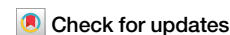


<https://doi.org/10.1038/s44303-024-00020-8>

Applications of artificial intelligence in the analysis of histopathology images of gliomas: a review



Jan-Philipp Redlich¹ ✉, Friedrich Feuerhake^{2,3}, Joachim Weis⁴, Nadine S. Schaadt², Sarah Teuber-Hanselmann⁵, Christoph Buck⁶, Sabine Luttmann⁷, Andrea Eberle⁷, Stefan Nikolin⁴, Arno Appenzeller⁸, Andreas Portmann⁹ & André Homeyer¹

In recent years, the diagnosis of gliomas has become increasingly complex. Analysis of glioma histopathology images using artificial intelligence (AI) offers new opportunities to support diagnosis and outcome prediction. To give an overview of the current state of research, this review examines 83 publicly available research studies that have proposed AI-based methods for whole-slide histopathology images of human gliomas, covering the diagnostic tasks of subtyping (23/83), grading (27/83), molecular marker prediction (20/83), and survival prediction (29/83). All studies were reviewed with regard to methodological aspects as well as clinical applicability. It was found that the focus of current research is the assessment of hematoxylin and eosin-stained tissue sections of adult-type diffuse gliomas. The majority of studies (52/83) are based on the publicly available glioblastoma and low-grade glioma datasets from The Cancer Genome Atlas (TCGA) and only a few studies employed other datasets in isolation (16/83) or in addition to the TCGA datasets (15/83). Current approaches mostly rely on convolutional neural networks (63/83) for analyzing tissue at 20x magnification (35/83). A new field of research is the integration of clinical data, omics data, or magnetic resonance imaging (29/83). So far, AI-based methods have achieved promising results, but are not yet used in real clinical settings. Future work should focus on the independent validation of methods on larger, multi-site datasets with high-quality and up-to-date clinical and molecular pathology annotations to demonstrate routine applicability.

Gliomas account for roughly a quarter of all primary non-malignant and malignant central nervous system (CNS) tumors, and 81% of all malignant primary CNS tumors¹. Generally, gliomas can be differentiated into circumscribed and diffuse gliomas^{2,3}. Whereas circumscribed gliomas are usually benign and potentially curable with complete surgical resection, diffuse gliomas are characterized by diffuse infiltration of tumor cells in the brain parenchyma and are mostly not curable, with an inherent tendency for malignant progression and recurrence^{2,3}. In particular, glioblastoma, the most common and most aggressive malignant CNS tumor, makes up 60% of

all diagnosed gliomas and caused 131,036 deaths in the US between 2001 and 2019 alone¹. Although the introduction of now well-established standard therapy protocols, involving surgical resection, radiation, and chemotherapy⁴ as well as guidance by MGMT promoter methylation status⁵, has substantially improved clinical outcomes, 5-year survival in patients with glioblastoma has been largely constant at only 6.9%¹.

The current international standard for the diagnosis of gliomas is the fifth edition of the WHO Classification of Tumors of the Central Nervous System, published in 2021⁶. Based on advances in the understanding of CNS

¹Fraunhofer Institute for Digital Medicine MEVIS, Max-von-Laue-Straße 2, 28359 Bremen, Germany. ²Hannover Medical School, Carl-Neuberg-Straße 1, 30625 Hannover, Germany. ³Institute of Neuropathology, Medical Center—University of Freiburg, Breisacherstraße 64, 79106 Freiburg, Germany. ⁴Institute of Neuropathology, RWTH Aachen University Hospital, Pauwelsstrasse 30, 52074 Aachen, Germany. ⁵Department of Neuropathology, Center for Pathology, Klinikum Bremen-Mitte, Sankt-Jürgen-Straße 1, 28205 Bremen, Germany. ⁶Leibniz Institute for Prevention Research and Epidemiology—BIPS, Achterstraße 30, 28359 Bremen, Germany. ⁷Bremen Cancer Registry, Leibniz Institute for Prevention Research and Epidemiology—BIPS, Achterstraße 30, 28359 Bremen, Germany. ⁸Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Fraunhoferstraße 1, 76131 Karlsruhe, Germany. ⁹German Heart Center Berlin, Augustenburger Platz 1, 13353 Berlin, Germany. ✉e-mail: jan-philipp.redlich@mevis.fraunhofer.de

tumors, the WHO classification system has been constantly evolving since its first issue in 1979, with new versions released every 7–9 years⁷. Whereas previous versions greatly relied on histological assessment via light microscopy, the current classification system incorporates both histopathological features and molecular alterations (Fig. 1), allowing for a more uniform delineation of disease entities^{6,7}. The basic principle of classification according to histological typing, histogenesis, and grading, however, remains fundamental and underlines the role of conventional histological assessment as an important element of timely, standardized, and globally consistent diagnostic workflows⁶.

The rapidly evolving field of computational pathology, in particular, the analysis of digital whole-slide images (WSIs) of tumor tissue sections using artificial intelligence (AI), offers new opportunities to support current diagnostic workflows. Promising applications range from automation of time-consuming routine tasks, such as subtyping and grading, to tasks that cannot be accurately performed by human observers, such as predicting molecular markers or survival directly from hematoxylin and eosin (H&E)-stained routine sections of formalin-fixed and paraffin-embedded (FFPE) tumor tissue^{8–10}. Essential concepts of AI and AI-based analysis of WSIs are briefly explained in Tables 1 and 2 and Fig. 2.

This review provides a comprehensive overview of the current state of research on AI-based analysis, including both traditional machine learning

(ML) and deep learning (DL), of whole-slide histopathology images of human gliomas. Motivated by a significant increase in corresponding publications (Fig. 3), this review focuses on the diagnostic tasks of subtyping, grading, molecular marker prediction, and survival prediction. Studies primarily addressing image segmentation^{11–15}, image retrieval¹⁶, or tumor heterogeneity^{17–22} have been considered out of scope. The reviewed studies are examined with regard to diagnostic tasks and methodological aspects of WSI processing, as well as discussed by addressing limitations and future directions. Prior related reviews on AI-based glioma assessment addressed other imaging modalities, such as computed tomography or magnetic resonance imaging (MRI)^{23–29}, and specific diagnostic tasks, such as grading^{30,31} or survival prediction^{32–34}.

Results

Diagnostic tasks

The reviewed studies covered the diagnostic tasks of subtyping (23/83), grading (27/83), molecular marker prediction (20/83), and survival prediction (29/83). All studies focused on the assessment of adult-type diffuse gliomas (Fig. 1) and nearly all were based on H&E-stained tissue sections (only two studies considered other histological staining techniques^{35,36}).

The majority of the studies (52/83) were based on two public datasets from The Cancer Genome Atlas, which originated from the Glioblastoma^{37,38} (TCGA-GBM) and Low-Grade Glioma³⁹ (TCGA-LGG)

Fig. 1 | Schematic diagnostic criteria for adult-type diffuse glioma as defined by the 2021 WHO Classification of Tumors of the Central Nervous System.

The 2021 WHO classification system reorganizes gliomas into adult-type diffuse gliomas, pediatric-type diffuse low-grade and high-grade gliomas, circumscribed astrocytic gliomas, and ependymal tumors⁵. Adult-type diffuse gliomas comprise Glioblastoma, IDH-wildtype, CNS WHO grade 4, Astrocytoma, IDH-mutant, CNS WHO grade 2–4, and Oligodendroglioma, IDH-mutant and 1p/19q-codeleted, CNS WHO grade 2–3⁶. Identified adult-type diffuse glioma are stratified from left to right according to molecular alterations, including mutations in isocitrate dehydrogenase 1/2 (IDH1/2) genes and whole-arm codeletion of chromosomes 1p and 19q, as well as histological features, including increased mitoses, necrosis and/or microvascular proliferation. “or” is understood to be non-exclusive.

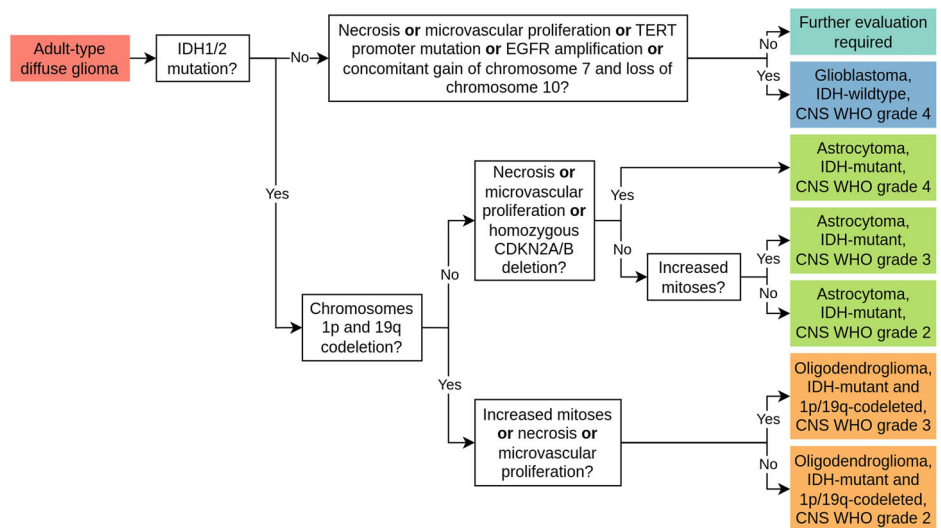


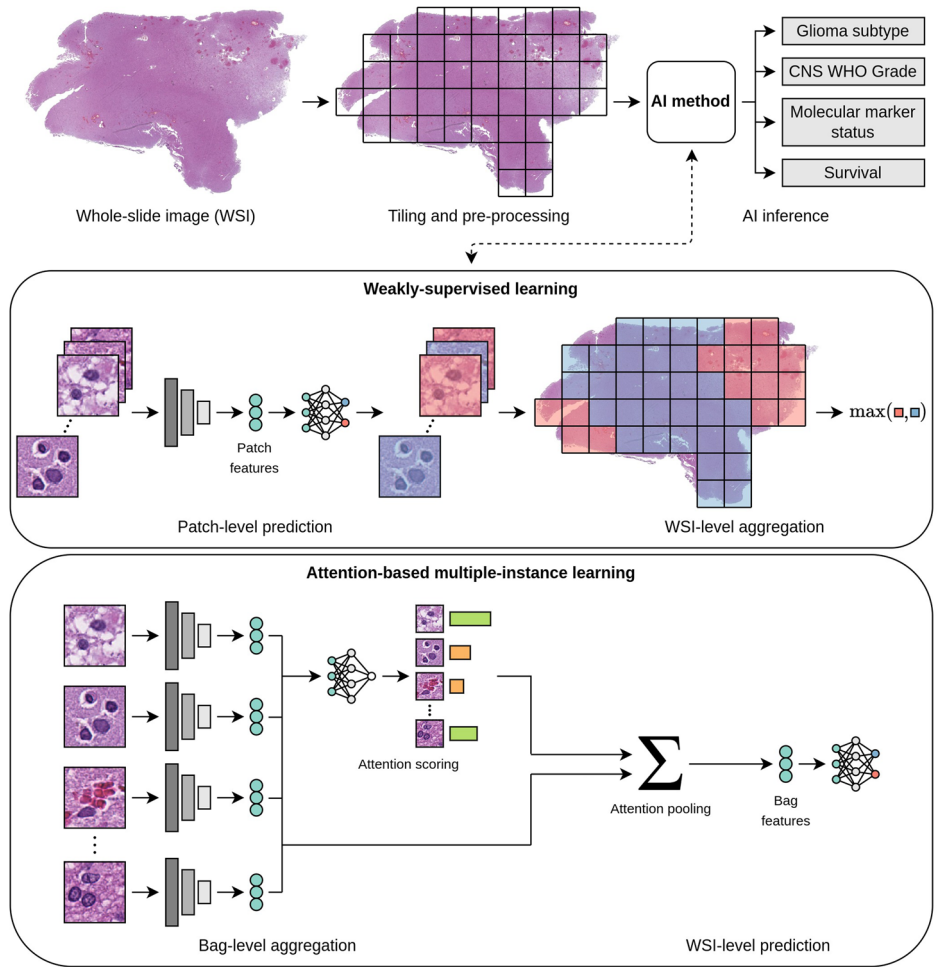
Table 1 | Essential concepts in artificial intelligence

Artificial intelligence (AI)	The broad field of AI encompasses the research and development of computer systems that are capable of solving complex tasks typically attributed to human intelligence.
Machine learning (ML)	ML is a subfield of AI that uses algorithms that automatically learn to recognize patterns from data. In this way, computer systems can be enabled to perform complex cognitive tasks that are difficult to express in the form of human-understandable rules. ML algorithms are mainly used for tasks such as classification (e.g., predicting tumor types), regression (e.g., predicting survival) or segmentation.
Training and Testing	The development of ML models is typically split into a training and testing phase. During training, an ML algorithm deduces a ML model from a given training dataset. During testing, the model’s performance is assessed using a test dataset. The test dataset comprises data which the model has not seen during training, which is crucial for meaningful performance estimates.
Performance metrics	The performance of ML models is measured by metrics such as accuracy (ACC) or area under the operator receiver characteristic curve (AUC) for classification tasks (e.g., subtyping, grading, molecular marker prediction) or concordance index (C-index) ³² for survival prediction. Metrics closer to 1.0 (resp. 0.5) indicate higher (resp. random) predictive performance.
Features	To apply ML algorithms to image data, images are usually encoded into a set of features. Traditionally, these features were “handcrafted” by human researchers based on prior knowledge or assumptions to quantify relevant pixel patterns or the morphology or spatial distribution of segmented cell or tissue structures.
Deep learning (DL)	DL is an advanced form of ML in which deep artificial neural networks are used to automatically learn features from the training data. In this way, more complex cognitive tasks can often be solved than is possible with handcrafted features. Important types of neural network, so called “architectures”, for DL-based image analysis are convolutional neural networks (CNNs) ^{102–107} and Vision Transformers (ViTs) ¹³² . The training of DL models usually requires very large datasets, which can be mitigated by techniques such as transfer learning ^{14,133} or self-supervised learning (SSL) ^{108,134} .

