Received: 8 January 2024 | Revised: 5 April 2024 | Accepted: 10 April 2024

DOI: 10.1111/nan.12981

REVIEW

Neuropathology and Applied Neurobiology
JOURNAL OF THE BRITISH NEUROPATHOLOGICAL SOCIETY
WILEY

Artificial intelligence in histopathological image analysis of central nervous system tumours: A systematic review

Melanie P. Jensen[1,2] | Zekai Qiang[3] | Danyal Z. Khan[4,5] | Danail Stoyanov[5] | Stephanie E. Baldeweg[6,7] | Zane Jaunmuktane[8,9,10] | Sebastian Brandner[8,9] | Hani J. Marcus[4,5]

[1]Pathology Department, Charing Cross Hospital, Imperial College Healthcare NHS Trust, London, UK

[2]Briscoe Lab, The Francis Crick Institute, London, UK

[3]School of Medicine and Population Health, University of Sheffield Medical School, Sheffield, UK

[4]Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, University College London Hospitals NHS Foundation Trust, London, UK

[5]Department of Computer Science, University College London, London, UK

[6]Department of Diabetes and Endocrinology, University College London Hospitals, London, UK

[7]Centre for Obesity and Metabolism, Department of Experimental and Translational Medicine, Division of Medicine, University College London, London, UK

[8]Division of Neuropathology, National Hospital for Neurology and Neurosurgery, University College London Hospitals NHS Foundation Trust, London, UK

[9]Department of Neurodegenerative Disease, University College London Queen Square Institute of Neurology, London, UK

[10]Department of Clinical and Movement Neurosciences, University College London Queen Square Institute of Neurology, London, UK

Correspondence
Melanie P. Jensen, Pathology Department, Charing Cross Hospital, Imperial College Healthcare NHS Trust, London, UK.
Email: m.jensen@cantab.net

Funding information
This work was supported by the National Institute for Health and Care Research Academic Clinical Fellowships to MPJ and DZK, and Wellcome/EPSRC Centre for Interventional and Surgical Sciences (203145/Z/16/Z) and UCLH/UCL Biomedical Research Centre (BRC) Neuroscience to HJM.

Abstract

The convergence of digital pathology and artificial intelligence could assist histopathology image analysis by providing tools for rapid, automated morphological analysis. This systematic review explores the use of artificial intelligence for histopathological image analysis of digitised central nervous system (CNS) tumour slides. Comprehensive searches were conducted across EMBASE, Medline and the Cochrane Library up to June 2023 using relevant keywords. Sixty-eight suitable studies were identified and qualitatively analysed. The risk of bias was evaluated using the Prediction model Risk of Bias Assessment Tool (PROBAST) criteria. All the studies were retrospective and preclinical. Gliomas were the most frequently analysed tumour type. The majority of studies used convolutional neural networks or support vector machines, and the most common goal of the model was for tumour classification and/or grading from haematoxylin and eosin-stained slides. The majority of studies were conducted when legacy World Health Organisation (WHO) classifications were in place, which at the time relied predominantly on histological (morphological) features but have since been superseded by molecular advances. Overall, there was a high risk of bias in all studies analysed. Persistent issues included inadequate transparency in reporting the number of patients and/or images within the model development and testing cohorts, absence of external validation, and

MPJ and ZQ shared first co-authorship. SB and HJM shared senior co-authorship.

insufficient recognition of batch effects in multi-institutional datasets. Based on these findings, we outline practical recommendations for future work including a framework for clinical implementation, in particular, better informing the artificial intelligence community of the needs of the neuropathologist.

KEYWORDS
artificial intelligence, central nervous system tumours, digital pathology, histopathology, medical image analysis

INTRODUCTION

Benign and malignant tumours of the central nervous system (CNS) encompass over 100 distinct entities. CNS tumours (both malignant and non-malignant) are the most common tumour site in children (0–15 years), and the second most common tumour site in adolescents and young adults (15–39 years).[1] The diagnostic pathway for CNS tumours involves multidisciplinary input, with the integration of clinical, demographic, imaging and pathological parameters. Pathological assessment, in particular, is the gold standard for precise, evidence-based classification of CNS tumours, with the 2021 World Health Organisation (WHO) Classification of Tumours of the CNS acting as the current reference for taxonomic classification.[2]

The emergence of artificial intelligence (AI) has the potential to provide tools for automated, rapid analysis of medical data, improving diagnostic workflow efficiency. AI refers to the use of machines (computers) to solve complex tasks that typically require human cognition and analysis. Within the diagnostic pathway for CNS tumours, the application of AI to radiological image analysis has been reviewed, with demonstrable benefits in predicting tumour grade and molecular profile.[3] Similarly, DNA methylation profiling by AI-based classifiers (machine learning algorithms) has become a well-established tool for classification based on epigenetic parameters.[2,4] However, the potential benefits of AI in interpreting histopathological features on slides of CNS tumour specimens remain unclear. In other solid organ tumours, AI-based algorithms have successfully detected breast, prostate and oesophageal cancer in histopathological image analysis; subtyped lung and kidney cancers; and classified cancers of unknown origin.[5–10] Advances have also been made in histopathological tasks where interobserver variation exists, such as Gleason grading of prostate cancer and in time-consuming tasks, such as determining and counting mitotic figures in tumour cells.[6,11] Indeed, some of these capabilities are available as FDA-approved products (e.g. Paige AI for prostate cancer detection).[12] Unique challenges, however, exist in CNS tumour classification from slide image analysis algorithms, namely the large number of tumour subtypes and the frequent overlap of morphological phenotypes across diagnostic entities, in particular in many low-grade glial and glioneuronal tumour types. It remains unclear whether these unique challenges have been accounted for in the existing literature.

A systematic analysis of AI-based histopathological image analysis of CNS tumours is lacking despite a growing body of relevant literature. The objective of this study is to survey the scope of AI employed in histopathological slide image analysis of CNS tumours, with the goal of identifying future directions in this field.

Key points

- This review explores the use of AI for image analysis of central nervous system tumour slides.
- The field is at an early stage and poorly aligned with current diagnostic challenges.
- Practical recommendations for future work are outlined.

METHODS

The review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines and prospectively registered with the International Prospective Register of Systematic Reviews (PROSPERO) database of systematic reviews (registration ID: CRD42023434059).[13]

We systematically interrogated the EMBASE, Medline and Cochrane Library databases up to June 2023 to identify studies utilising AI in the histopathological image analysis of CNS tumour tissue. A combination of MeSH terms and relevant keywords were used in the search strategy, including AI, machine learning, deep learning, brain neoplasms, pathology and computer-assisted image processing (Table S1). We limited the scope of the review to include studies focussing on conventional, clinically well-established histopathological image analysis (i.e. haematoxylin and eosin (H&E) and/or immunohistochemically stained tissue) and excluding studies exploring experimental (currently unvalidated) techniques such as Raman spectroscopy. We excluded studies not published as full-text articles in English.

Full-text articles meeting the inclusion criteria were independently assessed by two investigators (MPJ and ZQ). Information extracted from each study included the following: publication year; study stage; purpose of the AI algorithm; tumour type studied; use of H&E staining and/or immunohistochemical markers; characteristics and source of the training and testing datasets; data pre-processing techniques; details of internal and external validation; feature extraction and dimensionality reduction techniques; code availability;

summary of the AI algorithm and model architecture; interpretability considerations; and AI model outcome measures.

Risk of bias assessment was performed by two investigators (MPJ and ZQ) using the Prediction model Risk of Bias Assessment Tool (PROBAST).[14] A narrative synthesis was conducted to provide a comprehensive summary of the study characteristics, AI techniques employed, and key findings.

RESULTS

The literature search identified 68 studies meeting the eligibility criteria for inclusion (Figure 1).[15–82] All studies were retrospective and preclinical (Tables 1 and S2, and Figure 2). Studies were published between 1995 and 2023, half of which were published from 2020 onwards (Table 1 and Figure 2).

CNS tumour types

Gliomas were the most frequently analysed tumour type (52 studies) (Table 1 and Figure 2). Although glioblastoma was analysed in 33 studies, only eight out of 28 studies published post-2016 specified isocitrate dehydrogenase (IDH) gene mutation status (as per recommended classification systems).[2] Eight studies examined medulloblastoma in the paediatric population.[26,27,29,31,51–53,73] Nine studies investigated meningiomas.[16,33–35,41,46,58,61,62] Brain metastases from the breast, lung or melanoma were analysed in four studies.[39,43,58,60] Ependymomas (subtype not specified) were investigated in three studies.[34,43,73] CNS lymphoma was investigated in one study.[43] The exact CNS tumour type studied was unclear in one study.[65]

Dataset characteristics

One study utilised a mouse model of disseminated malignancy, and all other studies utilised human tissue.[60] The studies utilising human tissue covered adult and paediatric populations, ranging in size from 4 to 1185 patients and 10–97,252 digitised images (Figure 2). All studies were retrospective and cross-sectional (i.e. samples were analysed at a single point in time; rather than over several points in time as in longitudinal analyses). The most commonly used dataset was derived from The Cancer Genome Atlas, used in model development for 31 studies and external validation for two studies (Table 1 and

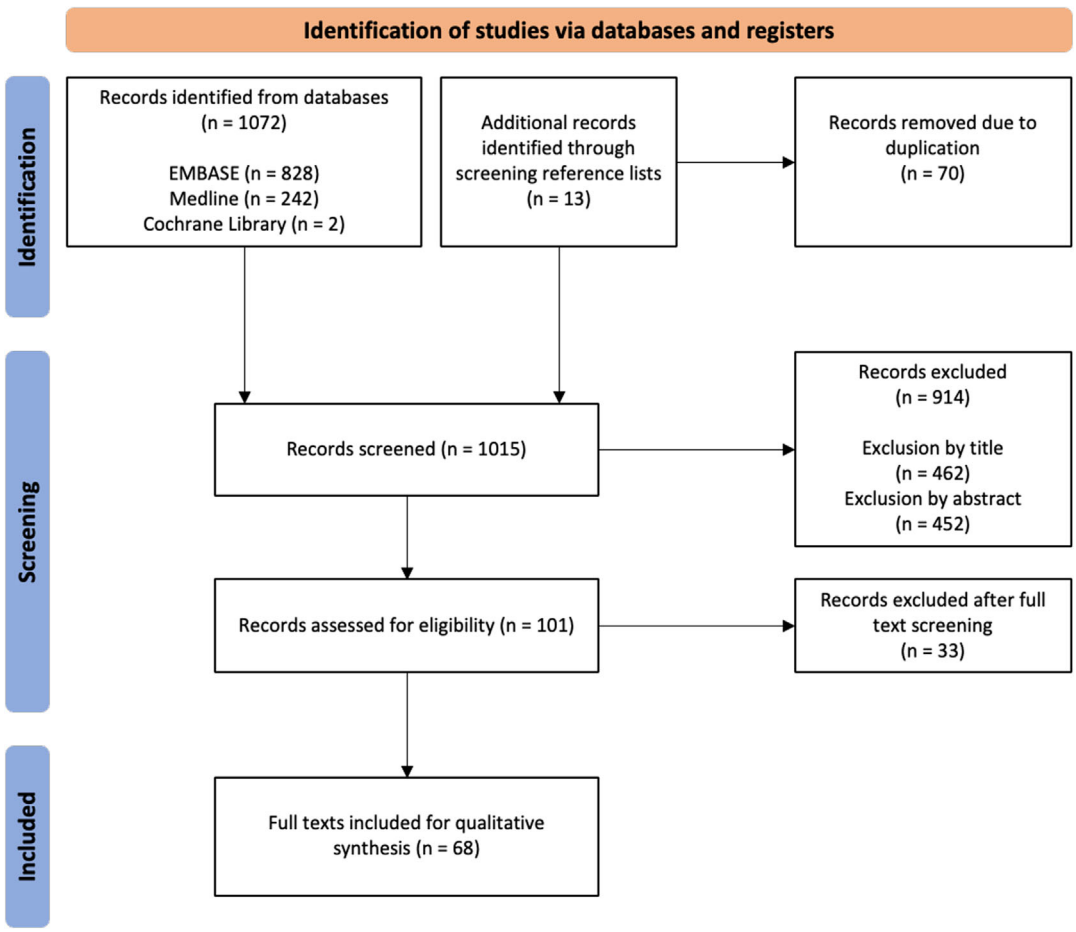

FIGURE 1 Preferred reporting items for systematic reviews and meta-analysis (PRISMA) flow diagram outlining study selection process. The primary search strategy yielded 1072 results, of which 68 studies were suitable for inclusion in the systematic review.

