Computers in Biology and Medicine 186 (2025) 109642

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Brain tumour histopathology through the lens of deep learning: A systematic review

Chun Kiet Vong [a,b], Alan Wang [a,b,d], Mike Dragunow [b,c], Thomas I-H. Park [b,c], Vickie Shim [a,*]

[a] Auckland Bioengineering Institute, The University of Auckland, New Zealand
[b] Centre for Brain Research, The University of Auckland, New Zealand
[c] Department of Pharmacology, The Faculty of Medical and Health Sciences, The University of Auckland, New Zealand
[d] Faculty of Medical and Health Sciences, The University of Auckland, New Zealand

ARTICLE INFO

Keywords:
Glioblastoma
Machine learning
Deep learning
Review
Hematoxylin and eosin
Bioinformatics

1 ABSTRACT

Problem: Machine learning (ML)/Deep learning (DL) techniques have been evolving to solve more complex diseases, but it has been used relatively little in Glioblastoma (GBM) histopathological studies, which could benefit greatly due to the disease's complex pathogenesis.
Aim: Conduct a systematic review to investigate how ML/DL techniques have influenced the progression of brain tumour histopathological research, particularly in GBM.
Methods: 54 eligible studies were collected from the PubMed and ScienceDirect databases, and their information about the types of brain tumour/s used, types of -omics data used with histopathological data, origins of the data, types of ML/DL and its training and evaluation methodologies, and the ML/DL task it was set to perform in the study were extracted to inform us of trends in GBM-related ML/DL-based research.
Results: Only 8 GBM-related studies in the eligible utilised ML/DL methodologies to gain deeper insights into GBM pathogenesis by contextualising histological data with -omics data. However, we report that these studies have been published more recently. The most popular ML/DL models used in GBM-related research are the SVM classifier and ResNet-based CNN architecture. Still, a considerable number of studies failed to state training and evaluative methodologies clearly.
Conclusion: There is a growing trend towards using ML/DL approaches to uncover relationships between biological and histopathological data to bring new insights into GBM, thus pushing GBM research forward. Much work still needs to be done to properly report the ML/DL methodologies to showcase the models' robustness and generalizability and ensure the models are reproducible.

1. Introduction

Glioblastoma (GBM) is the most common malignant brain tumour [1]. GBM is almost invariably fatal, with the patient's median overall survival being approximately one year [2]. The standard care for many newly diagnosed GBM patients involves surgical tumour resection and radio-chemotherapy [3]. However, GBM tumours often recur in functional brain areas and are less sensitive to therapy than the original tumour, thus preventing another surgical resection [4]. A distinct feature of GBM is its heterogeneity, which plays a significant role in the difficulty of treating GBM [5,6]. Therefore, much research has investigated the nature of the heterogeneity of GBM and its contribution to treatment difficulties and fatality.

The consensus amongst The Cancer Genome Atlas (TCGA) Research Network and others is that GBM can be categorised into four subtypes: proneural, neural, classical and mesenchymal [7,8]. Verhaak and colleagues noticed that subtypes could be differentiated using distinct molecular signatures. They reported that 100 % of the classical subtype exhibited a chromosome 7 amplification paired with chromosome 10 loss, with 97 % of the classical subtype having high levels of epidermal growth factor receptor (EGFR) gene amplifications [8] Verhaak and colleagues further noted that mesenchymal subtype predominantly featured lower NF1 expression levels, while proneural subtypes had alterations platelet-derived growth factor alpha (PDGFRA) receptor and isocitrate dehydrogenase (IDH) 1 as their significant features [8] This shows that although GBM is heterogenous, patterns were observed.

* Corresponding author.
E-mail address: v.shim@auckland.ac.nz (V. Shim).

https://doi.org/10.1016/j.compbiomed.2024.109642
Received 2 July 2024; Received in revised form 26 December 2024; Accepted 27 December 2024
Available online 8 January 2025

Based on this, potential treatments can be more effective and personalised to improve patient outcomes.

More recently, the World Health Organization (WHO) broadly classified GBM into two types: IDH-wildtype (wt) GBM and IDH-mutant (mt) GBM [9]. Louis and colleagues further noted GBM variants under the two classifications mentioned above. They classified epithelioid GBM (Ep-GBM) under the IDH-wt group. It featured large epithelioid cells with abundant eosinophilic cytoplasm, vesicular nuclei and large melanoma-like nucleoli, all of which can be distinguished histopathologically [9]. Louis and colleagues further categorised GBM with primitive neuronal component (GBM-PNC) to have highly cellular nodules in an otherwise diffuse astrocytoma in MRI screenings. They can be histopathologically identified by the Homer Wright rosettes along with large cell/anaplastic features [9]. Small cell GBM can be characterised by uniform, small neoplastic cells and often exhibiting EGFR amplifications, while granular cell GBM exhibits granular to macrophage-like, lysosomic-rich tumour cells [9]. Overall, the histopathological/morphological classification of GBM types and the shared genes identified give rise to the potential of correlating image data and molecular markers in GBM for better diagnosis and prognosis.

Recent advancements in machine learning (ML) and deep learning (DL) have diversified its use, particularly in clinical use. ML/DL can extract patterns from examples at its core, and it has proven to be more beneficial with more enormous datasets, especially for more intricate patterns, which are usually too complex for human analyses. In the field of brain tumour research, many of the ML/DL applications have revolved around the segmentation and volumetric analyses of brain tumours from MRI and CT screenings [10–13]. This was often in the form of segmentation of the tumour volume in MRI images using various convolutional neural networks (CNN) [14,15]. Classification tasks and improving accuracy are also common objectives in ML/DL-based MRI research. Studies would often develop novel pipelines and asses their accuracies in classifying various types of brain tumours [16,17]. These tasks would otherwise be incredibly difficult to achieve without the ML/DL's capabilities to recognise underlying patterns that are too complex for humans to understand.

Emerging research has also utilised ML/DL tools to recognise patterns in histopathological images, either as a segmentation tool or to classify the tumours [18,19]. In segmentation, CNNs are standard tools used to segment tumour features [19–22]. ML/DL has also been used to identify and correlate patterns in brain tumour histopathologic images with their corresponding molecular statuses. Notably, Cui and colleagues extracted features from histopathology images of glioma patients and correlated them with their IDH mutation statuses via multiple-instance learning [18]. In another example, Fatima and colleagues extracted features from segmented nuclei of meningioma histology images to classify different subtypes of meningioma using a support vector machine (SVM) ensemble ML model [23].

It is evident that in more recent brain tumour histopathological studies, ML/DL has been used more often. However, our recent literature survey suggests that relatively few works consolidate bio-information with image data, and fewer pieces of literature focused their investigations on GBM histopathology. This is important because GBM diagnoses are primarily based on expensive multi-omics assays on tissue biopsies [9]. These are more closely related to cheaper and affordable histopathological work than relatively more expensive and less accessible MRI/CT scans. Therefore, it is logical to infer that the contextualisation of bio-information acquired from biopsies and state-of-the-art research using ML/DL methods would result in a better understanding of GBM's microenvironment, as well as the development of accessible and standardised diagnostic and prognostic tools for GBM. Thus, a knowledge gap exists that utilises ML/DL to meaningfully combine GBM bio-information and its corresponding histopathology data. We propose that a review of ML/DL utilisation in brain tumour research as a whole can inform us about what and how we can improve multi-scale ML/DL-based GBM research.

Therefore, this systematic review consolidates the various ML/DL models and their purpose in brain tumour histopathological research. The primary contributions of this review are as follows.

- Conduct a meta-analysis of various ML/DL-based studies on brain tumours to investigate the current trend of ML/DL use in brain tumour histopathological studies. From the current trends, identify the common ML/DL models with a focus on the task they perform in brain tumour histopathology studies. Uncovering their role in understanding brain tumour pathology is key here.
- Suggest novel ML/DL utilisation avenues in GBM histopathology studies based on the current trends and gaps identified in the meta-analysis.

The remainder of the article is split into four sections. The methodology section (Section 3) describes how the articles were searched, filtered and analysed. Section 4 consists of the result section, where trends in types of brain tumours used, types of datasets and dataset availability, and the ML/DL models together with their purpose in each study were reported. Section 5 discusses the findings of the results, while Section 6 concludes the article.

2. Methods

2.1. Search strategy, eligibility screening and data extraction

This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [24]. Original articles were sourced from the PubMed and Science Direct electronic databases published up to March 2022 (from 2000-) to review the recent advancements in the field. After the initial search, duplicated articles were identified and removed manually. Reviews, non-English articles and conference articles were excluded. The three main categorical keywords used to identify relevant studies to this systematic review were: 1.) Histology or histopathology, 2.) Machine learning or deep learning (DL), and 3.) Brain tumour or glioblastoma. The studies that include the three main categorical keywords were screened using Boolean operations in the databases' advanced search function. Table 1 details the search parameters of PubMed and Science Direct.

The studies were further screened for eligibility in this systematic review. During the initial search, studies involving MRI and CT with minimal to no histology work were included in the search results, which were beyond the scope of this review. Thus, an additional NOT Boolean operation was implemented in the search to exclude non-relevant MRI/CT studies, along with the inclusion of the "histology OR histopathology" search terms (Table 1). Furthermore, the glioblastoma search term was returning very sparse results. Hence, an additional brain tumour search term with the OR operator was added to find more relevant studies (Table 1). The "segmentation OR classification" and "deep learning OR machine learning" search terms were included to constrain the studies towards ML/DL-based brain tumour research. Then, a further manual screening was applied to remove the remaining non-relevant

Table 1
Search terms used in the advanced search feature from ScienceDirect and PubMed databases.

Database	Search
ScienceDirect	("brain tumour" OR "glioblastoma") AND ("segmentation" OR "classification) AND ("histology" OR "histopathology") AND ("deep learning" OR "machine learning") NOT "magnetic resonance imaging"
PubMed	((glioblastoma) OR (glioma) OR (brain tumour)) AND ((segmentation) OR (classification)) AND ((histology) OR (histopathology)) AND ((machine learning) OR (neural network)) NOT ((magnetic resonance) OR (computer tomography) OR (positron emission))

studies that did not include all 3 of the categorical keywords.

2.2. Risk of bias assessment

The screened articles were assessed for their methodological quality by conducting a risk of bias assessment. Based on the QualSyst Assessment Tool for quantitative studies [25], ten questions were asked for each of the screened articles within the scope of this review. The questions were as follows: 1.) Was a clear objective stated? 2.) Was the study design appropriate to answer the study question/goal? 3.) Was the subject selection or information source described appropriate? 4.) Were subject characteristics or input variables clearly described and appropriate? 5.) Was the methodology and outcome clearly described? 6.) Was there an appropriate sample/training size for the study? 7.) Were there analyses and validation appropriate for the study? 8.) Was there an estimate of variance, if applicable? 9.) Were the result details appropriate and sufficient? 10.) Did the result support the conclusion made in the study? The questions were answered either yes, partial or no, denoted under the scores of 2,1 and 0, respectively. The scores across the ten questions were summated to acquire the summary score, and the final score was calculated by dividing the summary score by the highest possible score for the study. Articles with a score of 0.78 or higher were deemed eligible for data extraction. The author, C.V., initially assessed the risk, which was then cross-checked with the corresponding author, V.S.

2.3. Data extraction and analyses

After the quality assessment, specific data was extracted from each study. ML or DL methods used in these studies were identified with a particular focus on how they were used in relation to histopathological work. Five data criteria were extracted from each article and investigated: 1.) The type of brain tumour used in the study, 2.) What histopathological type, and if applicable, what combination of other data types and the histopathological data, were used in the study? 3.) Where the data was collected 4.) The type of ML/DL tools, and its training and evaluative methodologies conducted in the study, 5.) How were the ML/DL tools used to achieve the study's research goals? These first and third criteria were developed to understand the landscape of brain tumour studies and the available data, while the three other criteria ascertain the current trends in brain tumour histopathological studies and how ML/DL is utilised in brain tumour studies. Within the fourth criterion, ML/DL training and testing methodologies, along with its evaluation methodologies (when applicable), were collected to investigate the reproducibility of the studies' models and to observe the trends in recent ML/DL model design. The resultant extracted criteria would highlight the innovative trends that can push ML/DL-based brain tumour research in a new direction.

3. Results

3.1. Search results

A total of 388 articles were initially retrieved from the search. Of the 388 articles, 324 articles were deemed irrelevant for this systematic review and 8 duplicate articles were removed. Interestingly, many irrelevant studies used MRI/CT-based datasets. 2 articles were removed as they failed to meet our risk of bias assessment criteria (Supplementary Table 1). Notably, most articles included in this systematic review were published in recent years (Fig. 1). This is indicative of the nascent nature of this scope of research.

As the scope of this review is brain tumour studies that used ML/DL techniques in analyzing histopathological work for integration with available bioinformation, 54 eligible articles were selected (Fig. 2). We identified the following features – 1.) brain tumour types, 2.) image stain types, 3.) databases used to obtain histopathology work or additional data, 4.) how ML/DL methods were used. Tables 2 and 3 detail the culmination of the analyses completed on the 54 articles, with Table 2 providing a general summary of how the ML/DL tools were utilised in the group, while Table 3 provides a detailed overview of the data sources and ML/DL training and evaluation methodologies.

3.2. Brain tumour types and data sources

The purpose of the collection and analysis of types of brain tumours and the data sources used in the 54 studies is to give us an insight into the landscape of which brain tumour is more studied and the data availability of the studies, respectively. To this end, each study noted the main types of tumours used and which data centres or hospitals they collected their specimens or data. From the review, 37 studies were reported to be GBM-related, while the remaining 17 were not GBM-related. Of the GBM-related studies, 12 focused on GBM only, while 18 focused on gliomas including GBM, and the remaining 7 studies either studied human-derived GBM cell lines or studied various brain tumours including GBM (Fig. 3). In GBM-unrelated studies, the most common brain tumour studied was gliomas, excluding GBM (6 studies;

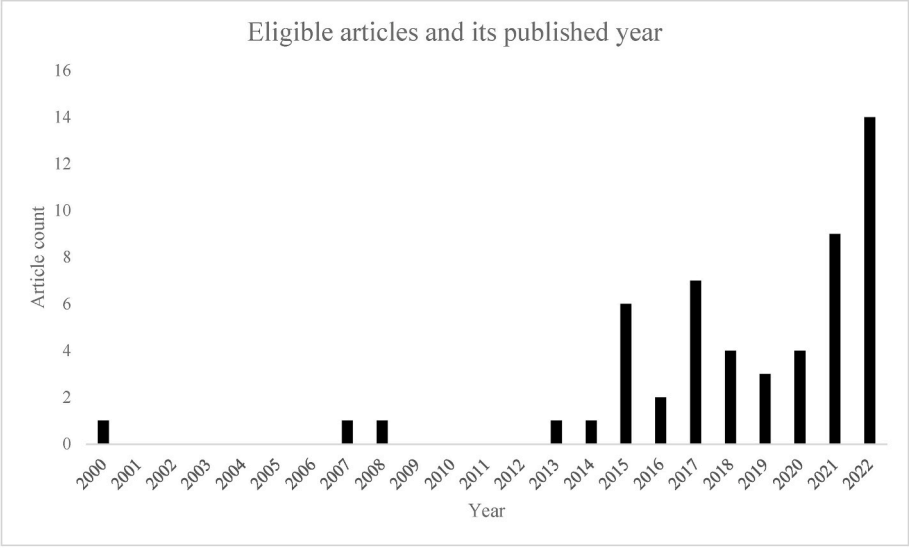

Fig. 1. Histogram of final selection of articles sorted based on the year it is publish from ScienceDirect and PubMed.

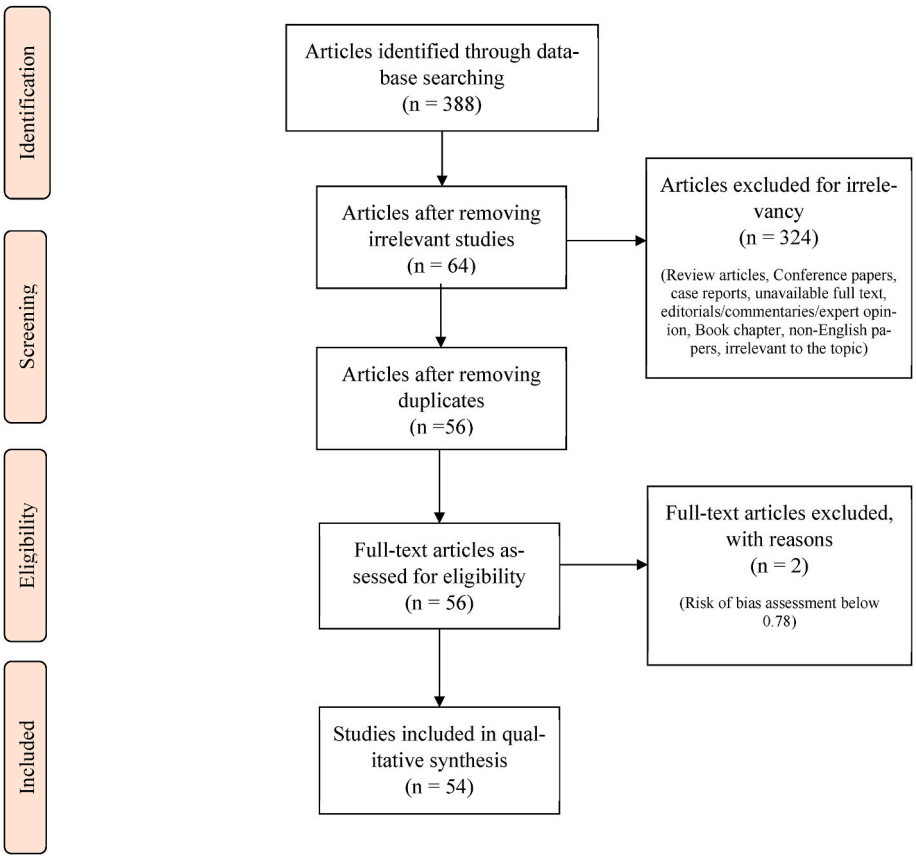

Fig. 2. Flowchart of how the articles from the search was screened for relevancy and assessed for eligibility.