Table 2 | Essential concepts of artificial intelligence-based analysis of WSIs

Whole-slide images (WSIs)	WSIs are digitized microscopic images of whole tissue sections. As they are scanned with a very high resolution of up to 0.25 micrometers per pixel, WSIs can be several gigapixels in size. To enable faster access, WSIs are usually stored as image pyramids consisting of several magnification levels, which are labeled according to the comparable objective magnifications of an analog microscope (e.g., 5x, 10x, 20x, or 40x). The large size and the great variability and complexity of depicted morphological patterns make the analysis of WSI very challenging.
Regions of interest (ROIs)	As tumor tissue sections naturally include different tissue types, some ML models require a delineation of ROIs of relevant tumor tissue. Usually, this delineation is a laborious task that requires expert knowledge.
Patch-based processing of WSIs	Due to their gigapixel size, WSIs and ROIs are usually processed by partitioning them into smaller image patches (or tiles) and processing patches individually. This greatly reduces the memory requirements and facilitates parallelization.
Weakly-supervised learning (WSL)	WSL enables the training of ML models on entire WSIs, eliminating the need for annotating ROIs. The classical approach to WSL assumes that all patches inherit the label (e.g., the tumor type) of the respective WSI. Using these pairs of patches and patch-level labels the ML model is trained to predict a label for each patch. The patch-level predictions are then aggregated (e.g., by averaging) to obtain a WSI-level prediction ¹³⁵ (Fig. 2).
Multiple-instance learning (MIL)	MIL is a special type of WSL and thus also functions without ROIs. A common approach to MIL groups all patches from a WSI in a “bag” and assigns the label of the WSI to the bag instead of to each individual patch. The ML model is thus trained to predict labels for bags instead of individual patches, enabling the modeling of complex interactions between patches ¹³⁵ . Examples of MIL include attention-based MIL ¹³⁶ (Fig. 2), and clustering-constrained attention MIL ¹³⁷ . In the rest of this paper, the term “WSL” only refers to the aforementioned classical WSL approach.
Multi-modal fusion	Fusing WSIs with other modalities of potentially complementary information, such as MRI, clinical, or omics data, poses great opportunities for improving diagnostic and prognostic precision ¹³⁸ . Common approaches can be differentiated by the extent to which given modalities are processed in combination before predictions are inferred: Early fusion concatenates modalities at the onset; late fusion aggregates individual uni-modal predictions, e.g., by averaging; and intermediate fusion infers final predictions from intermediate multi-modal representations ^{138,139} .

Fig. 2 | Typical workflow for artificial intelligence-based analysis of WSIs. Most approaches partition WSIs into smaller image patches (“Tiling”). Using AI methods, the patches are processed individually and then aggregated to obtain a prediction for the entire WSI. AI methods are often based on weakly-supervised learning or attention-based multiple-instance learning¹³⁶, each of which are depicted in greater detail.



projects. Both contain information on diagnosis, molecular alterations, and survival for approximately 600 and 500 patients, respectively. Patients in both datasets had been diagnosed before 2013 and 2011^{37–39} and therefore according to versions of the WHO classification system before 2016. The

TCGA-GBM dataset includes primary glioblastomas (WHO grade IV), and the TCGA-LGG dataset includes astrocytomas, anaplastic astrocytomas, oligodendrogliomas, oligoastrocytomas, and anaplastic oligoastrocytomas (all WHO grade II or III). An independent research project later annotated

regions of interest (ROIs) of relevant tumor tissue for both datasets and made these publicly available⁴⁰.

Other studies employed other datasets in isolation (16/83) or in addition to the TCGA datasets (15/83). Diagnoses in these datasets corresponded either to the current version of the WHO classification from 2021^{41–43}, or previous versions from 2016^{44–46} or earlier. Table 3 presents an overview of all publicly available datasets utilized by the reviewed studies.

It should be noted that the 2021 WHO Classification of Tumors of the Central Nervous System changed the notation of tumor grades from Roman to Arabic numerals and endorsed the use of the term “CNS WHO grade”³⁶. Throughout this review, these changes in notation are used only for diagnoses that were reported according to the 2021 WHO classification system.

Subtyping. Subtyping gliomas is of fundamental importance for the diagnosis and subsequent treatment of patients^{6,7}. Before 2016, the WHO classification system incorporated only histological features. Oligodendrogliomas, for instance, were characterized by features including round nuclei and a “fried egg” cellular appearance; astrocytomas by more irregular nuclei with clumped chromatin, eosinophilic cytoplasm, and branched cytoplasmic processes; and glioblastomas by an astrocytic phenotype combined with the presence of necrosis and/or microvascular proliferation⁷. The latest revisions of the WHO classification system published in 2016 and 2021 reflect important advances in the

understanding of the molecular pathogenesis of gliomas by incorporating molecular alterations in an integrated diagnosis (Fig. 1). For instance, since 2021, the diagnosis of glioblastoma has been limited to malignant non-IDH-mutated (i.e., IDH-wildtype) gliomas. Furthermore, the diagnosis of oligoastrocytomas has been removed from the 2021 WHO classification system. Until 2016, oligoastrocytomas were considered a distinct tumor entity with mixed oligodendroglial and astrocytic histological features, and between 2016 and 2021, they could only be diagnosed in cases with unknown 1p/19q and IDH status (these cases were designated as “oligoastrocytoma, NOS (Not Otherwise Specified)” and diagnosis was not recommended unless molecular testing was unavailable)^{6,7}. Table 4 provides an overview of all corresponding studies grouped according to the investigated subtyping tasks.

Whereas one study aligned IHC- (for IDH1-R132H and ATRX) and H&E-stained tissue sections to differentiate gliomas along IDH mutation status and astrocytic lineage³⁵, all other studies predicted glioma types only from H&E-stained tissue. Prediction was usually performed in an end-to-end manner, i.e., without prior identification of established histological or molecular markers. Many approaches were based on weakly-supervised learning (WSL), in which predictions of patch-level convolutional neural networks (CNNs) were aggregated to patient-level diagnosis^{46–49}. Aggregation of patch-level to slide-level results was either performed by majority voting^{46,49} or traditional ML methods fitted to histograms of patch-level predictions^{47,48}.

Two studies focused on the analysis of intraoperative frozen sections (IFSs): Shi et al.⁴⁶ proposed two weakly supervised CNNs for diagnosing intracranial germinomas, oligodendrogliomas, and low-grade astrocytomas from H&E-stained and intraoperative frozen sections (IFSs), respectively⁴⁶. They compared the diagnostic accuracy of three pathologists with and without the assistance of their proposed models and reported an average improvement of 40% and 20%, respectively. Nasrallah et al.⁴² used a hierarchical Vision Transformer (ViT) architecture and three glioma patient cohorts to support several diagnostic tasks during surgery, including the identification of malignant cells, as well as the prediction of histological grades, glioma types according to the 2021 WHO classification system, and numerous molecular markers⁴².

Three other recent studies also leveraged emerging ViT architectures for subtyping^{41,45,50}. Notably, Li et al., employed TCGA data and a proprietary patient cohort for three tasks of brain tumor classification, including glioma subtyping, as well as prediction of IDH1, TP53, and MGMT status in gliomas⁴⁵. Their proposed model featured an ImageNet pre-trained ViT for encoding image patches and a graph- and Transformer-based aggregator for modeling relations among patches. It achieved high performance in all tasks and automatically learned histopathological features, including blood vessels, hemorrhage, calcification, and necrosis.

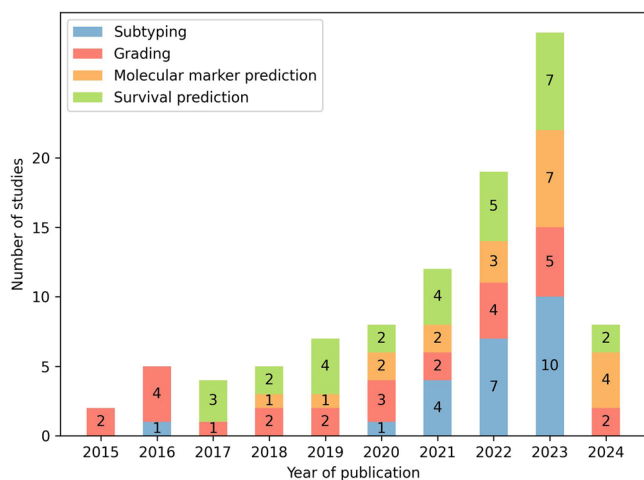


Fig. 3 | Number of studies by year of publication and diagnostic task. All 70 studies included in this review are shown. Studies published in 2024 are considered up until March 18, 2024.

Table 3 | Overview of all publicly available datasets utilized by reviewed studies

Dataset	Tumor types	WHO classification version	#Patients	#WSIs	Tissue preparation	Tissue staining	Other information
CPTAC-GBM ¹⁴⁰	Glioblastoma	n.s.	178	462	n.s.	H&E	CT, MRI, omics, clinical
Digital Brain Tumor Atlas ¹⁴¹	126 distinct brain tumor types	2016	2880	3115 ^a	FFPE	H&E	Clinical
Ivy Glioblastoma Atlas Project ¹⁴²	Glioblastoma	n.s.	41	~34,500	Frozen	ISH, H&E	CT, MRI, omics, clinical, annotations
TCGA-GBM ^{37,38}	Glioblastoma	2000, 2007	617	2053	FFPE, frozen	H&E	Omics, clinical
TCGA-LGG ³⁹	Astrocytoma, anaplastic astrocytoma, oligodendroglioma, oligoastrocytoma, anaplastic oligoastrocytoma	2000, 2007	516	1572	FFPE, frozen	H&E	Omics, clinical
UPENN-GBM ¹⁴³	Glioblastoma	2016	34	71	n.s.	H&E	MRI, omics, clinical

Other datasets that were utilized but were not made publicly available are not listed.

FFPE formalin-fixed and paraffin-embedded, H&E hematoxylin and eosin staining, ISH in situ hybridization, CT computed tomography, MRI magnetic resonance imaging, n.s. not specified

^aincluding 47 non-tumor WSIs.