TABLE 1 Overview of the included studies.

Purpose of the artificial intelligence framework	Author and publication year	Brain tumour type	Data source for model development	Artificial intelligence algorithms employed	The clinical problem
Morphology recognition	Bao 2021[15]	Glioblastoma	Multi-centre (Australian Genomics and Clinical Outcomes of Glioma tissue bank)	Deep learning (Convolutional Neural Networks)	Identify six key morphological features of glioblastoma: palisading necrosis, microvascular proliferation, histologically normal-appearing blood vessels, geographic necrosis, brain tissue, and tumour background. In addition, simultaneously map CD276 expression, a prognostic marker linked to microvascular proliferation
	Lessman 2007[16]	WHO Grade I meningioma	Single-centre (Bethel Department of Neurosurgery, Bielefeld, Germany)	Deep learning (Self-Organising Maps)	Feature visualisation in meningioma histopathological images
	Li 2020[17]	Glioblastoma	Multi-centre (TCGA)	Deep learning (Analysis-Synthesis Learning With Shared Features)	Detect morphological features associated with each tumour type. In glioblastoma, microvascular proliferation regions are considered
	Li 2019[18]	Astrocytoma, oligodendroglioma, glioblastoma	Single-centre (Huashan Hospital of Fudan University)	Deep learning (Convolutional Neural Networks)	Detect and quantify the micro-vascularity in glioma and relate to clinical (survival) features
	Prokop 2022[19]	Glioblastoma	Not reported	Deep learning (Convolutional Neural Networks)	Quantify intra-tumour heterogeneity in glioblastoma
	Roy 2018[20]	Grade III astrocytomas	Multi-centre (TCGA)	Classical machine learning (Ensemble Classifiers: Logistic Regression, Random Forest, AdaBoost, Naive Bayes, Quadratic Discriminant Analysis and Neural Net; k-Nearest Neighbours)	Uncover the phenotypic factors distinguishing cells across different molecular groups, such as IDH wild type versus IDH mutant tumours
	Xu 2021[21]	Glioblastoma	Previous study by Xu et al. 2015	Classical machine learning (Tissue Cluster Level Graph Cut)	Recognise tumour and non-tumour regions on histological images
	Zhong 2017[22]	Glioblastoma	Multi-centre (TCGA)	Classical machine learning (Support Vector Machines)	Recognise necrotic from non-necrotic regions
	Nayak 2013[23]	Glioblastoma	Multi-centre (TCGA)	Classical machine learning and deep learning (Autoencoder, Support Vector Machines)	Recognise necrotic, transition into necrosis and viable tissues from whole slide images

TABLE 1 (Continued)

Purpose of the artificial intelligence framework	Author and publication year	Brain tumour type	Data source for model development	Artificial intelligence algorithms employed	The clinical problem
	Wen 2017[24]	Grade II glioma	Multi-centre (TCGA)	Classical machine learning (Random Forest, Support Vector Machines)	Evaluate nuclei segmentation performance
	Chang 2013[25]	Glioblastoma	Multi-centre (TCGA)	Classical machine learning (Support Vector Machines)	Recognise three morphological features of glioblastoma: tumour, necrosis and transition to necrosis
Tumour classification and/or grading	Attallah 2021[26]	Medulloblastoma	Single-centre (The Guwahati Medical College and Hospital and Guwahati Neurological Research Centre)	Classical machine learning and deep learning (Convolutional Neural Networks, Support Vector Machines, k-Nearest Neighbours, Linear Discriminant Analysis, Ensemble Subspace Discriminant)	Differentiate between normal and abnormal cases in paediatric medulloblastoma. Subsequently, identify the subclass of paediatric medulloblastoma, including classic, desmoplastic, large cell, and nodular subtypes
	Attallah 2021[27]	Medulloblastoma	Single-centre (The Guwahati Medical College and Hospital and Guwahati Neurological Research Centre)	Deep learning (Long Short-Term Memory Network)	As above
	Cevik 2021[28]	51 diagnostic entities obtained from the WHO Classification of CNS Tumours	Single-centre (Simulated Population)	Classical machine learning and deep learning (k-Nearest Neighbour, U-Net)	Classify brain cancer types and elucidate the roles of clinical, histological and molecular data in diagnostic processes
	Cruz-Roa 2012[29]	Medulloblastoma	Single-centre (St. Jude Children's Research Hospital)	Classical machine learning (k-Nearest Neighbours)	Differentiate between anaplastic and non-anaplastic medulloblastoma using whole slide images
	Decaestecker 1998[30]	Astrocytomas	Single-centre (Erasmus Hospital)	Classical machine learning (Decision Trees)	Develop an algorithm for improved astrocytic tumour grading incorporating morphological information
	Das 2021[31]	Medulloblastoma	Single-centre (The Guwahati Medical College and Hospital and Guwahati Neurological Research Centre)	Deep learning (Convolutional Neural Networks)	Differentiate between paediatric medulloblastoma and normal histopathological images, then further classify paediatric medulloblastoma according to WHO subtypes (nodular, classical, large cell, desmoplastic) at both the architectural and cellular levels.

(Continues)

TABLE 1 (Continued)

Purpose of the artificial intelligence framework	Author and publication year	Brain tumour type	Data source for model development	Artificial intelligence algorithms employed	The clinical problem
	Ertosun 2015[32]	Glioblastoma and low-grade glioma	Multi-centre (TCGA)	Deep learning (Convolutional Neural Networks)	Distinguish between glioblastoma and low-grade glioma, and further subclassify low-grade glioma into grades II and III from whole slide images
	Fatima 2017[33]	Meningioma	Single-centre (Bethel Department of Neurosurgery, Bielefeld, Germany)	Classical machine learning (Support Vector Machines)	Classify four subtypes of meningioma using nuclear morphology data
	Ghosh 2020[34]	Astrocytoma, ependymoma, and meningioma	Single-centre (Department of Neurosurgery, Bangur Institute of Neurosciences, Institute of Post Graduate Medical Education and Research, India)	Classical machine learning (Support Vector Machines)	Classify three types of brain tumour using a selection of histopathological features
	Grala 2009[35]	Meningioma	Multi-centre (Pathology Department of the Military Institute of Health Services, Warsaw, Poland and Pathomorphology Department of the Medical University of Lodz, Poland)	Classical machine learning (Support Vector Machines)	Estimate the Ki-67 labelling index in meningiomas to assist in histological grading
	Im 2021[36]	Grade II, III, and IV gliomas	Single-centre (Catholic University of Korea Yeouido St. Mary's Hospital)	Deep learning (Convolutional Neural Networks)	Classify the glioma histological subtypes and grade using H&E-stained whole slide images
	Jin 2021[37]	Oligodendroglioma, anaplastic oligodendroglioma, astrocytoma, anaplastic astrocytoma, glioblastoma	Single-centre (Central Nervous System Disease Biobank, Huashan Hospital, Fudan University, Shanghai)	Deep learning (Convolutional Neural Networks)	Classify subtypes of glioma using histopathological images
	Jose 2022[38]	Astrocytoma, oligodendroglioma and glioblastoma	Multi-centre (TCGA)	Deep learning (Convolutional Neural Networks)	Classify subtypes of glioma using histopathological images
	Jungo 2023[39]	Astrocytoma, oligodendroglioma and brain metastasis of small cell lung cancer, breast carcinoma and melanoma	Single-centre (Institutional Data)	Unknown	Classify histopathology images into gliomas versus brain metastases, distinguish between astrocytomas and astrocytosis, and predict 1p19q co-deletion status in IDH mutant tumours
	Bukhari 2020[40]	Astrocytoma	Single-centre (University of Lahore, Islamabad Campus)	Deep learning (Convolutional Neural Networks)	Differentiate between astrocytoma and normal brain tissue from digitised H&E pathology images

TABLE 1 (Continued)

Purpose of the artificial intelligence framework	Author and publication year	Brain tumour type	Data source for model development	Artificial intelligence algorithms employed	The clinical problem
	Kostopoulos 2015[41]	Astrocytoma, oligodendroglioma, and meningioma	Single-centre (University Hospital of Patras, Greece)	Deep learning (Probabilistic Neural Network)	Classify brain tumours into low and high grade based on histopathology images
	Kurc 2020[42]	Oligodendroglioma and astrocytoma	Multi-centre (MICCAI 2018 CPM Challenge)	Classical machine learning and deep learning (Convolutional Neural Networks, Support Vector Machines)	Segment nuclei from whole slide images of gliomas and distinguish between oligodendroglioma and astrocytoma
	Ma 2023[43]	Oligodendroglioma, astrocytoma, ependymoma, lymphoma, metastasis, background and non-tumoral tissue	Single-centre (Huashan Hospital's BioBank of Central Nervous System Diseases)	Deep learning (Multiple Instance Learning)	Classify central nervous system tumours based on histopathological images
	Mohan 2022[44]	Glioblastoma and low-grade glioma	Multi-centre (TCGA)	Classical machine learning (k-Nearest Neighbours, Support Vector Machines, Naïve Bayes, Logistic Regression)	Determine whether tumour tissue is glioblastoma or low-grade glioma based on histopathology images
	Qiu 2023[45]	Glioblastoma, low-grade glioma	Multi-centre (TCGA)	Deep learning (Autoencoder, Graph Convolutional Networks, Self-Normalising Networks)	Classify cancer using histopathology images and genomic data
	Qureshi 2008[46]	Meningioma	Not reported	Classical machine learning (Support Vector Machines, k-Nearest Neighbours, Naïve Bayes)	Subtype meningiomas based on histopathology images
	Wang 2019[47]	Grade II, III gliomas, glioblastoma	Single-centre (Shandong Provincial Hospital)	Classical machine learning and deep learning (Random Forest, Gradient Boosting Decision Tree, Support Vector Machines, Neural Networks)	Grade gliomas according to tissue whole slide image morphological features and Ki67 staining
	Xu 2017[82]	Glioblastoma and low-grade glioma	Multi-centre (MICCAI 2014 Brain Tumour Digital Pathology Challenge and Department of Pathology of Zhejiang University, China)	Classical machine learning and deep learning (Convolutional Neural Networks, Support Vector Machines)	Distinguish between glioblastoma and low-grade glioma using histological images, and recognise necrosis from non-necrosis regions
	Mousavi 2015[48]	Oligodendroglioma, anaplastic oligodendroglioma, astrocytoma, anaplastic astrocytoma, glioblastoma	Multi-centre (TCGA)	Classical machine learning (Decision Tree)	As above
	Reza 2016[49]	Low-grade glioma, glioblastoma	Multi-centre (TCGA)	Classical machine learning and deep learning (k-Nearest Neighbours, Multilayer Perceptron)	Classify glioma into glioblastoma or low-grade glioma based on morphological features of cell nuclei

(Continues)

TABLE 1 (Continued)

Purpose of the artificial intelligence framework	Author and publication year	Brain tumour type	Data source for model development	Artificial intelligence algorithms employed	The clinical problem
	Hou 2016[50]	Oligodendroglioma, anaplastic oligodendroglioma, astrocytoma, anaplastic astrocytoma, glioblastoma	Multi-centre (TCGA)	Classical machine learning and deep learning (Convolutional Neural Networks, Support Vector Machines, Random Forest)	Classify glioma into six subtypes: glioblastoma, oligodendroglioma, oligoastrocytoma, diffuse astrocytoma, anaplastic astrocytoma, anaplastic oligodendroglioma
	Das 2018[51]	Medulloblastoma	Multi-centre (The Guwahati Medical College and Hospital and Guwahati Neurological Research Centre)	Classical machine learning (Decision Tree, Support Vector Machines, k-Nearest Neighbour, Logistic Regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis)	Classify the different subtypes of paediatric medulloblastoma
	Das 2020[52]	Medulloblastoma	Multi-centre (The Guwahati Medical College and Hospital and Guwahati Neurological Research Centre)	Classical machine learning (Linear Discriminant Analysis, Quadratic Discriminant Analysis, Support Vector Machines, k-Nearest Neighbour, Decision Tree)	Classify the four subtypes of paediatric medulloblastoma
	Galaro 2011[53]	Medulloblastoma	Single-centre (St. Jude Children's Research Hospital)	Classical machine learning (k-Nearest Neighbours)	As above
	Yonekura 2016[54]	Oligodendroglioma, anaplastic oligodendroglioma, astrocytoma, anaplastic astrocytoma, glioblastoma	Multi-centre (TCGA)	Deep learning (Convolutional Neural Networks)	Classify glioma into low-grade glioma versus glioblastoma
	Truong 2020[55]	Astrocytoma, oligoastrocytoma, oligodendroglioma, glioblastoma	Multi-centre (TCGA)	Deep learning (Convolutional Neural Networks)	Classify glioma into glioblastoma and non-glioblastoma, and distinguish grade II from grade III glioma
	Su 2023[56]	Grade II and III gliomas	Multi-centre (TCGA)	Classical machine learning and deep learning (Convolutional Neural Networks, Ensemble Models: Naïve Bayes, Logistic Regression, Linear Discriminant, Decision Tree)	Distinguish between grades II and III glioma
Cell detection and/or quantification	Alzoubi 2022[57]	Glioblastoma	Multi-centre (Australian Genomics and Clinical Outcomes of Glioma tissue bank)	Deep learning (Convolutional Neural Networks)	Automatically detect, count, and identify cells labelled with CD276 immunohistochemically, a marker potentially representing glioblastoma stem cells