Fig. 3), followed by meningioma brain tumours (Fig. 3). This indicates that much ML/DL-histopathological research focuses on GBM and LGG. Interestingly, the surveyed glioma-GBM studies tended to compare the LGG against GBM tumour types against each other. For instance, Barker and colleagues developed a pipeline to accurately distinguish between GBM and LGG histopathology slides [26]. What is also interesting to note is that although GBM has been a keen area of study in this systematic review in the past two decades, the general consensus is that the median overall survival rate of patients has only improved from approximately 6 months–10 months in the last two decades [27,28]. This indicates that despite the keen interest in studying GBM, our understanding of the disease remains nascent and inadequate to improve patient outcomes.

Looking at the datasets used in all studies in our analysis, we report that The Cancer Genome Atlas, a publicly available dataset, is the most used public dataset in our review (21 articles; Fig. 4). We report that the other two popular publicly available datasets used were the Computational Precision Medicine (CPM) dataset and the Medical Image Computing and Computer Assisted Intervention (MICCAI) dataset, with 3 articles and 2 articles reported to use these datasets, respectively (Fig. 4). However, 25 articles used non-public datasets (data collected from hospitals, research centres or in-house cell lines) in their studies (Fig. 4). In GBM-related studies, a similar trend is observed in the ratio between the most popular publicly available dataset and the non-public datasets used for studies. It is concerning to note that 4 of the 54 articles did not clearly specify where their data was collected from (Fig. 4), which would prove to be difficult for others to replicate its results. Overall, there is a clear interest in using publicly available datasets in brain tumour-related studies, but a similar amount of interest in curating the data from research centres, hospitals or cell lines.

3.3. Stain types and analyses

This section explores primarily why the ML/DL-based tools were used in the studies. To understand why, we first report what type of histopathological data they are working with and whether they were integrated with other data types (e.g. -omics data, patient data, clinical diagnostic data) in their studies' ML/DL models. We also investigated how these data were analysed, categorising them as studies that only segment or processed images: Studies that primarily predict patient survival and tumour grades using only histopathological data are categorised under typical analysis, whilst studies that integrate histopathological data with other data types (i.e. additional -omics data) for their analysis were categorised as unique analysis. The results from this section give us insight into what types of data are being combined with histopathology in ML/DL-based brain tumour research and what novel ways the data are being combined to enhance brain tumour research.

The eligible studies used hematoxylin and eosin (H&E) stained images as their data type, with 41 of the 54 studies combining various other data types with H&E data for more complex analyses and investigations (Fig. 5). Interestingly, 3 studies utilised DAPI stained images, stimulated Raman histology (SRH) and immunohistochemical (IHC) stained images, and it was included in the relevant studies as we believe that the methodologies in these studies are similar enough to other H&E-based studies and can be transferable to H&E-based applications. Orzan and colleagues utilised IHC staining with their respective transcriptomic data to train their random forest (RF) classifier to discriminate between different GBM subtypes (proneural, mesenchymal and classical). Wollman and colleagues used nuclei segmentations from DAPI-stained images to train their model to accurately segment nuclei [19]. Hollon and colleagues trained their model to classify and segment SRH images of brain tumours; it was able to accurately classify various types of brain tumours, including GBM, diffuse LGG subtypes (oligodendroglioma and

Table 2
Table summarising how ML/DL tools were used to achieve the task/s of each eligible paper in the systematic review.

Authors	Publication Date (Year)	Brain Tumour specified	Data type	Utilisation of ML/DL
[43].	2007	Meningioma (Meningotheliomatous, Fibroblastic, Transitional, Psammomatous)	Hematoxylin and eosin (H&E) images and clinical data (tumour subtypes)	Authors used SOM to cluster and classify meningioma subtypes. They utilised a class separation measure to show that a subset of features can accurately distinguish the four meningioma subtypes.
[30]	2022	Glioblastoma (GBM) and Lower-Grade Glioma (LGG)	H&E images, genomic and clinical data (survival data and tumour subtype)	presented a multi-modal fusion framework based on multi-task correlation learning (MultiCoFusion) for survival analysis and cancer grade classification using whole slide images (WSIs) (trained under the ResNet-152 model) and mRNA expression (trained under SCGN) data. As a result, the multi-modal framework was able to learn better representations than traditional feature extraction models for accurate survival prediction.
[66]	2015	Meningioma (Fibroblastic, Meningothelial, Psammomatous, Transitional)	H&E images and clinical data (meningioma subtypes)	The H&E image patch features were used to train their custom fractal model and extract optimal features for meningioma classification. Once the optimal features were extracted, the SVM, Bayesian and kNN algorithms were used to classify them with their known subtypes. The fractal-model-design-based SVM, Bayesian, and kNN classifiers achieved an overall classification accuracy of 94.12 %, 92.50 % and 79.70 %, respectively, which outperformed classical energy-based selection approaches (86.31 %, 83.19 % and 51.63 % for the co-occurrence matrix, and 76.01 %, 73.50 % and 50.69 % for the energy texture signatures; respectively).
[67]	2021	GBM	H&E and Ki-67-stained histological images	Propose an unsupervised tissue cluster level graph cut (TisCut) method for segmenting histological images into meaningful compartments. TisCut could segment necrotic sections accurately in the dataset, with comparable performance to trained models such as TL-InceptionV3, U-Net, U-Net-MA, and U-Net-AS.
[26]	2016	GBM and LGG	H&E images and clinical data (tumour subtype)	Extraction of spatially localised features of shape, colour and texture from tiled regions covering the slide. The tiles were clustered based on their features via Kmeans++ algorithm. The various feature types from the clusters were normalized and passed through the Elastic Net linear regression model to classify the clusters, and weighted voting aggregate the final decision value (GBM vs. LGG) based on the classification of the clusters. Their method was evaluated on 203 brain cancer cases with an accuracy of 93.1 % and achieved a 100 % accuracy in the 2014 MICCAI Pathology Classification Challenge.
[68]	2008	Astrocytomas (low-grade and high-grade)	H&E images and clinical data (tumour subtype)	Extracted image features were fed into FCM to grade tumours (low vs. high). The FCM was designed based on the histopathologists' expertise in the Department of Pathology, University Hospital of Patras, Greece, which was then used to accurately grade the brain tumours at the time (90.26 % and 93.22 % accuracy on low and high grade, respectively).
[69]	2022	Gliomas (with focus on IDH mutation status)	H&E images	Used the dataset to train the CAGAN to normalise stains. The aim is to create a network that can consistently normalise stains so that further training of images for other purposes may be more effective. They achieved better performance of benchmark algorithms by 5–10 % compared to baselines not using normalisation.
[70]	2021	GBM and LGG	H&E images	present a lightweight slice-wise CS-Net building on novel hierarchical dimension-decomposed (HDD) convolutions and a novel instance-aware loss for 2D and 3D microscopy image segmentation. In the case of the brain tumour images, they were utilised to train the network for nuclei segmentation. As a result, the network's nuclei segmentation was able to outperform some of the more common models (including variations of U-Net, HoVer-Net, Mask-RCNN and DCAN)
[52]	2017	GBM and LGG	H&E images	Output accurate probability maps of histological objects but also depict clear contours simultaneously for separating clustered object instances, further boosting the segmentation performance. GBM and LGG images and ground truths were used to train and

(*continued on next page*)

Table 2 (*continued*)

Authors	Publication Date (Year)	Brain Tumour specified	Data type	Utilisation of ML/DL
				validate the network's performance. The network was able to outperform most other methods in the 2015 MICCAI gland segmentation challenge and the 2015 MICCAI nuclei segmentation challenge
[40]	2022	GBM	H&E images and associated gene data from spatial transcriptomics	Using their trained RESEPT pipeline, they could segment and identify a prominent area in a WSI containing elongated nuclei characteristic of infiltrating GBM cells. These areas matched with the infiltrating tumour marker genes, and RESEPT was able to recognise the tumour, non-tumour and infiltrating tumour architecture.
[55]	2022	GBM and LGG	H&E images and clinical data (tumour subtype)	They developed a new characterisation approach, where evolutionary algorithms were used to generate a small set of features that accurately represent each tissue. The aim was to optimize the output of pre-trained deep networks to mitigate computational bottlenecks in the form of irrelevant/redundant features. GBM and LGG H&E images were used to train the DNN in this pipeline, and the optimised features achieved 93 % classification accuracy.
Komosin'ski et al., 2000 [71]	2000	Various brain tumours including GBM	microscopic images and clinical data (tumour subtype)	Proposed a pipeline that uses the GA and kNN classifiers to select and weight features in the brain tumour images to improve the predictive accuracy (astrocytic vs. glial tumours). The best-achieved classification accuracy exceeded 80 %, which was considered the minimum accuracy needed for medical diagnosis
[72]	2021	Human-derived Glioblastoma stem cells (GSC)	H&E images and biomarker data	The H&E tumour images were used to identify tumour regions, which were then trained in an unspecified ML model using Iba-1, F4/80 and CD45 markers. Incorrectly classified images in the article were removed from the final analysis.
[36]	2017	LGG	H&E images and overall survival (OS), molecular biomarkers and clinical factors data	SVM was used to classify the long and short OS based on the H&E image features via the bag-of-words (BOW) methodology. The BOW methodology used K-means clustering to cluster statistical features extracted from the H&E image patches and created a "dictionary" for the SVM to use and classify. The study showed that it was able to discriminate OS with a predictive area under the receiver operating characteristic curve (AUC) of 0.76 with the dictionary alone, 0.82 when supplemented with molecular markers, and 0.89 when further supplemented with other clinical factors
[46]	2022	GBM and LGG	H&E images	Proposed meta-learning and multi-task learning on the U-Net model. This allowed the model to learn more generalised features and increase its generalisation and generalizability to unseen data. The H&E images were used to train and test the model. Meta-MTL U-Net was able to maintain improved performance in nuclei segmentation over decreasing training samples compared to some of the common models (HoVer-Net and TripleU models)
[73]	2019	GBM and LGG	H&E images	They presented a CNN model, Micro-Net, to segment cells, nuclei and glands in fluorescence and histology images. LGG and GBM H&E images were used to train and evaluate the model. It was shown that it was able to outperform other recent models (U-Net, SAMS-Net, FCN8, etc.) in cell and nuclei segmentation
[74]	2022	GBM and LGG	H&E images and tumour samples for genomics analysis	They developed a MIL model that predicts tumour purity in H&E images. The images in the TCGA brain tumour were used to train and evaluate the model's performance. The model was able to predict the tumour purity in 8 TCGA cohorts successfully, and the predictions correlated very well with the subsequent genomic tumour purity values
[54]	2022	GBM and LGG	H&E images and matching CNV, mutation and RNA-Seq data	Developed the PORPOISE workflow, a weakly supervised MMF DL algorithm to integrate WSI and molecular profile data to perform tasks such as survival analysis. GBM and LGG samples were used as training, but additional LGG samples were used to evaluate their performance. Overall, it was able to accurately predict survival outcomes and discover features that correlate with poor and favorable outcomes.
[42]	2022	Neuroblastoma	H&E images, classification and prognostic evaluation data	To extract their features, they segmented neuroblastoma nuclei via the HoVer-Net model (pre-

(*continued on next page*)

Table 2 (*continued*)

Authors	Publication Date (Year)	Brain Tumour specified	Data type	Utilisation of ML/DL
				trained on the PanNuke Dataset). The nucleus and patient-level features were then fed into the K-means algorithm to learn visual words (i.e., clustering centres). The features were then further optimised before it would predict survival and thus calculate the degree of risk to the pathological prognosis. It achieved an AUC of 0.946 in the training/test dataset and an AUC of 0.938 in the independent validation dataset, thus showing good generalizability.
[75]	2018	Human-derived GSC	H&E, magnetic resonance (MR) and fluorescence images	The software segments images into classes of interest, labelling them into objects of interest. These segmented images were further processed in ImageJ to extract labelled area measurements. The labelled areas were used to investigate the invasive pattern of tumours
[31]	2021	GBM	H&E images and proteomics analysis data	They used the models for 3 different prediction tasks, including G-CIMP phenotype, immune response and telomere length. The images were trained in various DL CNN models to find the best performance for their use case. Once trained, the model was applied to discover histological features associated with G-CIMP phenotype, telomere length and immune response. NMF was used to perform unsupervised clustering of tumour samples and to identify proteogenomic features.
[34]	2022	Medulloblastoma	H&E images and associated clinical and molecular data	Nuclei in the images were segmented via watershed segmentation, and their features were extracted. The top-performing features were fed into SVM classifiers to distinguish molecular subtypes and disease-specific outcomes. The overall pipeline was able to distinguish the different molecular subtypes with an AUC of 0.7, and survival within Group 3 tumours was predicted with an AUC of 0.92. They proposed that this model can be used to study medulloblastoma genetic expression phenotypes as it could distinguish meaningful features of disease pathology
[48].	2022	GBM	H&E images, classification, single-cell data and spatial transcriptomics data	Trained VGG16 with the H&E images and classification (infiltrating, necrosis, necrotic edge, cellular, vascular) ground truths. The model was then used to predict spotwise histological phenotype. Transcriptional data were integrated into the segmented image to corroborate with histologic estimation. This was used to infer the relationships between genotypic data with phenotypic observations.
[76]	2018	LGG	Immunohistochemical (IHC) images	They used an ML algorithm that scores CD31 positivity by segmenting DAB-positive pixels. Once trained, it was used to segment CD31-positive objects (e.g., microvessels) and was summed in each ROI.
[77]	2022	GBM and LGG	H&E images and clinical data (tumour subtype)	The aim is to use ML algorithms such as SNN to distinguish GBM and LGG tumours in WSIs. Features extracted from the segmented nuclei in the H&E images were clustered (100 clusters) via k-means clustering, and the cluster centres were then fed into SVM, kNN and SNN to compare their performance in predicting GBM vs. LGG. The proposed SNN model outperformed kNN and SVM models (97.21 % accuracy compared to 93.04 % and 95.13 %, respectively).
[78]	2017	GBM	H&E images	Incorporated StackedPSD-KSPM and AlexNet-KSPM DL methods for feature extraction and FE-KSPM ML model for three-category classification (tumour, necrosis and transition to necrosis). The H&E image patches were used to train and test the DL models. To this end, the authors found that sparse feature encoders and feature extraction strategies based on DL techniques consistently improved the performance of tissue histology classification.
[79]	2022	GBM	IHC images	A bifocal CNN is used that takes in 2 patches of different sizes, and they are each fed into one of 2 convolutional sub-nets; one serves as a feature concatenation module, and the other is a classification layer. This allowed the trained model to assign the positivity of CD276 expression by "halo cells", and a heat map was produced. It achieved an accuracy of 97.7 %, outperforming other models, such as Resnet-50.

(*continued on next page*)

Table 2 (*continued*)

Authors	Publication Date (Year)	Brain Tumour specified	Data type	Utilisation of ML/DL
[44]	2021	Medulloblastoma	H&E images and clinical data (i.e. subtypes)	Created a pipeline where medulloblastoma images were trained on ResNet-50, MobileNet, and DenseNet-201 to extract features. The discrete wavelet transform (DWT) method enhanced feature extractions by giving spatial-time-frequency representations. This was fed into DCT and PCA methods to fuse and drop redundant features. Finally, features were used to classify medulloblastoma subtypes (normal, classic, desmoplastic, large cell, and nodular) via LDA and ESD classifiers (99.4 % accuracy).
[47]	2021	Medulloblastoma	H&E images and clinical data (i.e. subtypes)	Similar to the MB-AI-His pipeline by the same author. Various CNN models were used to train with pediatric medulloblastoma H&E images for feature extraction, and optimal features were selected to be classified via Bi-LSTM. The classification technique is capable of binary (normal vs. medulloblastoma) and multi-classification of medulloblastoma subtypes (normal, classic, desmoplastic, large cell, and nodular) (100 % and 99.35 % accuracy for binary- and multi-classification respectively).
[35]	2022	GBM	IHC images, overall survival and associated miRNA data	Using their IHC images, they trained the VGG16 network to extract various features optimised for PCA analysis. Together with the associated miRNA and overall survival, the features were classified with their unspecified classifiers to determine low and high-risk GBM groups (AUC = 95 %).
[18]	2020	Gliomas	H&E images and associated IDH status	The tissue images were trained in a MIL CNN model, where the model takes small patches of the images as instances of the bag and calculates their scores. These scores were aggregated to generate the classification result of the IDH1 mutation (positive vs. negative). They were able to achieve an optimum accuracy of around 78 %.
[29]	2015	GBM and LGG	H&E images and clinical data (tumour subtype)	They proposed an ensemble CNN pipeline, where the CNNs have a LeNeT-like architecture. The first CNN will distinguish between GBM and LGG, and any LGG classification results will be fed into the 2nd CNN to determine between tumour grade II or III. They achieved 96 % accuracy in distinguishing GBM and LGG while achieving 71 % accuracy in distinguishing between LGG tumour grades II and III.
[80]	2014	Meningioma (Meningotheliomatous, Fibroblastic, Transitional, Psammomatous)	H&E images and clinical data (tumour subtype)	Authors utilised k-means clustering to segment nuclei and used shape-based and texture-based (multilayer perceptron) to classify meningioma subtypes. They achieved 92.50 % classification accuracy with their hybrid classifier technique.
[45]	2018	gliomas (GBM-specified)	H&E and IHC images	Utilised CNN-VGG19 to organise histomorphologic information in brain tumour H&E images to visualise meningioma subtype classifications. t-SNE was used to discretise the relationship between classes, and their distribution was used to create statistically driven classifications. After tuning, they reduced errors in multi-class classifications to about 4 %.
[23]	2017	Meningioma (Meningotheliomatous, Fibroblastic, Transitional, Psammomatous)	H&E images and clinical data (tumour subtype)	The authors investigated various feature extraction techniques, feature selection techniques and classifiers to achieve accurate classification of the various meningioma subtypes. Together with GA evolution, they could classify meningioma subtypes with 94.88 % accuracy in their dataset.
[20]	2020	Various brain tumours including GBM	Stimulated Raman histology (SRH) images and clinical data (tumour subtype)	Used ResNet-v2 to extract features and predict the 13 diagnostic classes. T-SNE was used to discretise the relationship between each class based on the extracted features. They also implemented semantic segmentation by mapping each image patch's tumour class probability maps into the WSI. Overall, they were able to achieve a 94.6 % classification accuracy (Classes included malignant glioma (glioblastoma and diffuse midline glioma, diffuse lower-grade gliomas (oligodendrogliomas and diffuse astrocytomas), pilocytic astrocytoma, ependymoma, lymphoma, metastatic tumours, medulloblastoma, meningioma, pituitary adenoma, gliosis/reactive astrocytosis/ treatment effect, white matter, grey matter and nondiagnostic tissue).