Table 4 | Overview of all studies related to subtyping

Study	Method Features	Learning paradigm		Multi-modal fusion		Dataset		Performance		
		Modality	Fusion Strategy	TCGA-GBM	TCGA-LGG	Other	#Patients	#WSIs	Metric	Result
Astrocytoma vs. Oligodendroglioma										
Jin et al. ⁵⁶	CNN (DenseNet)	SL	H&E	X	X	n.s.	733	n.s.	ACC	0.920
Kurc et al. ¹³	CNN (DenseNet)	WSL	H&E, MRI	Late	X	52	n.s.	n.s.	ACC	0.900
Astrocytoma IDH-mutant vs. Oligodendroglioma IDH-mutant, 1p/19q codeleted										
Kim et al. ⁵⁴	CNN (ResNet50)	WSL	H&E		X	673	673	673	AUC	0.837
Astrocytoma IDH-mutant vs. Astrocytoma IDH-wildtype vs. Oligodendroglioma IDH-mutant, 1p/19q codeleted										
Faust et al. ⁵⁵	CNN (VGG19)	SL	H&E, IHC	Late	X	n.s.	1013	1013	ACC	1.0
Astrocytoma IDH-mutant vs. Oligodendroglioma IDH-mutant, 1p/19q codeleted vs. Glioblastoma IDH-wildtype										
Hewitt et al. ¹¹	CNN, VIT (CTransPath)	MIL	H&E		X	2741	n.s.	n.s.	AUC	0.840*, 0.910*, 0.900*
Nasrallah et al. ⁵²	VIT	WSL	IFS		X	1524	2334	2334	AUC	0.900*, 0.880*, 0.930*
Wang et al. ⁴³	CNN (ResNet50)	WSL	H&E		X	2624	2624	2624	AUC	0.941*, 0.973*, 0.988*
Wang et al. ⁵⁵	VIT	MIL	H&E		X	940	2633	2633	ACC	0.773
Jose et al. ⁴⁹	CNN (ResNet50)	WSL	H&E		X	700	926	926	AUC	0.961
Astrocytoma vs. Oligodendroglioma vs. Glioblastoma										
Hsu et al. ⁵¹	CNN (ResNet50)	WSL	H&E, MRI	Late	X	329	n.s.	n.s.	balanced ACC	0.654
Majlya et al. ⁵³	CNN	MIL	H&E, MRI	KD	X	n.s.	221	221	balanced ACC	0.752
Suman et al. ¹⁴	CNN	WSL, MIL	H&E		X	230	n.s.	n.s.	balanced ACC	0.730
Wang et al. ⁵²	CNN (EfficientNet, SEResNeX101)	WSL	H&E, MRI, clinical	Late	X	378	n.s.	n.s.	balanced ACC	0.889
Lu et al. ¹⁰⁹	CNN (ResNet50)	SSL, MIL	H&E	Intermediate	X	700	700	700	ACC	0.886
Astrocytoma IDH-mutant vs. Astrocytoma IDH-wildtype vs. Oligodendroglioma IDH-wildtype vs. Glioblastoma IDH-mutant vs. Glioblastoma IDH-wildtype										
Chitrinis et al. ¹¹³	CNN (KimiNet)	MIL	H&E		X	791	866	866	AUC	0.969
Astrocytoma vs. Anaplastic Astrocytoma vs. Oligodendroglioma vs. Anaplastic Oligodendroglioma vs. Glioblastoma										
Li et al. ⁴⁵	VIT (ViT-L-16)	MIL	H&E		X	749	1487	1487	AUC	0.932
Jin et al. ⁴⁸	CNN (DenseNet)	WSL	H&E		X	323	323	323	ACC	0.865
Xing et al. ¹⁰⁰	CNN (DenseNet)	WSL	H&E		X	n.s.	440	440	ACC	0.700
Astrocytoma vs. Anaplastic Astrocytoma vs. Oligodendroglioma vs. Anaplastic Oligodendroglioma vs. Oligoastrocytoma vs. Glioblastoma										
Hou et al. ⁴⁷	CNN	WSL	H&E		X	539	1064	1064	ACC	0.771
Astrocytoma vs. Oligodendroglioma vs. Intracranial Germinoma										
Shi et al. ⁴⁶	CNN (ResNet152)	WSL	H&E, IFS		X	346	832	832	ACC	0.769, 0.820*
Astrocytoma vs. Oligodendroglioma vs. Ependymoma vs. Lymphoma vs. Metastasis vs. Non-tumor										
Ma et al. ⁵⁵	CNN (DenseNet)	SL	H&E		X	n.s.	1038	1038	ACC	0.935
Astrocytic vs. Oligodendroglial vs. Ependymal										
Yang et al. ¹⁰⁸	CNN (ResNet18)	SSL, MIL	H&E		X	n.s.	935	935	weighted F1-score	0.780
Oligodendroglioma vs. Non-Oligodendroglioma										
Im et al. ⁴⁴	CNN (ResNet50)	ROIs	H&E		X	369	n.s.	n.s.	balanced ACC	0.873

Studies are organized into specific subtasks and sorted by year of publication. Columns "TCGA-GBM", "TCGA-LGG" and "Other" indicate whether TCGA-GBM, TCGA-LGG and/or other public or proprietary datasets were employed. "#Patients" and "#WSIs" state the reported size of the dataset including subsets for training and validation. "Performance" states a subset of the reported performance metrics. CNN convolutional neural network, VIT Vision Transformer, ROI region of interest, SL supervised learning, WSL weakly-supervised learning, SSL self-supervised learning, MIL multiple-instance learning, IHC immunohistochemistry, IFS intraoperative frozen sections, ACC accuracy, AUC area under the operator receiver curve, n.s. not specified.

*for astrocytoma.
 *for oligodendroglioma.
 *for glioblastoma.
 *for model based on intraoperative frozen sections.

Some studies aimed to improve the accuracy of subtyping by integration of MRI. A particularly large number of such studies emerged from the CPM RadPath Challenges 2018, 2019 and 2020, which provided paired magnetic resonance and whole-slide image datasets for the differentiation of astrocytomas, oligodendrogliomas and glioblastomas^{13,51,52}. The best performing approaches used weakly-supervised CNNs (or an ensemble thereof⁵²) for processing WSIs and fused WSI- and MRI-derived predictions by an average pooling⁵², a max pooling¹³ or by favoring WSI-derived predictions⁵¹. The improvement in balanced accuracy compared to using WSI alone was up to 7.8%⁵². Related to the integration of MRI, Mallya et al. utilized the same CPM-RadPath dataset and proposed a knowledge distillation framework to transfer the knowledge from WSIs to a model that differentiates gliomas based solely on MRI⁵³.

In addition to the aforementioned end-to-end approaches, some studies first predicted established histological or molecular markers, and then derived glioma types based on these predictions^{41,50,54–56}. In a series of studies, Ma et al., for instance, proposed a classification model for the differentiation of astrocytoma, oligodendroglioma, ependymoma, lymphoma, metastasis, and healthy tissue, based on the presence of respective cell types, and glioma grades based on the presence of microvascular proliferation and necrosis^{48,55}. They later assessed their model's performance using a multi-center dataset comprising 127 WSIs from four institutions⁵⁶. Interestingly, Hewitt et al. compared approaches for predicting the 2021 WHO glioma subtypes, and reported that first predicting IDH, 1p/19q, and ATRX status, and then inferring subtypes from these predictions performed overall better than an end-to-end approach⁴¹. In addition, they conducted further interesting experiments, such as comparing the predictability of 2016 versus 2021 WHO glioma subtypes, and validated all results on independent patient cohorts.

Grading. Complementing tumor types, tumor grades differentiate gliomas according to predicted clinical behavior and range from CNS WHO grade 1 to 4, with CNS WHO grade 1 indicating the least malignant behavior^{2,6}. Similar to subtyping, grading is currently based on both histological and molecular markers (Fig. 1), whereas in the past only histological markers, including cellular pleomorphism, proliferative activity, necrosis, and microvascular proliferation were considered⁷. Table 5 provides an overview of all corresponding studies grouped according to the investigated grading tasks.

All studies predicted glioma grades from WSIs of H&E-stained tissue sections in an end-to-end manner. Exceptions from this include one earlier study that inferred glioma grade from detected necrosis and microvascular proliferation⁵⁷, and a recent study that analyzed human leukocyte antigen (HLA)-stained tissue microarrays and inferred glioma grade by assessing the infiltration of myeloid cells in the tumor microenvironment³⁶.

Earlier studies mostly employed handcrafted features and traditional ML methods^{15,57–62}. In a typical study, Barker et al. proposed a coarse-to-fine analysis based on handcrafted features extracted from segmented nuclei. In this manner, they first selected discriminative patches and subsequently predicted grades using a linear regression model⁶¹.

All other studies employed CNNs for predicting grades from image patches. Three such approaches decomposed II vs. III vs. IV grading as stepwise binary II and III vs. IV followed by II vs. III classifications and agreed on greater difficulty in differentiating grades II vs. III^{63–65}. Su et al. focused on this challenge and reported considerable performance improvement by employing an ensemble of 14 weakly-supervised CNN classifiers whose predictions were aggregated by a logistic regression model⁶⁶. Whereas this approach seems computationally intensive, Momeni et al. proposed a deep recurrent attention model, for II and III vs. IV grading, which achieved state-of-the-art results while only analyzing 144 image patches per WSI⁶⁷.

Two recent studies examined the newly adopted approach of grading gliomas within types rather than across types, as endorsed by the 2021 WHO classification system^{43,56}. Most notably, Wang et al. proposed a clustering-based CNN model for this purpose (similar to that of Zhao

et al.⁶⁸), which they validated on two external cohorts of 305 and 328 patients, respectively⁴³.

Several studies integrated additional modalities such as genomics^{63,69,70}, age and sex⁶⁰ or proliferation index manually assessed from Ki-67 staining⁵⁸. Most notably, Qiu et al. adopted a self-training strategy to address the effects of label noise and proposed an attention-based feature guidance aiming to capture bidirectional interactions between WSIs and genomic features⁶⁹. They demonstrated the superiority of multi-modal over uni-modal predictions (AUC 0.807 (WSI), 0.804 (genomics), 0.872 (WSI+genomics)) as well as the effectiveness of modeling bidirectional interactions between modalities for glioma grading and lung cancer subtyping.

Addressing the problem of possibly limited availability of genomics in clinical practice, Xing et al. proposed a knowledge distillation framework to transfer the privileged knowledge of a teacher model, trained on WSIs and genomics, to a student model which subsequently predicts grade solely from WSIs⁷¹. Remarkably, they reported superior performance of their uni-modal student model over a state-of-the-art multi-modal model⁷⁰ on the same validation cohort.

Molecular marker prediction. Molecular alterations in gliomas have diagnostic, predictive, and prognostic value critical for effective treatment selection, e.g., the presence of MGMT promoter methylation was associated with a beneficial sensitivity to alkylating agent chemotherapy, resulting in prolonged survival in glioblastoma patients^{2,5}. Standards of practice for accurate, reliable detection of molecular markers include immunohistochemistry, sequencing, or fluorescent in situ hybridization (FISH)². Table 6 provides an overview of all corresponding studies grouped according to the investigated molecular markers.

Although there are no established criteria for determining molecular markers from histology yet, a growing body of research has suggested their predictability from WSIs of H&E-stained tissue sections with considerable accuracy. Among others, pan-cancer studies predicted various molecular markers across multiple cancer types^{72,73} and with regard to glioblastoma reported an externally validated accuracy of over 0.7 for 4 out of 9 investigated markers⁷².

Most studies predicted IDH mutation status. Similar approaches based on weakly-supervised CNNs were augmented by either the incorporation of additional synthetically generated image patches (AUC 0.927 vs. 0.920)⁷⁴ or by the integration of MRI (ACC from WSI: 0.860; from MRI: 0.780; from both: 0.900)⁷⁵. As patient age and IDH mutation status are highly correlated (e.g., the median age at diagnosis was reported to be 37 years for IDH-mutant astrocytomas and 65 years for IDH-wildtype astrocytomas¹), integration of patient age into multi-modal predictions further improved performance^{74–76}.

Comparing their performance in predicting IDH status to pathologists, Liechty et al. proposed a multi-magnification ensemble that averaged predictions from multiple weakly-supervised CNN classifiers, each processing individual magnification levels⁷⁷. On an external validation cohort of 174 WSIs, their model did not surpass human performance (AUC 0.881 vs. 0.901), but averaged predictions from pathologists and their model performed on par with the consensus of two pathologists (AUC 0.921 vs. 0.920).

Recently, Zhao et al. proposed a clustering-based hybrid CNN-ViT model to predict IDH mutation in diffuse glioma⁶⁸. They trained and assessed their model based on two large and independent patient cohorts from two hospitals, and, motivated by the changes in the 2021 WHO classification system, investigated subgroups with particularly challenging morphologies. Of note, their study provided a clear explanation of how the ground truth IDH status was determined through sequencing for all patients in both cohorts.