TABLE 1 (Continued)

Purpose of the artificial intelligence framework	Author and publication year	Brain tumour type	Data source for model development	Artificial intelligence algorithms employed	The clinical problem
	Lee 2023[58]	Meningioma, glioblastoma, small cell neuroendocrine metastatic cancer and oligodendroglioma	Not reported	Deep learning (Convolutional Neural Networks)	Quantify Ki67 immunohistochemistry staining on histological slides of brain tumours
	Lopez 2012[59]	Glioblastoma and anaplastic oligoastrocytoma	Not reported	Classical machine learning (Hierarchical Cluster Analysis)	Quantitatively characterise Ki67 hotspots in histopathological images
	Sikpa 2019[60]	Breast cancer brain metastasis	Single-centre (Institutional Dataset)	Classical machine learning (Random Forest)	Quantify breast metastatic disease burden to the brain from histological images
	Swiderska-Chadaj 2016[61]	Meningioma	Single-centre (Department of Pathomorphology, Military Institute of Medicine, Warsaw, Poland)	Classical machine learning (Support Vector Machines)	Detect Ki67 proliferation marker hotspots in meningioma
	Wirjadi 2016[62]	Meningioma	Multi-centre (The Broad Bioimage Benchmark Collection and SIMCEP dataset)	Deep learning (Convolutional Neural Networks)	Identify and segment meningioma nuclei under Ki67 staining
	Kong 2011[63]	Glioma	Not reported	Classical machine learning (Mean-Shift Clustering)	Detect nuclei from histopathology images
	Nalisnik 2017[64]	Low-grade glioma	Multi-centre (TCGA)	Classical machine learning (Random Forest)	Identify and segment vascular endothelial cell nuclei
	Xing 2016[65]	Not reported	Not reported	Deep learning (Convolutional Neural Networks)	Segment the nuclei on histopathology images of brain tumours
Molecular characterisation	Cooper 2010[66]	Glioblastoma	Multi-centre (TCGA)	Classical machine learning (Support Vector Machines)	Correlate pathology image data such as nuclear morphometry with molecular data
	Lietchy 2022[67]	Glioblastoma and low-grade glioma	Multi-centre (TCGA)	Deep learning (Convolutional Neural Networks)	Predict IDH molecular status from H&E histology slides
	Liu 2020[68]	Glioma	Multi-centre (TCGA and Yeditepe University Hospital, Istanbul, Turkey)	Deep learning (Generative Adversarial Networks, Convolutional Neural Networks)	Predict IDH molecular status from H&E histology slides
	Saldanha 2023[69]	Glioblastoma	Multi-centre (TCGA and CPTAC)	Deep learning (Convolutional Neural Networks)	Predict genomic alterations from cancer histology
Survival and outcome prognostication	Decaestecker 1995[70]	Astrocytomas	Single-centre (Erasmus Hospital)	Classical machine learning (Decision Trees)	Using DNA ploidy features to stratify patients by prognosis

(Continues)

TABLE 1 (Continued)

Purpose of the artificial intelligence framework	Author and publication year	Brain tumour type	Data source for model development	Artificial intelligence algorithms employed	The clinical problem
	Rathore 2019[71]	Glioblastoma	Multi-centre (TCGA)	Deep learning (Convolutional Neural Networks)	Predict the overall survival and molecular status using digital pathology images
	Rathore 2021[72]	Glioblastoma and low-grade glioma	Multi-centre (TCGA)	Classical machine learning (Support Vector Machines)	Predict the overall survival using MRI and digital pathology images
	Steyaert 2022[73]	Glioblastoma, low-grade glioma, high-grade astrocytoma, high-grade ependymoma and high-grade medulloblastoma	Multi-centre (Adult Cohort from TCGA and CPTAC; Paediatric Cohort from the Gabriella Miller Kids First Data Resource)	Deep learning (Convolutional Neural Networks)	Derive prognosis using pathology and genomic data
	Zadeh Shirazi 2020[74]	Glioblastoma	Multi-centre (TCGA)	Deep learning (Convolutional Neural Networks)	Derive prognosis using histopathology data
	Powell 2017[75]	Low-grade glioma	Multi-centre (TCGA)	Classical machine learning (Support Vector Machines)	Predict overall survival in a mixed histology and grade cohort of lower-grade glioma patients and identify their corresponding features
	Chen 2020[76]	Glioblastoma and low-grade glioma	Multi-centre (TCGA)	Classical machine learning and deep learning (Convolutional Neural Networks, Graph Convolutional Networks, k-Nearest Neighbours, Self-Normalising Networks)	Determine survival outcome of gliomas using histology and genomic data
	Hao 2020[77]	Glioblastoma	Multi-centre (TCGA)	Deep learning (Convolutional Neural Networks)	Combine pathological, genomic and demographic information for survival analysis
	Luo 2023[78]	Glioma	Single-centre (Xiangya Hospital)	Deep learning (Convolutional Neural Networks)	Predict the recurrence and overall survival in glioma patients using extracted histopathological features on H&E-stained images combined with clinical information
	Mobadersany 2018[79]	Grade II, III and IV gliomas	Multi-centre (TCGA)	Deep learning (Convolutional Neural Networks)	Predict recurrence and overall survival in glioma patients using extracted histopathological features on H&E-stained images combined with clinical information

TABLE 1 (Continued)

Purpose of the artificial intelligence framework	Author and publication year	Brain tumour type	Data source for model development	Artificial intelligence algorithms employed	The clinical problem
Image generation	Levine 2020[80]	Low-grade glioma	Multi-centre (TCGA)	Deep learning (Generative Adversarial Networks)	Synthesise pathology images, which may be used for education and quality assurance purposes
	Ozyoruk 2022[81]	Glioblastoma and low-grade glioma	Multi-centre (TCGA)	Deep learning (Generative Adversarial Networks)	Subtype meningiomas based on histopathology images

Figure 2). The number of cases ranged from 52 to 1185, whereas the number of images varied from 200 to 3611. The Guwahati Neurological Research Centre was another recurrently used dataset, albeit constrained by smaller sample sizes, with a maximum of 204 images or 20 patients included.[26,27,31,51,52] Six studies did not report the source of their datasets.[19,46,58,59,63,65] One study used a dataset with a simulated population derived from published literature.[28] Only two studies conducted exploratory analyses to examine the impact of sample size on the predictive performance of the model, aiming to address the challenge of requiring extensive labelled data for model training. Among them, only one study discussed methodologies for sample-size determination, employing inverse power law functions.[36]

AI algorithm usage

AI algorithms can be classified into classical machine learning and deep learning. Classical machine learning algorithms tend to be computationally simpler and advantageous when dealing with structured data, such as tabular data. Deep learning algorithms are computationally complex and are suitable for analysing complex data such as images and natural language. In Figure 3, we summarise key algorithm types used by included studies, and whether they fall under the classical machine learning or deep learning type. The most frequently employed classical machine learning algorithms were support vector machines, which identify the best margin of separation between data points of different classes in high-dimensional space (Figure 3), featured in 21 studies. The most frequently employed deep learning algorithms were convolutional neural networks, which employ hierarchical operations to process data and identify important features in an image (Figure 3), and featured in 30 studies. Classical machine learning algorithms dominated the landscape in earlier years, being the choice for 90% of studies published before 2013 (Table 1 and Figure 2). In contrast, deep learning algorithms were more frequently (67.2%) used in studies published after 2013.

Data pre-processing

Data pre-processing pipelines help render raw data suitable for training AI models. Eighteen studies, especially those utilising publicly available datasets, implemented quality control measures such as removing images with inferior resolution or processing artefacts. Image augmentation describes the technique of artificially expanding the training dataset to enhance model generalisability and mitigate class imbalances. This was implemented in 20 studies through a range of techniques, including flipping, rotating and geometric transformations, with some benefits for model performances.[38,68] Image normalisation, a process whereby image pixel values are standardised to a common scale to ensure model training efficiency, was described in 17 studies, using a variety of methods including contrast adjustment, colour adjustment and normalisation techniques to overcome inconsistencies in the staining process.[4,17,23,29,38,42,45,48,59,63,64,68,72,76–80] Four studies used pre-developed, open-source image pre-processing pipelines, two of which

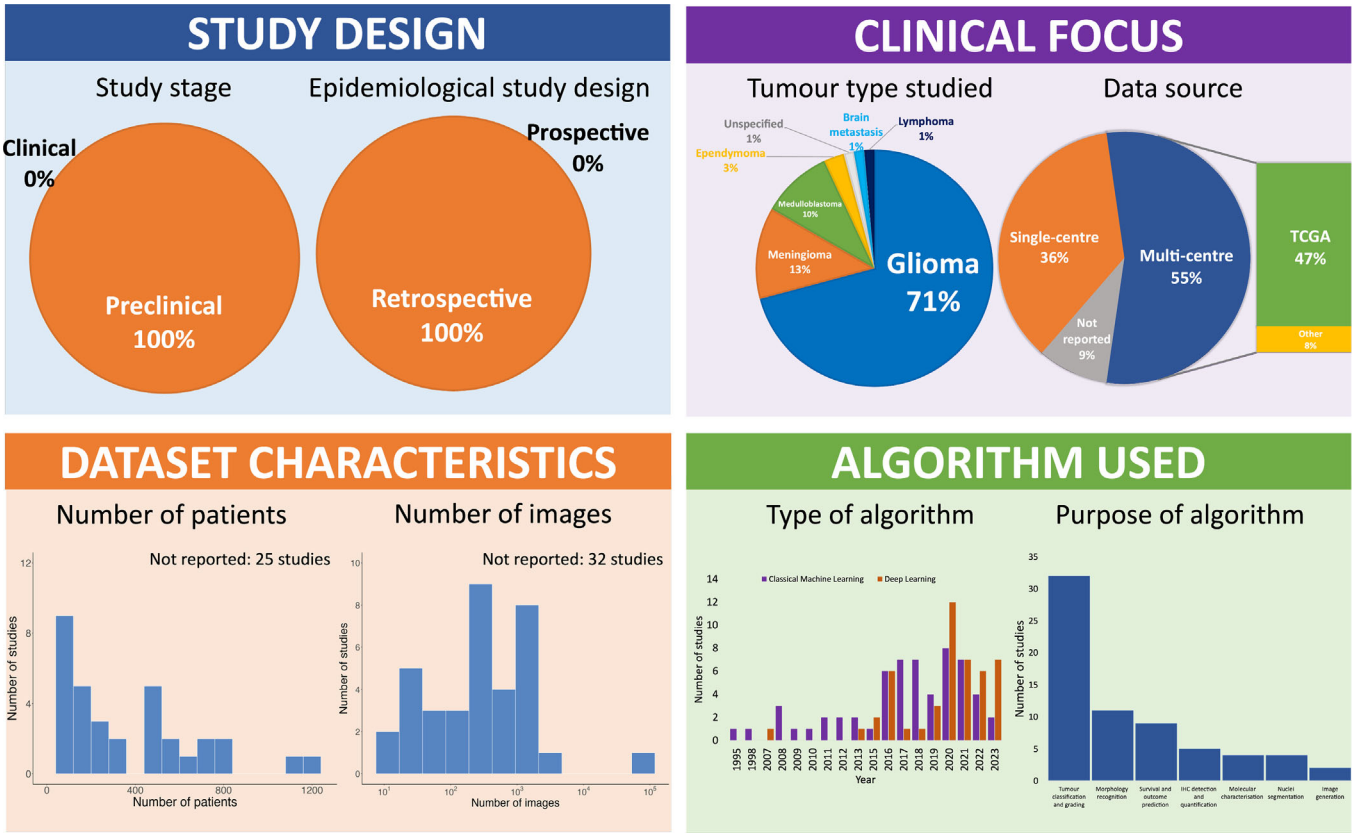

FIGURE 2 Summary results of included studies, including study design, clinical and dataset characteristics, and artificial intelligence algorithm type and goal.

used the Python Whole Slide Image (WSI) Pre-processing pipeline from https://github.com/deroneriksson/python-wsi-preprocessing, which performs a range of manoeuvres including colour correction, image tiling and tissue identification.[38,64,68,78] Furthermore, dimensionality reduction, the technique of reducing input features whilst retaining essential information from the training data, was primarily utilised in studies adopting classical machine learning algorithms. This was carried out to enhance training efficiency and reduce the risk of overfitting variables (whereby a model performs well on the training dataset but this is not recapitulated on an independent external dataset). Deep learning typically does not involve explicit dimensionality reduction because of its intrinsic capacity to learn hierarchical features from raw data. Therefore, dimensionality reduction was only performed in one study utilising a deep learning algorithm.[78]

Image analysis goal

The reviewed studies encompassed a range of image analysis goals (Figure 2). For each goal, we describe the performance metrics used and whether model interpretability was considered. Model interpretability involves discerning the model's primary contributing features to comprehend the model's decision-making process. It is crucial for trusted clinical integration, protecting against errors during model training and potentially revealing new insights through the recognition of previously undiscovered patterns.