(*continued on next page*)

Table 2 (*continued*)

Authors	Publication Date (Year)	Brain Tumour specified	Data type	Utilisation of ML/DL
[81].	2016	Various gliomas including GBM	H&E images and clinical data (tumour subtype)	Presented a patch-based (MIL CNN) together with t-SNE based classification. With this pipeline, they were able to segment and classify various glioma types (GBM, oligodendroglioma (ODG), oligoastrocytoma (OC), diffuse astrocytoma (DA), anaplastic astrocytoma (AAC), anaplastic oligodendroglioma (AODG), LGG with the best accuracy of 77.1 %
[33]	2021	LGG (grade II and III) and GBM (grade IV)	H&E images, clinical data (tumour subtypes) and associated molecular data	Trained various glioma images on a range of CNNs to classify ODG vs. non-ODG (identified initially with associated molecular data) and grading diffuse gliomas; ResNet50V2, Inception V4, Xception, DenseNet201 were used to classify ODG vs. non-ODG via majority voting of patch classifications, while MnasNet, EfficientNet-B4, EfficientNet-B5 and DenseNet201 CNNs were used to classify diffuse glioma grading (grade II-IV).
[82]	2021	GBM, anaplastic astrocytoma (AAC), astrocytoma (AC), oligodendroglioma (ODG), anaplastic oligodendroglioma (AODG)	H&E images	Proposed the SD-Net-WCE model to classify pathology images into 6 types (ODG, AODG, AC, AAC, GBM and non-tumour). The H&E images were used to train and evaluate its performance, and the model was compared against DenseNet and Inception-FCN models. The SD-Net-WCE model outperformed the 2 popular models with an accuracy of 87 %.
[56]	2019	AC, ODG and GBM	H&E images, clinical (tumour subtype, survival data) and genotyping data	Mainly used H&E images from both databases to train the FCN-GoogLeNet model to train and segment the microvasculature of the H&E images. The microvessels' nuclei were further segmented by thresholding. Then its features, such as microvascular proliferation, microvascular area, microvessel density, was calculated. Overall, they used this to correlate with glioma subtype (i.e., increased microvascular density correlated well with GBM subtypes and TERT-mut cases). Additionally, their survival analysis showed that microvascular features can be used to cluster cases into 2 groups with different survival periods.
[37]	2015	Gliomas, but GBM is not specified	H&E images	Segmented endothelial and non-endothelial cells in glioma brain tumours were fed into the trained random forest classifier to evaluate its classification performance. This paper aimed to showcase their pipeline's ability to classify endothelial nuclei accurately and propose that it can be used to allude to the angiogenesis in brain tumours, which can signal disease progression and a negative prognostic factor.
[83]	2017	LGG	H&E images and genotyping data	Similar to their older paper [37], they set out to use a random forest classifier to phenotype microvascular structures in LGG H&E images to predict survival. Additionally, they implemented molecular data to explore pathways associated with these phenotypes.
[38]	2013	GBM	H&E images	They used the sparse auto-encoder to learn unsupervised intrinsic features in H&E image patches. Once extracted features were used to classify necrosis, transition to necrosis, and viable regions via a multi-class regularised SVM. They were able to map the classified patches onto the WSI, which achieved a classification accuracy of 84 %.
[84]	2020	GBM	IHC and RNA-seq data	Generated a panel of selected biomarkers that can discriminate between different GBMs (proneural, mesenchymal, classical) using RNA-seq data. They generated a molecular panel via IHC staining with the panel and used it with their transcriptomic data to train the random forest classifier.
[32]	2021	LGG and GBM	H&E images and molecular data	Used U-Net to segment nuclei in H&E images. The cellularity from the nuclei segmentations was fed into the ResNet CNN with the molecular information and the feature extracted from the same H&E image to classify glioma subtypes (LGG II, LGG III, HGG). They achieved 93.81 % accuracy when classifying between LGG and GBM and a 73.95 % accuracy in classifying LGG II and LGG III.
[85]	2015	Not specified	H&E images	The authors developed a pipeline where a learned dictionary of cell shapes was used to detect cells in the H&E image via sparse reconstruction. Then, the stacked denoising autoencoder, trained with the original data and their structured labels, is used to

(*continued on next page*)

Table 2 (*continued*)

Authors	Publication Date (Year)	Brain Tumour specified	Data type	Utilisation of ML/DL
				segment the cells. They could accurately segment heavily clustered and abnormally shaped nuclei that would have been very difficult to segment otherwise.
[86]	2016	GBM	H&E images	Using learned dictionaries in their DL method allows for good classification with sparse training data. They could distinguish between a microvascular proliferative (MVP) vessel and a non-MVP vessel in GBM H&E images. Additionally, the authors showed that DFDL exhibits a more graceful decay in classification accuracy as training images become sparser.
[39]	2017	LGG	H&E images	Trained SVM and random forest classifiers to distinguish under or over-segmented nuclei in LGG II H&E images. The goal is to develop a pipeline that can assess nuclei segmentation quality. The authors were able to achieve an 84.7 % classification accuracy
[19]	2019	GBM	DAPI-stained nucleic images	Proposed a modified version of U-Net, which implements gated recruitment units (GRU) to allow for iterative refinement of feature maps. The DAPI-stained GBM nuclei segmentations were used to train this model. Their model was able to achieve better segmentation performance while using much fewer parameters than the competing models (i.e. U-Net and Deconv network)
[87]	2015	Not specified	H&E images	A CNN was used to segment nuclei in the brain tumour H&E images. It was able to outperform other models such as SVM, random forest and deep belief network
[22]	2017	GBM and LGG	H&E images and clinical data (tumour subtype)	The author proposed a CNN -SVM pipeline that can distinguish GBM from LGG H&E images. AlexNet was used to extract features from the H&E images, and they were then pooled, selected for, and classified using the SVM method. Furthermore, they extended their framework to segment the images by modifying the SVM to classify positive and negative image patches based on extracted features. A segmentation/heat map was produced as a result. They achieved accurate classification and segmentation results in the MICCAI 2014 challenge.
[49]	2018	LGG and GBM	H&E images and clinical data (tumour subtype)	The authors created a Deep CNN model to train and classify GBM and LGG H&E images in the TCGA dataset. Their model outperformed (96.5 % classification accuracy) their benchmark (LeNet, ZFNet, VGGNet).
[41]	2020	Mixed glioma, ODG, AC and GBM	H&E images, genotypic and clinical data (survival data)	Proposed the DeepSurvNet model, which predicts the survival class (class I = 0–6 months, II = 6–12 months, III = 12–24 months, IV = >24 months survival after diagnosis) based on H&E image data. TCGA dataset was used to train the various CNNs (VGG19, GoogleNet, ResNet50, InceptionV3 and MobileNetV2) and they were evaluated based on their classification performance with locally derived datasets. GoogleNet outperformed the other CNNs in the classification of survival classes. Additionally, they correlated various mutated genes with the associated survival classes and suggested that these genetic fingerprints are associated with patient survival.
[53]	2021	GBM	H&E images, RNA-seq data	The authors trained a DenseNet model to semantically segment leading edge (LE), infiltrating tumour (IT), cellular tumour (CT), cellular tumour microvascular proliferation (CTmvp), cellular tumour pseudopalisading region around necrosis (CTpan), cellular tumour perinecrotic zones (CTpnz), cellular tumour necrosis (CTne) and background of the H&E images from the Ivy Gap dataset. These segmented features were then correlated with genetic signatures associated with the segmented image. This was evaluated on the TCGA datasets. The authors found that the GBM tumour sample had different gene signatures that were driven by different cell types in the tumour microenvironment.

Table 3
Detailed overview of data origin in training and testing datasets, proposed ML/DL model, training-testing-validation parameters, evaluation metrics,and what other models the proposed models are being compared to.

Authors	Publication Date (Year)	Training Dataset	Proposed model	Data Processing	training/test/validation parameters	Accuracy evaluation metric	Testing Dataset?	What models are being compared?
[43].	2007	Meningioma tissue from neurosurgical resections at a hospital in Bielefeld, Germany	Self-Organizing Maps (SOM) combined with Discrete Wavelet Transform (DWT) for feature selection and clustering.	80H&E images, subdivided into 1280 image tiles (256x256 pixels each).	Four-patients-left-out cross-validation, where one patient per class was left out during training. Classification was based on majority vote of the winner node in the SOM.	Correct classification as a percentage, average of 79 %	Same dataset (four-patients-left-out cross-validation)	N/A
[30]	2022	Histopathological images and genomic data from TCGA-LGG and TCGA-GBM projects	MultiCoFusion: a multi-modal fusion framework using ResNet-152 and sparse graph convolutional network (SGCN)	953 samples for 469 patients from TCGA, including 1024x1024 ROI histopathological images and mRNA expression data	80 % training, 20 % testing, repeated 15 times with alternate training (cross-validation) for survival and grade classification tasks	C-index of 0.857 for survival analysis (a generalisation of AUC for censored data), micro-AUC of 0.923 for grade classification. Model performance was improved with the inclusion of genomic mRNA data	Same dataset (20 % held-out for testing)	Compared with Pathomic Fusion and traditional methods (LASSO-Cox, MLP, Logistic Regression, PWMK)
[66]	2015	Four subtypes of grade I meningioma tissue biopsies from neurosurgical resections at the Bethel Department of Neurosurgery, Bielefeld, Germany	Fractal dimension-based model with wavelet packet (WP) decomposition and machine learning models (SVM, Bayesian, k-NN)	320 images subdivided into 512x512 pixel sub-images after truncation	Leave-one-patient-out cross-validation across 20 patients	Classification accuracy as a percentage: SVM: 94.12 %, Bayesian: 76.47 %, k-NN: 82.35 %	Same dataset (leave-one-patient-out cross-validation)	Compared with Bayesian and k-Nearest Neighbors classifiers
[67]	2021	Three datasets: brain histological images (GBM), skin melanoma images, and lung necrosis images. Unspecified where it was collected	TisCut (Tissue Cluster Level Graph Cut) for unsupervised segmentation	H&E stained brain images for necrosis detection (500x5000 pixels)	No training, unsupervised method with graph cut partitioning	Brain tumour-specific accuracy: Jaccard-Index-coefficient (JIC): 70.24 %, Dice similarity coefficient (DSC): 80.64 % for TisCut; BTA-SVM: JIC: 55.87 %, DSC: 68.79 %; TL-InceptionV3: JIC: 65.05 %, DSC: 77.05 %; U-Net: JIC: 67.22 %, DSC: 78.58 %	Same dataset (unsupervised segmentation applied to the brain dataset)	Compared with BTA-SVM, TL-InceptionV3, U-Net models
[26]	2016	604 images from TCGA (364 GBM, 240 LGG)	Coarse-to-fine profiling with Elastic Net classifier for GBM vs LGG classification	H&E stained images, 1024x1024 tiles used for fine profiling after coarse clustering	5-fold cross-validation on TCGA dataset (364 GBM, 240 LGG)	Classification accuracy on TCGA dataset: 93.1 %, AUC: 0.96 for Elastic Net, and 100 % classification accuracy on MICCAI Pathology Challenge models; Bueno et al.: Accuracy: 98.1 %, Chang and Parvin: 85.83 %, Xu et al.: Accuracy: 97.8 %	Same dataset (TCGA) for 5-fold cross-validation; additional 45 tissue slices used during comparison with MICCAI challenge models	Compared to manual subsetting methods (e.g., Bueno et al., Xu et al.)
[68]	2008	100 Hematoxylin-Eosin stained biopsies, classified into low-grade (41 cases) and high-grade (59 cases), from the Department of Pathology of the University Hospital of Patras, Greece	Fuzzy Cognitive Maps (FCM) with Activation Hebbian Learning (AHL) algorithm	Qualitative assessments of eight histopathological features provided by pathologists (e.g., cellularity, mitoses) of H&E images	The model was evaluated on 100 cases, and the Activation Hebbian Learning (AHL) algorithm was used to optimize the FCM. Tumor classification was performed after four interaction cycles.	Classification accuracy Low-grade: 90.26 %, High-grade: 93.22 %; FCM Grading Tool overall accuracy: 92 %; Compared with ID3 Decision Tree: 80 %, J48 Decision Tree: 85.71 %,	Same dataset (100 samples)	ID3 Decision Tree, J48 Decision Tree, Fuzzy Decision Tree

(*continued on next page*)

Table 3 (*continued*)

Authors	Publication Date (Year)	Training Dataset	Proposed model	Data Processing	training/test/validation parameters	Accuracy evaluation metric	Testing Dataset?	What models are being compared?
						Fuzzy Decision Tree: 93 %		
[69]	2022	Four public histopathology image datasets: TCGA-IDH (1494 whole-slide images from 921 glioma patients), CAMELYON16 (400 slides), CAMELYON17 (1000 slides), and BreakHis (7909 images)	Colour Adaptive Generative Network (CAGAN) for stain normalisation using both supervised and unsupervised learning, ResNet34 classification model to classify after stain normalisation	Stain augmentation of H&E images with supervised and unsupervised stain normalisation using a GAN. 1191 samples from TCGA-IDH, into 1024x1024 pixel tiles	Dual-decoder structure, supervised (target domain) and unsupervised (source domain), with histogram loss. 1191 slides for training and 154 for validation and 149 for testing for classification model	SSIM (Structural Similarity Index Measure): 0.984 for CAGAN, 0.870 for Macenko, 0.948 for Vahadane, PSNR (Peak Signal-to-Noise Ratio): 32.86 for CAGAN, 23.41 for Macenko, 26.14 for Vahadane. Classification accuracy, F1 and AUC: 0.981, 0.973 and 0.981 for CAGAN method, 0.938, 0.899 and 0.885 for Macenko method, 0.908, 0.910 and 0.921 for Vahadane.	Same datasets used for testing, with cross-domain testing on CAMELYON17 after training on CAMELYON16	Compared to other stain normalisation methods like Macenko, Reinhard, Vahadane, StainGAN, STST, Tellez et al.
[70]	2021	Two datasets for mitochondria segmentation (EPFL dataset and Kasthuri++ dataset) and the CPM-17 dataset for nuclei segmentation	Authors proposed CS-Net (Derived from U-Net), a lightweight deep network for cellular segmentation using hierarchical dimension-decomposed (HDD) convolutions, with two variants: CS-Net (2D) and CS-Net (2.5D++)	40 pathological slides into 128x128 image tiles, randomly cropped	26 train and 14 test split	DSC and Aggregated Jaccard Index (AJI): 88.3 and 71.1 for CS-Net (2D), 86.9 and 70.5 for HoVer-Net, 85.6 and 59.4 for SegNet + WS	Same dataset (14 test)	Compared to a large range of models, some state-of-the arts include Mask-RCNN, U-Net, FCN, SegNet, HoVer-Net
[52]	2017	Two datasets: 2015 MICCAI Gland Segmentation (GLaS) Challenge dataset and 2015 MICCAI Nuclei Segmentation Challenge dataset, containing images from glioblastoma and lower-grade glioma tissues	Used DCAN (Deep Contour-Aware Network), a multi-task learning framework combining object and contour detection for instance segmentation of histology images.	Manually segmented nuclei from the H&E images in 15 image tiles and 18 images	500 manually segmented nuclei in training, 18 images used for evaluation	Dice coefficient (D1) and object-level Dice coefficient (D2): 0.876 and 0.748 for proposed model, 0.826 and 0.694 for Team 2's model, 0.792 and 0.642 for Team 3's model for the 2015 MICCAI nuclei segmentation challenge	Same dataset (18 images held-out)	Compared to other teams' model for the 2015 MICCAI nuclei segmentation challenge
[40]	2022	Multiple datasets including 10x Genomics Visium spatial transcriptomics datasets on the human and mouse cortex, and in-house Alzheimer's disease (AD) samples.	RESEPT (REconstructing and Segmenting Expression mapped RGB images based on sPatially resolved Transcriptomics), which uses deep learning (based on the ResNet101 architecture) for characterising tissue architecture by converting spatial transcriptomics into RGB images and performing segmentation with convolutional neural networks (CNN).	The datasets were processed by converting spatial transcriptomics data into RGB images using 10x Genomics Visium datasets.	16-fold Leave-One-Out Cross-Validation (LOOCV). The final model was selected based on the Moran's I autocorrelation index for the testing data	Adjusted Rand Index (ARI) and Fowlkes–Mallows index (FM). RESEPT achieved ARI: 0.706 and FM: 0.780 for segmentation accuracy.	Same datasets used for training and testing, with Leave-One-Out Cross-Validation (LOOCV) strategy	Compared to Seurat, BayesSpace, SpaGCN, stLearn, STUtility, HMRF, and Giotto

(*continued on next page*)

Table 3 (*continued*)