Five studies were dedicated to the prediction of the codeletion of chromosomes 1p and 19q. In particular, Kim et al. predicted 1p/19q fold change values from H&E-stained WSIs and compared their CNN-based method to the traditional method of detecting 1p/19q status via FISH⁵⁴. Based on a cohort of 288 patients, whose 1p/19q status were confirmed

Table 5 | Overview of all studies related to grading

Study	Method		Multi-modal fusion		Dataset					Performance	
	Features	Learning paradigm	Modalities	Fusion Strategy	TCGA-GBM	TCGA-LGG	Other	#Patients	#WSIs	Metric	Result
WHO grade II vs. WHO grade III											
Su et al. ⁶⁶	CNN (ResNet18)	WSL	H&E			X		507	n.s.	ACC	0.801
CNS WHO grade 2 vs. CNS WHO grade 3 vs. CNS WHO grade 4											
Jin et al. ⁵⁶	CNN (DenseNet)	SL	H&E				X	n.s.	733	ACC	0.730
Wang et al. ⁴³	CNN (ResNet50)	WSL	H&E				X	2624	2624	AUC	0.939 ^a , 0.930 ^b , 0.990 ^c , 0.948 ^d , 0.967 ^e
WHO grade II vs. WHO grade III vs. WHO grade IV											
Qiu et al. ⁶⁹	CNN (ResNet)	ROIs	H&E, omics	Intermediate	X	X		683	683	AUC	0.872
Zhang et al. ¹¹⁰	CNN (EfficientNet)	SSL, MIL	H&E		X	X		499	n.s.	ACC	0.790 ^f , 0.750 ^g
Xing et al. ⁷¹	CNN (ResNet18)	ROIs	H&E, omics	KD	X	X		736	1325*	AUC	0.924
Pei et al. ⁶³	CNN (ResNet)	WSL	H&E, omics	Early	X	X		n.s.	549	ACC	0.938 ^h , 0.740 ⁱ
Chen et al. ⁷⁰	CNN (VGG19)	ROIs	H&E, omics	Intermediate	X	X		769	1505*	AUC	0.908
Truong et al. ⁶⁴	CNN (ResNet18)	WSL	H&E		X	X		1120	3611	ACC	0.730 ^h , 0.530 ⁱ
Wang et al. ⁵⁸	Handcrafted		H&E, PI	Early			X	146	n.s.	ACC	0.900
Ertosun et al. ⁶⁵	CNN	WSL	H&E		X	X		n.s.	37	ACC	0.960 ^h , 0.710 ⁱ
WHO grade II and III vs. WHO grade IV											
Jiang et al. ¹¹²	CNN (ResNet18)	SSL, MIL	H&E		X	X		n.s.	n.s.	AUC	0.964
Brindha et al. ¹²⁰	Handcrafted		H&E		X	X		n.s.	1114	ACC	0.972
Mohan et al. ⁵⁹	Handcrafted		H&E		X	X		n.s.	310	AUC	0.974
Im et al. ⁴⁴	CNN (MnasNet)	ROIs	H&E				X	468	n.s.	balanced ACC	0.580
Rathore et al. ⁶⁰	Handcrafted	ROIs	H&E, clinical	Early	X	X		735	n.s.	AUC	0.927
Momeni et al. ⁶⁷	CNN	WSL	H&E		X	X		710	n.s.	AUC	0.930
Yonekura et al. ¹²¹	CNN	WSL	H&E		X	X		n.s.	200	ACC	0.965
Xu et al. ¹⁴	CNN (AlexNet)	WSL	H&E				X	n.s.	85	ACC	0.975
Barker et al. ⁶¹	Handcrafted		H&E		X	X	X	n.s.	649	AUC	0.960
Hou et al. ⁴⁷	CNN	WSL	H&E		X	X		539	1064	ACC	0.970
Reza et al. ⁶²	Handcrafted		H&E		X	X		n.s.	66	AUC	0.955
Fukuma et al. ¹⁵	Handcrafted		H&E		X	X		n.s.	300	ACC	0.996
WHO grade I vs. WHO grade II vs. WHO grade III vs. WHO grade IV											
Elazab et al. ¹⁴⁴	CNN (ResNet50)	WSL	H&E		X	X		445	654	ACC	0.972
WHO grade I and II vs. WHO grade III and IV											
Mousavi et al. ⁵⁷	Handcrafted		H&E		X	X		138	n.s.	ACC	0.847
Non-tumor vs. WHO grade I vs. WHO grade II vs. WHO grade III vs. WHO grade IV											
Pytlarz et al. ³⁶	CNN (DenseNet)	SSL, WSL	HLA				X	n.s.	206***	ACC	0.690
Non-tumor vs. WHO grade II and III vs. WHO grade IV											
Ker et al. ¹³³	CNN (InceptionV3)	WSL	H&E				X	n.s.	154**	F1-score	0.991

Studies are organized into specific subtasks and sorted by year of publication. Columns “TCGA-GBM”, “TCGA-LGG” and “Other” indicate whether TCGA-GBM, TCGA-LGG and/or other public or proprietary datasets were employed. “#Patients” and “#WSIs” state the reported size of the dataset including subsets for training and validation. “Performance” states a subset of the reported performance metrics.

*not WSIs but ROIs, **probably not gigapixel WSIs but smaller histopathology images, ***tissue microarrays.

CNN convolutional neural network, ROI region of interest, SL supervised learning, WSL weakly-supervised learning, SSL self-supervised learning, MIL multiple-instance learning, KD knowledge distillation, H&E hematoxylin and eosin staining, HLA human leukocyte antigen staining, PI proliferation index, ACC accuracy, AUC area under the operator receiver curve, n.s. not specified.

^{a/b/c}for astrocytoma CNS WHO grade 2/3/4.

^{d/e}for oligodendroglioma CNS WHO grade 2/3.

^ffor frozen sections.

^gfor FFPE sections.

^hfor II + III vs. IV.

ⁱfor II vs. III.

Table 6 | Overview of all studies related to molecular marker prediction

Study	Method		Multi-modal fusion		Dataset			Performance			
	Features	Learning paradigm	Modalities	Fusion Strategy	TCGA-GBM	TCGA-LGG	Other	#Patients	#WSIs	Metric	Result
IDH mutation											
Zhao et al. ⁶⁸	CNN (ResNet50), ViT	WSL	H&E				X	2275	n.s.	AUC	0.953
Liechty et al. ⁷⁷	CNN (DenseNet)	WSL	H&E		X	X	X	513	975	AUC	0.881
Jiang et al. ⁷⁶	CNN (ResNet18)	WSL	H&E, clinical	Early		X		490	843	AUC	0.814
Wang et al. ⁷⁵	Handcrafted		H&E, MRI	Late	X	X	X	217	n.s.	ACC	0.900
Liu et al. ⁷⁴	CNN (ResNet50)	WSL	H&E, clinical	Early	X	X	X	266	n.s.	AUC	0.931
MGMT promoter methylation											
He et al. ¹¹⁵	ViT (Swin Transformer V2)	ROIs, WSL	H&E		X		X	224	657	AUC	0.860
Yin et al. ¹⁴⁵	Spiking neural P system		H&E		X			110	110	ACC	0.700
Dai et al. ¹⁴⁶	Spiking neural P system		H&E		X			110	110	ACC	0.681
IDH mutation, MGMT promoter methylation											
Krebs et al. ¹⁰¹	CNN (ResNet18)	SSL, MIL	H&E		X			325	196 ^a , 216 ^b	ACC	0.912 ^a , 0.861 ^b
IDH mutation, MGMT promoter methylation, TP53 mutation											
Li et al. ⁴⁵	ViT (ViT-L-16)	MIL	H&E		X	X	X	1005 ^a , 257 ^b , 877 ^c	n.s.	AUC	0.960 ^a , 0.845 ^b , 0.874 ^c
IDH mutation, MGMT promoter methylation, 1p/19q codeletion											
Momeni et al. ⁶⁷	CNN	WSL	H&E		X	X		710	n.s.	AUC	0.860 ^a , 0.750 ^b , 0.760 ^d
1p/19q fold change values											
Kim et al. ⁵⁴	CNN (RetCCL)	WSL	H&E		X	X	X	673	673	AUC	0.833 ^a , 0.837 ^f
IDH mutation, 1p/19q codeletion											
Rathore et al. ¹⁴⁷	CNN (ResNet)	ROIs	H&E		X	X		663	n.s.	AUC	0.860 ^a , 0.860 ^d
IDH mutation, 1p/19q codeletion, homozygous deletion of CDKN2A/B											
Wang et al. ⁵⁰	ViT	MIL	H&E		X	X		940	2633	AUC	0.920 ^a , 0.881 ^d , 0.772 ^g
IDH mutation, 1p/19q codeletion, homozygous deletion of CDKN2A/B, ATRX mutation, EGFR amplification, TERT promoter mutation											
Hewitt et al. ⁴¹	CNN, ViT (CTransPath)	MIL	H&E		X	X	X	2840	n.s.	AUC	0.900 ^a , 0.870 ^c , 0.730 ^g , 0.790 ^h , 0.850 ⁱ , 0.600 ^j
IDH mutation, 1p/19q codeletion, homozygous deletion of CDKN2A/B, ATRX mutation, EGFR amplification, TP53 mutation, CIC mutation											
Nasrallah et al. ⁴²	ViT	WSL	IFS		X	X	X	1524	2334	AUC	0.820 ^a , 0.820 ^d , 0.800 ^g , 0.790 ^h , 0.710 ⁱ , 0.870 ^c , 0.790 ^k
DNA methylation											
Zheng et al. ¹⁴⁸	Handcrafted		H&E, clinical	Early	X	X		327	n.s.	AUC	0.740
Pan-cancer studies											
Saldanha et al. ⁷²	CNN (RetCCL)	MIL	H&E		X		X	493	n.s.	AUC	0.840 ^a , 0.700 ^c , 0.700 ^h
Arslan et al. ⁸⁰	CNN (ResNet34)	WSL	H&E		X	X	X	894	1999	AUC	0.669 ⁱ
Loeffler et al. ⁷³	CNN (ShuffleNet)	ROIs	H&E		X	X		680	n.s.	AUC	0.764 ^a , 0.787 ^c , 0.726 ^h

Studies are organized into specific subtasks and sorted by year of publication. Columns “TCGA-GBM”, “TCGA-LGG” and “Other” indicate whether TCGA-GBM, TCGA-LGG and/or other public or proprietary datasets were employed. “#Patients” and “#WSIs” state the reported size of the dataset including subsets for training and validation. “Performance” states a subset of the reported performance metrics.

CNN convolutional neural network, ViT Vision Transformer, SL supervised learning, ROI region of interest, WSL weakly-supervised learning, SSL self-supervised learning, MIL multiple-instance learning, H&E hematoxylin and eosin staining, IFS intraoperative frozen sections, MRI magnetic resonance imaging, ACC accuracy, AUC area under the operator receiver curve, n.s. not specified.

^afor IDH mutation.

^bfor MGMT promoter methylation.

^cfor TP53 mutation.

^dfor 1p/19q codeletion.

^efor 1p.

^ffor 19q.

^gfor CDKN2A/B.

^hfor ATRX.

ⁱfor EGFR.

^jfor TERT.

^kfor CIC.

^laverage performance.

through next generation sequencing, they ascribed a superior predictive power to their proposed method. They further used their model to classify IDH-mutant gliomas into astrocytomas and oligodendrogliomas and validated all results on an external cohort of 385 patients from TCGA.

As various molecular markers show complex interactions and should not be considered as independent variables (e.g., “essentially all 1p/19q co-deleted tumors are also IDH-mutant, although the converse is not true”⁹²), Wang et al. proposed a multiple-instance learning (MIL)- and ViT-based model for predicting IDH mutation, 1p/19q codeletion, and homozygous deletion of CDKN2A/B as well as the presence of necrosis and microvascular proliferation, while simultaneously recognizing interactions between predicted markers⁵⁰. They validated their system in ablation studies and outperformed popular state-of-the-art MIL frameworks and CNN architectures in all tasks, with improvements in accuracy of up to 6.3%.

Survival prediction. Accurate prediction of survival and a comprehensive understanding of prognostic factors in gliomas are crucial for appropriate disease management⁷⁸. Median survival in adult-type diffuse gliomas ranges from 17 years in oligodendrogliomas to only 8 months in glioblastomas and generally decreases with older age¹. Besides type, grade, and age, other prognostic factors include tumor size and location, extent of tumor resection, performance status, cognitive impairment, status of molecular markers such as IDH mutation, 1p/19q codeletion, MGMT promoter methylation and treatment protocols, including temozolomide and radiotherapy^{2,78}. Table 7 provides an overview of all corresponding studies grouped according to the investigated aspects of prognostic inference.

Almost all reviewed studies investigated the prediction of overall survival from WSIs of H&E-stained tissue sections in an end-to-end manner. One exception inferred predictions by first classifying vascular endothelial cells and quantifying microvascular hypertrophy and hyperplasia⁷⁹ and two studies additionally investigated disease-free survival^{80,81}. Most studies predicted risk scores or survival times, and only four studies conducted survival prediction by classifying patients into either low- or high-risk groups^{82–84} or four predefined risk groups⁸⁵.

Given the high variability of disease manifestation and co-morbidities, as well as the stochastic nature of the immediate cause of death, prognosis can be regarded as the most challenging task. In particular, prediction of overall survival from histology has been considered inherently more challenging than tasks such as subtyping and grading^{70,86–89}, as complex interactions of heterogeneous visual concepts, such as immune cells in general⁹⁰, or more specifically lymphocyte infiltrates⁹¹ in the tumor microenvironment, have been recognized as potentially relevant factors.

Zhu et al. proposed WSISA, the earliest end-to-end system for survival prediction from WSIs, which leveraged a priorly proposed CNN-based survival model⁹² to first compute clusters of survival-discriminate image patches and second aggregate cluster-based features for subsequent risk score regression⁸⁹.

Potentially due to the aforementioned complexity, only a minority of the reviewed studies proposed uni-modal approaches based on handcrafted features or weakly-supervised CNNs. The majority of the studies utilized either more intricate architectures, or the integration of additional modalities of information.

To this end, Li et al. and Chen et al. set out to capture context-aware representations of histological features by employing graph convolutional neural networks, in which vertices corresponded to patch-level feature vectors and edges were defined by either Euclidean distance between feature vectors or adjacency of patches^{93,94}. They both compared their models, DeepGraphSurv and Patch-GCN, respectively, on multiple cancer types, where DeepGraphSurv improved on WSISA by 6.5% and Patch-GCN on DeepGraphSurv by 2.6% in terms of C-index.