GOAL 1: IMAGE GENERATION

Tissue image generation was the focus of two studies, aiming to develop tools for dataset augmentation and education.[80,81] Both studies adopted Turing tests (i.e. asking pathologists to assess whether the images were artificially generated or real) to show that distinguishing real from synthetic images was somewhat challenging (in both studies just over half of the images were deemed 'real').

GOAL 2: MORPHOLOGY RECOGNITION

Eleven studies focussed on the identification and analysis of morphological features, particularly microvascularity quantification and necrosis detection in glioblastomas.[15–25] Microvascular characteristics such as vessel circularity and area were considered in one study; however, determining whether vessels were normal or pathological was not explicitly performed.[18]

The area under the receiver operating characteristic curve (AUC) was the most commonly used performance measure (see Figure 3 for

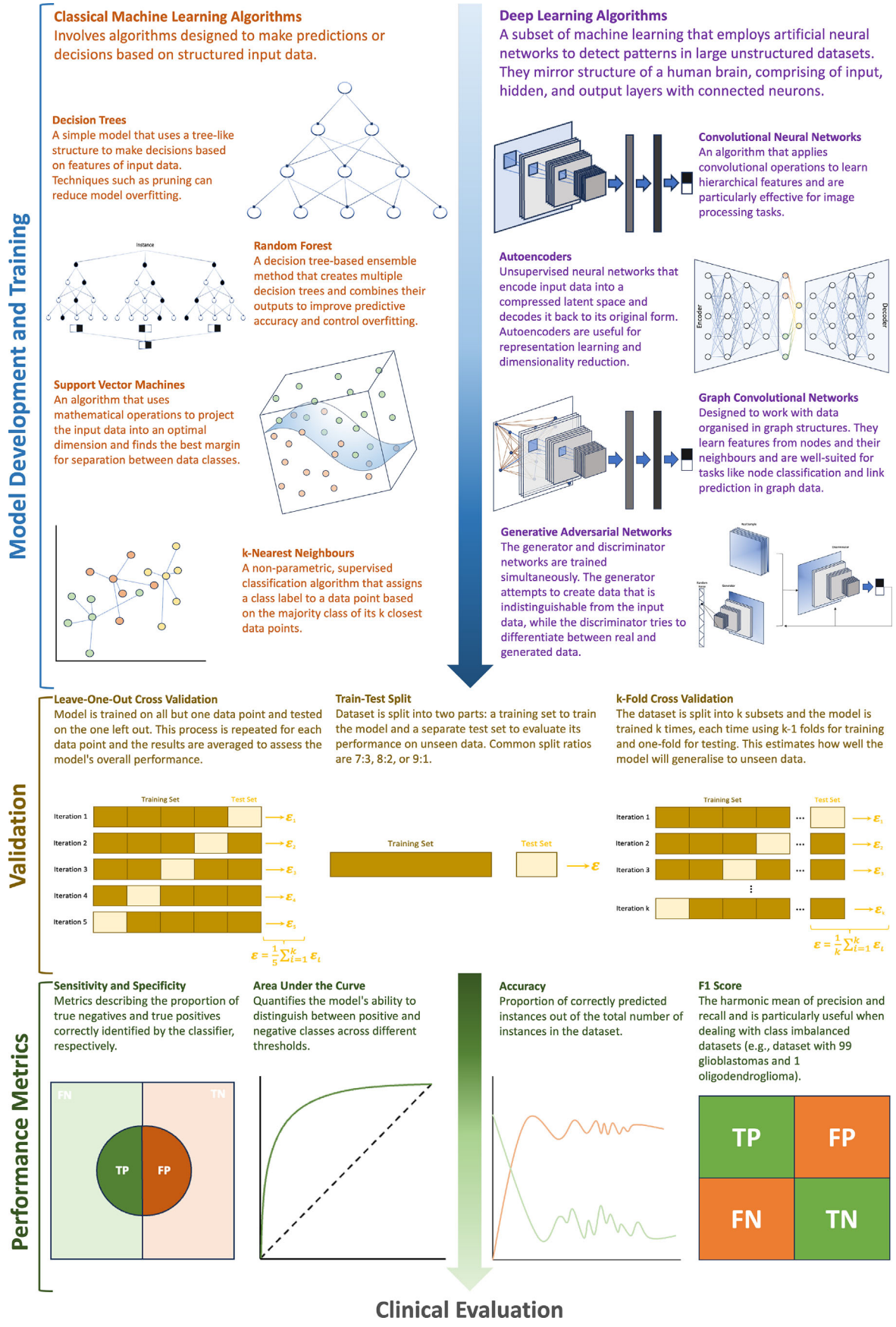

FIGURE 3 Development pipeline for artificial intelligence in digital histopathology, with relevant definitions (not an exhaustive list; see Goodfellow et al. 2016 for a detailed review).

a detailed definition and graphic of AUC).[15–18,22–25] Two studies did not report performance measures.[19,20] In one study, the AI model performed similar to human observers, particularly in detecting microvascular proliferation (AUC 0.994), geographic necrosis (AUC 0.994) and palisading necrosis (AUC 0.964).[15] However, given the rapidity at which pathologists can screen slides for these features and the relatively short time it takes to diagnose common CNS tumours (such as glioblastoma, meningioma, and most instances of ependymoma, astrocytoma and oligodendroglioma), the time- and cost-benefit analysis of implementing AI for this purpose is debatable. Interpretability was investigated in five studies, most of which involved direct visualisation of learned imagery features.[15–17,20,24] However, only two of these studies conducted comparisons with human pathologist's opinions.[15,24]

Features guiding the model were identified, including the observation that cells in IDH-mutant cases were larger and more circular versus wild-type counterparts; however, the clinical relevance of these features was not explored in the context of existing literature.[20]

GOAL 3: IMMUNOHISTOCHEMISTRY DETECTION AND QUANTIFICATION

Six studies focussed on AI-mediated quantification of immunohistochemical staining.[35,57–59,61,62] Among them, five studies quantified cellular proliferation hotspots using Ki67 immunohistochemistry and performed grading as per the WHO 2007 classification system.[35,58,59,61,62] In one study, AI was used to quantify CD276 immunohistochemically labelled cells, a putative glioblastoma stem cell marker.[57] The algorithm's intricacy demanded a labour-intensive training process, involving the manual labelling of 31,947 cells across eight WSIs. Subsequent external validation using an independent cohort revealed a quoted accuracy of 97.7%; however, the cohort was small relative to the number of cells in the training process (12,211 CD27-stained cells only). As such, the clinical applicability (and general utility) of the model is highly questionable, given the extensive human labelling process required to capture sufficient variance in the data.

Model outputs were commonly compared with that of human pathologists or conventional image analysis software, and concordance was demonstrated using measures of correlation such as Spearman's rho.[58,59,61] The AI model was demonstrated to have less variability compared to manual annotations between pathologists for Ki67 quantification in only one of these studies.[58] In this study, the algorithm was adopted to align Ki67-stained WSIs to H&E staining, facilitating automated region of interest selection and reducing inter-observer variability for Ki67 quantification.[58]

GOAL 4: NUCLEUS SEGMENTATION

Nucleus detection was performed in four studies.[60,63–65] Sikpa and others applied nucleus detection to quantify breast metastatic disease in the brain using an animal model with disseminated cancer spread, serving as an indicator of disease burden.[60] However, whether the results would be translatable to humans is unclear; the model used (representing hundreds of micrometastasis in the mouse brain) is not representative of the typical human counterpart (a single large metastasis). In Nalisnik et al., an AI nucleus detection model was employed to quantitatively characterise glioma microvascular structures, such as hypertrophy and hyperplasia.[64] Increased hyperplasia was found to be associated with higher grades within each molecular subtype (IDH-wild-type astrocytoma, IDH-mutant astrocytoma and oligodendroglioma). A regression analysis model was trained using these phenotypes across 781 WSIs, revealing a concordance index of 0.76, demonstrating some ability to rank patient survival based on these phenotypes. However, this is unsurprising as these phenotypes are those chosen by the WHO classification as prognostically relevant; hence, the conclusions are somewhat circular. Generalisation to other datasets was not performed and would be necessary for clinical validation. Meanwhile, Xing et al. proposed a generalisable model of nucleus detection applicable across multiple staining and tissue preparation methods, in an attempt to address the problem of batch effect in multicentre datasets.[65]

Model outputs for nucleus segmentation were generally in agreement with manual annotations or simpler computational techniques, as demonstrated through statistical analyses such as Pearson's correlations and false-positive area ratios.[60,63] Segmentation margins were examined in all four studies to assess interpretability.

GOAL 5: TUMOUR CLASSIFICATION AND GRADING

Thirty-two studies focussed on tumour classification or grading directly from H&E-stained tissue sections. Eighteen of these studies focussed on grading gliomas, the majority of which aimed to distinguish glioblastoma from lower-grade counterparts.[32,36,40,42,44,45,48,49,54,55,82] Several studies did not specify the subtype of tumour classified (e.g. astrocytoma subtype unspecified, oligodendroglioma/astrocytoma subtype unspecified), thus their inclusion criteria and therefore clinical utility are questionable.

Most of the studies (12 studies) were published in or before 2021 and therefore classified gliomas according to the 2007 or 2016 editions of the WHO classification (before molecular classifications were introduced in the 2021 edition).[21,32,34,40–42,47–50,54,55] Three studies (all published in 2022 or 2023) adopted the latest WHO integrated classification for gliomas as per new molecular markers.[38,39,43]

Jose et al. successfully differentiated IDH wild-type glioblastoma from IDH-mutant and 1p/19q-codeleted oligodendroglioma, and IDH-mutant astrocytoma,[38,83] whereas Jungo et al. distinguished IDH-mutant astrocytomas from astrocytosis.[39] Both models relied solely on H&E-stained WSIs and achieved accuracies of 91.7% and 96.7%, respectively. However, neither study reported on the specific morphological features enabling these predictions, preventing assessment regarding whether the model could reveal subvisual features unapparent to the pathologist.

Ma et al., however, employed a two-step algorithm to categorise tumours based on cell type and histological grading.[43] Subsequently, molecular parameters were imputed to formulate an integrated diagnosis using a decision tree classification algorithm, a simple classical machine learning method (see Figure 3). Although this approach acknowledges the significance of both morphological characteristics and molecular features, it did not exhibit discernible enhancements when compared to the established pathology pipeline.