Authors	Publication Date (Year)	Training Dataset	Proposed model	Data Processing	training/test/validation parameters	Accuracy evaluation metric	Testing Dataset?	What models are being compared?
[55]	2022	TCGA dataset, containing 32,072 whole-slide images (WSIs) across 32 primary tumor types, including GBM (35 WSIs) and LGG (39 WSIs)	A deep neural network (DNN) model with evolutionary feature selection method that compresses deep feature vectors extracted from gigapixel histopathology images, reducing the size of these vectors by 11,000 times while maintaining classification accuracy	WSIs were preprocessed by extracting patches (1000x1000 pixels) from each WSI	Feature selection is performed by using an evolutionary algorithm with multi-objective optimization.	F1-score for both LGG and GBM: 97 % and 97 % for proposed model, 93 % and 94 % for PCA, 86 % and 87 % for Autoencoder model, 91 % and 90 % for ANN-based feature selectors	Same dataset used for both training and testing (TCGA WSIs)	Compared to PCA, Autoencoder, and ANN-based feature selectors
Komosin'ski et al., 2000	2000	Various CNS tumours were collected from the Department of Pathology, University School of Medicine in Poznan'.	Evolutionary algorithm with genetic algorithm optimization technique for feature selection and weighting to improve the classification accuracy (using kNN classifier) for CNS tumours based on microscopic images.	Microscopic images of neuroepithelial tumours that are segmented into approximately 1300 regions (512x512 pixels)	Genetic algorithm with uniform crossover and bit-flipping mutation to optimize the feature weights and improve classification accuracy, with two-level cross-validation	Classification accuracy: 83.43 for astrocytic tumours and 77.83 for glial tumours	Same dataset (two-level cross-validation)	N/A
[72]	2021	In-house glioblastoma stem cells (GSCs) derived from human and mouse models. Serial transplantation into immunocompetent BL6 mouse models was used to mimic the tumor microenvironment (TME).	"Built-in" machine learning (ML) algorithm - model not specified. Part of the Inform software package	H&E stained and GFP stained images, and 5 randomly selected regions from the tumour area. Cell markers Iba-1, F4/80 and CD45 used to train the algorithm into phenotyping these cells	N/A	N/A	N/A	N/A
[36]	2017	H&E stained slides from the LGG cohort of The Cancer Genome Atlas (TCGA). The dataset includes Grade II and III tumours from 53 patients, split into two groups based on short and long overall survival (OS)	Using a bag-of-words (BoW) machine learning approach to create a visual dictionary of image-derived features associated with overall survival (OS). A SVM model would then be used to predict OS categories	H&E images were colour separated to H and E components using Pipeline Pilot, H component was used to segment the nuclei and the image was tiled into 256x256 patches with 50-pixel overlap. The segmentations had their feature extracted. Various clinical data were also extracted per patient	SVM model was trained using a bag-of-words representation of image patches, clustered into 100 visual words using K-means clustering. The optimal model was determined through 10-fold cross-validation.	AUC and F1 score: 0.76 AUC with machine learned dictionary alone, but 0.82 when supplemented with molecular biomarkers and 0.89 when also supplemented with clinical factors.	Same dataset was used for training and testing, with 10-fold cross-validation.	N/A
[46]	2022	Two public nuceli segmentation dataset were used. Monuseg: 30 histopathology images from 7 tissue types (breast, kidney, liver, prostate, bladder, colon, and stomach). CPM-17: 64 histopathology images, including images from glioblastoma multiforme (GBM), lower-grade glioma (LGG)	Meta Multi-Task Learning (Meta-MTL) model (Based off U-Net architecture) for nuclei segmentation that uses a model-agnostic meta-learning (MAML) algorithm to improve model generalization with fewer training samples	H&E images cropped into 256x256 pixels and data augmentation applied: elastic transformation, scaling, shift, rotation and flipping	No explicit statement on how the dataset was used to train and split models and whether cross-validation was used	DICE coefficient and AJI for over all segmentation quality. Panoptic Quality (PQ) to evaluate nuclei segmentation of overlapping regions: 0.8756 DICE, 0.7081 AJI and 0.6791 PQ in ablation studies, and proposed model outperformed the state-of-the-art models in sparser datasets while maintaining near	Same dataset used for evaluation	Various state-of-the-art models compared including: HoVer-Net, TripleU, DCAN, Mask R-CNN, and others

(*continued on next page*)

Table 3 (*continued*)

Authors	Publication Date (Year)	Training Dataset	Proposed model	Data Processing	training/test/validation parameters	Accuracy evaluation metric	Testing Dataset?	What models are being compared?
						identical performance at larger datasets		
[73]	2019	Three datasets were used in this article. An in-house dataset containing 10 multiplexed fluorescent microscopy images. CPM 2017 challenge: 64 images with various tumours including GBM and LGG. The third dataset used was the Warwick-QU dataset from the GLaS challenge	Micro-Net, a model with multi-resolution input and bypass layers	H&E images from CPM dataset were cropped into 300x300 pixels image tiles. Various data augmentation applied, such as gaussian filter, rotation and flipping.	32 images were used for training and another 32 images were used for testing	DICE, F1, Object Dice (OD), Pixel Accuracy and Object Hausdorff (OH) respectively: 82.43, 71.79, 74.12, 63.53, 27.53 for the proposed model. State-of-the-arts like U-Net performed with 78.39, 66.43, 67.35, 80.28, 40.49	Same dataset used (32 images reserved in CPM for testing)	Compared against U-Net, FCN8, and DCAN
[74]	2022	Two datasets were used. 10 cohorts from the TCGA dataset which included various brain tumours including GBM and LGG. Histologic sections from the East Asian cohort from Singapore were also used, which included 179 lung adenocarcinoma patients	A Multiple instance learning (MIL) model (Derived from a ResNet18 architecture) designed to predict tumour purity from H&E-stained histopathology slides. The model can learn spatial tumour purity distributions without requiring pixel-level annotations, based on corresponding genomic sequencing data.	H&E images cropped into non-overlapping 512x512 pixel patches and the corresponding tumour purity level in the patch was extracted from adjacent slide, to be used as ground truth labels	60 % training, 20 % validation and 20 % testing distribution	Mean absolute error between purity values and model predictions, and between purity values and pathologists' % tumour nuclei estimates: GBM and LGG predictions achieved <0.15 mean absolute error compared to pathologists' estimates while maintaining <0.2 when compared to genomic purity values	Same dataset (20 % held-out for testing)	N/A
[54]	2022	6592 gigapixel WSIs from 5720 patient samples across 14 cancer types from the TCGA (including LGG and GBM)	Pathology-Omics Research Platform for Integrative Survival Estimation (PORPOISE), which uses a multimodal fusion (MMF) algorithm integrating WSIs and molecular profile features (mutation status, RNA-seq, and copy-number variations)	512x512 pixel patches were extracted from the WSIs, and genomic data was processed from RNA sequencing and mutation profiles.	5-fold cross-validation on the paired WSI-molecular datasets from the 14 cancer types (GBM inclusive)	Survival AUC and cross-validated concordance index (c-index). MMF (WSI and molecular features combined) model: 0.662 AUC and 0.644 c-index. AMIL (WSIs only): 0.615 AUC and 0.578 c-index. SNN (molecular features only): 0.588 AUC and 0.606 c-index	Same dataset (5-fold cross validation; GBM excluded)	N/A
[42]	2022	WSIs of neuroblastoma were retrospectively obtained from the Affiliated Children's Hospital of Xi'an Jiaotong University. In total, 563 WSIs from 107 patients with neuroblastoma collected	Logistic regression approach to classify neuroblastoma patients into favorable histology (FH) and unfavorable histology (UH) groups	HoVer-Net model was used to segment nuclei and had their features extract. 13 cell-level morphological features and 36 intensity features were extracted from the nuclei in various color spaces (RGB, Lab, HSV, H&E). Patient-level data was also collected and used as features	Bootstrap resampling training, with 80 % training 20 % test, and another 20 % withheld for independent validation	AUC, Classification accuracy, and Mathhews Correlation Coefficient (MCC) respectively. Proposed model: 0.938, 0.865 and 0.630. SVM: 0.913, 0.814 and 0.599. Siamese-kNN: 0.910, 0.909, 0.,828.	Same dataset (20 % held-out for testing)	SVM and siamese kNN neural netowrk
[75]	2018	In-House GBM animal models: Patient-derived xenograft (PDX) mouse model and cdkn2a−/	The ilastilk software, used to segment image data using supervised random forest algorithms	N/A	N/A	N/A	N/A	N/A

(*continued on next page*)

Table 3 (*continued*)

Authors	Publication Date (Year)	Training Dataset	Proposed model	Data Processing	training/test/validation parameters	Accuracy evaluation metric	Testing Dataset?	What models are being compared?
		−PDGFRB lentivirus-induced mouse GBM model						
[31]	2021	Histopathology slides with associated quantified features (cellularity, necrosis, tumor nuclei, age, tumour weight) from the Cancer Imaging Archive (TCIA) used to trained the model	Trained various convolutional neural networks (InceptionV1 to V4, InceptionResNetV1 and V2, and self-designed simple CNNs) to find the best performing to use in their analysis	Histopathology images cropped into 299x299 pixel patches, with 49 pixels of overlap from edge to edge. Patches with >30 % background were not used. Genotypic information were associated with the histopathology images.	70 % training 15 % validation and 15 % testing at patient level	N/A	N/A	InceptionV1 to V4, InceptionResNetV1 and V2, and self-designed simple CNNs were tested against each other
[34]	2022	69H&E medulloblastoma formalin-fixed paraffin embedded whole-slide images obtained from patients under the age of 18 from the Children's Hospital Los Angeles and digitized at University Hospital	Tested various classifier algorithms trained on nuclear histomorphometric features to distinguish various molecular subgroups	Nuclei in 2000x2000 image tiles segmented with a watershed algorithm, and nuclear morphological features (texture, shape, architectural rearrangement) were extracted. Molecular subtypes and survival was recorded for ground truth labelling	100 iterations of per-patient 3-fold cross-validation training.	AUC for molecular subtype prediction (0.7 - LDA when SHH and WNT vs. Group 3 and 4 tumours) and AUC for survival prediction (0.92 - neural network for Group 3 survival)	Same dataset (3-fold cross-validation)	4 models were tested and compared (Random Forest, Neural network, SVM and LDA)
[48].	2022	Patients at the Department of Neurosurgery of the Medical Center, University of Freiburg (Freiburg, Germany)	A customised pipeline that utilises various models for various tasks in the article (machine-learning-based nuclei segmentation with ilastik, artificial neural network to predict tumour content and a pre-trained VGG16 CNN model that will be trained to predict various histological morphology)	No clear data processing for the cell detection model. Performed single-cell sequencing, and corresponding spatially resolved transcriptomics from the same donor to train the ANN. 500x500 pixel image patches of H&E secitons at random positions across samples from all patients were used to train the CNN model.	No clear training parameter provided for the cell detection model. 2000 training spots with 19 different cell numbers were outlined for the prediciton of tumour content. No clear training parameters were outlined for the histomorphological prediction model.	F-score: Ilastik + Cellprofiler cell detection 0.841, overall accuracy of 95 % in predicting tumour content and significant correlation between training and validation datasets. Refined necrosis morphological segmentation with an F-score of 0.893 when overlapped to the necrosis segmentation based on transcriptional data.	Same dataset	N/A
[76]	2018	In-house TS603 subcutaneous xenograft tumour tissues, which are mouse glioma models	Machine-learning algorithm in the inForm software was used in this article	N/A	N/A	N/A	N/A	N/A
[77]	2022	TCGA dataset, containing H&E WSIs of 599 GBM and 515 LGG cases, totalling into 239,600 patches for GBM and 206,000 patches for LGG tissue	Proposed the shallow neural network (SNN) model	H&E WSIs cropped into 128x128 pixel tiles, and their disease classifcation was collected for each tile	5-fold cross-validation method was applied	Classification accuracy, Precision, Recall, Specificity, F1 score, respectively. SNN: 97.21, 0.9782, 0.9667, 0.9787, 0.9714. SVM: 95.13, 0.9194, 0.8903, 0.9241, 0.9036	Same dataset (5-fold cross validation)	SVM and kNN models

(*continued on next page*)

Table 3 (*continued*)

Authors	Publication Date (Year)	Training Dataset	Proposed model	Data Processing	training/test/validation parameters	Accuracy evaluation metric	Testing Dataset?	What models are being compared?
[78]	2017	The TCGA dataset was used in this article. GBM and kidney renal clear cell carcinomas were the two specified diseases used in the TCGA dataset	Proposed models include both human-engineered features and unsupervised feature learning. The study used combinations such as Cellular Morphometric Features (CMF) and Predictive Sparse Decomposition (PSF) with Kernel-based Spatial Pyramid Matching (KSPM) and Sparse Feature Encoding (SFE), and using a linear SVM classifier	H&E stained histology images were processed into 1000x1000 pixels, and feature extraction includes CMF from segmented nuclei and Dense/Salient SIFT for other types of features	Cross-validation with 10 iterations, with 160 images per category used for training	Correct classification rate, reported as the mean and standard error. CMF-LCDL-LSPM SPE model reported near 94 % classification rate while a pre-trained Alex-Net based network can achieve similar performance as well	Same dataset used (From cross-valdiation splits)	Dense SIFT (DSIFT), Dense Color-Texture (DCT), and deep learning models like AlexNet
[79]	2022	WSIs from Australian Genomics and Clinical Outcomes of Glioma (AGOG) tissue bank, as well as a second cohort from the Sydney Brain Tumour Bank	Bifocal Convolutional Neural Network (BCNN), used for automatic detection, profiling, and counting of cells in whole-slide images	H&E and immunochemically labelled for CD276, cropped into 32x32 and 64x64 pixel patches. Data augmentation (rotation, contrast, sharpness) was applied to improve model generalization	50,714 image patches, split into 36,734 for training and 14,005 for testing	Accuracy, Precision, Recall, and F1-Score, respectively. BCNN achieved 97.7 for all metrics. Resnet-50 achieved 94.0 % for all metrics	Same dataset (14,005 for testing)	ResNet-50 model
[44]	2021	Pediatric medulloblastoma was collected from Guwahati Medical College and Hospital (GMCH) and Guwahati Neurological Research Centre (GNRC). The dataset consists of images from 15 children, with a total of 204 images.	MB-AI-His, a computer-aided diagnosis system that combines deep learning (DL) and textural analysis techniques. It utilises three pre-trained CNNs (ResNet-50, DenseNet-201, and MobileNet) for spatial feature extraction and integrates these with Discrete Wavelet Transform (DWT) and Discrete Cosine Transform (DCT)	Histopathological images were resized to 224x224x3 pixels image patches. Spatial features were extracted using CNNs, and textural features were added using DWT	5-fold cross-validation was used	The accuracy for binary classification reached nearly 100 % for ResNet-50, DenseNet-201 and MobileNet. For multi-class classification, the highest accuracy was 95.74 % for ResNet-50, 94.54 % for MobileNet, and 97.16 % for DenseNet-201	Same dataset	ResNet-50, DenseNet-201 and MobileNet models were compared against each other
[47]	2021	Pediatric medulloblastoma was collected from Guwahati Medical College and Hospital (GMCH) and Guwahati Neurological Research Centre (GNRC). The dataset consists of images from 15 children, with a total of 204 images.	CoMB-Deep, a deep learning pipeline that combines features from 10 different CNNs (DenseNet-201 + Inception V3 + ResNet-50 + MobileNet + DarkNet-53 + NasNetMobile) with the Discrete Wavelet Transform (DWT) and uses Bi-Directional Long Short-Term Memory (Bi-LSTM) for classification	Histopathological images resized appropriately as input to the CNNs for feature extraction, and features were added using DWT. Data augmentation (e.g., flipping, translation, scaling, and shearing) wasn applied prior to training	10-fold cross-validation were used	Classification accuracy. ResNet-50 and InceptionResNet are capable of 100 % accuracy in binary classification, but the combination of CNNs for feature extraction in the CoMB-Deep model achieved 99.35 % multi-class classification accuracy, which superior than end-end single CNN-based classification	Same dataset (10-fold cross validation)	ShuffleNet, ResNet-50, Squeeze, MobileNet, Inception V3, DenseNet-201, Inception-ResNet V2, Xception, NasNetMobile, DarkNet-53
[35]	2022	37 Tissue samples from GBM patients from the General University Hospital of Patras, Greece	Hybrid deep learning system for risk stratification of GBM patients. The system used machine learning algorithms for feature extraction (VGG16 CNN) and data analysis,	RNA extraction from FFPE tissues, followed by qRT-PCR to quantify miRNA expression levels. IHC images cropped into 224x224x3 image patches	10 iterations of bootstrapping, and training and validation were performed on the miRNA expression data	Classification accuracy and F1-score. Random forest algorithm achieved the highest performance, with an accuracy of 94.32 %, an F1 score of	Same dataset	Models include LDA, SVM, kNN and naïve Bayes

(*continued on next page*)

Table 3 (*continued*)

Authors	Publication Date (Year)	Training Dataset	Proposed model	Data Processing	training/test/validation parameters	Accuracy evaluation metric	Testing Dataset?	What models are being compared?
			including random forest, LDA, SVM, and kNN	to be fed into the VGG16 CNN model for feature extraction		92.82 %, and an AUC of 97 % in the validation data. Sensitivity for high-risk patients was 95.36 %, and specificity for low-risk patients was 90.67 %.		
[18]	2020	1121 histopathology images of glioma patients, with clinical information collected from TCGA dataset. 682 images with the IDH1 mutation and 439 images without the mutation	Multiple-Instance Learning-based Convolutional Neural Network (MIL-based CNN), a CNN to extract features from histopathology images and integrates MIL to classify the presence of the IDH1 mutation	Histopathological images were segmented into smaller patches of different sizes (32x32, 64x64, 128x128, 256x256, 512x512, and full-size) before being input into the CNN	10-fold cross-validation technique used to train and evaluate the model	Classification accuracy, precision, recall, AUC. The model achieved an accuracy of 78.95 %, a precision of 82.23 %, and a recall of 83.43 %. The area under the curve (AUC) was 0.84	Same dataset (10-fold cross validation)	N/A
[29]	2015	Digital pathology images of LGG and GBM from the TCGA dataset	Ensemble of CNN to grade gliomas	WSI images were cropped into tiles (1024x1024 pixels) and only tiles with tissue occupying at least 90 % of the tile were selected for further analysis. Nuclei were segmented but not cropped individually to leave it in place in the tile. Patches of 256x256 were then used for the model	80 % training and 20 % testing split.	Classification accuracy, precision, sensitivity, specificity. The model achieved a 96 % accuracy in classifying GBM and LGG, with 0.94, 0.98 and 0.94 precision, sensitivity and specificity in predicting the GBM class.	Same dataset (20 % held-out for testing)	N/A
[80]	2014	320 images of meningioma tumor samples obtained from the Bethel Department of Neurosurgery, Bielefeld, Germany	Proposed a Hybrid classification technique that integrates texture and shape characteristics for the classification of four meningioma subtypes. It uses Grey-Level Co-occurrence Matrix (GLCM) textural features and a Multilayer Perceptron (MLP) classifier for transitional and psammomatous subtypes, while nuclei shape-based features are used for meningothelial and fibroblastic subtypes	H&E images were converted into grayscale images, and nuclei were segmented via K-means clustering for feature extraction. The shape of the nuclei is used in Phase 1, and GLCM textural features are extracted for Phase 2 of the classification process.	64 images per subtype used for training and 16 images for testing	Classification accuracy. The Hybrid classifier achieved an average accuracy of 92.50 % across five test trials, which outperformed other feature extraction techniques such as RADWPT (88 % overall accuracy) and WPFD (90.31 % overall accuracy)	Same Dataset (16 testing)	N/A
[45]	2018	WSIs for their own neuropathology service was used, although where the data comes from is not specified. The data includes gliomas, metastatic carcinomas, meningiomas, lymphomas, and schwannomas	The proposed model is a multi-class Convolutional Neural Network (CNN) classifier based on VGG19, using t-distributed stochastic neighbour embedding (t-SNE) for dimensionality reduction and visualisation of high-dimensional data.	H&E images cropped into 1024 x 1024-pixel tiles. Each tile is classified into one of the 13 neuropathological categories, and dimensionality reduction is performed using t-SNE for further analysis	85 % of the dataset used for training and 15 % for testing	Classification accuracy and AUC: The model achieved an accuracy of 94.8 % for the classification of 13 tissue and lesion types. The AUC of the proposed model is 0.99. Using t-SNE dimensional reductionality technique before classifcaiton	Same dataset (15 % validation set)	N/A