Three recent studies leveraged the Transformer architecture for modeling histological patterns across patches, either by employing multiple feature detectors to aggregate distinct morphological patterns^{87,95} or by

fusing information from multiple magnification levels⁸⁶. All studies reported state-of-the-art performance on multiple cancer types, including gliomas, from the TCGA-GBM and TCGA-LGG datasets. Most notably, Liu et al. reported a 13.2% and a 4.3% C-index performance improvement compared to DeepGraphSurv and Patch-GCN on TCGA-LGG, respectively, and Wang et al. reported a 5.5%, a 4.5%, and a 3.5% C-index performance improvement compared to DeepGraphSurv, Patch-GCN, and an extension of Patch-GCN via a variance pooling⁹⁶ on TCGA-GBM and TCGA-LGG^{86,87}.

Contrasting the above mentioned uni-modal approaches, half of all studies stated prognosis as a multi-modal problem by integrating clinical or omics data. Simple approaches to such integration included fusing WSI-derived risk scores and variables from other modalities in Cox proportional hazards models^{76,81,97} or by concatenating image features and variables prior to subsequent risk score prediction, as proposed by Mobadersany et al.⁴⁰. In particular, they proposed two CNN-based models, SCNN, which predicted survival from priorly delineated ROIs, and GSCNN, which improved SCNN by concatenating IDH mutation and 1p/19q codeletion status as additional prognostic variables (C-index 0.801 vs. 0.754). They further reported the superiority of concatenation vs. Cox models as strategies for multi-modal fusion (C-index 0.801 vs. 0.757).

More advanced approaches fused distinct uni-modal feature representations into an intermediate feature tensor, using the Kronecker product, while controlling the expressiveness of each modality via a gating-based attention mechanism^{70,98,99}. Most notably, by integrating 320 genomics and transcriptomics features in this manner, Chen et al. surpassed GSCNN on the same ROIs with a C-index performance improvement of 5.8%⁷⁰. The same authors later expanded on this approach by adopting MIL and applying it to 14 cancer types⁹⁸. They conducted extensive analysis regarding the interpretability of their model and reported that the presence of necrosis and IDH1 mutation status were the most attributed features in gliomas from the TCGA-LGG dataset, which is in line with the current WHO classification. They further developed an interactive research tool, to drive the discovery of new prognostic biomarkers.

Methodological aspects of WSI processing

Across all diagnostic tasks, most of the considered studies processed WSIs in a patch-based manner. The predictive performance of such approaches depends on choices for patch size and magnification, methods for encoding image patches, and approaches for learning relations among patches. In the following, the studies are examined in more detail with regard to these aspects.

Patch sizes and magnifications. The size of image patches and the magnification at which they are extracted not only determine the number of patches stemming from given WSIs, but also influence the degree to which either cellular details or more distributed characteristics in tissue architecture are captured. However, patch size and magnification were not consistently reported (63/83 and 50/83, respectively) and only three studies compared performance across multiple sizes and magnifications, e.g., patches of size 256 × 256 pixels, and 2.5x, 5x, 10x, and 20x magnification for IDH mutation prediction (AUC 0.80, 0.85, 0.88, and 0.84, respectively)⁷⁷; highest performance in grading by employing 672 × 672 pixels patches at 40x magnification¹⁴; and highest performance for survival risk group classification consistent across multiple CNN architectures by utilizing 256 × 256 pixels patches at 20x magnification⁸⁵. Apart from these comparisons, smaller patches, i.e., 224 × 224 and 256 × 256 pixels, and 20x magnification were most frequently employed (Fig. 4).

In addition to single-magnification approaches, few studies processed multiple magnifications^{47,75,77,86,100,101}, e.g., by concatenating single-magnification features at the onset¹⁰¹, averaging of single-magnification predictions⁷⁷ or modeling magnifications using graphs¹⁰⁰ or cross-attention⁸⁶.

Table 7 | Overview of all studies related to survival prediction

Study	Method		Multi-modal fusion		Dataset			Performance			
	Features	Learning paradigm	Modalities	Fusion Strategy	TCGA-GBM	TCGA-LGG	Other	#Patients	#WSIs	Metric	Result
Risk score/survival time regression											
Yin et al. ¹⁴⁵	Spiking neural P system		H&E		X			110	110	PCC	0.541
Dai et al. ¹⁴⁶	Spiking neural P system		H&E		X			110	110	PCC	0.515
Jiang et al. ⁹⁵	CNN (ResNet18)	MIL	H&E			X		490	843	C-index	0.714
Jiang et al. ¹¹²	CNN (ResNet18)	SSL, MIL	H&E			X		n.s.	n.s.	C-index	0.685
Liu et al. ⁸⁶	CNN (ResNet50)	MIL	H&E			X		486	836	C-index	0.702
Luo et al. ⁸¹	CNN (ResNet50)	SL	H&E, clinical	Early			X	162	n.s.	C-index	0.928**
Wang et al. ⁸⁷	CNN (ResNet50)	MIL	H&E		X	X		1,041	n.s.	C-index	0.861
Carmichael et al. ⁹⁶	CNN (ResNet18)	MIL	H&E		X	X		872	n.s.	C-index	0.738
Chen et al. ⁷⁰	CNN (VGG19)	ROIs	H&E, omics	Intermediate	X	X		769	1,505*	C-index	0.826
Chunduru et al. ⁹⁷	CNN (ResNet50)	ROIs	H&E, omics, clinical	Late	X	X		766	1,061	C-index	0.840
Braman et al. ⁹⁹	CNN (VGG19)	ROIs	H&E, omics, clinical, MRI	Intermediate	X	X		176	372*	C-index	0.788
Chen et al. ⁸⁸	CNN (ResNet50)	MIL	H&E, omics		X	X		1,011	n.s.	C-index	0.817
Chen et al. ⁹⁴	CNN (ResNet50)	MIL	H&E		X	X		1,011	n.s.	C-index	0.824
Jiang et al. ⁷⁶	CNN (ResNet18)	WSL	H&E, omics, clinical	Early		X		490	843	C-index	0.792
Hao et al. ¹³⁰	CNN	WSL	H&E, omics, clinical	Late	X			447	n.s.	C-index	0.702
Rathore et al. ¹⁴⁷	CNN (ResNet)	ROIs	H&E		X	X		663	n.s.	C-index	0.820
Tang et al. ¹⁴⁹	Capsule network	WSL	H&E		X			209	424	C-index	0.670
Li et al. ⁹³	CNN (VGG16)	MIL	H&E		X			365	491	C-index	0.622
Mobadersany et al. ⁴⁰	CNN (VGG19)	ROIs	H&E, omics	Late	X	X		769	1,061*	C-index	0.801
Nalisnik et al. ⁷⁹	Handcrafted		H&E, clinical	Late		X		n.s.	781	C-index	0.780
Zhu et al. ⁸⁹	CNN	WSL	H&E		X			126	255	C-index	0.645
Survival time distribution modeling											
Liu et al. ¹⁵⁰	CNN (ResNet50)	MIL	H&E			X		486	836	C-index	0.642
Risk group classification											
Baheti et al. ⁸²	CNN (ResNet50)	MIL	H&E, omics, clinical	Late	X	X		188	188	AUC	0.746
Shirazi et al. ⁸⁵	CNN (Inception)	ROIs	H&E		X	X	X	454	858*	AUC	1.0
Zhang et al. ⁸³	Handcrafted		H&E, omics, clinical	Late	X			606	1,194	AUC	0.932
Powell et al. ⁸⁴	Handcrafted		H&E, omics, clinical	Early		X		53	n.s.	AUC	0.890
Pan-cancer studies											
Arslan et al. ⁸⁰	CNN (ResNet34)	WSL	H&E		X	X		705	1,489	n.s.	n.s.
Chen et al. ⁹⁸	CNN (ResNet50)	MIL	H&E, omics	Intermediate		X		479	n.s.	C-index	0.808
Cheerla et al. ¹¹¹	CNN (SqueezeNet)	SSL	H&E, omics, clinical	Intermediate		X		n.s.	n.s.	C-index	0.850

Studies are organized into specific subtasks and sorted by year of publication. Columns “TCGA-GBM”, “TCGA-LGG” and “Other” indicate whether TCGA-GBM, TCGA-LGG and/or other public or proprietary datasets were employed. “#Patients” and “#WSIs” state the reported size of the dataset including subsets for training and validation. “Performance” states a subset of the reported performance metrics.

CNN convolutional neural network, ROI region of interest, SL supervised learning, WSL weakly-supervised learning, SSL self-supervised learning, MIL multiple-instance learning, H&E hematoxylin and eosin staining, MRI magnetic resonance imaging, PCC Pearson correlation coefficient, C-index concordance index, AUC area under the operator receiver curve, n.s. not specified.

*not WSIs but ROIs, **probably not gigapixel WSIs but smaller histopathology images.

Encoding of image patches. The reviewed studies employed handcrafted features (13/83), CNNs (63/83) (including ResNet¹⁰², VGG¹⁰³, DenseNet¹⁰⁴, EfficientNet¹⁰⁵, Inception¹⁰⁶, and AlexNet¹⁰⁷), capsule networks (1/83), and ViTs (6/83) to encode image patches. While handcrafted features were mostly employed by earlier approaches, ViT

emerged as an alternative to CNNs in recent years. Few studies compared different CNN architectures^{36,44,49,74,85} and mostly agreed on the superiority of ResNet over other popular architectures^{44,49,74}. Apart from these comparisons, ResNet architectures pre-trained on ImageNet were most commonly employed (33/63) (Fig. 5). Self-supervised learning

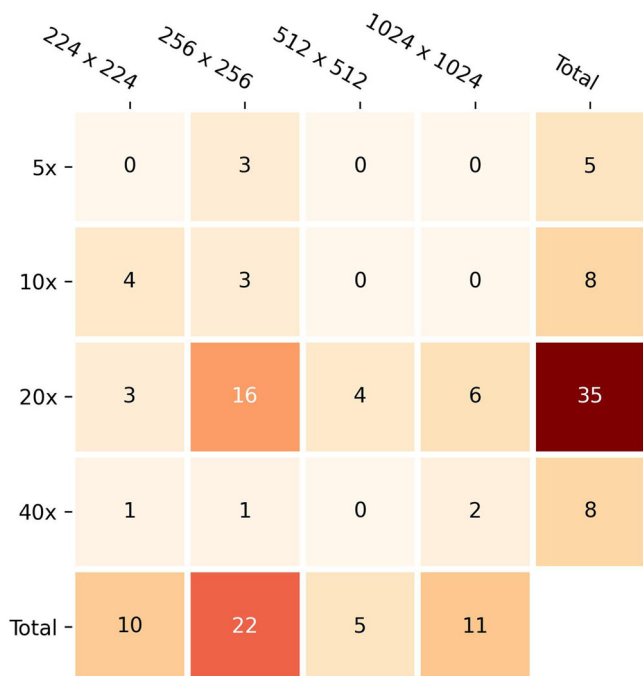


Fig. 4 | Patch sizes and magnifications employed by studies. Studies were taken into account if at least one of the two pieces of information was specified. Other patch sizes and magnifications employed by single studies (e.g., 150 × 150 pixels, 4x magnification) are not shown.

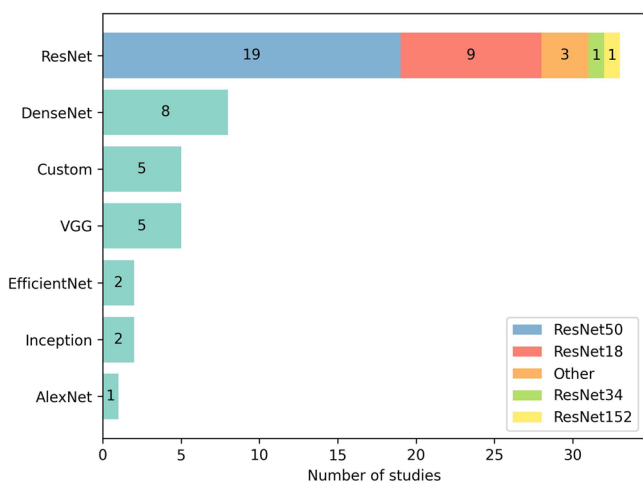


Fig. 5 | Convolutional neural network architectures employed by studies. With only a few exceptions all convolutional neural networks were pre-trained using the ImageNet dataset. Except for ResNet architectures exact variants of stated architectures are not shown. “Custom” refers to custom (i.e., self-configured) architectures.

approaches for pre-training CNNs or ViTs using histopathological images leveraged pre-text tasks, including contrastive learning^{36,101,108–111}, masked pre-training¹¹² or cross-stain prediction¹⁰⁸.

