conference on Symposium on Biomedical Imaging: From Nano to Macro* 1107–1110 (IEEE Press, 2009).
29. Wang X, Yang S, Zhang J, et al. Transformer-based unsupervised contrastive learning for histopathological image classification. *Med. Image Anal.* **81**, 102559 (2022).
30. Niehues JM, Quirke P, West NP, et al. Generalizable biomarker prediction from cancer pathology slides with self-supervised deep learning: A retrospective multi-centric study. *Cell Rep Med* **4**, 100980 (2023).
31. Wang LB, Karpova A, Gritsenko MA, et al. Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer Cell* **39**, 509–528.e20 (2021).
32. Mandrekar JN. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *J. Thorac. Oncol.* **5**, 1315–1316 (2010).
33. Stacke K, Eilertsen, G, Unger J, Lundström C. A Closer Look at Domain Shift for Deep Learning in Histopathology. *arXiv [cs.CV]* (2019). <https://doi.org/10.48550/arXiv.1909.11575>
34. Brown NF, Ottaviani D, Tazare J, et al. Survival Outcomes and Prognostic Factors in Glioblastoma. *Cancers* **14**, (2022).

Figure and Table Captions

Table 1: Table 1 provides information on the number of patients within each cohort for which genetic alteration data was collected. In the TCGA cohort, the low grade glioma (LGG) and high grade glioma (GBM) datasets were combined. As the subtype according to WHO CNS5 was not available in the TCGA dataset, molecular alteration data for *IDH*, 1p19q and *ATRX* was used to formulate the WHO CNS5 diagnosis. This process was completed with guidance from a Neuropathologist. Consort charts for each cohort can be found in Suppl. Fig. 2. The CPTAC cohort was only used in the *CDKN2A/B* external validation experiment and thus case numbers for the other mutations were not applicable.

Abbreviations: N/d = no data available. N/a = data not applicable. Ch+7/-10 indicates trisomy of chromosome 7 with monosomy of chromosome 10. For gene amplifications, altered represents amplification and unaltered no amplification. For chromosomal alterations, altered indicates deletion and/or gain and unaltered normal ploidy. * subtype was formulated using molecular alteration data provided with the cohort.

Table 2: Results of Internal and External Validation Experiments. This table provides results for all experiments and details of the cohorts used for each experiment. The molecular alteration experiments were run as individual experiments. The subtype experiments were run as one experiment but results are listed for the prediction of each subtype individually. Experiments 1 - 13 were run in an internal validation setting. A five-fold internal cross-validation experiment was conducted, where the data is split randomly into five parts, with four parts being used for training the network and the remaining part used to test the model performance. This was repeated five times with the test data being rotated each time. Experiments 14 - 25 were external validation experiments, with training and deployment on the cohorts listed. For external validation, all data from one cohort was used to train the model and then deployed on an independent test

cohort. We used UCL as a training cohort and TCGA and/or CPTAC as test sets, with the exception of *CDKN2A/B*, as we only had data within one cohort.

Experiments where the target is a 2016 subtype refers to gliomas that were diagnosed according to the 2016 WHO diagnostic guidelines. Likewise, experiments where the target is a 2021 subtype refers to gliomas that were diagnosed according to the 2021 WHO diagnostic guidelines. Further details on the diagnostic approaches can be found in Fig. 1 A&B.

Abbreviations: Exp. No. = experiment number, AUROC = area under the receiver operating curve. The rate at which a model can correctly predict the target, 95% CI = 95% confidence interval. A range of the AUROC +/- amount indicated by 95% CI. We can be 95% certain that the true AUROC falls within this range. Target = the feature that the network was aiming to predict in that experiment. N/a = not applicable.

Figure 1: Overview of our Experimental Approach. The flow-chart figures in 1A and 1B outline the 2016 and 2021 WHO diagnostic algorithm for diffuse-type adult gliomas, for targets we included in our experiments. In the 2016 system, gliomas were tested for *IDH* mutation status before the morphological features were assessed to determine grade. Both *IDHmut* and *IDHwt* tumors with high-grade (grade IV) features were designated glioblastoma. Tumors with lower grade (grade II or III) morphology were designated as astrocytoma, unless *1p19q* co-deletion was present. Lower grade tumors with *IDHmut* and *1p19q* co-deletion were designated oligodendrogliomas. In the 2021 system, molecular alteration status determines the subtype of glioma. Astrocytomas are *IDHmut* and *ATRXmut*. If *CDKN2A* or *CDKN2B* homozygous deletion is additionally present in an astrocytoma, this automatically upgrades the tumor to grade 4. Absence of *CDKN2A/B* homozygous deletion and absence of high grade morphology indicates grade 2 or 3. Oligodendrogliomas are *IDHmut*, *1p19q* co-deleted (complete loss of both arms), *TERTmut* and *ATRXwt*. Grade 4 oligodendroglioma are not recognized by the 2021 system. Glioblastomas are *IDHwt* with either classic morphology on histology (microvascular proliferation and/or necrosis) or at least one of