(*continued on next page*)

Table 3 (*continued*)

Authors	Publication Date (Year)	Training Dataset	Proposed model	Data Processing	training/test/validation parameters	Accuracy evaluation metric	Testing Dataset?	What models are being compared?
						yielded better results than using the PCA technique, but no other model was used as comparison.		
[23]	2017	Four subtypes of grade-I benign meningioma from the Bethel Department of Neurosurgery, Bielefeld, Germany. Total of 320 images collected	The proposed model is an evolutionary classification framework based on nuclear spatial and spectral features. The framework applies a SVM classifier combined with the genetic algorithm to select optimal features and classifier parameters for meningioma subtype classification	Nuclei isolated using adaptive color thresholding and watershed transformation. Various morphological, intensity, and texture features were computed in the RGB color space. Features were extracted from the isolated nuclei to be used for training the model	5-fold cross validation training and testing, and features were selected using Genetic Algorithm	The proposed framework achieved an overall classification accuracy of 94.88 %, while other classifiers like the linear-SVM and MLP only achieved an overall classification accuracy of 88.75 % and 85.63 %, respectively	Same dataset (5-fold cross validation)	Linear SVM, RBF SVM, quadratic SVM, MLP, Random Forest, kNN, and Naive Bayes
[20]	2020	415 patients from two datasets: University of Michigan (UM) images from a prototype clinical Stimulated Raman Histology (SRH) microscope; and UM images from one NIO Imaging System	Proposed an Inception-ResNet-v2 architecture classification model	Sliding window algorithm to crop images into 300x300 pixels patches with 100 pixel step size. Image patches undergo contrast enhancement and transformation to create three-channel input for the CNN model. Data augmentation is also applied to mitigate class imbalance, including transformations like rotation, shift, and reflection	415 patient training group with 16 patient validation set. Performance evaluated with University of Miami and Columbia University datasets	Classification accuracy and AUC: The proposed model achieved an overall diagnostic accuracy of 94.6 % in the experimental arm of the trial, compared to 93.9 % accuracy using conventional H&E histology interpreted by pathologists. The model also achieved an area under the curve (AUC) of 97 %	Testing dataset from two other image datasets: Columbia University images from a second NIO Imaging System; and University of Miami images from a third NIO Imaging System	N/A
[81].	2016	H&E WSIs of gliomas (including GBM) and Non-Small-Cell Lung Carcinoma (NSCLC) cases from the TCGA dataset	Proposed a patch-based Convolutional Neural Network (CNN) combined with an Expectation-Maximization (EM)-based model.	Image patches from large H&E images were generated, and patches with <30 % or too bloody were discarded. Data augmentation (rotation, mirror, adjusting H and E component of image) was applied to each patch	80 % training 20 % testing split	Classification accuracy: EM-CNN with logistic regression (LR) achieved the best accuracy of 77.1 % for gliomas, whereas other combinations of CNN + classifier, like the CNN-SVM, only achieved 0.697.	Same dataset (20 % withheld for testing)	Compared various CNN + classifier combos: CNN-Vote, CNN-SMI, CNN-Fea-SVM, and Pretrained CNN-Fea-SVM
[33]	2021	468H&E glioma slides from Catholic University of Korea Yeouido St. Mary's Hospital from 2017 to 2019 during routine clinical 1p/19q fluorescence in situ hybridization (FISH) test	The proposed model is a deep transfer learning model based on ResNet50V2 for the classification task.	ROI sections of the H&E images were tiled into 1024x1024 pixels and then cropped into 224x224 pixels when inputted into the model. Random image augmentation (e.g., flipping, rotation, scaling, and Gaussian noise) was applied. Molecular information from FISH was processed together with associated images for	70 % training, 10 % validation, 20 % testing	The proposed model achieved a balanced accuracy of 0.8727 using the majority voting technique for glioma subtype classification (ODG vs. non-ODG). For grading, the best-balanced accuracy was 0.5801 using the MnasNet model for grade II, III, and IV gliomas	Same dataset (20 % testing)	ResNet50V2, InceptionV4, Xception, DenseNet201, and MnasNet

(*continued on next page*)

Table 3 (*continued*)

Authors	Publication Date (Year)	Training Dataset	Proposed model	Data Processing	training/test/validation parameters	Accuracy evaluation metric	Testing Dataset?	What models are being compared?
				classifying oligodendroglioma and non-oligodendroglioma				
[82]	2021	97,252 histopathological images and molecular data from 323 glioma patients, collected from the Central Nervous System Disease Biobank at Huashan Hospital, Fudan University, Shanghai. The dataset includes various brain tumour tissues and background (non-tumour tissues)	"AI Neuropathologist for Glioma" platform is proposed, which is a Squeeze-and-Excitation block DenseNet (SD-Net-WCE), which uses deep convolutional neural networks (CNNs) for the classification of six glioma subtypes. Using additional molecular data, the subtyped cases can be further classified with logical algorithms similar to decision trees	H&E images were preprocessed by resizing them to 512 x 384 pixels and undergoing data augmentation (e.g., flipping, rotation, and random changes in brightness and contrast)	6-fold cross-validation method	SD-Net-WCE shows the best classification accuracy at 87 % compared to DenseNet and Inception-FCN. It achieved a patch-level accuracy of 86.5 % and a patient-level accuracy of 87.5 % for classifying glioma subtypes. Sensitivity and specificity across different subtypes are also reported, with glioblastoma achieving 74.6 % and 97.60 % sensitivity and specificity, respectively.	Same dataset (6-fold cross-validation)	DenseNet and Inception-FCN
[56]	2019	350 adult patients who were pathologically diagnosed with glioma between 2012 and 2017 at the Huashan Hospital of Fudan University. This includes H&E sections with associated genotypic data	Three types of FCNs (FCN-GoogLeNet preferred) were tested as a model for segmentation.	H&E images were cropped into various patches, with the smallest patch being 204x220 pixels, the largest patch being 4594 × 3718, with the average being 763 × 815 pixels. Data augmentation techniques like random rotation and color disruption were applied. These images were associated with their corresponding genotypic information	58 WSIs for training and 292 WSIs for testing	Pixel accuracy. FCN-GoogLeNet achieved the highest segmentation performance, with a pixel accuracy (PA) of 95.7 %, a mean pixel accuracy (MPA) of 54.9 %, and a Dice coefficient of 95.4 % for microvessel regions	292 testing sets from the same dataset and 195 patches from the independent TCGA dataset were used as an independent validation set	FCN-GoogLeNet, FCN-VGG, and U-Net
[37]	2015	WSIs from The Cancer Genome Atlas (TCGA), specifically focusing on glioma brain tumours.	An interactive machine learning framework that enables users to rapidly build classifiers for histological entities using a browser-based system. The model employs active learning to iteratively improve classifier performance by selecting the most ambiguous samples for labeling by the user. In this article, the random forest classifier is used in the pipeline	WSI images were segmented to isolate cell nuclei, and a total of 48 descriptive features were extracted from each cell to describe their size, shape, and texture. These features were used to build the classifier	18 sampling iterations, 169 samples used for training and 321 independent testing set (cross-validated)	The classifier achieved an AUC of 0.902. The abundance of detected endothelial cells was positively correlated with PECAM-1 expression (associated genomic data from TCGA)	Same dataset (321 withheld for testing)	N/A
[83]	2017	781 (464 tumours) images from the TCGA LGG dataset were used.	Proposes HistomicsML, an interactive machine-learning system for histopathology image analysis, which uses active learning to efficiently classify histologic objects. The	WSIs were tiled into 4096x4096 tiles. Nuclei were segmented from these tiles and 48 histomic features describing shape, intensity and texture	N/A	The VECN classifier (based from the random forest classifier) achieved an AUC of 0.964	Same dataset (67 slide used to validate)	N/A

(*continued on next page*)

Table 3 (*continued*)

Authors	Publication Date (Year)	Training Dataset	Proposed model	Data Processing	training/test/validation parameters	Accuracy evaluation metric	Testing Dataset?	What models are being compared?
			model involves iterative feedback between users and the system to train classification rules					
[38]	2013	Images of glioblastoma multiforme (GBM) and clear cell kidney carcinoma (KIRC) were obtained from TCGA. The GBM dataset contained 1400 images, and the KIRC dataset contained 2500 images	The proposed model uses a variation of the Restricted Boltzmann Machine (RBM) with added sparsity constraints for unsupervised feature learning. A SVM classifier is then trained using the learned features to classify different tissue types	WSIs split into 1000x1000 pixel tiles, and 25x25 pixel patches were extracted. These patches would be used to learn features and then classified for various tissue classes	4000 patches per tissue class, with cross-validation repeated 100 times	An overall classification accuracy of 84.3 % was achieved when distinguishing between necrosis, transition to necrosis, and tumour in GBM	Same dataset (cross-validation testing sets)	N/A
[84]	2020	197 glioblastoma (GBM) patients obtained from the Institutional Database of Spedali Civili of Brescia. Additionally, RNA sequencing analysis was performed on 51 cases	They propose an integrated molecular and immunohistochemical (IHC) approach combined with a machine-learning algorithm (random forest classifier) to predict glioblastoma transcriptional subtypes (GliTS)	RNA sequencing was first done to validate gene expression profile, and IHC staining was used to quantify these biomarkers. Hierarchical cluster analysis was performed using Kendall correlation coefficient as similarity metric and Ward criterion, and these features will be learned by the model	39 of 51 cases for testing, (assumed 12 cases used to train classifier)	Concordance between IHC classification and transcriptional data was used as a metric of accuracy. The predictive model achieved 79.5 % overall concordance, with a concordance of 90 % for the mesenchymal (MES) subgroup. The classical (CL) subgroup showed 81.3 % concordance, while the proneural (PN) subgroup had a lower concordance of 69.2 %	Same dataset (39 testing set)	N/A
[32]	2021	549 patient cases from TCGA dataset. This includes LGG and HGG (GBM). WSIs and molecular data were used	The proposed model integrates pathology images with molecular data using a deep neural network (DNN), including a regular DNN and the ResNet architecture. The model can incorporate cellularity features and molecular information for better classification	ROIs extracted from H&E images had their nuclei segmented using an Unet model trained on the MonuSeg dataset. Cellularity features were calculated based on these nuclei and passed to the model together with the colour-normalized stained tissue samples and associated molecular information	80 % training and 20 % testing using a 5-fold cross-validation	When classifying between HGG and LGG, the proposed model achieved 90.16 % classification accuracy with the pathology image alone, but when combined with genetic and cellularity features, the accuracy increased to 93.81 %. This improvement was also seen in grading LGG II and III (from 70.69 % to 73.95 %). This performed better than most other models, given that the performance of this model was trained and tested with a much larger dataset compared to the other models trained and tested by other authors	Same dataset (5-fold cross-validation)	CNN, SVM, ResNet, ElasticNet classifier, Decision tree

(*continued on next page*)

Table 3 (*continued*)

Authors	Publication Date (Year)	Training Dataset	Proposed model	Data Processing	training/test/validation parameters	Accuracy evaluation metric	Testing Dataset?	What models are being compared?
[85]	2015	Two data sets including about 2000 lung tumor cells and 1500 brain tumour cells. Where the data is collected is unspecified.	Used sparse reconstruction and stacked denoising autoencoders (sDAE) to build a cell segmentation model	Each cell in in the histopathological images was centralised in a 45x45 pixel patch for cell detection, and a 28x28 pixel patch for cell segmentation. Data augmentation is applied (rotation and random translation)	No clear mention on how the training and testing data was split	Primary measure of accuracy is F1 score. The proposed model achieved 0.96 F1 score for cell detection and 0.85 F1 score in segmentation performance in the brain tumour data, which outperformed other state-of-the-art methods like Laplacian-of-Gaussian (LoG), Iterative Radial Voting (IRV), ITCN, and Single-Pass Voting (SPV)	Same dataset (not clearly specified)	Laplacian-of-Gaussian (LoG), Iterative Radial Voting (IRV), ITCN, and Single-Pass Voting (SPV)
[86]	2016	190 GBM images from the TCGA dataset	Discriminative Feature-oriented Dictionary Learning (DFDL) method that learns class-specific dictionaries for automatic feature discovery	Patches were extracted from the images at 20x20 pixels, which would be vectorized for dictionary learning.	20 images of 190 images (3000x3000 pixel size) randomly picked for training	Overall accuracy for the TCGA dataset (containing GBM images) was 92.85 %, which is competitive compared to other dictionary learning methods like LC-KSVD and FDDL	Same dataset (remaining images used to test)	WND-CHARM, SRC, SHIRC, LC-KSVD, and FDDL
[39]	2017	Images from The Cancer Genome Atlas (TCGA), specifically focusing on WHO Grade II lower-grade glioma (LGG) and lung adenocarcinoma (LUAD) cases	A machine-learning-based semi-automated workflow to assess the quality of nucleus segmentation in histopathology images using texture features. The methodology uses a classification model trained on labelled image patches from segmentation results	512x512 pixel patches extracted from labelled regions, and a set of texture and intensity features is computed from the red, green, and blue channels. A stepwise variable selection is applied to reduce redundant features	10-fold cross-validation technique used to train and evaluate the model	F1 score. Random forest classifier achieved F1 scores of 84.71 %, 95.49 % and 73.76 % in good, under and over cases, respectively. It was able to outperform the SVM classifier in the testing set	Same dataset (10-fold cross validation)	Random forest and SVM classifiers were tested against each other to see which one performs better
[19]	2019	50 maximum intensity projection tissue images of glioblastoma cells	Proposed GRUU-Net, which combines the base U-net architecture with Gated Recurrent Units (GRU), which is capable of multi-scale feature aggregation through the CNN and iterative refinement using GRUs	Nuclei microscopy images were segmented by two experts by drawing contours around them. Data augmentation was performed	25 training, 5 validation and 20 test images	object-wise Jaccard similarity index (SEG), Dice and Hausdorff distance. GRUU-Net consistently outperformed (0.886 Dice and 0.648 SEG) in nuclei segmentation in the testing datasets compared to other state-of-the-art models like U-Net and UP-PT	Own 20 test images and 22 microscopy images from the Cell Tracking Challenge, which includes DAPI stained cell nuclei in GBM tissue	U-Net and UP-PT, ASPP-Net
[87]	2015	Unclear specification of where the dataset was collected from. Dataset include brain tumours, neuroendocrine tumours (NET) and breast cancer images	Proposed a deep convolutional neural network (CNN) framework to segment nuclei combined with a selection-based sparse shape model and a repulsive deformable model to separate individual nuclei	Small patches were cropped from the images (55x55x3 pixels), and the YUV colour space was used for the image representation. Augmentation techniques like rotation were applied to the patches	No clear mention on how the training and testing data was split	Dice coefficient (DSC), Hausdorff Distance (HD), and Mean Absolute Distance (MAD). The proposed model was able to segment better than more traditional segmentation techniques like marker-based watershed and graph-cut	Not clearly specified	Mean shift (MS), isoperimetric graph partition (ISO), superpixel (SUP), marker-based watershed (MWS), graph-cut and colouring (GCC), and repulsive level set (RLS)

(*continued on next page*)

Table 3 (*continued*)

Authors	Publication Date (Year)	Training Dataset	Proposed model	Data Processing	training/test/validation parameters	Accuracy evaluation metric	Testing Dataset?	What models are being compared?
						colouring, achieving a superior DSC of 0.85, HD of 5.06 and competitive MAD of 3.26		
[22]	2017	Two histopathology image datasets: the MICCAI 2014 Brain tumour Digital Pathology Challenge and a colon cancer dataset. The MICCAI dataset consisted of 23 GBM and 22 LGG images	Proposed a model in which a CNN (based on the AlexNet architecture) is used to extract necessary features to be pooled, selected and classified via an SVM classifier for segmentation and classification tasks	Overlapping image patches of 336x336 pixels were generated from H&E images. Patches with majority white background was excluded from training, and then further resized to 224x224 pixels.	5-fold cross-validation was performed for classification, and leave-one-out cross-validation for segmentation	The proposed model achieved 97.5 % classification accuracy in the MICCAI 2014 Brain tumour Digital Pathology Challenge for distinguishing GBM from LGG. For segmentation, the model achieved 84 % accuracy in detecting necrosis vs. non-necrosis regions	Same dataset	Compared against other state-of-the-art methods in the training data from MICCAI 2014 challenge.
[49]	2018	GBM and LGG histopathological images from TCGA, with 100 images per glioma subtype being collected for this study	Proposed model is a deep CNN that has been to train to classify the glioma subtypes	H&E images were cropped into 1000x1000 patches, and 100 image patches were generated per image.	4-fold cross-validation technique to train and test model	Classification accuracy. The proposed model achieved 96.5 % mean classification accuracy, which is comparable to VGGNet (97 %) and better than ZFNet (95.2 %) and LeNet (79.4 %).	Same dataset (4-fold cross-validation)	LeNet, ZFNet, and VGGNet
[41]	2020	WSIs and clinical data of 490 brain cancer patients from the TCGA dataset was used.	DeepSurvNet, a deep CNN (Based on GoogleNet) that is designed to classify brain cancer patient survival rates based on histopathological images	ROIs from H&E images had their patches extracted at 256x2566, 512x512 and 1024x1024 pixels.	80 % training, 18 % validation, and 2 % testing sets, trained and test with 3 different testing folds	The model achieved precision and AUC of 0.99 and 1 with 256x256 patches and used the GoogleNet architecture, which outperformed other state-of-the-art models like InceptionV3 and ResNet50. The best model also achieved an average precision and AUC of 0.8 and 0.96, respectively, with the independent dataset.	Same dataset (2 % testing set) and independet dataset of tissue samples collected from 9 patients at SA Pathology	VGG19, GoogleNet, ResNet50, InceptionV3, and MobileNetV2
[53]	2021	The Ivy GAP dataset, which includes 32 GBM patients diagnosed by primary surgery type with a total of 805 WSIs	Proposed a DenseNet segmentation model (based on the tiramisu design) to segment various regions in the tumour.	H&E images were resized to 1024x1024 pixels, and patches of size 512x512 pixels were extracted for training. Data augmentation techniques such as random crops and vertical flips were applied	Training, validation and test sets were split into 14:1:1 ratio.	Accuracy score (Rand index). The model achieved an overall accuracy across ~70 % across all eight classes.	Same dataset. The trained model was applied to the TCGA dataset for further analysis with the various segmented regions	N/A