Five studies subtyped paediatric medulloblastoma into classic, nodular, desmoplastic or large cell.[26,27,31,51,52] Two studies delineated anaplastic from non-anaplastic medulloblastoma.[29,53] However, given that molecular stratifications in medulloblastoma are becoming increasingly important, the diagnostic value of such histological classification in the absence of integration with molecular parameters is debatable.[2] Nonetheless, anaplasia in medulloblastoma is still regarded as a high-risk feature, and whilst its significance is diminishing in certain molecular subtypes, such an algorithm would be helpful if clinically validated. Two studies focussed on tissue feature subtyping of meningiomas into meningothelial, fibroblastic, transitional, or psammomatous.[33,46] Although this may demonstrate the ability of image recognition algorithms to discern distinct features, again the diagnostic value is limited as these subtypes are of less importance and have been superseded by molecular stratification algorithms.[84] Three studies performed a broad classification of CNS tumours, including meningioma, astrocytoma, ependymoma and oligodendroglioma.[28,34,41] However, all of these classification models were based on morphological categories with no clear demonstration of time–cost benefit relative to pathologist review nor comparison of accuracy relative to the final molecular diagnosis, making unclear their ability to offer additional clinical and prognostic utility. This is particularly relevant to tumour types, for example, meningiomas, in which current classifications are primarily at the genomic and epigenomic level.[2]

The most commonly utilised performance metrics were accuracy, sensitivity, specificity and *F*1 score (see Figure 3 for definitions of these performance metrics). Studies reported variable accuracy rates ranging from 85% to 100%; however, none conducted comparative analyses against human pathologist assessment (indeed 85% would be considered poor performance relative to the accuracy required in clinical practice). Only eight studies investigated interpretability.[29,32,34,36,44,45,47,48] This included the use of representation spaces to illustrate morphological features learned during training, such as edges, nuclear stains and cellular orientations, and visualisations with limited apparent clinical utility.[29,32] Other studies generated probabilistic heatmaps to highlight the model's attention during the decision-making process, which included tumour cell clusters, suggesting the plausibility of the proposed models.[37,45]

GOAL 6: MOLECULAR CHARACTERISATION

Four studies aimed to predict the molecular status of tumours based on H&E-stained tissue sections.[66–69] One study used nuclear morphology to predict the transcriptional profile of glioblastoma: classical, proneural, neural and mesenchymal.[66] However, this classification has been superseded by other systems because of emerging evidence, including the IDH status. Jungo et al. predicted the 1p19q co-deletion status of IDH-mutant tumours, reporting an accuracy of 88.6%, arguably lower than that acceptable in clinical practice[39] and probably even inferior to the morphological examination by an experienced neuropathologist. Another study sought to predict mutational status in glioblastoma and scored AUC metrics over 0.7 in four genes of interest (*IDH1*, *ATRX*, *TP53* and *RB1*).[69] Lietchy et al. and Liu et al. focussed on predicting IDH status from H&E stained slides.[67,68] Although Lietchy et al.'s model did not outperform human pathologists when assessed using the AUC metric when combining decisions made by both humans and the AI model within a man–machine hybrid framework, the model achieved superior performance compared to the consensus of two expert neuropathologists.[67]

Two studies assessed interpretability.[66,67] For example, human-recognisable features deterministic of IDH mutational status were revealed using methods to make predictions understandable through dimensionality reduction of complex datasets. These characteristics included oligodendroglial cytomorphology and the extent of pleomorphism.[67] However, during external validation, the model showed reduced performance (accuracy 0.809 vs 0.936 at internal testing), suggesting failure to generalise to independent datasets. The value of AI-based prediction of molecular status needs to be justified where relatively rapid cost-efficient methods already exist (e.g. widely utilised immunohistochemical tests for IDH mutations).

GOAL 7: SURVIVAL AND OUTCOME PREDICTION

Nine studies focussed on predicting patient prognosis directly from histopathological images.[70–79] Most studies adopted a multi-modal approach, integrating histological data with other modalities such as radiological, genomic or clinical data. Patients were stratified into survival probability groups or derived survival predictions through regression analysis. Evaluation metrics involved accuracy, AUC and concordance index. AI models improved performance when considering data from multiple modalities compared to histopathological data alone.[72,77–79] No studies explicitly showed that histopathology data alone performed better or similar to multimodal data.

Model interpretation was attempted in five studies.[70,73,75–77,79] Factors such as the percentage of hypertriploid nuclei and small, dense chromatin clump frequency were found to be relevant in stratifying anaplastic astrocytoma patients into prognostic outcomes.[70] Three studies considered interpretability by defining molecular pathways and genetic expression features linked to survival.[73,76,77] However, the histopathological features associated with survival were mainly demonstrated using representative images from the long and short survival groups, without explicit evaluation of which morphological features guided AI decision-making.

Internal and external validation

Internal validation refers to reserving a proportion of the original dataset to assess AI model reliability. Internal validation plays a crucial role in selecting the optimal model among candidate models and estimating whether the model will be able to generalise on unseen data. Robust but computationally expensive methods such as k-fold cross-validation were used in 37 studies, and leave-one-out cross-validation was utilised in four studies.[16,34,41,82] Seven studies relied solely on the train-test split approach, which is computationally simple but less representative of the model's true generalisability.[18,19,28,32,64,65,72] Nine studies did not provide details about internal validation.[35,39,45,58,60,61,63,78,81] See Figure 3 for detailed definitions of internal validation techniques employed.

External validation evaluates model performance using entirely new and independent data that were not part of the model's training or validation process. It is essential in determining a model's reproducibility and applicability in real-world clinical settings. Only seven studies conducted external validation.[18,28,32,57,67,74,81] Only three studies within this subset reported model performance on the corresponding unseen datasets.[18,67,74] See Figure 3 for detailed definitions of external validation methods employed.

Risk of bias assessment

Using the PROBAST evaluation tool, a significant proportion of studies displayed high risks of bias (61 studies) and limited applicability (66 studies) overall (Table S3 and Figure 4). In this context, risk of bias refers to flaws in the study's design, execution, or analysis that may result in systematically skewed assessments of a model's predictive accuracy. Applicability refers to whether the model will be representative of the population to which it will ultimately be applied. Forty-six studies scored a high risk of bias in the 'Participants' domain. This was largely attributed to (39 studies) sourcing of participant data from pre-existing datasets, where data are typically collected for a purpose other than model development or validation and often without an appropriate protocol.[14] Six studies did not provide clear information regarding the data source used.[19,46,58,59,63,65] Concerning the 'Predictors' domain, a high risk of bias was identified in 16 studies because of the use of manual annotation for ground truth labelling. This can result in inter-observer bias, as manual techniques may vary across observers. Within the 'Outcomes' domain, although the risk of bias was infrequent, a majority (56 studies) of studies demonstrated low applicability because of a lack of accessible published source code. The majority of studies (43 studies) scored a high risk of bias in the 'Analysis' domain. This was typically attributed to (37 studies) lack of reporting of the number of patients and/or images within the development and testing cohorts, impeding assessment of whether an adequate number of participants with the investigated outcome were included and whether the analysis covered all enrolled participants. Only two studies described methods for handling missing data.[72,76] Four studies did not provide any model performance information.[19,20,22,28] Except for the seven studies that conducted external validation, the risk of model overfitting on training data was largely overlooked.

DISCUSSION

Summary of findings

This review highlights the status of AI-driven histopathology image analysis in neuro-oncology. This is an evolving field, with half of the 68 reviewed studies published after 2020. The field is in its early stage; all of the studies were in the preclinical phase, retrospective in nature, and most failed to conduct direct comparisons with human pathologist assessment and to validate their outcomes with molecular tests. Moreover, all studies displayed a high risk of bias and/or limited applicability and thus potential clinical utility. Persistent issues included inadequate reporting of dataset characteristics (including the number of patients and/or images used for model development/validation and describing the methods for handling missing data), absence of external validation, insufficient recognition of batch effects in multi-institutional datasets or normalisation approaches for batch effects, and lack of published source code. Together, such issues preclude testing of model performance in independent patient cohorts by different research groups, critical in judging a model's safety, reliability and generalisability.

AI-driven image analysis for CNS tumour histopathology lags behind several other disciplines. For example, the capacity of AI-driven histopathology image analysis to achieve diagnostic accuracies on par with human pathologists has been prospectively demonstrated in other cancer types, such as gastric and colonic cancer.[85,86] The use of AI in prostate cancer grading is already at clinical evaluation stages.[87] In the field of neuro-oncology, AI applied to radiomic and tumour DNA methylation data is also at a more advanced stage. For example, AI algorithms applied to magnetic resonance imaging (MRI) images of pituitary neuroendocrine tumours to predict Ki67 proliferation indices have been tested in clinical settings.[88]

Challenges facing AI-driven image analysis of CNS tumours

This review reveals an absence of clinical integration of the AI image analysis algorithms. Achieving accurate CNS tumour classification through AI algorithms presents a multifaceted challenge. In contrast to many somatic tumours, CNS tumours, particularly low-grade gliomas, encompass a broad spectrum of subtypes, with either considerable morphological heterogeneity even within a single tumour type or considerable morphological overlap between distinct molecular subtypes.[89] There is often a poor correlation between morphological features and molecular precision diagnosis, particularly in low-grade gliomas and glioneuronal tumours. This presents challenges in curating large-scale databases for model training, as it necessitates the

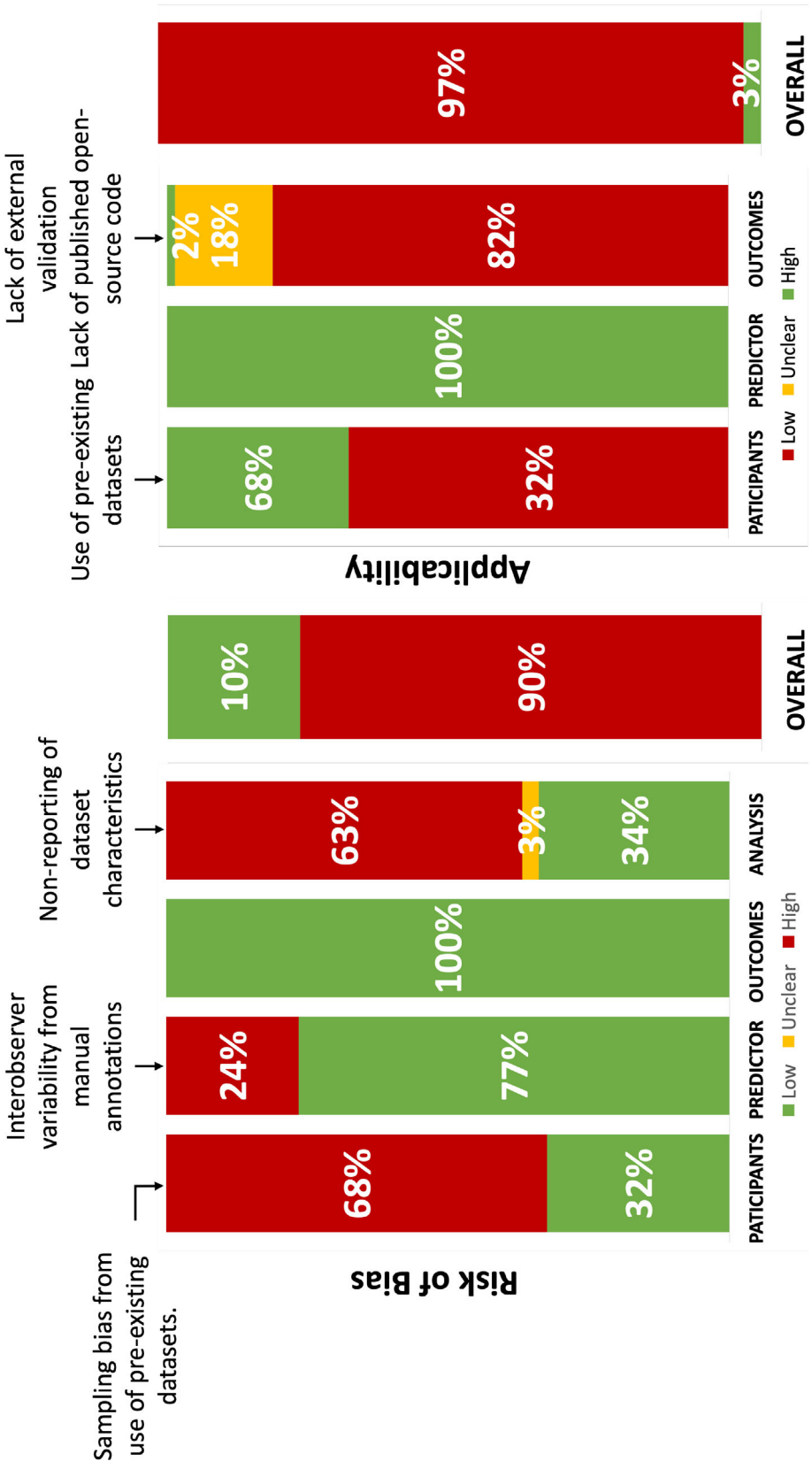

FIGURE 4 Bar chart summary of risk of bias and applicability assessment of included studies according to the PROBAST tool, and the key factors contributing to poor scores in each domain. Any study rated as a high risk of bias/concerns regarding applicability in one domain (analysis/outcomes/predictor/participants) is rated as a high risk of bias/concerns regarding applicability overall. The majority of studies (61 studies) displayed a high risk of bias, frequently attributed to (in 39 studies) sourcing of participant data from pre-existing datasets, where data are collected for a purpose other than for model development or validation. Virtually, all studies (66 studies) displayed high concerns regarding applicability, largely attributed to (in 56 studies) lack of accessible published source code (leading to a high concern regarding applicability score in the 'outcomes' domain).[14] PROBAST, Prediction model Risk of Bias Assessment Tool.

inclusion of numerous tumour categories whilst ensuring sample size comparability between classes. For instance, IDH-wild-type glioblastoma can comprise multiple histological variants, including giant cell, epithelioid or sarcomatous types.[2] This complexity can make it challenging for computational models to perform effective representation learning from histopathology images and derive accurate predictions.