*TERT*mut, *EGFR* amplification, +/-10 genotype. Abbreviations: MUT - mutated form, WT - wildtype form. ND - not deleted.

The doughnut graphs in figures 1C and 1D show the data split by both 2016 and 2021 subtype diagnosis for the UCL and TCGA cohorts respectively. Subtype diagnosis was available for both 2016 and 2021 criteria for the UCL dataset. In the TCGA dataset, the molecular alteration status for *IDH*, 1p19q and/or *ATRX* was used to formulate the 2021 subtype. Figure 1E provides an overview of our Deep Learning pipeline. The first step is pre-processing where digital WSIs are tessellated into tiles and features extracted for each tile. These features are then given to our network and used for either training, testing or deployment, depending on the type of experiment being run. Further details of our Deep Learning methods can be found in Supplementary Methods.

Figure 2: Results for Subtype Experiments. This figure shows Receiver Operating Characteristic (ROC) Curves and Confusion Matrices (CM) for both 2016 and 2021 subtype experiments. Subtype experiments were run as a single experiment. The AUROCs visualize results for each subtype individually, whereas the CMs and additional statistics in Fig. 4 relate to overall model performance. CMs and additional statistics were calculated with a threshold of 0.5. In each ROC plot (A-C, E-G), the thin lines indicate ROC curves for internal validation experiments. Internal validation was performed as five-fold cross-validation. The dark blue line indicates the external validation results and shaded area around this line indicates the CI. The $AUC \pm$ bootstrap CI is given in the bottom right of each plot. Please note, AUC refers to the area under the ROC curve, and is thus the same as AUROC.

D and H are heatmap confusion matrices for the 2016 and 2021 subtype experiments respectively. The confusion matrices are constructed from the model prediction output for the external validation experiments i.e. the class with the highest probability in external validation was selected as the predicted class.

Figure 3: Results for Molecular Alteration Experiments. A-D show Receiver Operating Characteristic (ROC) Curves for the molecular alteration experiments. In each ROC plot, the thin lines indicate ROC curves for internal validation experiments. Internal validation was performed as five-fold cross-validation. The dark blue line indicates external validation results and the shaded area around this line indicates the bootstrap CI. The $AUC \pm$ bootstrap CI is also given in the bottom right of each plot. Please note, AUC refers to the area under the ROC, and is thus the same as AUROC. E-H are confusion matrices (CMs) for each molecular alteration experiment. The CMs are constructed from the model prediction output for the external validation experiments i.e. the class with the highest probability in external validation was selected as the predicted class. CMs were calculated with a threshold of 0.5.

Abbreviations: MUT = mutant, WT = wildtype, ALT = altered, UA = unaltered.

Figure 4: Statistical Heatmaps. Figure 4 is a heatmap of the further statistical analysis for our results. A shows results for the direct prediction of the 2021 subtypes. B shows results for the final prediction following the sequential approach. Here, the 2021 subtype was calculated by stacking the final predictions for the *IDH*, *ATRX* and 1p19q experiments. Results for *IDH* prediction were considered first, followed by *ATRX* and 1p19q which were assessed together. Statistics were calculated with a threshold of 0.5.

Table 1.

		UCL n=1882			TCGA n=864			CPTAC n=99		
		Altered	Un-altered	N/d	Altered	Un-altered	N/d	Altered	Un-altered	N/d
Core Diagnostic Alterations	<i>IDH</i>	1116	755	11	410	318	136	N/a	N/a	N/a
	<i>ATRX</i>	584	887	411	207	521	136	N/a	N/a	N/a
	1p19q	316	683	883	162	542	160	N/a	N/a	N/a
Additional Diagnostic Alterations	<i>TERT</i>	524	277	1081	7	721	136	N/a	N/a	N/a
	<i>EGFR</i>	454	1115	313	207	650	7	N/a	N/a	N/a
	Ch+7/-10	N/d	N/d	N/d	247	461	156	N/a	N/a	N/a
	<i>CDKN2A/B</i>	7	54	1821	266	592	6	56	40	3
WHO Subtype 2016	Astrocytoma	600		7	180		280	N/a		
	Glioblastoma	822			231					
	Oligodendroglioma	453			173					
WHO Subtype 2021	Astrocytoma	666		7	276*		148	N/a		
	Glioblastoma	756			279*					
	Oligodendroglioma	453			161*					