Learning paradigms. The majority of all 70 DL-based studies employed either WSL (29/83), MIL (21/83), or previously delineated ROIs (11/83) to process WSIs (the methodological differences of the three approaches are explained in Table 2). In recent years, MIL was most commonly used (Fig. 6) and approaches mostly utilized ResNets for encoding image patches (14/21) as well as intricate architectures based on graphs^{93,94}, Transformers^{45,50,86–88,95,98,112,113} or other attention

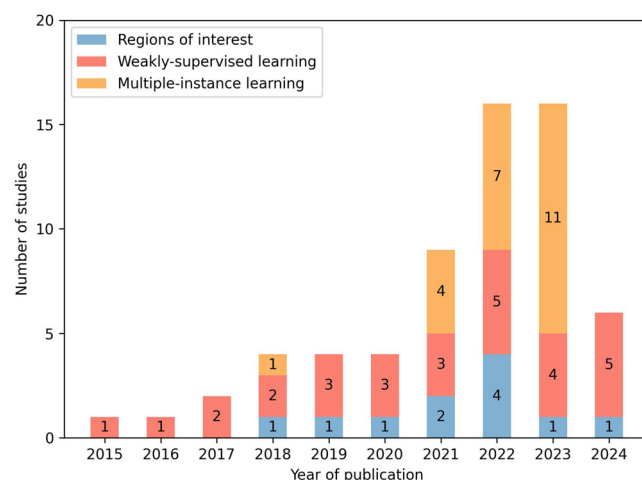


Fig. 6 | Learning paradigms employed by studies. Usage of regions of interest, weakly-supervised learning and multiple-instance learning by all 70 deep learning-based studies included in this review distributed by year of publication. The methodological differences of the three approaches are explained in Table 2.

mechanisms^{72,82,100,108–110,114} (often combinations thereof) to model histological patterns among patches.

Few studies compared approaches^{43,45,50}. Li et al. conducted insightful comparisons across three subtyping tasks and reported an overall superiority of ViT-L-16 over ResNet50 (both pre-trained on ImageNet) and a worse performance of WSL in all three tasks, with an especially large performance gap in a more fine-grained subtyping task including 11 brain tumor types⁴⁵. Wang et al., however, reported that their clustering-based CNN model outperformed four state-of-the-art MIL frameworks in several subtyping and grading tasks⁴³.

Discussion

As described in the previous sections, a large body of research on the application of AI to various aspects of glioma tissue assessment has been published in recent years. Considering the difficulty of the diagnostic tasks, each study provides important insights for the future application of AI-based methods in real clinical settings. However, there remain serious limitations.

Limitations of current research

Only 13 of the included studies evaluated the performance of their proposed method on independent test datasets from external pathology departments that were not used during the development of the method^{41–43,45,46,54,61,68,72,77,80,85,115}; 10 of these studies were published in 2023 or 2024. External validation, however, is crucial for meaningful performance estimates¹¹⁶. Moreover, the performance metrics reported in the solely TCGA-based studies might be inflated, as none of these studies ensured that the data used for model development and performance estimation originated from distinct tissue source-sites^{117,118}. As the TCGA datasets were originally compiled for different purposes^{37–39}, it must be assumed that they are not representative of the intended use of the proposed analysis methods and contain biases. Given these limitations, the performance metrics reported by the reviewed studies should be interpreted with caution and should not be generalized to practical applications¹¹⁹.

Relatedly, all TCGA-based studies utilized different subpopulations of the two TCGA cohorts without specifying which selection criteria were employed. When studies used other datasets, these were generally not made publicly available. This makes it very hard to reproduce the results and compare results between studies.

Another important aspect is the repeated revision of the WHO classification system, which makes it necessary to revise the now outdated ground truth labels in the TCGA datasets with the current subtypes and grades.

Only three studies utilized the current version from 2021^{41–43}; two other studies made an effort to reclassify the TCGA cohorts according to the current 2021 criteria based on available molecular information^{49,50}, and another three studies considered the previous version from 2016^{44–46,100}. All other studies were either based on versions from 2007 or earlier, or did not report the version used, which seriously limits their comparability with the current diagnostic standards, and largely disregards important advances in the understanding of glioma tumor biology. Notably, this concerns important advances regarding the molecular pathogenesis of gliomas with an impact on classification and prognostication, as reflected by the revisions from 2016 and 2021⁶⁷.

A commonality of many studies in the field is their emphasis on technical aspects of AI approaches, with less focus on the clinical applicability of their proposed methods. The design of several studies suggests that the investigations were primarily based on the availability of data, while a clearly defined clinically relevant research question was not immediately apparent. Examples include studies that mixed distinct glioma types and grades without properly substantiating the motivation for their respective approaches, for instance, when differentiating patients in the TCGA-GBM cohort from those in the TCGA-LGG cohort^{15,47,59–62,67,112,120,121}. Moreover, not all studies accurately specified if FFPE or frozen sections were considered for histological assessment. In many studies, it remained unclear for which specific practical diagnostic tasks the proposed methods were envisioned to be relevant.

Future directions

Considering that the majority of currently available research results are based largely on data from the public TCGA-GBM and TCGA-LGG projects, the acquisition of new datasets with high-quality and up-to-date clinical and molecular pathology annotations is of utmost importance to progress toward clinical applicability. Such new datasets should address the aforementioned limitations, that is, they should consider the latest version of the WHO classification system, be representative of the intended use¹¹⁹, and include multiple tissue source sites to enable independent external validation¹¹⁶.

Future work should focus on applications of AI that are clinically relevant, but have received little attention to date. Analysis of intraoperative frozen sections may be valuable for clinicians to improve the robustness of preliminary diagnoses and guide the surgery^{42,46,122}. Similarly unexplored is the automated preselection of representative tissue areas for DNA-extraction to speed up the molecular pathology workflow. In general, future work should incorporate a wider variety of staining techniques, including IHC, and further explore their value for AI-based approaches to improve diagnostic tasks or to support predictive and/or prognostic biomarker analysis. In particular, the automated quantification of relevant IHC markers, such as Ki-67 and TP53, and the assessment of their spatial heterogeneity has not yet been investigated specifically for glioma tissue. Besides these applications, future investigations should also consider glioma types other than adult-type diffuse glioma⁶.

While the current research applies a wide range of AI methods, some promising approaches remain largely unexplored. The emergence of foundation models for computational pathology^{123,124}, developed using large amounts of histopathological data, may be an improvement over the current reliance on the ResNet CNN architecture and the ImageNet dataset for pre-training. With their multi-task capabilities, these models could unify many of the reviewed approaches which usually include a specific model for each particular diagnostic task, and potentially open avenues for new clinical use cases¹²⁵. Generative AI models allow for the synthesis of novel histopathological image data, providing new opportunities for AI-based WSI analysis. One of the reviewed studies predicting IDH mutation status has used generative models to synthesize additional training data⁷⁴. Their positive results encourage further investigation. Other promising applications of generative models such as the normalization of staining variations resulting from different image acquisition or tissue staining protocols^{126,127}, or the synthesis of virtual stainings¹²⁸, have not yet been explored in the context of glioma diagnosis.

Most current DL-based methods for WSI analysis predict endpoints directly from pixel data in an end-to-end manner. However, methods considering intermediate, human-understandable features, for example of segmented tissue structures and their spatial relations, can greatly facilitate the verification and communication of results and should remain an active area of work^{41,50,55}. Deep learning can also be used to discover such features and reliably segment relevant structures^{13,129}.

Regarding multi-modal fusion, all studies, across all diagnostic tasks, reported better predictive performance when clinical data, omics data, MRI, and WSIs were used in combination rather than individually. So far, the additional value of WSIs for multi-modal predictive performance has been consistently low^{70,81,82,98,130}. Moreover, good predictive performance can often be achieved with much simpler models. For instance, Cox proportional hazards or logistic regression models based on a few clinical variables, such as patient age and sex, can perform on par with WSI-based predictions for survival or IDH mutation status^{40,70,76,97,98,131}. Further research should therefore be conducted to better understand and improve the added value of AI-based analysis of WSIs for glioma.

Above all, the reviewed literature clearly indicates a need for future studies with a stronger focus on clinically relevant scientific questions. This requires a truly interdisciplinary approach from the outset, involving neuropathology, neuro-oncology, and the broad expertise of computational science and medical statistics.

Methods

The literature search for this review was last updated on March 18, 2024 and conducted as follows: The query “(glioma OR glioblastoma OR brain tumor) AND (computational pathology OR machine learning OR deep learning OR artificial intelligence) AND (whole slide image OR WSI)” was used to search PubMed, Web of Science, and Google Scholar. The results from all three searches were then screened for inclusion. Results met the inclusion criteria, if they were original research articles that proposed at least one AI-based method (including traditional ML and DL) for at least one of the diagnostic tasks of subtyping, grading, molecular marker prediction, or survival prediction from whole-slide histopathology images of human gliomas. Studies which incorporated other modalities, e.g., MRI, omics or clinical data, in addition to whole-slide histopathology images were also included. Results were excluded if they were not original research articles (e.g., reviews or book chapters), did not examine glioma or whole-slide histopathology images, primarily investigated image segmentation or retrieval methods, focused on tumor heterogeneity, were conference papers or preprints that were subsequently published as journal articles (in these cases, only journal articles were included), or were not written in English.

The PubMed search yielded exactly 100 results; the search on Web of Science yielded exactly 90 results; and the search on Google Scholar yielded approximately 23,300 results. After screening the first 191 results from Google Scholar (using the setting “Sort by relevance”), no further relevant results were identified. For this reason, it was decided to stop screening after the 300th result and concluded that all further results would likely be irrelevant. From these combined 490 results, 96 duplicates were removed and the remaining 394 results were screened for inclusion. Additional studies that were identified through references or recommendations were also screened for inclusion. As a result, 83 studies, comprising journal articles, conference papers, and preprints, were included in this review.

Received: 5 February 2024; Accepted: 27 May 2024;

Published online: 01 July 2024

References

- Ostrom, Q. T. et al. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2016–2020. *Neuro-Oncol* **25**, iv1–iv99 (2023).
- Weller, M. et al. Glioma. *Nat. Rev. Dis. Primer* **1**, 15017 (2015).
- Yang, K. et al. Glioma targeted therapy: insight into future of molecular approaches. *Mol. Cancer* **21**, 39 (2022).