Furthermore, none of the studies conducted an assessment of the time and financial cost-effectiveness of implementing AI models within the existing pathological workflow (in particular, in scenarios where digital images are not routinely generated), especially when compared to the expertise of neuropathologists. Future studies should delve into these aspects to provide convincing evidence for evaluation in the clinical setting.

The scarcity of prospective validation trials raises concerns given their pivotal role in evaluating clinical utility and safety of AI models.[90] Such trials are essential because changes in data characteristics between AI training and deployment stages can lead to performance degradation, a phenomenon known as 'data shift.'[91] Creating appropriate clinical studies for AI-based analysis is subject to numerous methodological challenges including lack of necessary expertise to translate these tools into practice in clinical pathology diagnostics, alongside the need to integrate human factors, estimate generalisability across sites and populations, and account for user variability.[90] Moreover, our review highlights that within the field of neuropathological image analysis, the absence of prospective validation trials likely relates to the lack of evidence that existing AI-based algorithms match the accuracy of neuropathologists in preclinical models, alongside limited comprehension of how such models may integrate into clinical practice.

Over one third of studies failed to report the number of patients evaluated and/or the number of images evaluated, with only one study discussing sample-size determination methodologies and subsequently testing the effect of sample size on AI model performance. AI studies with small test datasets risk overfitting of data and finding spurious correlations between confounding variables (e.g. scanner type, scanning settings, such as resolution and file compression parameters, slide origin, staining and slide quality) and target variables (e.g. tumour type).[92] Conversely, excessively large test datasets may not result in significant improvements in model accuracy despite increased time and cost. Finding an optimal balance between these requirements is a challenge for AI studies and merits greater exploration. Established criteria for evaluating sample size in AI studies are lacking, but potential methodologies are increasingly being proposed, including relatively simple confidence interval-based sample size calculations.[92]

Clinical recommendations

Currently, the literature appears skewed towards using AI to classify gliomas into morphological subtypes which are no longer listed in the 2021 WHO Classification (and have been superseded by molecular classifications), so it is unclear how they could assist current clinical workflows. Indeed, genetic and epigenetic parameters have now superseded the importance of histological subtyping in low-grade glioneuronal tumours, as they show considerable morphological overlap which may not be addressed with histological image analysis alone. The use of AI for image analysis in CNS tumour histopathology requires application to tasks which could be more usefully integrated into existing diagnostic workflows. For example, specific labour-intensive tasks, including determining mitotic and Ki67 indices to inform prognosis and stratify aggressive subtypes, have demonstrated convincing performances when executed by AI algorithms compared to human counterparts.[93] These tasks require significant time investment and are prone to interobserver disagreement and human error. Historically, these tasks have been difficult to automate (i.e. using rule-based software which operates on a set of predefined rules) and may benefit from AI assistance (which can iteratively improve by learning from data and making predictions based on new data).[94,95] AI-guided image analysis may also help inform and/or streamline requests for molecular testing based on a preliminary morphological diagnosis, although again this would require demonstrable time- and/or cost-benefit relative to neuropathologist review. Finally, AI could be used to 'mine' histopathological imaging data for 'subvisual' morphological features useful in diagnosis/prognostication unapparent to the pathologist. This may be particularly helpful in cases deemed unsolvable after assessment by pathologist review and available molecular testing, including DNA methylation arrays and genome sequencing.[96] Relevant to prognostication, AI has been used to predict the survival of breast cancer patients from H&E-stained slides, with greater accuracy than standard pathologist grading, based on stromal morphological structures previously unrecognised as prognostically relevant.[97] Similarly, AI models have been shown to extract prognostic information and make molecular predictions from tissue morphology in colorectal and bladder cancer, with greater accuracy than pathologists.[98,99] Improved communication between clinicians and engineers is imperative to achieve these advancements given the unique challenges in developing AI models for image analysis of CNS tumours.

Moreover, an essential prerequisite for the implementation of any AI algorithm on CNS tumour histopathology is the availability of a clinically validated digital pathology workflow integrated within the neuropathology department. This should include dedicated scanners for routine real-time digitisation of WSIs, image management software, and real-time access of AI algorithms to digitised images. Whilst the requirement for dedicated equipment imposes financial hurdles, access to external image analysis systems to stored histology datasets imposes data privacy and logistical hurdles.

Engineering recommendations

Studies to date are largely of low quality, with a high risk of bias and limited applicability. Key issues include inadequate documentation of dataset attributes and the handling of missing data. A critically small number of studies are externally validated, which is essential for demonstrating a model's ability to generalise on unseen datasets. Only a limited number of studies share their model source code, a practice which enhances research reproducibility, facilitates collaboration efforts and enables peer validation. Finally, AI model evaluation should be evaluated using clinically relevant appropriate metrics (e.g. relevant online tools).[100]

Several multi-centre datasets are utilised in the current literature, but this can cause batch effects (non-biological factors that create

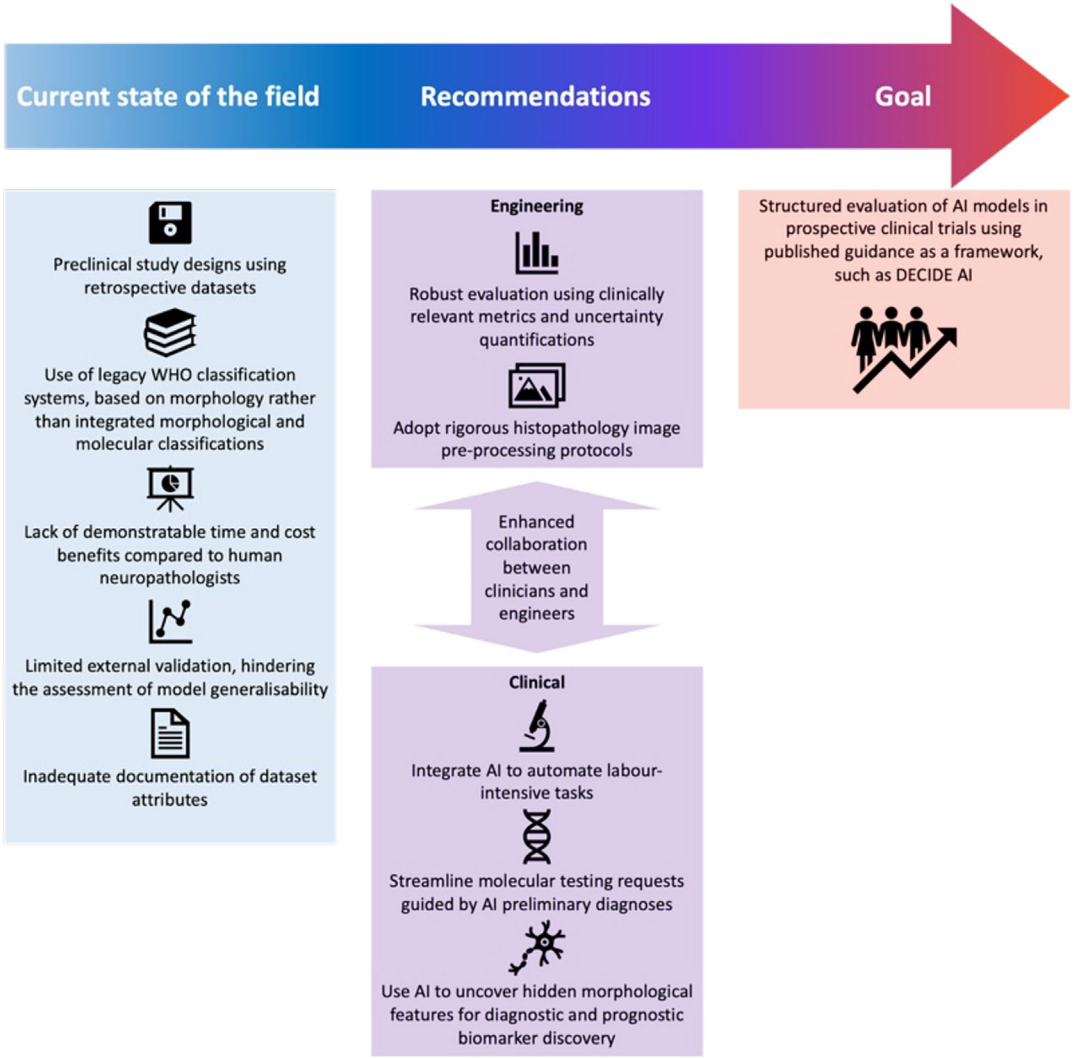

FIGURE 5 Recommendations for the clinical and engineering communities to help bridge the gap between preclinical studies (the current state of the field) and clinical implementation in the field of AI-driven histopathology image analysis of CNS tumours.

variation in the data) at various stages, from tissue collection to image digitisation. This could cause AI models to focus on the unique WSI signatures of individual sites, rather than inherent biological attributes.[101] Recommendations have been made for studies utilising multi-centre datasets, including reporting variations in outcomes observed across sites and implementing various pre-processing steps, including stain normalisation.[101] These steps are often omitted in the reviewed studies and should be considered. Comprehensive, freely available single-centre histopathology datasets (e.g. The Digital Brain Tumour Atlas) could be exploited for AI analysis whilst overcoming some of the issues associated with batch effects.[102]

Strengths and limitations

Through a systematic review of the literature, the present study offers an up-to-date exploration of AI-driven applications for the analysis of CNS tumour histopathology image analysis. The findings are critically evaluated in the context of clinical utility, with the provision of practical recommendations (Figure 5). However, certain limitations should be acknowledged. Although the identification of studies was comprehensive, it was constrained to the search strategies employed. Only full-text articles in the English language were considered, which could result in the omission of certain studies. Whilst an array of databases in the biomedicine domain have been examined, future investigations could encompass databases within computer science and related disciplines, including resources such as the IEEE Xplore Digital Library.

CONCLUSION

We present a systematic review of the literature concerning the use of AI for the analysis of neuro-oncological histopathological images. Despite a growing body of relevant literature, the field remains at an early stage; all of the studies were retrospective and preclinical, and poorly aligned with current diagnostic neuropathology workflows. A

high risk of bias was identified across the majority of studies; persistent issues identified included an absence of external validation and inadequate reporting of study characteristics. Based on these findings, we propose specific clinical and engineering recommendations, including adopting up-to-date integrated classification systems, improved reporting transparency of the number of patients and/or images within the model training and testing cohorts, rigorous external validations, and better considerations of model interpretability. We suggest that implementations of such changes, alongside better cross-disciplinary collaborations among clinicians, computer scientists, image analysts and engineers, are needed for the creation of robust AI models able to transition from preclinical models into clinical trials, with structured evaluation as per published guidance (e.g. DECIDE AI, CONSORT-AI).[68,90]

AUTHOR CONTRIBUTIONS

Melanie P Jensen and Zekai Qiang contributed to the conception and design of the work; the acquisition, analysis, and interpretation of the data; and drafted the work. Danyal Z Khan, Danail Stoyanov, Stephanie E Baldeweg, Zane Jaunmuktane and Sebastian Brandner reviewed the work critically for important intellectual content. Hani J Marcus contributed to the conception and design of the work and reviewed the work critically for important intellectual content. All authors gave the final approval of the version to be published.