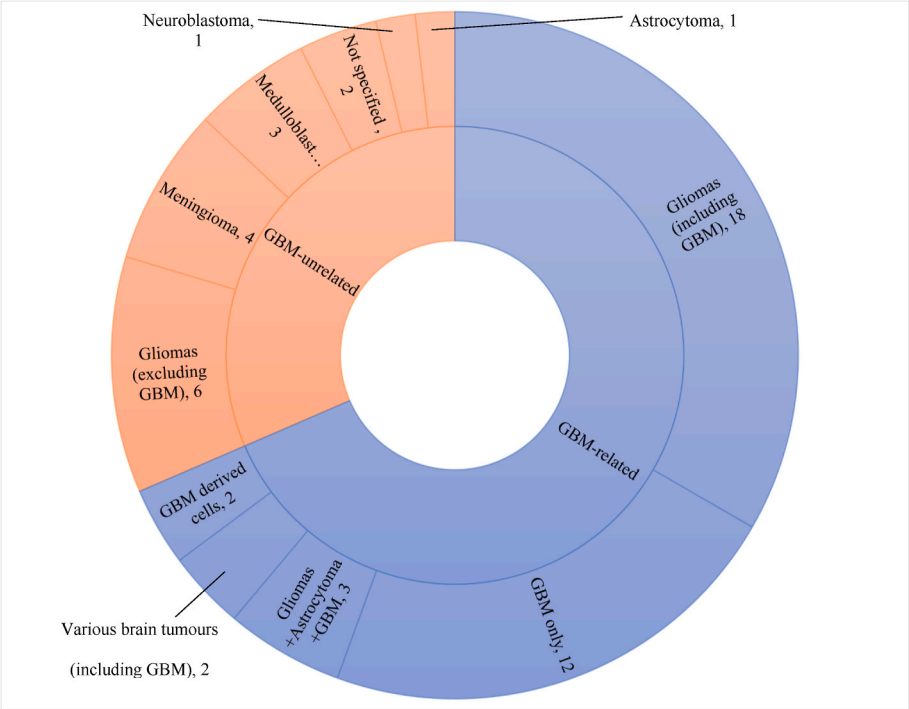

Fig. 3. Distribution of what brain tumour types the studies focused on. Studies were grouped either under GBM-related or GBM-unrelated studies, and further stratified to the specified types of brain tumours that has been specified. Data derived from Tables 2 and 3

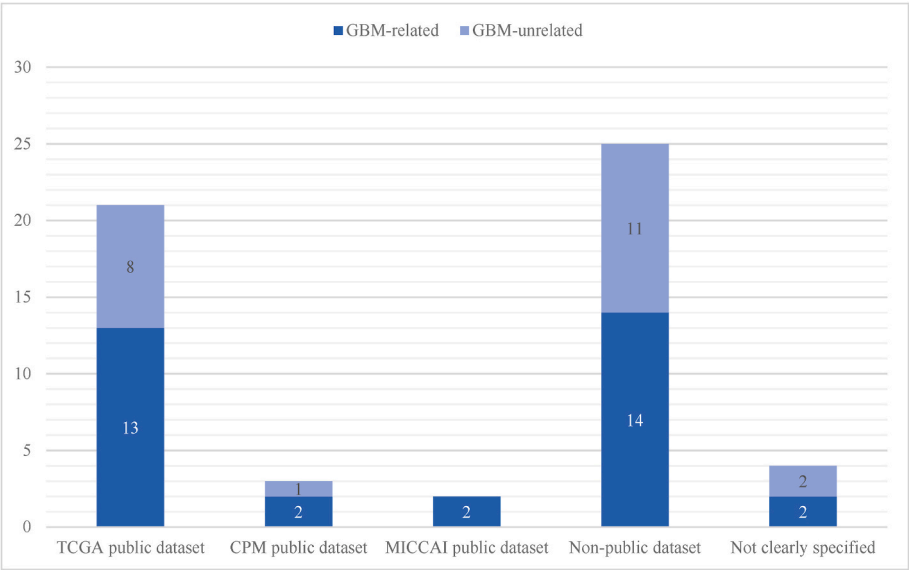

Fig. 4. Studies were distributed by the dataset they used. The top 3 public dataset were compared against the number of articles that collected their data from non-public dataset, and against articles that did not clearly specify where they collected their data. Abbreviations: Computational Precision Medicine (CPM), Genomic Data Commons (GDC), Glioblastoma (GBM), Medical Image Computing and Computer Assisted Intervention (MICCAI), The Cancer Genome Atlas (TCGA). Data is derived from Table 3.

astrocytomas) and more [20]. Much of the same trend is observed in GBM-related studies, suggesting that the studies tended to investigate GBM concerning clinical and biological data.

Known clinical data, such as overall survival and tumour subtypes, were typically paired with H&E data. Studies utilising the survival + H&E data typically model their ML/DL pipelines to predict patient survival based on H&E images. In contrast, tumour subtypes/grades + H&E data were used for tumour subtype/grade classifications. However, while these studies provide a good tool for accessible and affordable prognosis/diagnosis, they lack additional insights into GBM disease pathology as they lack additional context from biological data. These analyses were referred to as "Typical" analyses for the rest of the article. This category represents the majority of studies, as 26 articles in this review have been reported to perform typical analysis (Fig. 6). Ertosun and Rubin illustrate the methodology of "typical" analysis as they have built an accurate model that could distinguish GBM from LGG, and even grade LGG based on H&E-stained images only [29].

However, several articles (12 out of 54; Fig. 6) performed more

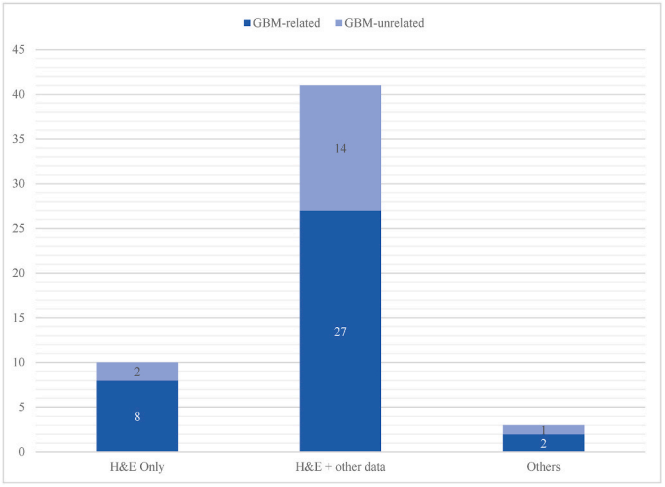

Fig. 5. The data types used in their studies distributed the number of articles. Three categories that the studies were distributed into are: Studies only using H&E (H&E only), studies combining H&E with various data types (H&E + other data types) and studies using different histological data types altogether (Other). Data is derived from Tables 2 and 3

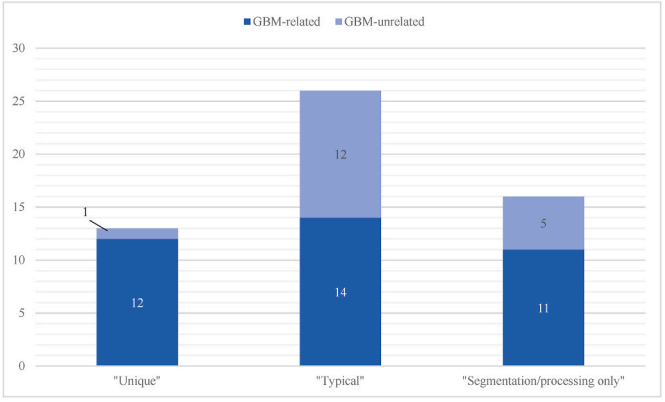

Fig. 6. The number of studies distributed by the types of ML/DL-aided analyses. Typical analyses included H&E + tumour subtype/grade or H&E + survival data. Unique analyses include H&E data combined with various other data types (genomic, proteomic, etc.) for their analyses. The remaining articles only used ML/DL techniques to investigate segmentation or processing performance. Data is derived from Tables 2 and 3

unique analyses using additional data sets, such as genomic or proteomic data, to provide biological context to the analysis/prediction, thus gaining insight into the disease, particularly for GBM. For example, Tan and colleagues used H&E images and corresponding survival and mRNA expression datasets to classify GBM and LGG grades and predict their overall survival [30]. Similarly, Wang and colleagues utilised the Glioma CpG island methylator phenotype (G-CIMP), telomere length and immune response profiles to associate with various histopathological features in H&E images. This has provided some much-needed context in the overall classification task, as it allowed Wang and colleagues to find associations in the various immune profile with histologic features in their predictive model [31]. In addition, several articles in the "unique analysis" have also shown marked improvements in model prediction tasks when biological context is given. For instance, Pei and colleagues designed a model that would predict HGG and LGG and found that by combining the genetic features from genomic profiling and H&E-based cellularity features, they were able to enhance the prediction accuracy of the model, compared to a model without genetic features [32]. Similarly, Tan and colleagues reported improved performance of their model in predicting long- and short-term overall survival in LGG patients by integrating associated mRNA biomarkers and H&E features collected from the TCGA dataset, although the selected biomarkers were not specified [30]. This suggests a relationship exists between the H&E-derived feature set and mRNA biomarker-derived feature set that would improve predictive capabilities, which the authors can further explore.

We also report that the number of articles is spread relatively evenly between the three categories of analyses within the GBM-related studies, indicating (12 "uniquely analysed" studies, 14 "typically analysed" and 11 "segmentation/processing" studies; Fig. 6). Furthermore, most studies categorised as "Unique" are GBM-related (Fig. 6), indicating that context and biological insight-driven studies in ML/DL-aided GBM research are more critical than other brain tumour research. Interestingly, the oldest article published in the "unique analysis" category is in 2020, indicating that this type of analysis is relatively novel.

We further stratified the articles within the "Unique analysis" category to investigate how many of them used biological data effectively to explore relationships between it and histological data. We found that only 8 of 12 GBM-related studies within the "Unique analysis" category explored the relationship between biological features and histological features to gain insights into GBM pathology, while the other 4 GBM-related studies only used the biological features as contextual data to improve their prediction model (Fig. 7).

Overall, studies conducting "unique analysis" are relatively novel. Integrating H&E data, additional -omics data, and clinical data has proven critical in better understanding GBM pathology and enhancing the predictive capabilities of ML/DL models.

3.4. ML/DL utilisation

In this section, the overarching aim is to ascertain what model architectures and how these models are commonly used. To achieve this, we collected the type of preferred model architecture used and the task it performed. Moreover, we collected the training parameters, evaluation metrics and the common model architectures used to compare with their preferred models (Table 3). This would clearly indicate how standardised the training parameters and evaluation methodology are in GBM-related studies.

The eligible studies utilised ML/DL techniques for predictive classification and segmentation tasks. Typically, ML classifier and clustering algorithms were used on histopathological data and various clinical data to cluster and predict classes, whether it be: (1.) tumour subtypes or tumour grades [29,33,34], (2.) predefined survival classes [34–36], (3.) nuclei types or feature types in the tissue [37,38], and in one case, it was

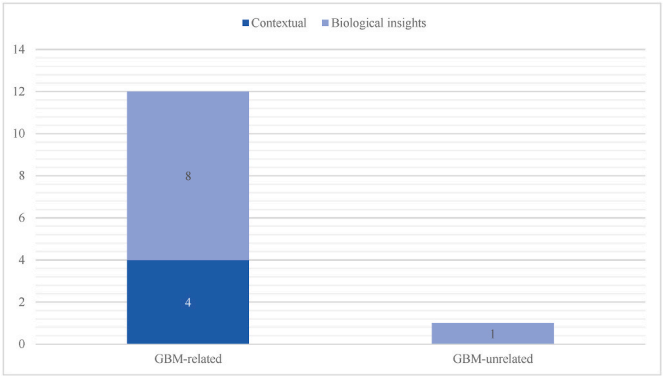

Fig. 7. Distribution of articles within the "Unique analysis" category that either used biological purely as a "Contextual" information without further analyses, or deeply explored the relationship between to biological data and histological data to offer additional "Biological insights". Data derived from Table 3.

used to determine the quality of nuclei segmentations [39](Table 3). Moreover, DL CNN models were commonly used to extract features from histopathological images for both segmentation and classification tasks. For instance, Chang and colleagues developed the RESEPT network containing the ResNet101 backbone network to segment and identify GBM's characteristic elongated nuclei. They could match these infiltrative tumour-marking genes, thus allowing them to distinguish tumour, non-tumour and infiltrating tumour zones [40]. In another example, Shirazi and colleagues used various CNNs to be trained using H&E data and corresponding survival data so that it can accurately predict their survival classes [41]. However, the common trend observed from the literature survey was that the CNN architectures were combined with classifiers to create robust models. Liu and colleagues implemented HoVer-Net CNN to first segment neuroblastoma nuclei and have their features extracted. They then implemented K-means clustering, a classifier algorithm, to cluster the features and learn visual words (e.g., cluster centres). These were used to predict patient survival and thus predict the risk to the pathological prognosis [42].

In the analysed studies, some classifier algorithms were used more often than others. Classifier/clustering algorithms include SVM, k nearest neighbour (kNN)/k-Means clustering and random forest (RF) classifiers. In Fig. 8-A, the most common classifiers used in the preferred models were derivatives of SVM (6 articles overall; Fig. 8-A), kNN/K-Means clustering (3 articles overall; Fig. 7-A) and RF classifiers (3 articles overall; Fig. 8-A). In the GBM-related studies, the SVM remained the most popular classifier (3 articles; Fig. 8-A).

Of note are the studies utilising unique techniques to solve their classification problems. In one case, Lessmann and colleagues used the self-organizing map (SOM) ML algorithm to cluster certain features in meningioma H&E images and classify their subtypes (Meningothelial, Fibroblastic, Transitional, Psammomatous)[43]. Several other studies utilised less popular algorithms, such as the linear and quadratic discriminant analysis (LDA and QDA, respectively), naïve Bayes, and t-distributed Stochastic Neighbour Embedding (t-SNE) to solve their classification problems [20,23,35,44,45].

Another popular DL technique is the CNN architecture. In the relevant studies, many authors used the deep residual networks, better known as ResNet, and many of its derivatives, as the primary architecture for their CNN models. 6 GBM-related studies utilised many versions of the ResNet model or built their own derivative model, while the 2nd most popular model used in GBM-related studies was the U-Net model and its derivatives; 3 studies featured derivatives of U-Net as their main model (Fig. 8-B). This was followed closely by the VGG model; 4 overall studies, with 2 GBM-related studies using VGG-derived models as their preferred model (Fig. 8-B). Moreover, the U-Net model and its derivatives are used primarily for segmentation tasks [19,32,46]. Models such as the ResNet and VGG derived models tended to be used to extract features from the images, and with the model's decoding feature or with the help of classifiers, it was used for classification tasks [31,45,47,48] Interestingly, several studies opted for simpler architectures as their model of choice. For example, Yonekura and colleagues developed a custom but simple CNN network consisting of 7 convolutional layers. They trained their model on GBM and LGG H&E images to discriminate between the two types of gliomas [49].

We report that SVM and kNN-based classifiers are also often used as benchmark classifiers (Fig. 9-A). At the same time, the ResNet, U-Net and Inception-based CNN architectures remain the most common state-of-the-art benchmarking model in GBM-related studies (Fig. 9-B). This solidifies our findings that many of the models used in the GBM-related research have revolved around the refinement of existing and state-of-the-art models in the past 2 decades.

When investigating how the models were trained and evaluated for their performance, we found that most of the train-evaluation strategies the studies used were variations of the cross-fold validation strategy (23 articles with 13 articles being GBM-related; Fig. 10-A). The cross-validation strategy is considered a gold standard in evaluating the robustness of a model, as it builds validation metrics on several models trained on multiple sub-parts of a training set, known as "folds" [50]. Although the gold standard train-evaluation strategy represents a large portion of GBM-related studies, the number of studies reporting the usage of single train/test or train/validation/test splits is concerning (17 out of 37 articles; Fig. 10-A). Even more concerning, 5 out of 37 GBM-related articles did not specify how their model's performance was evaluated. This suggests that model robustness may not be at the forefront of the ML/DL model design ethos in GBM-related studies.

We also observed that many GBM-related studies used the same dataset to evaluate their model, as 30 of 37 GBM-related studies have reserved a portion of their training dataset as their testing dataset to evaluate their models' performance (Fig. 10-B). However, 4 GBM-related studies collected an independent testing dataset to evaluate their performance. For instance, Shirazi and colleagues collected histopathological data from 9 patients from a locally derived dataset to evaluate the performance of their model, which was trained on the TCGA public dataset. They achieved competitive performance in predicting overall survival classes in the locally derived dataset compared to the TCGA dataset [41]. This showcased good generalizability in their model and further strengthened their model design.

In Fig. 11, we report the evaluation metrics employed by studies to benchmark their segmentation and classification tasks. We report that the classification accuracy and area under the receiver operating characteristic curve (AUC) evaluation metric (16 and 7 GBM-related studies, respectively; Fig. 11) is most used in evaluating classification performance, and we observe that the two metrics are often paired together. Similarly, the two most common evaluation metric in segmentation tasks, the DICE score and Accuracy score (also known as the Rand Index) (8 and 7 GBM-related studies found to use these metrics, respectively; Fig. 11), are often paired together as well. This suggests that the standard pairing of evaluation metrics for segmentation tasks and

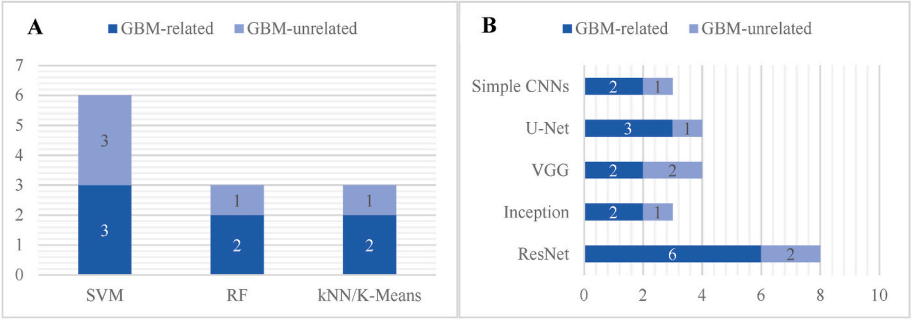

Fig. 8. Distribution of popular classifiers (A) and convolutional neural networks (CNN)(B) used either as the main architecture or a base-architecture that has been built upon, in the relevant studies. Abbreviations: Support vector machine (SVM), k nearest neighbour (kNN), random forest (RF). Data is derived from Table 3.