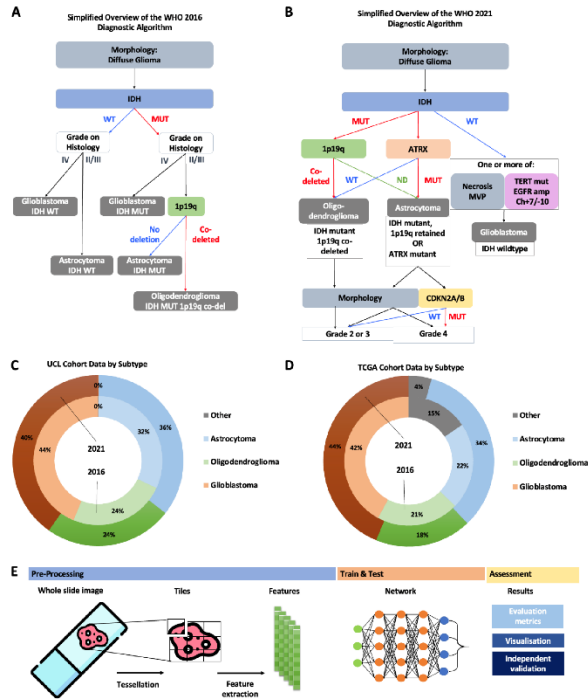
Table 2.

	Exp. No.	Target	Train Cohort	Deploy Cohort	AUROC	95% CI
Internal Validation Experiments	1	2016 Astrocytoma	UCL	n/a	0.89	0.02
	2	2016 Glioblastoma	UCL	n/a	0.94	0.01
	3	2016 Oligodendroglioma	UCL	n/a	0.93	0.01
	4	2021 Astrocytoma	UCL	n/a	0.92	0.03
	5	2021 Glioblastoma	UCL	n/a	0.95	0.02
	6	2021 Oligodendroglioma	UCL	n/a	0.93	0.02
	7	<i>IDH</i>	UCL	n/a	0.95	0.02
	8	<i>ATRX</i>	UCL	n/a	0.91	0.03
	9	1p19q	UCL	n/a	0.80	0.03
	10	<i>TERT</i>	UCL	n/a	0.74	0.03
	11	<i>EGFR</i>	UCL	n/a	0.83	0.04
	12	Ch+7/-10	TCGA	n/a	0.91	0.03
	13	<i>CDKN2A/B</i>	TCGA	n/a	0.82	0.02
	Exp. No.	Target	Train Cohort	Deploy Cohort	AUROC	95% CI
External Validation Experiments	14	2016 Astrocytoma	UCL	TCGA	0.86	0.03
	15	2016 Glioblastoma	UCL	TCGA	0.91	0.03
	16	2016 Oligodendroglioma	UCL	TCGA	0.86	0.04
	17	2021 Astrocytoma	UCL	TCGA	0.84	0.03
	18	2021 Glioblastoma	UCL	TCGA	0.90	0.02
	19	2021 Oligodendroglioma	UCL	TCGA	0.91	0.03
	20	<i>IDH</i>	UCL	TCGA	0.90	0.02

21	<i>ATRX</i>	UCL	TCGA	0.79	0.04
22	1p19q	UCL	TCGA	0.87	0.03
23	<i>TERT</i>	UCL	TCGA	0.60	0.27
24	<i>EGFR</i>	UCL	TCGA	0.85	0.03
25	<i>CDKN2A/B</i>	TCGA	UCL + CPTAC	0.73	0.07