CONFLICT OF INTEREST STATEMENT

The Editors of Neuropathology and Applied Neurobiology are committed to peer-review integrity and upholding the highest standards of review. As such, this article was peer-reviewed by independent, anonymous expert referees, and the authors (including SB and ZJ) had no role in either the editorial decision or the handling of the paper.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/nan.12981.

DATA AVAILABILITY STATEMENT

No new data was generated or analysed in support of this research.

ORCID

Melanie P. Jensen https://orcid.org/0000-0002-2027-5023
Zekai Qiang https://orcid.org/0000-0001-8271-8507
Sebastian Brandner https://orcid.org/0000-0002-9821-0342

REFERENCES

1. Ostrom QT, Price M, Neff C, et al. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2016-2020. *Neuro Oncol*. 2023;25(Supplement_4):iv1-iv99. doi:10.1093/neuonc/noad149
2. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro Oncol*. 2021;23(8):1231-1251. doi:10.1093/neuonc/noab106
3. Abdel Razek AAK, Alksas A, Shehata M, et al. Clinical applications of artificial intelligence and radiomics in neuro-oncology imaging. *Insights Imaging*. 2021;12(1):152. doi:10.1186/s13244-021-01102-6
4. Capper D, Jones DTW, Sill M, et al. DNA methylation-based classification of central nervous system tumours. *Nature*. 2018;555(7697):469-474. doi:10.1038/nature26000
5. Sandbank J, Bataillon G, Nudelman A, et al. Validation and real-world clinical application of an artificial intelligence algorithm for breast cancer detection in biopsies. *NPJ Breast Cancer*. 2022;8(1):129. doi:10.1038/s41523-022-00496-w
6. Bulten W, Pinckaers H, van Boven H, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol*. 2020;21(2):233-241. doi:10.1016/S1470-2045(19)30739-9
7. Gehrung M, Crispin-Ortuzar M, Berman AG, O'Donovan M, Fitzgerald RC, Markowetz F. Triage-driven diagnosis of Barrett's esophagus for early detection of esophageal adenocarcinoma using deep learning. *Nat Med*. 2021;27(5):833-841. doi:10.1038/s41591-021-01287-9
8. Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med*. 2018;24(10):1559-1567. doi:10.1038/s41591-018-0177-5
9. Chanchal AK, Lal S, Kini J. Deep structured residual encoder-decoder network with a novel loss function for nuclei segmentation of kidney and breast histopathology images. *Multimed Tools Appl*. 2022;81(7):9201-9224. doi:10.1007/s11042-021-11873-1
10. Lu MY, Chen TY, Williamson DFK, et al. AI-based pathology predicts origins for cancers of unknown primary. *Nature*. 2021;594(7861):106-110. doi:10.1038/s41586-021-03512-4
11. Veta M, van Diest PJ, Jiwa M, Al-Janabi S, Pluim JP. Mitosis counting in breast cancer: object-level Interobserver agreement and comparison to an automatic method. *PLoS ONE*. 2016;11(8):e0161286. doi:10.1371/journal.pone.0161286
12. AI P. Paige Receives First Ever FDA Approval for AI Product in Digital Pathology. 2023. https://paige.ai/paige-receives-first-ever-fda-approval-for-ai-product-in-digital-pathology/. Accessed 12 October 2023.
13. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339(jul21 1):b2535. doi:10.1136/bmj.b2535
14. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med*. 2019;170(1):W1-W33. doi:10.7326/M18-1377
15. Bao G, Wang X, Xu R, et al. Pathofusion: an open-source AI framework for recognition of pathomorphological features and mapping of immunohistochemical data. *Cancer*. 2021;13(4):617. doi:10.3390/cancers13040617
16. Lessmann B, Nattkemper TW, Hans VH, Degenhard A. A method for linking computed image features to histological semantics in neuropathology. *J Biomed Inform*. 2007;40(6):631-641. doi:10.1016/j.jbi.2007.06.007
17. Li X, Monga V, Rao UKA. Analysis-synthesis learning with shared features: algorithms for histology image classification. *IEEE Trans Biomed Eng*. 2020;67(4):1061-1073. doi:10.1109/TBME.2019.2928997
18. Li X, Tang Q, Yu J, Wang Y, Shi Z. Microvascularity detection and quantification in glioma: a novel deep-learning-based framework. *Lab Invest*. 2019;99(10):1515-1526. doi:10.1038/s41374-019-0272-3
19. Prokop G, Ortl M, Fotteler M, et al. Quantifying heterogeneity in tumors: proposing a new method utilizing convolutional neuronal networks. *Stud Health Technol Inform*. 2022;289. doi:10.3233/SHTI210942
20. Roy M, Wang F, Teodoro G, Vega JV, Brat D, Kong J. Analysis of Cellular Feature Differences of Astrocytomas with Distinct Mutational

Profiles Using Digitized Histopathology Images. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*. 2018; 2018.
21. Xu H, Liu L, Lei X, Mandal M, Lu C. An unsupervised method for histological image segmentation based on tissue cluster level graph cut. *Comput Med Imaging Graph*. 2021;93:101974. doi:10.1016/j.compmedimag.2021.101974
22. Zhong C, Han J, Borowsky A, Parvin B, Wang Y, Chang H. When machine vision meets histology: a comparative evaluation of model architecture for classification of histology sections. *Med Image Anal*. 2017;35:35-543. doi:10.1016/j.media.2016.08.010
23. Nayak N, Chang H, Borowsky A, Spellman P, Parvin B. Classification of tumor histopathology via sparse feature learning. *Proc IEEE Int Symp Biomed Imaging*. 2013;2013.
24. Wen S, Kurc TM, Gao Y, Zhao T, Saltz JH, Zhu W. A methodology for texture feature-based quality assessment in nucleus segmentation of histopathology image. *J Pathol Inform*. 2017;8(1):38. doi:10.4103/jpi.jpi_43_17
25. Chang H, Borowsky A, Spellman P, Parvin B. Classification of tumor histology via morphometric context. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2013;2013.
26. Attallah O. CoMB-deep: composite deep learning-based pipeline for classifying childhood Medulloblastoma and its classes. *Front Neuroinform*. 2021;15:15. doi:10.3389/fninf.2021.663592
27. Attallah O. Article mb-ai-his: histopathological diagnosis of pediatric medulloblastoma and its subtypes via ai. *Diagnostics*. 2021;11(2):359. doi:10.3390/diagnostics11020359
28. Cevik L, Landrove MV, Aslan MT, et al. Information theory approaches to improve glioma diagnostic workflows in surgical neuropathology. *Brain Pathol*. 2022;32(5):e13050. doi:10.1111/bpa.13050
29. Cruz-Roa A, Gonzalez F, Galaro J, et al. A visual latent semantic approach for automatic analysis and interpretation of anaplastic medulloblastoma virtual slides. *Medical image computing and computer-assisted intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2012; 15(Pt 1).
30. Decaestecker C, Camby I, Nagy N, Brotchi J, Kiss R, Salmon I. Improving morphology-based malignancy grading schemes in astrocytic tumors by means of computer-assisted techniques. *Brain Pathol*. 1998;8(1):29-38. doi:10.1111/j.1750-3639.1998.tb00131.x
31. Das D, Mahanta LB. A comparative assessment of different approaches of segmentation and classification methods on childhood medulloblastoma images. *J Med Biol Eng*. 2021;41(3):379-392. doi:10.1007/s40846-021-00612-4
32. Ertosun MG, Rubin DL. Automated grading of gliomas using deep learning in digital pathology images: a modular approach with ensemble of convolutional neural networks. *AMIA Annual Symposium Proceedings AMIA Symposium*. 2015;2015:1899-1908.
33. Fatima K, Majeed H, Irshad H. Nuclear spatial and spectral features based evolutionary method for meningioma subtypes classification in histopathology. *Microsc Res Tech*. 2017;80(8):851-861. doi:10.1002/jemt.22874
34. Ghosh K, Chandra JK, Ghosh A. Brain tumour classification by machine learning applications with selected biological features: towards a newer diagnostic regime. *J Anal Oncol*. 2020;9:9-19. doi:10.30683/1927-7229.2020.09.02
35. Grala B, Markiewicz T, Kozlowski W, Osowski S, Slodkowska J, Papierz W. New automated image analysis method for the assessment of Ki-67 labeling index in meningiomas. *Folia Histochem Cytobiol*. 2009;47(4). doi:10.2478/v10042-008-0098-0
36. Im S, Hyeon J, Rha E, et al. Classification of diffuse glioma subtype from clinical-grade pathological images using deep transfer learning. 2021;21(10):3500. doi:10.3390/s21103500
37. Jin L, Shi F, Chun Q, et al. Artificial intelligence neuropathologist for glioma classification using deep learning on hematoxylin and eosin stained slide images and molecular markers. *Neuro Oncol*. 2020;23(1): 44-52. doi:10.1093/neuonc/noaa163
38. Jose L, Liu S, Russo C, et al. Artificial intelligence-assisted classification of gliomas using whole-slide images. *Arch Pathol Lab Med*. 2022; 147(8):916-924. doi:10.5858/arpa.2021-0518-OA
39. Jungo P, Hewer E. Code-free machine learning for classification of central nervous system histopathology images. *J Neuropathol Exp Neurol*. 2023;82(3):221-230. doi:10.1093/jnen/nlac131
40. Bukhari SUK, Bokhari SKA, Syed A, Syed SH, Syed UA, Shah SSH. The diagnostic accuracy of convolutional neural network architectures for the diagnosis of brain cancer. *Pak J Med Health Sci*. 2020; 14(3):1037-1039.
41. Kostopoulos S, Konstandinou C, Sidiropoulos K, et al. Assessing the performance of four different categories of histological criteria in brain tumours grading by means of a computer-aided diagnosis image analysis system. *J Microsc*. 2015;260(1):37-46. doi:10.1111/jmi.12264
42. Kurc T, Bakas S, Ren X, et al. Segmentation and classification in digital pathology for glioma research: challenges and deep learning approaches. *Front Neurosci*. 2020;14:27. doi:10.3389/fnins.2020.00027
43. Ma Y, Shi F, Sun T, et al. Histopathological auxiliary system for brain tumour (HAS-Bt) based on weakly supervised learning using a WHO CNS5-style pipeline. *J Neurooncol*. 2023;163(1):71-82. doi:10.1007/s11060-023-04306-6
44. Mohan G, Monica SM. Intelligent framework for brain tumor grading using advanced feature analysis. *Comput Meth Biomechanics Biomed Eng Imag Visual*. 2022;11(3):485-503. doi:10.1080/21681163.2022.2083017
45. Qiu L, Zhao L, Hou R, 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*. 2023;104:102176. doi:10.1016/j.compmedimag.2022.102176
46. Qureshi H, Sertel O, Rajpoot N, Wilson R, Gurcan M. Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification. *Medical image computing and computer-assisted intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2008; 11(Pt 2).
47. Wang X, Wang D, Yao Z, et al. Machine learning models for multiparametric glioma grading with quantitative result interpretations. *Front Neurosci*. 2019;13(JAN). doi:10.3389/fnins.2018.01046
48. Mousavi HS, Monga V, Rao G, Rao AU. Automated discrimination of lower and higher grade gliomas based on histopathological image analysis. *J Pathol Inform*. 2015;6(1):15. doi:10.4103/2153-3539.153914
49. Reza SM, Iftekharuddin KM. Glioma grading using cell nuclei morphologic features in digital pathology images. *Proc SPIE Int Soc Opt Eng*. 2016;9785.
50. Hou L, Samaras D, Kurc TM, Gao Y, Davis JE, Saltz JH. Patch-based convolutional neural network for whole slide tissue image classification. *Proc IEEE Comput Soc Conf Comput vis Pattern Recognit*. 2016; 2016:2424-2433.
51. Das D, Mahanta LB, Ahmed S, Baishya BK, Haque I. Study on contribution of biological interpretable and computer-aided features towards the classification of childhood Medulloblastoma cells. *J Med Syst*. 2018;42(8):151. doi:10.1007/s10916-018-1008-4
52. Das D, Mahanta LB, Ahmed S, Baishya BK. Classification of childhood medulloblastoma into WHO-defined multiple subtypes based on textural analysis. *J Microsc*. 2020;279(1):26-38. doi:10.1111/jmi.12893