4. Stupp, R. et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **352**, 987–996 (2005).
5. Hegi, M. E. et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N. Engl. J. Med.* **352**, 997–1003 (2005).
6. Louis, D. N. et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro-Oncol* **23**, 1231–1251 (2021).
7. Horbinski, C., Berger, T., Packer, R. J. & Wen, P. Y. Clinical implications of the 2021 edition of the WHO classification of central nervous system tumours. *Nat. Rev. Neurol.* **18**, 515–529 (2022).
8. Echle, A. et al. Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br. J. Cancer* **124**, 686–696 (2021).
9. Shmatko, A., Ghaffari Laleh, N., Gerstung, M. & Kather, J. N. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nat. Cancer* **3**, 1026–1038 (2022).
10. van der Laak, J., Litjens, G. & Ciompi, F. Deep learning in histopathology: the path to the clinic. *Nat. Med.* **27**, 775–784 (2021).
11. Zhang, P. et al. Effective nuclei segmentation with sparse shape prior and dynamic occlusion constraint for glioblastoma pathology images. *J. Med. Imaging* **6**, 017502 (2019).
12. Li, X., Wang, Y., Tang, Q., Fan, Z. & Yu, J. Dual U-Net for the segmentation of overlapping glioma nuclei. *IEEE Access* **7**, 84040–84052 (2019).
13. Kurc, T. et al. Segmentation and classification in digital pathology for glioma research: challenges and deep learning approaches. *Front. Neurosci.* **14**, 27 (2020).
14. Xu, Y. et al. Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. *BMC Bioinformatics* **18**, 281 (2017).
15. Fukuma, K., Surya Prasath, V. B., Kawanaka, H., Aronow, B. J. & Takase, H. A study on feature extraction and disease stage classification for Glioma pathology images. In *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* 2150–2156 <https://doi.org/10.1109/FUZZ-IEEE.2016.7737958> (2016).
16. Kalra, S. et al. Pan-cancer diagnostic consensus through searching archival histopathology images using artificial intelligence. *Npj Digit. Med.* **3**, 1–15 (2020).
17. Winkelmaier, G., Koch, B., Bogardus, S., Borowsky, A. D. & Parvin, B. Biomarkers of tumor heterogeneity in glioblastoma multiforme cohort of TCGA. *Cancers* **15**, 2387 (2023).
18. Yuan, M. et al. Image-based subtype classification for glioblastoma using deep learning: prognostic significance and biologic relevance. *JCO Clin. Cancer Inform.* **8**, e2300154, <https://doi.org/10.1200/CCI.23.00154> (2024).
19. Liu, X.-P. et al. Clinical significance and molecular annotation of cellular morphometric subtypes in lower-grade gliomas discovered by machine learning. *Neuro-Oncol* **25**, 68–81 (2023).
20. Song, J. et al. Enhancing spatial transcriptomics analysis by integrating image-aware deep learning methods. in *Biocomputing 2024* 450–463 (WORLD SCIENTIFIC, 2023). https://doi.org/10.1142/9789811286421_0035.
21. Zadeh Shirazi, A. et al. A deep convolutional neural network for segmentation of whole-slide pathology images identifies novel tumour cell-perivascular niche interactions that are associated with poor survival in glioblastoma. *Br. J. Cancer* **125**, 337–350 (2021).
22. Zheng, Y., Carrillo-Perez, F., Pizurica, M., Heiland, D. H. & Gevaert, O. Spatial cellular architecture predicts prognosis in glioblastoma. *Nat. Commun.* **14**, 4122 (2023).
23. Luo, J., Pan, M., Mo, K., Mao, Y. & Zou, D. Emerging role of artificial intelligence in diagnosis, classification and clinical management of glioma. *Semin. Cancer Biol.* **91**, 110–123 (2023).
24. Zadeh Shirazi, A. et al. The application of deep convolutional neural networks to brain cancer images: a survey. *J. Pers. Med.* **10**, 224 (2020).
25. Sotoudeh, H. et al. Artificial intelligence in the management of glioma: era of personalized medicine. *Front. Oncol.* **9**, 768 (2019).
26. Philip, A. K., Samuel, B. A., Bhatia, S., Khalifa, S. A. M. & El-Seedi, H. R. Artificial intelligence and precision medicine: a new frontier for the treatment of brain tumors. *Life* **13**, 24 (2023).
27. Jin, W. et al. Artificial intelligence in glioma imaging: challenges and advances. *J. Neural Eng.* **17**, 021002 (2020).
28. Cè, M. et al. Artificial intelligence in brain tumor imaging: a step toward personalized medicine. *Curr. Oncol.* **30**, 2673–2701 (2023).
29. Liu, Y. & Wu, M. Deep learning in precision medicine and focus on glioma. *Bioeng. Transl. Med.* **8**, e10553 (2023).
30. Bhatele, K. R. & Bhadauria, S. S. Machine learning application in Glioma classification: review and comparison analysis. *Arch. Comput. Methods Eng.* **29**, 247–274 (2022).
31. Muhammad, K., Khan, S., Ser, J. D. & Albuquerque, V. H. C. Deep learning for multigrade brain tumor classification in smart healthcare systems: a prospective survey. *IEEE Trans. Neural Netw. Learn. Syst.* **32**, 507–522 (2021).
32. Zhao, R. & Krauze, A. Survival prediction in gliomas: current state and novel approaches. *Exon Publ.* 151–169 <https://doi.org/10.36255/exonpublications.gliomas.2021.chapter9> (2021).
33. Alleman, K. et al. Multimodal deep learning-based prognostication in glioma patients: a systematic review. *Cancers* **15**, 545 (2023).
34. Wijethilake, N. et al. Glioma survival analysis empowered with data engineering—a survey. *IEEE Access* **9**, 43168–43191 (2021).
35. Faust, K. et al. Integrating morphologic and molecular histopathological features through whole slide image registration and deep learning. *Neuro-Oncol. Adv.* **4**, v04001 (2022).
36. Pytlarz, M., Wojnicki, K., Pilanc, P., Kaminska, B. & Crimi, A. Deep learning glioma grading with the tumor microenvironment analysis protocol for comprehensive learning, discovering, and quantifying microenvironmental features. *J. Imaging Inform. Med.* <https://doi.org/10.1007/s10278-024-01008-x> (2024).
37. McLendon, R. et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
38. Brennan, C. W. et al. The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
39. The Cancer Genome Atlas Research Network. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).
40. Mobadersany, P. et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl. Acad. Sci.* **115**, E2970–E2979 (2018).
41. Hewitt, K. J. et al. Direct image to subtype prediction for brain tumors using deep learning. *Neuro-Oncol. Adv.* **5**, v04139 (2023).
42. Nasrallah, M. P. et al. Machine learning for cryosection pathology predicts the 2021 WHO classification of glioma. *Med* **4**, 526–540.e4 (2023).
43. Wang, W. et al. Neuropathologist-level integrated classification of adult-type diffuse gliomas using deep learning from whole-slide pathological images. *Nat. Commun.* **14**, 6359 (2023).
44. Im, S. et al. Classification of diffuse glioma subtype from clinical-grade pathological images using deep transfer learning. *Sensors* **21**, 3500 (2021).
45. Li, Z. et al. Vision transformer-based weakly supervised histopathological image analysis of primary brain tumors. *iScience* **26**, 105872 (2023).
46. Shi, L. et al. Contribution of whole slide imaging-based deep learning in the assessment of intraoperative and postoperative sections in neuropathology. *Brain Pathol* **33**, e13160 (2023).
47. Hou, L. et al. Patch-based convolutional neural network for whole slide tissue image classification. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2424–2433 <https://doi.org/10.1109/CVPR.2016.266> (2016).
48. Jin, L. 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).

49. Jose, L. et al. Artificial intelligence–assisted classification of gliomas using whole slide images. *Arch. Pathol. Lab. Med.* **147**, 916–924 (2022).
50. Wang, X., Price, S. & Li, C. Multi-task Learning of Histology and Molecular Markers for Classifying Diffuse Glioma. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023* (eds. Greenspan, H. et al.) 551–561 (Springer Nature Switzerland, Cham, 2023). https://doi.org/10.1007/978-3-031-43990-2_52.
51. Hsu, W.-W. et al. A weakly supervised deep learning-based method for glioma subtype classification using WSI and mpMRIs. *Sci. Rep.* **12**, 6111 (2022).
52. Wang, X. et al. Combining radiology and pathology for automatic glioma classification. *Front. Bioeng. Biotechnol.* **10**, 841958 (2022).
53. Mallya, M. & Hamarneh, G. Deep multimodal guidance for medical image classification. in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022* (eds. Wang, L., Dou, Q., Fletcher, P. T., Speidel, S. & Li, S.) 298–308 (Springer Nature Switzerland, 2022). https://doi.org/10.1007/978-3-031-16449-1_29.
54. Kim, G. J., Lee, T., Ahn, S., Uh, Y. & Kim, S. H. Efficient diagnosis of IDH-mutant gliomas: 1p/19qNET assesses 1p/19q codeletion status using weakly-supervised learning. *Npj Precis. Oncol.* **7**, 1–9 (2023).
55. Ma, Y. et al. Histopathological auxiliary system for brain tumour (HAS-Bt) based on weakly supervised learning using a WHO CNS5-style pipeline. *J. Neurooncol.* **163**, 71–82 (2023).
56. Jin, L. et al. A multi-center performance assessment for automated histopathological classification and grading of glioma using whole slide images. *iScience* **26**, 108041 (2023).
57. Mousavi, H. S., Monga, V., Rao, G. & Rao, A. U. K. Automated discrimination of lower and higher grade gliomas based on histopathological image analysis. *J. Pathol. Inform.* **6**, 15 (2015).
58. Wang, X. et al. Machine learning models for multiparametric glioma grading with quantitative result interpretations. *Front. Neurosci.* **12**, 1046 (2019).
59. Mohan, G. et al. Intelligent framework for brain tumor grading using advanced feature analysis. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **11**, 485–503 (2023).
60. Rathore, S., Niazi, T., Iftikhar, M. A. & Chaddad, A. Glioma grading via analysis of digital pathology images using machine learning. *Cancers* **12**, 578 (2020).
61. Barker, J., Hoogi, A., Depeursinge, A. & Rubin, D. L. Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles. *Med. Image Anal.* **30**, 60–71 (2016).
62. Reza, S. M. S. & Iftikharuddin, K. M. Glioma grading using cell nuclei morphologic features in digital pathology images. in *Medical Imaging 2016: Computer-Aided Diagnosis* 9785 735–740 (SPIE, 2016).
63. Pei, L., Jones, K. A., Shboul, Z. A., Chen, J. Y. & Iftikharuddin, K. M. Deep neural network analysis of pathology images with integrated molecular data for enhanced glioma classification and grading. *Front. Oncol.* **11**, 668694 (2021).
64. Truong, A. H., Sharmanska, V., Limbäck-Stanic, C. & Grech-Sollars, M. Optimization of deep learning methods for visualization of tumor heterogeneity and brain tumor grading through digital pathology. *Neuro-Oncol. Adv.* **2**, vdaa110 (2020).
65. Ertosun, M. G. & Rubin, D. L. Automated grading of gliomas using deep learning in digital pathology images: a modular approach with ensemble of convolutional neural networks. *AMIA. Annu. Symp. Proc.* **2015**, 1899–1908 (2015).
66. Su, F. et al. Annotation-free glioma grading from pathological images using ensemble deep learning. *Heliyon* **9**, e14654 (2023).
67. Momeni, A., Thibault, M. & Gevaert, O. *Deep Recurrent Attention Models for Histopathological Image Analysis*. 438341 Preprint at <https://doi.org/10.1101/438341> (2018).
68. Zhao, Y. et al. Computational pathology for prediction of isocitrate dehydrogenase gene mutation from whole slide images in adult patients with diffuse glioma. *Am. J. Pathol.* **194**, 747–758, <https://doi.org/10.1016/j.ajpath.2024.01.009> (2024).
69. Qiu, L. et al. Hierarchical multimodal fusion framework based on noisy label learning and attention mechanism for cancer classification with pathology and genomic features. *Comput. Med. Imaging Graph.* **104**, 102176 (2023).
70. Chen, R. J. et al. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Trans. Med. Imaging* **41**, 757–770 (2022).
71. Xing, X. et al. Discrepancy and gradient-guided multi-modal knowledge distillation for pathological glioma grading. in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022* (eds. Wang, L., Dou, Q., Fletcher, P. T., Speidel, S. & Li, S.) 636–646 (Springer Nature Switzerland, 2022). https://doi.org/10.1007/978-3-031-16443-9_61.
72. Saldanha, O. L. et al. Self-supervised attention-based deep learning for pan-cancer mutation prediction from histopathology. *Npj Precis. Oncol.* **7**, 1–5 (2023).
73. Loeffler, C. M. L. et al. Predicting mutational status of driver and suppressor genes directly from histopathology with deep learning: a systematic study across 23 solid tumor types. *Front. Genet.* **12**, 806386 (2022).
74. Liu, S. et al. Isocitrate dehydrogenase (IDH) status prediction in histopathology images of gliomas using deep learning. *Sci. Rep.* **10**, 7733 (2020).
75. Wang, D. et al. Automated machine-learning framework integrating histopathological and radiological information for predicting IDH1 mutation status in glioma. *Front. Bioinforma.* **1**, 718697 (2021).
76. Jiang, S., Zanazzi, G. J. & Hassanpour, S. Predicting prognosis and IDH mutation status for patients with lower-grade gliomas using whole slide images. *Sci. Rep.* **11**, 16849 (2021).
77. Liechty, B. et al. Machine learning can aid in prediction of IDH mutation from H&E-stained histology slides in infiltrating gliomas. *Sci. Rep.* **12**, 22623 (2022).
78. Gorlia, T. et al. Nomograms for predicting survival of patients with newly diagnosed glioblastoma: prognostic factor analysis of EORTC and NCIC trial 26981-22981/CE.3. *Lancet Oncol* **9**, 29–38 (2008).
79. Nalisnik, M. et al. Interactive phenotyping of large-scale histology imaging data with HistomicsML. *Sci. Rep.* **7**, 14588 (2017).
80. Arslan, S. et al. A systematic pan-cancer study on deep learning-based prediction of multi-omic biomarkers from routine pathology images. *Commun. Med.* **4**, 1–15 (2024).
81. Luo, C., Yang, J., Liu, Z. & Jing, D. Predicting the recurrence and overall survival of patients with glioma based on histopathological images using deep learning. *Front. Neurol.* **14**, 1100933 (2023).
82. Baheti, B. et al. *Detecting Histologic & Clinical Glioblastoma Patterns of Prognostic Relevance*. Preprint at <https://doi.org/10.48550/arXiv.2302.00669> (2023).
83. Zhang, Y., Li, A., He, J. & Wang, M. A Novel MKL method for GBM prognosis prediction by integrating histopathological image and multi-omics data. *IEEE J. Biomed. Health Inform.* **24**, 171–179 (2020).
84. Powell, R. T. et al. Identification of histological correlates of overall survival in lower grade gliomas using a bag-of-words paradigm: a preliminary analysis based on hematoxylin & eosin stained slides from the lower grade glioma cohort of the cancer genome atlas. *J. Pathol. Inform.* **8**, 9 (2017).
85. Zadeh Shirazi, A. et al. DeepSurvNet: deep survival convolutional network for brain cancer survival rate classification based on histopathological images. *Med. Biol. Eng. Comput.* **58**, 1031–1045 (2020).
86. Liu, P., Fu, B., Ye, F., Yang, R. & Ji, L. DSCA: a dual-stream network with cross-attention on whole-slide image pyramids for cancer prognosis. *Expert Syst. Appl.* **227**, 120280 (2023).
87. Wang, Z. et al. Surformer: an interpretable pattern-perceptive survival transformer for cancer survival prediction from histopathology whole slide images. *Comput. Methods Programs Biomed.* **241**, 107733 (2023).