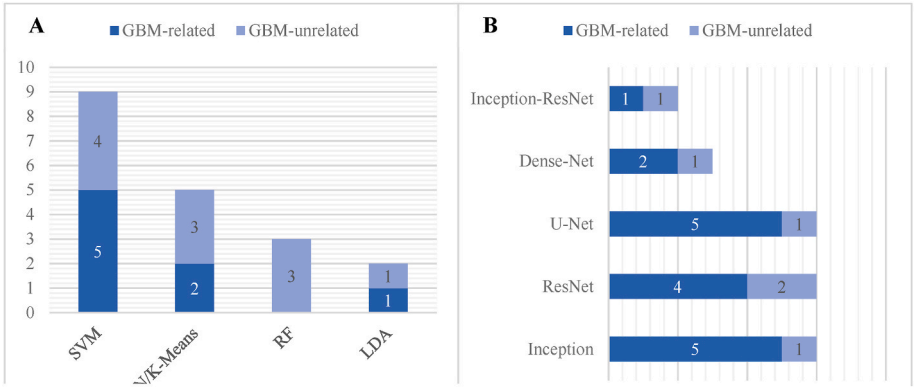

Fig. 9. Distribution of popular state-of-the-art classifiers (A) and convolutional neural networks (CNN)(B) used as a benchmark comparison in relevant studies. Abbreviations: Support vector machine (SVM), k nearest neighbour (kNN), random forest (RF). Data is derived from Table 3.

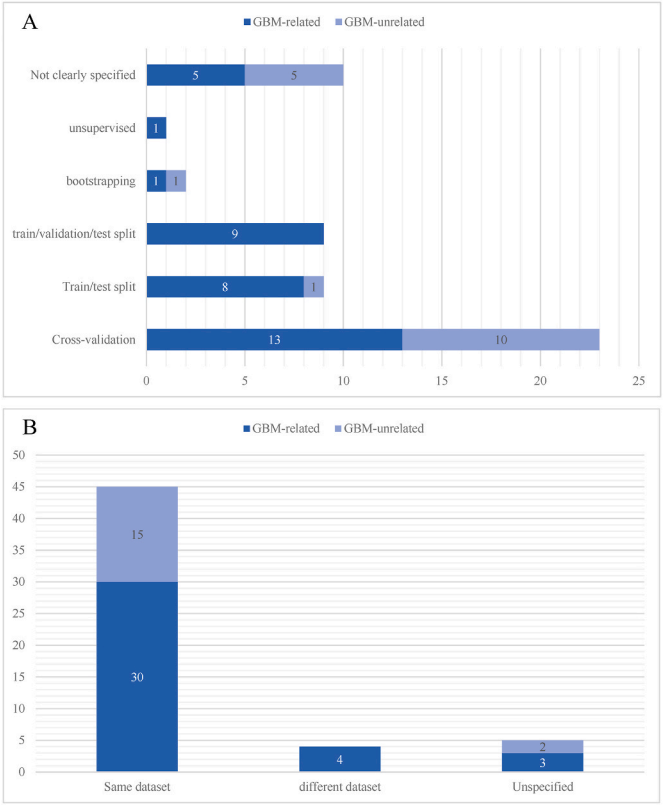

Fig. 10. Distribution of A.) various training-evaluation strategies of the relevant studies and B.) the type of dataset used to evaluate the ML/DL model performance. Data is derived from Table 3.

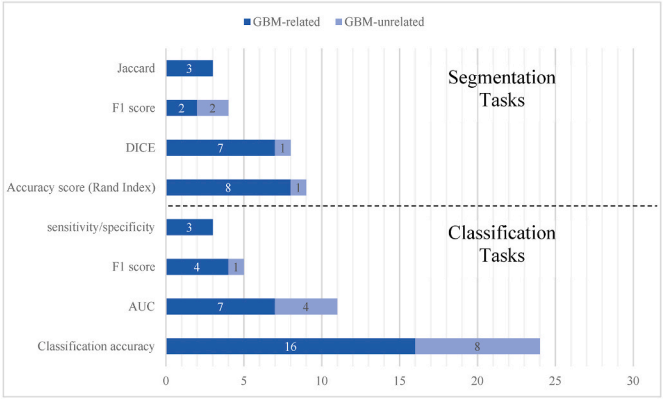

Fig. 11. Distribution of top 4 evaluation metrics used for segmentation and classification tasks. Studies that did not evaluate model performance were omitted from this analysis. Data is derived from Table 3.

classification tasks should be DICE + Rand Index and Classification Accuracy + AUC, respectively.

A notable observation made in the analysed studies was that many of the studies' objectives were to evaluate their model performances against other models. Still, more recent studies tended to apply their model for novel discoveries or complex biological analyses rather than evaluative purposes. For instance, Pei and colleagues utilised the U-Net-ResNet hybrid model to discern various types of gliomas and had their model performance evaluated. They used the segmented nuclei features, histopathologic features and IDH statuses for their classification tasks, and they evaluated its performance at 93.81 % classification accuracy between GBM and LGG, and a 73.95 % classification accuracy between LGG grade II and III [51]. On the other spectrum, Ravi and colleagues' approach to utilising ML/DL techniques differs, as they integrated their VGG classification model with various other data forms to create a complex analysis system that investigates how GBM can dynamically adapt to various environments. In Ravi's study, they used a VGG pre-trained model to classify tumour histologic microenvironments in GBM whole slide image (WSI) patches (infiltrating, necrosis, necrotic edge, cellular, and vascular microenvironments). The histologic phenotype classified by the model was integrated with transcriptomic, proteomic and metabolic data to contextualise the histologic phenotype and investigate recurring transcriptomic patterns spatially. They consequently characterised GBM at various molecular levels in a spatially resolved manner. They discovered that metabolic alterations such as hypoxia could lead to significant gene copy-number alterations. The data indicated that regional hypoxia metabolism ultimately represents drivers of microevolution that enable the evolution of therapy-resistant phenotypes [48]. The main critique is that although Ravi and colleagues used ML/DL tools in a novel way, their scope of studying hypoxic areas may be narrow, considering that GBM exhibits high inter- and intra-tumoural heterogeneity [5,6]. Overall, this could suggest that although the ML/DL integration into GBM research is nascent, it is slowly evolving and maturing to use these tools for more complex analyses and, thus, for potentially novel discoveries.

4. Discussion

Our literature review found that most relevant articles were published recently, suggesting that ML/DL utilisation in histopathological brain tumour research is nascent. Our literature review further

corroborates this, as many relevant articles mainly evaluated the ML/DL models' performance for various tasks [22,36,52]. Moreover, several studies have explored more complex analyses using ML/DL techniques to make potentially novel discoveries by integrating histological data from proteomic and genomic studies [41,48,53]. The additional bio-information has proven advantageous in improving model performance when provided in context with histological data [32,54]. Because of this, we envisage that studies using ML/DL-aided complex analyses will become the predominant form of study in the near future as the field matures. However, we also believe there is still considerable work to be done in the field of ML/DL-aided brain tumour histopathological research, as the ML/DL-aided histopathological research resulting in biological insights into GBM are few relative to the evaluative performance studies in our literature survey.

Furthermore, many of the relevant studies in our literature review investigated the general characterisation of brain tumour grades and/or subtypes [20,40,55], with some studies investigating further via characterisation of tumour histologic microenvironments in the WSIs [48,53,56]. We believe that ML/DL-aided investigations into GBM histologic tumour microenvironments will become more prominent moving forward, as GBM is known to be characterised by its high inter-tumoural and intra-tumoural heterogeneity [57,58], and this is expressed at various inter-related levels (histologic, cellular and molecular level). Notably, Phillips and colleagues showed that high-grade gliomas (HGG), which includes GBM, can be characterised into three subtypes using a panel of molecular gene expression profiles: proneural, proliferative and mesenchymal. They also found a correlation between the molecular subtypes and histologic grading of HGG. Moreover, they reported that proneural subtypes tended to have more prolonged survival than the proliferative and mesenchymal subtypes [7]. In a more recent study, Garofano and colleagues suggested that GBM can be subtyped based on biological traits of single cells and bulk tumours, and they were distributed along the neurodevelopmental and metabolic axis: proliferative/progenitor, neuronal, mitochondrial and glycolytic/plurimetabolic subtypes. Moreover, Garofano and colleagues found that mitochondrial GBM subtypes were particularly vulnerable to oxidative phosphorylation inhibitors. Finally, they postulate that this pathway-based classification of GBM can inform survival and enable more precise targeting of cancer metabolism [59]. Another study by Martinez-Lage and colleagues reported that immune heterogeneity exists in different subtypes of GBM and that mesenchymal GBMs showed the highest levels of immune infiltration. In contrast, proneural GBMs were observed to have the least infiltration. Likewise, they reported that various types of T-lymphocytes are heterogeneously involved in various subtypes, with higher percentages of CD163+ T-cells associated with worse prognosis [60]. This showcases that various aspects of GBM play a key role in its pathogenesis and that even within an aspect of GBM, its nature can be heterogeneous, thus culminating in many complex relationships in GBM that must be accounted for to understand its pathology fully.

Recent studies have given some hope for a better understanding of various aspects of GBM pathogenesis. Namely, Ravi and colleagues effectively integrate ML/DL segmentation and classification of histologic features with transcriptomic, proteomic and metabolic features to create a multi-layered and accurate characterisation of necrotic features of GBM [48]. Indeed, the mentioned study has constrained its scope to a smaller aspect of GBM, which would not be the best representation of GBM. However, we understand its limitations as the investigative efforts into other complex aspects of GBM would require a significant increase in expensive multi-omics and resources needed to complete such a study. This challenge seems to be understood by other studies undertaking similar integrative research in our literature review, hence the relatively low number of this type of research. Given this limitation, it would still be possible to integrate various "smaller" relationships that studies such as Garofano's and Martinez-Lage have found and integrate histologic features for the ML/DL models to learn and predict. With a more collaborative approach, these trained models for various constrained relationships can be integrated into a larger "macro" model. This approach would have to be an iterative approach, as newer discoveries will prompt the development of models to relate these insights to histologic features, which would then be embedded into the "macro" model. As a result, the "macro" model would provide the most feature-rich prediction and context to histopathological work, while maintaining a good representation of GBM's heterogeneous nature. Although, in practicality, this approach would require a substantial amount of collaborative effort and investments, we believe it to be a necessary step, in the long term, towards fully understanding GBM pathology and providing a comprehensive yet accessible tool for clinicians and researchers alike.

Our literature review also investigated the types of classifiers and CNN models used in various brain tumour research. We report that SVM, RF and kNN/k-Means are amongst the most common classifiers today for classification tasks specific to brain tumour research. ResNet and U-Net were the more common CNNs used in our literature review. Our brief literature inquiry suggests that recent studies used the same common ML classifiers we found in our systematic literature review in the scope of classifier tool utilisation. For example, Kang and colleagues used a collection of classifiers, including SVM, RF and kNN, to evaluate their performance in classifying MRI brain tumour images. They found that SVM is the optimum classifier for their use case. Interestingly, Kang and colleagues also considered the AdaBoost classifier, an ensemble of classifiers combined to produce more accurate classification outcomes, although it did not perform as well as SVM [61]. In another example, Saha and colleagues developed EMCNet, an automated COVID-19 diagnosis tool with an ensemble of ML classifiers to detect COVID-19 presence from patient X-ray images. They combined SVM, decision tree and AdaBoost into an ensemble classifier for a more accurate classification performance [62]. Our literature inquiry determined that ensemble classifiers seemed popular and accurate, and these types of classifiers were not considered as thoroughly in the brain tumour histopathology research field. Similarly, as illustrated by Kang and colleagues, ensemble CNNs can also combine feature extraction layers from various CNNs (including ResNet) to generate a more accurate classification performance overall [61]. Upon a literature inquiry into the CNNs available for biomedical applications, although we found studies that used derivations of ResNet and U-Net, we also observed a variety of models that combined ResNet or U-Net into other models. For example, Upschulte and colleagues proposed a contour proposal network (CPN), which can detect overlapping objects in an image, and it could implement a U-Net or ResNet backbone into their network. As a result, their model outperformed basic U-Net and ResNet models in detecting various objects, including neuronal cell bodies, U2OS cell nuclei and synthetic shapes [63]. Given the "macro" model approach that we have suggested, we suggest the implementation of deeper and more robust models such as the ResNet backboned models or ensemble models, so that more biological data can be integrated to improve performance and robustness of the final model.

The biggest concern noted in our literature review is that the number of studies that failed to report how the dataset was split for training and evaluation was larger than expected. Secondarily, a larger-than-expected number of studies forego the evaluation of their model robustness through cross-validation approaches, which is the gold standard in assessing the variance in accuracy predictions [50]. Furthermore, most of the articles in this literature review curate an evaluative set from the same dataset from which the training set was built, albeit the evaluative set does not contain any data that the model would have "seen" in the training set. This could still pose the issue of generalizability, especially when 14 of the 37 GBM-related studies collect their data from a single source of non-public dataset. Moreover, Thakkar and colleagues noted in their epidemiologic review of GBM that Asian Pacific Islanders have significantly better survival rates than whites and blacks at all time differences and that survival rates between

men and women differ as well [64]. These population-based differences are very important and are not as well represented when a mono-sourced dataset is used. Therefore, it is reasonable for clinicians to hold the accuracy variance and generalizability of the model to a very high standard, as the predictions from said model could determine the course of the patients' lives. We should uphold these standards so that the performance of our model can instil confidence in our clinicians to make the correct decisions when treating our patients. Lastly, we noted that several articles had poor reporting on where their data is sourced and how their models were trained and evaluated, which would result in poor model reproducibility. Thus, we suggest: 1.) using public datasets containing data from multiple institutions in combination with locally derived datasets to improve model generalizability, and 2.) A standard method of reporting includes where the data is sourced from and how many were sourced, how the data was split into cross-validation folds, and the results of the accuracy metrics (Classification accuracy and AUC for classification tasks, and DICE and Rand Index scores for segmentation tasks). This would improve reproducibility and ensure a standardised reporting of model robustness and generalizability, thus facilitating the collaborative efforts of designing the overarching "macro" model and guaranteeing the adoption of the model in the clinical setting.

We acknowledge that the scope has constrained the scope to only histopathological studies, but much of the literature in our initial survey utilised ML/DL tools in MRI/CT patient brain scans for brain tumour research. While important, there were relatively few ML/DL-aided histopathological studies compared to purely MRI/CT-based studies, as we found in our literature survey. We see this as a gap in the overall brain tumour research as MRI/CT data are known to be expensive to image and collect. In contrast, pathological slides are the standard and more cost-effective procedure to classify patient tumour type, grade and potential prognosis [65]. Hence, histopathological studies and analyses are more accessible than MRI/CT data in less privileged areas. Therefore, we believe that a wider usage of ML/DL techniques in histopathological studies is important in the future so that more ML/DL tools for analysing histopathological samples for prognostic and diagnostic purposes are made accessible for clinical use.

5. Conclusion

In conclusion, we have conducted a systematic review of the literature concerning ML/DL-based studies on brain tumours, particularly in GBM, and analysed the current trends in the how and why ML/DL models were used in histopathological studies. We sought to understand how these ML/DL-based studies uncovered biological insights towards GBM pathology. This would allow us to uncover the cutting-edge ways that studies have used ML/DL models to learn more about GBM pathology, and to find potential improvements in how the ML/DL-based GBM should be conducted.

We report that the majority of articles in this review focused on designing and improving model performance in classifying tumour grades/subtypes (i.e. classifying between GBM and LGG grades) or segmenting tumour microenvironments (necrotic region, tumour core etc.), while a small number (8 out of 37 GBM-related studies) implemented the models in a way that effectively integrates biological data with histopathological data to gain new insights into GBM pathogenesis. We also found that the two main types of models were utilised in this review were classifier models (SVM and kNN being the most popular classifiers) and CNNs (ResNet and U-Net based architectures are the most popular CNNs).

However, we also uncovered limitations in the articles in this review which concerned us. Namely, we found that a larger-than-expected number of articles failed to report where the data was sourced from which gives rise to reproducibility concerns. Furthermore, over half of the GBM-related studies did not use the gold standard cross-validation strategy and curated evaluation sets from the same dataset as the training set, which posed a concern towards model robustness and generalizability, respectively.

Overall, we postulate that ML/DL-based GBM research would evolve to be more collaborative in nature, to tackle the complexity of the disease. This would culminate to the design of a "macro" model that would integrate various smaller biology-contextualised and histology-based predictive models for various aspects of GBM. These suggestions below are the key actionable steps that would facilitate that evolution.

- Integration of genomic or proteomic data in the models with histological data, by extracting the -omics features (i.e. marker presence, shape, texture) and combining it with extracted histological features (i.e. nuclei shape, texture, cellularity), and training the model to predict the multi-omics features
- Designing ensemble models (i.e. SVM-kNN ensemble classifiers, or ResNet-Inception ensemble CNNs) and deeper CNN models (i.e. ResNet-backboned U-Net models) to achieve the task. This would facilitate the design of models that can integrate more biological data and provide more robust and accurate models overall
- Using multiple datasets, preferably one from a publicly available dataset like TCGA and a locally derived dataset (i.e. from a hospital and local research centre) to train and evaluate the ML/DL models. This would improve model generalizability and instil confidence in clinicians to adopt the models in a clinical setting
- Standardised training/testing/evaluation splits and standardised reporting of it:
  - o Clear description of where the datasets are acquired and how much data was collected
  - o How the data was split for cross-validation
  - o Clear description and reporting of the evaluation metric (Classification accuracy and AUC for classification tasks, and DICE and Rand Index scores for segmentation tasks)

CRediT authorship contribution statement

Chun Kiet Vong: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alan Wang:** Writing – review & editing, Supervision. **Mike Dragunow:** Writing – review & editing. **Thomas I-H. Park:** Writing – review & editing, Supervision. **Vickie Shim:** Writing – review & editing, Validation, Supervision.