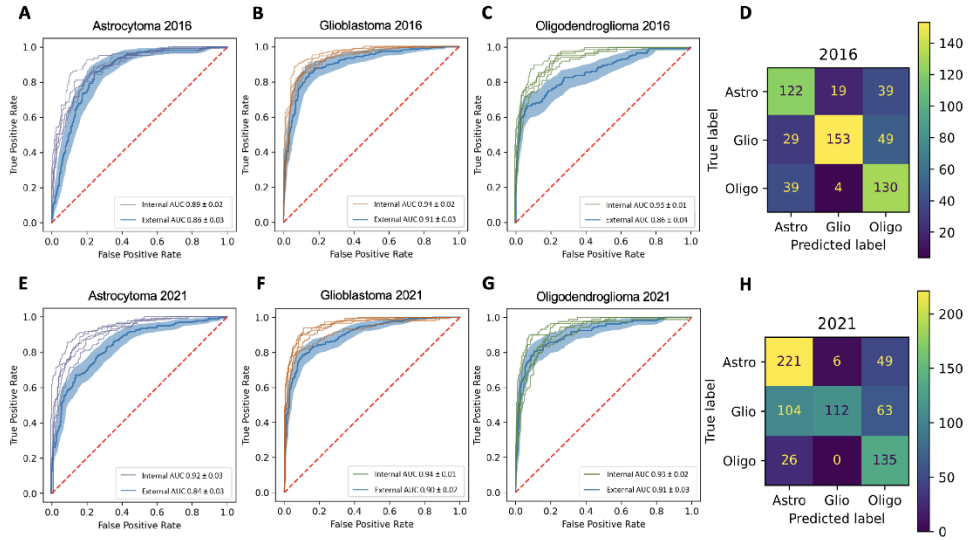
Accepted Manuscript

Figure 1



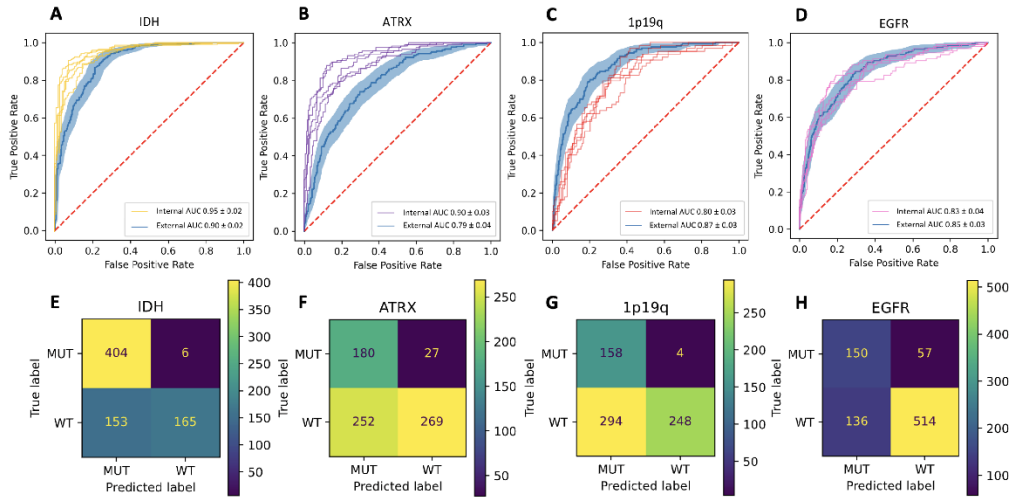
Accepted Manuscript

Figure 2



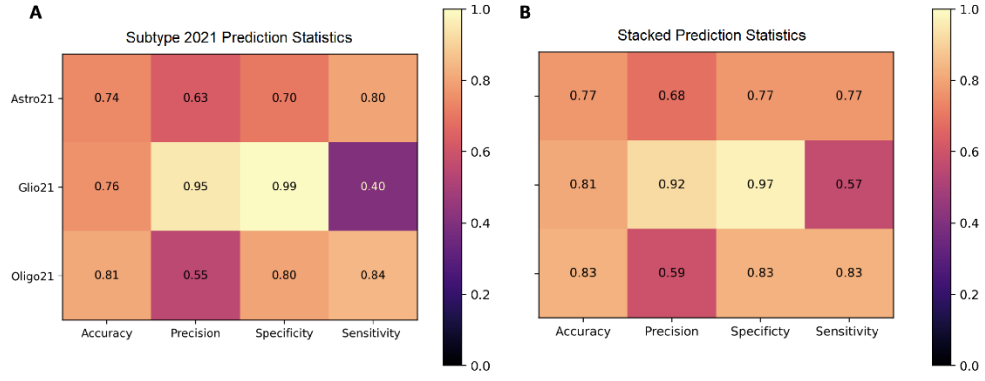
Accepted Manuscript

Figure 3



Accepted Manuscript

Figure 4



Accepted Manuscript