53. Galaro J, Judkins AR, Ellison D, Baccon J, Madabhushi A. An integrated texton and bag of words classifier for identifying anaplastic medulloblastomas. *Annu Int Conf IEEE Eng Med Biol Soc*. 2011;2011:3443-3446. doi:10.1109/IEMBS.2011.6090931
54. Yonekura A, Kawanaka H, Prasath VBS, Aronow BJ, Takase H. Automatic disease stage classification of glioblastoma multiforme histopathological images using deep convolutional neural network. *Biomed Eng Lett*. 2018;8(3):321-327. doi:10.1007/s13534-018-0077-0
55. Truong AH, 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. *Neurooncol Adv*. 2020;2(1):vdaa110. doi:10.1093/noajnl/vdaa110
56. Su F, Cheng Y, Chang L, et al. Annotation-free glioma grading from pathological images using ensemble deep learning. *Heliyon*. 2023;9(3):e14654. doi:10.1016/j.heliyon.2023.e14654
57. Alzoubi I, Bao G, Zhang R, et al. An open-source AI framework for the analysis of single cells in whole-slide images with a note on CD276 in glioblastoma. *Cancer*. 2022;14(14):3441. doi:10.3390/cancers14143441
58. Lee MKI, Rabindranath M, Faust K, et al. Compound computer vision workflow for efficient and automated immunohistochemical analysis of whole slide images. *J Clin Pathol*. 2022;76(7):480-485. doi:10.1136/jclinpath-2021-208020
59. Lopez XM, Debeir O, Maris C, et al. Clustering methods applied in the detection of Ki67 hot-spots in whole tumor slide images: an efficient way to characterize heterogeneous tissue-based biomarkers. *Cytometry Part A: J Int Soc Anal Cytol*. 2012;81(9):765-775. doi:10.1002/cyto.a.22085
60. Sikpa D, Fouquet JP, Lebel R, Diamandis P, Richer M, Lepage M. Automated detection and quantification of breast cancer brain metastases in an animal model using democratized machine learning tools. *Sci Rep*. 2019;9(1):17333. doi:10.1038/s41598-019-53911-x
61. Swiderska-Chadaj Z, Markiewicz T, Grala B, Lorent M. Content-based analysis of Ki-67 stained meningioma specimens for automatic hot-spot selection. *Diagn Pathol*. 2016;11(1):93. doi:10.1186/s13000-016-0546-7
62. Wirjadi O, Kim YJ, Stech F, Bonfert L, Wagner M. Bayesian model for detection and classification of meningioma nuclei in microscopic images. *J Microsc*. 2017;265(2):159-168. doi:10.1111/jmi.12471
63. Kong J, Cooper L, Kurc T, Brat D, Saltz J. Towards building computerized image analysis framework for nucleus discrimination in microscopy images of diffuse glioma. *Conference proceedings: … Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*. 2011.
64. Nalisnik M, Amgad M, Lee S, et al. Interactive phenotyping of large-scale histology imaging data with HistomicsML. *Sci Rep*. 2017;7(1):14588. doi:10.1038/s41598-017-15092-3
65. Xing F, Xie Y, Yang L. An automatic learning-based framework for robust nucleus segmentation. *IEEE Trans Med Imaging*. 2016;35(2):550-566. doi:10.1109/TMI.2015.2481436
66. Cooper LAD, Kong J, Gutman DA, et al. An integrative approach for in silico glioma research. *IEEE Trans Biomed Eng*. 2010;57(10):2617-2621. doi:10.1109/TBME.2010.2060338
67. Liechty B, Xu Z, Zhang Z, et al. Machine learning can aid in prediction of IDH mutation from H&E-stained histology slides in infiltrating gliomas. *Sci Rep*. 2022;12(1):22623. doi:10.1038/s41598-022-26170-6
68. Liu S, Shah Z, Sav A, et al. Isocitrate dehydrogenase (IDH) status prediction in histopathology images of gliomas using deep learning. *Sci Rep*. 2020;10(1):7733. doi:10.1038/s41598-020-64588-y
69. Saldanha OL, Loeffler CML, Niehues JM, et al. Self-supervised attention-based deep learning for pan-cancer mutation prediction from histopathology. *NPJ Precision Oncology*. 2023;7(1):35. doi:10.1038/s41698-023-00365-0
70. Decaestecker C, Salmon I, Camby I, et al. Identification of high versus lower risk clinical subgroups in a group of adult patients with supratentorial anaplastic astrocytomas. *J Neuropathol Exp Neurol*. 1995;54(3):371-384. doi:10.1097/00005072-199505000-00010
71. Rathore S, Iftikhar M, Nasrallah M, Gurcan M, Rajpoot N, Mourelatos Z. Prediction of overall survival, and molecular markers in gliomas via analysis of digital pathology images using deep learning. *Neuro Oncol*. 2019;21(Supplement 6):vi270. doi:10.1093/neuonc/noz175.1134
72. Rathore S, Chaddad A, Iftikhar MA, Bilello M, Abdulkadir A. Combining MRI and histologic imaging features for predicting overall survival in patients with glioma. *Radiol Imag Cancer*. 2021;3(4):e200108. doi:10.1148/rycan.2021200108
73. Steyaert S, Qiu YL, Zheng Y, Mukherjee P, Vogel H, Gevaert O. Multimodal data fusion of adult and pediatric brain tumors with deep learning. *medRxiv*. 2022.
74. Zadeh Shirazi A, Fornaciari E, Bagherian NS, Ebert LM, Koszyca B, Gomez GA. DeepSurvNet: deep survival convolutional network for brain cancer survival rate classification based on histopathological images. *Med Biol Eng Comput*. 2020;58(5):1031-1045. doi:10.1007/s11517-020-02147-3
75. Powell RT, Olar A, Narang S, 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*. 2017;8(1):9. doi:10.4103/jpi.jpi_43_16
76. Chen RJ, Lu MY, Wang J, et al. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Trans Med Imaging*. 2022;41(4):757-770. doi:10.1109/TMI.2020.3021387
77. Hao J, Kosaraju SC, Tsaku NZ, Song DH, Kang M. PAGE-net: interpretable and integrative deep learning for survival analysis using histopathological images and genomic data. *Pacific Symp Biocomput Pac Symp Biocomput*. 2020;25.
78. 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*. 2023;14. doi:10.3389/fneur.2023.1100933
79. Mobadersany P, Yousefi S, Amgad M, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci U S a*. 2018;115(13):E2970-E2979. doi:10.1073/pnas.1717139115
80. Levine AB, Peng J, Farnell D, et al. Synthesis of diagnostic quality cancer pathology images by generative adversarial networks. *J Pathol*. 2020;252(2):178-188. doi:10.1002/path.5509
81. Ozyoruk KB, Can S, Darbaz B, et al. A deep-learning model for transforming the style of tissue images from cryosectioned to formalin-fixed and paraffin-embedded. *Nature Biomed Eng*. 2022;6(12):1407-1419. doi:10.1038/s41551-022-00952-9
82. Xu Y, Jia Z, Wang L-B, et al. Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. *BMC Bioinform*. 2017;18(1):281. doi:10.1186/s12859-017-1685-x
83. Wang W, Zhao Y, Teng L, et al. Neuropathologist-level integrated classification of adult-type diffuse gliomas using deep learning from whole-slide pathological images. *Nat Commun*. 2023;14(1):6359. doi:10.1038/s41467-023-41195-9
84. Maas SLN, Stichel D, Hielscher T, et al. Integrated molecular-morphologic meningioma classification: a multicenter retrospective analysis, retrospectively and prospectively validated. *J Clin Oncol*. 2021;39(34):3839-3852. doi:10.1200/JCO.21.00784
85. Park J, Jang BG, Kim YW, et al. A prospective validation and observer performance study of a deep learning algorithm for pathologic diagnosis of gastric tumors in endoscopic biopsies. *Clin Cancer Res*. 2021;27(3):719-728. doi:10.1158/1078-0432.CCR-20-3159

86. Iizuka O, Kanavati F, Kato K, Rambeau M, Arihiro K, Tsuneki M. Deep learning models for histopathological classification of gastric and colonic epithelial Tumours. *Sci Rep*. 2020;10(1):1504. doi:10.1038/s41598-020-58467-9
87. Oxford TUO. Prostate cancer AI diagnosis tool begins evaluation in Oxford. 2023; https://www.ox.ac.uk/news/2023-03-10-prostate-cancer-ai-diagnosis-tool-begins-evaluation-oxford. Accessed 6 September, 2023.
88. Shu XJ, Chang H, Wang Q, et al. Deep learning model-based approach for preoperative prediction of Ki67 labeling index status in a noninvasive way using magnetic resonance images: a single-center study. *Clin Neurol Neurosurg*. 2022;219:107301. doi:10.1016/j.clineuro.2022.107301
89. Chen R, Smith-Cohn M, Cohen AL, Colman H. Glioma subclassifications and their clinical significance. *Neurotherapeutics*. 2017;14(2):284-297. doi:10.1007/s13311-017-0519-x
90. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med*. 2022;28(5):924-933. doi:10.1038/s41591-022-01772-9
91. Finlayson SG, Subbaswamy A, Singh K, et al. The clinician and dataset shift in artificial intelligence. *N Engl J Med*. 2021;385(3):283-286. doi:10.1056/NEJMc2104626
92. Homeyer A, Geissler C, Schwen LO, et al. Recommendations on compiling test datasets for evaluating artificial intelligence solutions in pathology. *Mod Pathol*. 2022;35(12):1759-1769. doi:10.1038/s41379-022-01147-y
93. La Rosa S. Diagnostic, prognostic, and predictive role of Ki67 proliferative index in neuroendocrine and endocrine neoplasms: past, present, and future. *Endocr Pathol*. 2023;34(1):79-97. doi:10.1007/s12022-023-09755-3
94. Balkenhol MCA, Tellez D, Vreuls W, et al. Deep learning assisted mitotic counting for breast cancer. *Lab Invest*. 2019;99(11):1596-1606. doi:10.1038/s41374-019-0275-0
95. Feng M, Deng Y, Yang L, et al. Automated quantitative analysis of Ki-67 staining and HE images recognition and registration based on whole tissue sections in breast carcinoma. *Diagn Pathol*. 2020;15(1):65. doi:10.1186/s13000-020-00957-5
96. Pickles JC, Fairchild AR, Stone TJ, et al. DNA methylation-based profiling for paediatric CNS tumour diagnosis and treatment: a population-based study. *Lancet Child Adolesc Health*. 2020;4(2):121-130. doi:10.1016/S2352-4642(19)30342-6
97. Beck AH, Sangoi AR, Leung S, et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci Transl Med*. 2011;3(108):108ra113(108). doi:10.1126/scitranslmed.3002564
98. Bychkov D, Linder N, Turkki R, et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep*. 2018;8(1):3395. doi:10.1038/s41598-018-21758-3
99. Woerl AC, Eckstein M, Geiger J, et al. Deep learning predicts molecular subtype of muscle-invasive bladder cancer from conventional histopathological slides. *Eur Urol*. 2020;78(2):256-264. doi:10.1016/j.eururo.2020.04.023
100. Maier-Hein L, Reinke A, Godau P, et al. Metrics reloaded: Recommendations for image analysis validation. 2022:arXiv:2206.01653. https://ui.adsabs.harvard.edu/abs/2022arXiv220601653M. Accessed June 01, 2022.
101. Howard FM, Dolezal J, Kochanny S, et al. The impact of site-specific digital histology signatures on deep learning model accuracy and bias. *Nat Commun*. 2021;12(1):4423. doi:10.1038/s41467-021-24698-1
102. Roetzer-Pejrimovsky T, Moser AC, Atli B, et al. The digital brain tumour atlas, an open histopathology resource. *Sci Data*. 2022;9(1):55. doi:10.1038/s41597-022-01157-0

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Jensen MP, Qiang Z, Khan DZ, et al. Artificial intelligence in histopathological image analysis of central nervous system tumours: A systematic review. *Neuropathol Appl Neurobiol*. 2024;50(3):e12981. doi:10.1111/nan.12981