88. Chen, R. J. et al. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In 4015–4025 (2021).
89. Zhu, X., Yao, J., Zhu, F. & Huang, J. WSISA: making survival prediction from whole slide histopathological images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 6855–6863 <https://doi.org/10.1109/CVPR.2017.725> (2017).
90. Klemm, F. et al. Interrogation of the microenvironmental landscape in brain tumors reveals disease-specific alterations of immune cells. *Cell* **181**, 1643–1660.e17 (2020).
91. Cordell, E. C., Alghamri, M. S., Castro, M. G. & Gutmann, D. H. T lymphocytes as dynamic regulators of glioma pathobiology. *Neuro-Oncol* **24**, 1647–1657 (2022).
92. Zhu, X., Yao, J. & Huang, J. Deep convolutional neural network for survival analysis with pathological images. In *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* 544–547 <https://doi.org/10.1109/BIBM.2016.7822579> (2016).
93. Li, R., Yao, J., Zhu, X., Li, Y. & Huang, J. Graph CNN for survival analysis on whole slide pathological images. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (eds. Frangi, A. F., Schnabel, J. A., Davatzikos, C., Alberola-López, C. & Fichtinger, G.) 174–182 (Springer International Publishing, 2018). https://doi.org/10.1007/978-3-030-00934-2_20.
94. Chen, R. J. et al. Whole slide images are 2D point clouds: context-aware survival prediction using patch-based graph convolutional networks. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021* (eds. de Bruijne, M. et al.) 339–349 (Springer International Publishing, 2021). https://doi.org/10.1007/978-3-030-87237-3_33.
95. Jiang, S., Suriawinata, A. A. & Hassanpour, S. MHAttnSur: multi-head attention for survival prediction using whole-slide pathology images. *Comput. Biol. Med.* **158**, 106883 (2023).
96. Carmichael, I. et al. Incorporating intratumoral heterogeneity into weakly-supervised deep learning models via variance pooling. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022* (eds. Wang, L., Dou, Q., Fletcher, P. T., Speidel, S. & Li, S.) 387–397 (Springer Nature Switzerland, 2022). https://doi.org/10.1007/978-3-031-16434-7_38.
97. Chunduru, P., Phillips, J. J. & Molinaro, A. M. Prognostic risk stratification of gliomas using deep learning in digital pathology images. *Neuro-Oncol. Adv.* **4**, vdac111 (2022).
98. Chen, R. J. et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell* **40**, 865–878.e6 (2022).
99. Braman, N. et al. Deep Orthogonal Fusion: Multimodal Prognostic Biomarker Discovery Integrating Radiology, Pathology, Genomic, and Clinical Data. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021* (eds. de Bruijne, M. et al.) 667–677 (Springer International Publishing, Cham, 2021). https://doi.org/10.1007/978-3-030-87240-3_64.
100. Xing, X. et al. A multi-scale graph network with multi-head attention for histopathology image diagnosis. In *Proceedings of the MICCAI Workshop on Computational Pathology* 227–235 (PMLR, 2021).
101. Krebs, O., Agarwal, S. & Tiwari, P. Self-supervised deep learning to predict molecular markers from routine histopathology slides for high-grade glioma tumors. In *Medical Imaging 2023: Digital and Computational Pathology* 12471 1247102 (SPIE, 2023).
102. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778. <https://doi.org/10.1109/CVPR.2016.90> (2016).
103. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations (ICLR 2015)* (2015).
104. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2261–2269. <https://doi.org/10.1109/CVPR.2017.243> (2017).
105. Tan, M. & Le, Q. V. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. Preprint at <https://doi.org/10.48550/arXiv.1905.11946> (2020).
106. Szegedy, C. et al. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 1–9. <https://doi.org/10.1109/CVPR.2015.7298594> (2015).
107. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25* (Curran Associates, Inc., 2012).
108. Yang, P. et al. CS-CO: a hybrid self-supervised visual representation learning method for H&E-stained histopathological images. *Med. Image Anal.* **81**, 102539 (2022).
109. Lu, M. et al. SMILE: sparse-attention based multiple instance contrastive learning for glioma sub-type classification using pathological images. In *Proceedings of the MICCAI Workshop on Computational Pathology* 159–169 (PMLR, 2021).
110. Zhang, L. et al. *Mutual Contrastive Low-rank Learning to Disentangle Whole Slide Image Representations for Glioma Grading*. Preprint at <https://doi.org/10.48550/arXiv.2203.04013> (2022).
111. Cheerla, A. & Gevaert, O. Deep learning with multimodal representation for pancancer prognosis prediction. *Bioinformatics* **35**, i446–i454 (2019).
112. Jiang, S., Hondelink, L., Suriawinata, A. A. & Hassanpour, S. Masked pre-training of transformers for histology image analysis. *J. Pathol. Inform.* 100386 <https://doi.org/10.1016/j.jpi.2024.100386> (2024).
113. Chitnis, S. R. et al. Domain-Specific Pre-training Improves Confidence in Whole Slide Image Classification. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* 1–4. <https://doi.org/10.1109/EMBC40787.2023.10340659> (2023).
114. Suman, S. & Prasanna, P. Multi-stage attention-based network for brain tumor subtype classification. In *Medical Imaging 2022: Digital and Computational Pathology* 12039 272–276 (SPIE, 2022).
115. He, Y., Duan, L., Dong, G., Chen, F. & Li, W. Computational pathology-based weakly supervised prediction model for MGMT promoter methylation status in glioblastoma. *Front. Neurol.* **15**, 1345687 (2024).
116. Kleppe, A. et al. Designing deep learning studies in cancer diagnostics. *Nat. Rev. Cancer* **21**, 199–211 (2021).
117. Howard, F. M. et al. The impact of site-specific digital histology signatures on deep learning model accuracy and bias. *Nat. Commun.* **12**, 4423 (2021).
118. Howard, F. M., Kather, J. N. & Pearson, A. T. Multimodal deep learning: an improvement in prognostication or a reflection of batch effect? *Cancer Cell* **41**, 5–6 (2023).
119. Homeyer, A. et al. Recommendations on compiling test datasets for evaluating artificial intelligence solutions in pathology. *Mod. Pathol.* **35**, 1759–1769 (2022).
120. Brindha, V., Jayashree, P., Karthik, P. & Manikandan, P. Tumor grading model employing geometric analysis of histopathological images with characteristic nuclei dictionary. *Comput. Biol. Med.* **149**, 106008 (2022).
121. Yonekura, A., Kawanaka, H., Prasath, V. B. S., Aronow, B. J. & Takase, H. Automatic disease stage classification of glioblastoma multiforme histopathological images using deep convolutional neural network. *Biomed. Eng. Lett.* **8**, 321–327 (2018).
122. Blokker, M. et al. Fast intraoperative histology-based diagnosis of gliomas with third harmonic generation microscopy and deep learning. *Sci. Rep.* **12**, 11334 (2022).
123. Chen, R. J. et al. Towards a general-purpose foundation model for computational pathology. *Nat. Med.* **30**, 850–862 (2024).
124. Lu, M. Y. et al. A visual-language foundation model for computational pathology. *Nat. Med.* **30**, 863–874 (2024).

125. Lu, M. Y. et al. *A Foundational Multimodal Vision Language AI Assistant for Human Pathology*. Preprint at <https://doi.org/10.48550/arXiv.2312.07814> (2023).
126. Grenko, C. M. et al. Towards population-based histologic stain normalization of glioblastoma. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (eds. Crimi, A. & Bakas, S.) 44–56 (Springer International Publishing, Cham, 2020). https://doi.org/10.1007/978-3-030-46640-4_5.
127. Madusanka, N., Jayalath, P., Fernando, D., Yasakethu, L. & Lee, B.-I. Impact of H&E stain normalization on deep learning models in cancer image classification: performance, complexity, and trade-offs. *Cancers* **15**, 4144 (2023).
128. Bai, B. et al. Deep learning-enabled virtual histological staining of biological samples. *Light Sci. Appl.* **12**, 57 (2023).
129. Hölscher, D. L. et al. Next-generation morphometry for pathomics-data mining in histopathology. *Nat. Commun.* **14**, 470 (2023).
130. Hao, J., Kosaraju, S. C., Tsaku, N. Z., Song, D. H. & Kang, M. PAGE-Net: interpretable and integrative deep learning for survival analysis using histopathological images and genomic data. in *Biocomputing 2020* 355–366 (WORLD SCIENTIFIC, 2019). https://doi.org/10.1142/9789811215636_0032.
131. Chen, L. et al. Predicting the likelihood of an isocitrate dehydrogenase 1 or 2 mutation in diagnoses of infiltrative glioma. *Neuro-Oncol* **16**, 1478–1483 (2014).
132. Dosovitskiy, A. et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. Preprint at <https://doi.org/10.48550/arXiv.2010.11929> (2021).
133. Ker, J., Bai, Y., Lee, H. Y., Rao, J. & Wang, L. Automated brain histology classification using machine learning. *J. Clin. Neurosci.* **66**, 239–245 (2019).
134. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. In Proceedings of the 37th International Conference on Machine Learning 1597–1607 (PMLR, 2020).
135. Ghaffari Laleh, N. et al. Benchmarking weakly-supervised deep learning pipelines for whole slide classification in computational pathology. *Med. Image Anal.* **79**, 102474 (2022).
136. Ilse, M., Tomczak, J. & Welling, M. Attention-based Deep Multiple Instance Learning. In Proceedings of the 35th International Conference on Machine Learning 2127–2136 (PMLR, 2018).
137. Lu, M. Y. et al. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* **5**, 555–570 (2021).
138. Boehm, K. M., Khosravi, P., Vanguri, R., Gao, J. & Shah, S. P. Harnessing multimodal data integration to advance precision oncology. *Nat. Rev. Cancer* **22**, 114–126 (2022).
139. Lipkova, J. et al. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell* **40**, 1095–1110 (2022).
140. Wang, L.-B. et al. Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer Cell* **39**, 509–528.e20 (2021).
141. Roetzer-Pejrimovsky, T. et al. The digital brain tumour atlas, an open histopathology resource. *Sci. Data* **9**, 55 (2022).
142. Puchalski, R. B. et al. An anatomic transcriptional atlas of human glioblastoma. *Science* **360**, 660–663 (2018).
143. Bakas, S. et al. The University of Pennsylvania glioblastoma (UPenn-GBM) cohort: advanced MRI, clinical, genomics, & radiomics. *Sci. Data* **9**, 453 (2022).
144. Elazab, N., Gab-Allah, W. A. & Elmogy, M. A multi-class brain tumor grading system based on histopathological images using a hybrid YOLO and RESNET networks. *Sci. Rep.* **14**, 4584 (2024).
145. Yin, X. et al. Deep synergetic spiking neural P systems for the overall survival time prediction of glioblastoma patients. *Expert Syst. Appl.* **245**, 123032 (2024).
146. Dai, J. et al. Hypergraph-based spiking neural P systems for predicting the overall survival time of glioblastoma patients. *Expert Syst. Appl.* **215**, 119234 (2023).
147. Rathore, S., Iftikhar, M. A. & Mourelatos, Z. Prediction of overall survival and molecular markers in gliomas via analysis of digital pathology images using deep learning. Preprint at <https://doi.org/10.48550/arXiv.1909.09124> (2019).
148. Zheng, H., Momeni, A., Cedoz, P.-L., Vogel, H. & Gevaert, O. Whole slide images reflect DNA methylation patterns of human tumors. *Npj Genomic Med* **5**, 1–10 (2020).
149. Tang, B., Li, A., Li, B. & Wang, M. CapSurv: capsule network for survival analysis with whole slide pathological images. *IEEE Access* **7**, 26022–26030 (2019).
150. Liu, P., Ji, L., Ye, F. & Fu, B. AdvMIL: adversarial multiple instance learning for the survival analysis on whole-slide images. *Med. Image Anal.* **91**, 103020 (2024).

Acknowledgements

Research reported in this publication was supported by the German Federal Ministry of Health based on a resolution of the German Bundestag (funding codes: ZMI5-2522DAT15A, ZMI5-2522DAT15B, ZMI5-2522DAT15C, ZMI5-2522DAT15D, ZMI5-2522DAT15E). F.F. received additional funding from ERACoSysMed and the German Federal Ministry of Education and Research (BMBF) under grant number FKZ 31L0237A (MiEDGE).

Author contributions

J.P.R. and A.H. conceived the manuscript. J.P.R. conducted the literature review and wrote the manuscript, with feedback from A.H., F.F., J.W., N.S.S., S.T.H., C.B., S.L., and A.E. The final version of this manuscript was reviewed and approved by all authors.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Jan-Philipp Redlich.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024