Ethics statement

No ethics required for this systematic review.

Declaration of competing interest

The authors declare that there are no competing financial and/or personal interests that would influence the work reported in this article.

Acknowledgements

We acknowledge The University of Auckland's Post-Graduate Scholarship for making this work possible.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2024.109642.

References

[1] H. Ohgaki, P. Kleihues, The definition of primary and secondary glioblastoma, Clin. Cancer Res. 19 (2013) 764–772, https://doi.org/10.1158/1078-0432.CCR-12-3002/85807/AM/THE-DEFINITION-OF-PRIMARY-AND-SECONDARY.

[2] C. Luo, K. Song, S. Wu, N.U.F. Hameed, N. Kudulaiti, H. Xu, Z.Y. Qin, J.S. Wu, The prognosis of glioblastoma: a large, multifactorial study, Br. J. Neurosurg. 35 (2021) 555–561, https://doi.org/10.1080/02688697.2021.1907306/SUPPL_FILE/IBJN_A_1907306_SM6507.DOCX.

[3] B. Campos, L.R. Olsen, T. Urup, H.S. Poulsen, A comprehensive profile of recurrent glioblastoma, Oncogene 35 (2016) 5819–5825, https://doi.org/10.1038/onc.2016.85.

[4] M. Weller, T. Cloughesy, J.R. Perry, W. Wick, Standards of care for treatment of recurrent glioblastoma—are we there yet? Neuro Oncol. 15 (2013) 4, https://doi.org/10.1093/NEUONC/NOS273.

[5] D.A. Nathanson, B. Gini, J. Mottahedeh, K. Visnyei, T. Koga, G. Gomez, A. Eskin, K. Hwang, J. Wang, K. Masui, A. Paucar, H. Yang, M. Ohashi, S. Zhu, J. Wykosky, R. Reed, S.F. Nelson, T.F. Cloughesy, C.D. James, P.N. Rao, H.I. Kornblum, J. R. Heath, W.K. Cavenee, F.B. Furnari, P.S. Mischel, Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA, Science 343 (2014) 72, https://doi.org/10.1126/SCIENCE.1241328.

[6] J.M. Stommel, A.C. Kimmelman, H. Ying, R. Nabioullin, A.H. Ponugoti, R. Wiedemeyer, A.H. Stegh, J.E. Bradner, K.L. Ligon, C. Brennan, L. Chin, R. A. DePinho, Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies, Science 318 (2007) 287–290, https://doi.org/10.1126/SCIENCE.1142946/SUPPL_FILE/STOMMEL.SOM.PDF, 1979.

[7] H.S. Phillips, S. Kharbanda, R. Chen, W.F. Forrest, R.H. Soriano, T.D. Wu, A. Misra, J.M. Nigro, H. Colman, L. Soroceanu, P.M. Williams, Z. Modrusan, B.G. Feuerstein, K. Aldape, Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis, Cancer Cell 9 (2006) 157–173, https://doi.org/10.1016/J.CCR.2006.02.019/ATTACHMENT/0B967CEF-1B18-4C7E-90C7-7AF2804EB01D/MMC7.XLS.

[8] R.G.W. Verhaak, K.A. Hoadley, E. Purdom, V. Wang, Y. Qi, M.D. Wilkerson, C. R. Miller, L. Ding, T. Golub, J.P. Mesirov, G. Alexe, M. Lawrence, M. O'Kelly, P. Tamayo, B.A. Weir, S. Gabriel, W. Winckler, S. Gupta, L. Jakkula, H.S. Feiler, J. G. Hodgson, C.D. James, J.N. Sarkaria, C. Brennan, A. Kahn, P.T. Spellman, R. K. Wilson, T.P. Speed, J.W. Gray, M. Meyerson, G. Getz, C.M. Perou, D.N. Hayes, An integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR and NF1, Cancer Cell 17 (2010) 98, https://doi.org/10.1016/J.CCR.2009.12.020.

[9] D.N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O.D. Wiestler, P. Kleihues, D.W. Ellison, The 2016 World Health organization classification of tumors of the central nervous system: a summary, Acta Neuropathol. 131 (2016) 803–820, https://doi.org/10.1007/s00401-016-1545-1.

[10] P. Blanc-Durand, A. Van Der Gucht, N. Schaefer, E. Itti, J.O. Prior, Automatic lesion detection and segmentation of 18F-FET PET in gliomas: a full 3D U-Net convolutional neural network study, PLoS One 13 (2018) e0195798, https://doi.org/10.1371/journal.pone.0195798.

[11] A.P. Nanthagopal, R.S. Rajamony, A region-based segmentation of tumour from brain CT images using nonlinear support vector machine classifier, J. Med. Eng. Technol. 36 (2012) 271–277, https://doi.org/10.3109/03091902.2012.682638.

[12] J. Unkelbach, T. Bortfeld, C.E. Cardenas, V. Gregoire, W. Hager, B. Heijmen, R. Jeraj, S.S. Korreman, R. Ludwig, B. Pouymayou, N. Shusharina, J. Söderberg, I. Toma-Dasu, E.G.C. Troost, E. Vasquez Osorio, The role of computational methods for automating and improving clinical target volume definition, Radiother. Oncol. 153 (2020) 15–25, https://doi.org/10.1016/j.radonc.2020.10.002.

[13] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, Y. Fan, A deep learning model integrating FCNNs and CRFs for brain tumor segmentation, Med. Image Anal. 43 (2018) 98–111, https://doi.org/10.1016/j.media.2017.10.002.

[14] M. Ali, S.O. Gilani, A. Waris, K. Zafar, M. Jamil, Brain tumour image segmentation using deep networks, IEEE Access 8 (2020) 153589–153598, https://doi.org/10.1109/ACCESS.2020.3018160.

[15] C.G.B. Yogananda, B.R. Shah, M. Vejdani-Jahromi, S.S. Nalawade, G. K. Murugesan, F.F. Yu, M.C. Pinho, B.C. Wagner, K.E. Emblem, A. Bjørnerud, B. Fei, A.J. Madhuranthakam, J.A. Maldjian, A fully automated deep learning network for brain tumor segmentation, Tomography 6 (2020) 186, https://doi.org/10.18383/J.TOM.2019.00026.

[16] V. Rajinikanth, S. Kadry, R. Damasevicius, R.A. Sujitha, G. Balaji, M. A. Mohammed, Glioma/glioblastoma detection in brain MRI using pre-trained deep-learning scheme, in: Proceedings of the 2022 3rd International Conference on Intelligent Computing, Instrumentation and Control Technologies: Computational Intelligence for Smart Systems, ICICICT 2022, 2022, pp. 987–990, https://doi.org/10.1109/ICICICT54557.2022.9917904.

[17] N.A. Zebari, C.N. Mohammed, D.A. Zebari, M.A. Mohammed, D.Q. Zeebaree, H. A. Marhoon, K.H. Abdulkareem, S. Kadry, W. Viriyasitavat, J. Nedoma, R. Martinek, A deep learning fusion model for accurate classification of brain tumours in Magnetic Resonance images, CAAI Trans Intell Technol 9 (2024) 790–804, https://doi.org/10.1049/CIT2.12276.

[18] D. Cui, Y. Liu, G. Liu, L. Liu, A multiple-instance learning-based convolutional neural network model to detect the *IDH1* mutation in the histopathology images of glioma tissues, J. Comput. Biol. 27 (2020) 1264–1272, https://doi.org/10.1089/cmb.2019.0410.

[19] T. Wollmann, M. Gunkel, I. Chung, H. Erfle, K. Rippe, K. Rohr, GRUU-Net: integrated convolutional and gated recurrent neural network for cell segmentation, Med. Image Anal. 56 (2019) 68–79, https://doi.org/10.1016/j.media.2019.04.011.

[20] T.C. Hollon, B. Pandian, A.R. Adapa, E. Urias, A.V. Save, S.S.S. Khalsa, D. G. Eichberg, R.S. D'Amico, Z.U. Farooq, S. Lewis, P.D. Petridis, T. Marie, A. H. Shah, H.J.L. Garton, C.O. Maher, J.A. Heth, E.L. McKean, S.E. Sullivan, S. L. Hervey-Jumper, P.G. Patil, B.G. Thompson, O. Sagher, G.M. McKhann, R. J. Komotar, M.E. Ivan, M. Snuderl, M.L. Otten, T.D. Johnson, M.B. Sisti, J.N. Bruce, K.M. Muraszko, J. Trautman, C.W. Freudiger, P. Canoll, H. Lee, S. Camelo-Piragua, D.A. Orringer, Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks, Nat. Med. 26 (2020) 52–58, https://doi.org/10.1038/s41591-019-0715-9.

[21] T. Kurc, S. Bakas, X. Ren, A. Bagari, A. Momeni, Y. Huang, L. Zhang, A. Kumar, M. Thibault, Q. Qi, Q. Wang, A. Kori, O. Gevaert, Y. Zhang, D. Shen, M. Khened, X. Ding, G. Krishnamurthi, J. Kalpathy-Cramer, J. Davis, T. Zhao, R. Gupta, J. Saltz, K. Farahani, Segmentation and classification in digital pathology for glioma research: challenges and deep learning approaches, Front. Neurosci. 14 (2020) 27, https://doi.org/10.3389/fnins.2020.00027.

[22] Y. Xu, Z. Jia, L.-B. Wang, Y. Ai, F. Zhang, M. Lai, E.I.-C. Chang, Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features, BMC Bioinf. 18 (2017) 281, https://doi.org/10.1186/s12859-017-1685-x.

[23] K. Fatima, H. Majeed, H. Irshad, Nuclear spatial and spectral features based evolutionary method for meningioma subtypes classification in histopathology, Microsc. Res. Tech. 80 (2017) 851–861, https://doi.org/10.1002/jemt.22874.

[24] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, D. Altman, G. Antes, D. Atkins, V. Barbour, N. Barrowman, J.A. Berlin, J. Clark, M. Clarke, D. Cook, R. D'Amico, J. J. Deeks, P.J. Devereaux, K. Dickersin, M. Egger, E. Ernst, P.C. Gøtzsche, J. Grimshaw, G. Guyatt, J. Higgins, J.P.A. Ioannidis, J. Kleijnen, T. Lang, N. Magrini, D. McNamee, L. Moja, C. Mulrow, M. Napoli, A. Oxman, B. Pham, D. Rennie, M. Sampson, K.F. Schulz, P.G. Shekelle, D. Tovey, P. Tugwell, Preferred reporting Items for systematic reviews and meta-analyses: the PRISMA statement, PLoS Med. 6 (2009), https://doi.org/10.1371/JOURNAL.PMED.1000097.

[25] L.M. Kmet, L.S. Cook, R.C. Lee, Standard quality assessment criteria for evaluating primary research papers from a variety of fields. https://doi.org/10.7939/R37M04F16, 2004.

[26] J. Barker, A. Hoogi, A. Depeursinge, D.L. Rubin, Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles, Med. Image Anal. 30 (2016) 60–71, https://doi.org/10.1016/j.media.2015.12.002.

[27] S. Mohammed, M. Dinesan, T. Ajayakumar, Survival and quality of life analysis in glioblastoma multiforme with adjuvant chemoradiotherapy: a retrospective study, Rep. Practical Oncol. Radiother. 27 (2022) 1026, https://doi.org/10.5603/RPOR.A2022.0113.

[28] M.J. Tait, V. Petrik, A. Loosemore, B.A. Bell, M.C. Papadopoulos, Survival of patients with glioblastoma multiforme has not improved between 1993 and 2004: analysis of 625 cases, Br. J. Neurosurg. 21 (2007) 496–500, https://doi.org/10.1080/02688690701449251.

[29] M.G. Ertosun, D.L. Rubin, 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 (2015) 1899–1908.

[30] K. Tan, W. Huang, X. Liu, J. Hu, S. Dong, A multi-modal fusion framework based on multi-task correlation learning for cancer prognosis prediction, Artif. Intell. Med. 126 (2022), https://doi.org/10.1016/j.artmed.2022.102260.

[31] L.B. Wang, A. Karpova, M.A. Gritsenko, J.E. Kyle, S. Cao, Y. Li, D. Rykunov, A. Colaprico, J.H. Rothstein, R. Hong, V. Stathias, M. Cornwell, F. Petralia, Y. Wu, B. Reva, K. Krug, P. Pugliese, E. Kawaler, L.K. Olsen, W.W. Liang, X. Song, Y. Dou, M.C. Wendl, W. Caravan, W. Liu, D. Cui Zhou, J. Ji, C.F. Tsai, V.A. Petyuk, J. Moon, W. Ma, R.K. Chu, K.K. Weitz, R.J. Moore, M.E. Monroe, R. Zhao, X. Yang, S. Yoo, A. Krek, A. Demopoulos, H. Zhu, M.A. Wyczalkowski, J.F. McMichael, B. L. Henderson, C.M. Lindgren, H. Boekweg, S. Lu, J. Baral, L. Yao, K.G. Stratton, L. M. Bramer, E. Zink, S.P. Couvillion, K.J. Bloodsworth, S. Satpathy, W. Sieh, S. M. Boca, S. Schürer, F. Chen, M. Wiznerowicz, K.A. Ketchum, E.S. Boja, C. R. Kinsinger, A.I. Robles, T. Hiltke, M. Thiagarajan, A.I. Nesvizhskii, B. Zhang, D. R. Mani, M. Ceccarelli, X.S. Chen, S.L. Cottingham, Q.K. Li, A.H. Kim, D. Fenyö, K. V. Ruggles, H. Rodriguez, M. Mesri, S.H. Payne, A.C. Resnick, P. Wang, R.D. Smith, A. Iavarone, M.G. Chheda, J.S. Barnholtz-Sloan, K.D. Rodland, T. Liu, L. Ding, A. Agarwal, M. Amin, E. An, M.L. Anderson, D.W. Andrews, T. Bauer, C. Birger, M. J. Birrer, L. Blumenberg, W.E. Bocik, U. Borate, M. Borucki, M.C. Burke, S. Cai, A. P. Calinawan, S.A. Carr, S. Cerda, D.W. Chan, A. Charamut, L.S. Chen, D. Chesla, A. M. Chinnaiyan, S. Chowdhury, M.P. Cieślik, D.J. Clark, H. Culpepper, T. Czernicki, F. D'Angelo, J. Day, S. De Young, E. Demir, S.M. Dhanasekaran, R. Dhir, M. J. Domagalski, B. Druker, E. Duffy, M. Dyer, N.J. Edwards, R. Edwards, K. Elburn, M.J. Ellis, J. Eschbacher, A. Francis, S. Gabriel, N. Gabrovski, L. Garofano, G. Getz, M.A. Gillette, A.K. Godwin, D. Golbin, Z. Hanhan, L.I. Hannick, P. Hariharan, B. Hindenach, K.A. Hoadley, G. Hostetter, C. Huang, E. Jaehnig, S.D. Jewell, N. Ji, C.D. Jones, A. Karz, W. Kaspera, L. Kim, R.B. Kothadia, C. Kumar-Sinha, J. Lei, F. D. Leprevost, K. Li, Y. Liao, J. Lilly, H. Liu, J. Lubínski, R. Madan, W. Maggio, E. Malc, A. Malovannaya, S. Mareedu, S.P. Markey, A. Marrero-Oliveras, N. Martinez, N. Maunganidze, J.E. McDermott, P.B. McGarvey, J. McGee, P. Mieczkowski, S. Migliozzi, F. Modugno, R. Montgomery, C.J. Newton, G. S. Omenn, U. Ozbek, O.V. Paklina, A.G. Paulovich, A.M. Perou, A.R. Pico, P. D. Piehowski, D.G. Placantonakis, L. Polonskaya, O. Potapova, B. Pruetz, L. Qi, S. Ramkissoon, A. Resnick, S. Richey, G. Riggins, K. Robinson, N. Roche, D. C. Rohrer, B.R. Rood, L. Rossell, S.R. Savage, E.E. Schadt, Y. Shi, Z. Shi, Y. Shutack, S. Singh, T. Skelly, L.J. Sokoll, J. Stawicki, S.E. Stein, J. Suh, W. Szopa, D. Tabor, D. Tan, D. Tansil, R.R. Thangudu, C. Tognon, E. Traer, S. Tsang, J. Tyner, K.S. Um, D.R. Valley, S. Vasaikar, N. Vatanian, U. Velvulou, M. Vernon, W. Wan, J. Wang, A. Webster, B. Wen, J.R. Whiteaker, G.D. Wilson, Y. Zakhartsev, R. Zelt, H. Zhang, L. Zhang, Z. Zhang, G. Zhao, J. Zhu, Proteogenomic and metabolomic characterization of human glioblastoma, Cancer Cell 39 (2021) 509–528.e20, https://doi.org/10.1016/j.ccell.2021.01.006.

[32] Linmin Pei, K.A. Jones, Z.A. Shboul, J.Y. Chen, K.M. Iftekharuddin, Deep neural network analysis of pathology images with integrated molecular data for enhanced glioma classification and grading, Front. Oncol. 11 (2021) 668694, https://doi.org/10.3389/fonc.2021.668694.
[33] S. Im, J. Hyeon, E. Rha, J. Lee, H.-J. Choi, Y. Jung, T.-J. Kim, Classification of diffuse glioma subtype from clinical-grade pathological images using deep transfer learning, Sensors 21 (2021) 3500, https://doi.org/10.3390/s21103500.
[34] J. Whitney, L. Dollinger, B. Tamrazi, D. Hawes, M. Couce, J. Marcheque, A. Judkins, A. Margol, A. Madabhushi, Quantitative nuclear histomorphometry predicts molecular subtype and clinical outcome in medulloblastomas: preliminary findings, J. Pathol. Inf. 13 (2022) 100090, https://doi.org/10.1016/j.jpi.2022.100090.
[35] V. Bafiti, S. Ouzounis, C. Chalikiopoulou, E. Grigorakou, I.M. Grypari, G. Gregoriou, A. Theofanopoulos, V. Panagiotopoulos, E. Prodromidi, D. Cavouras, V. Zolota, D. Kardamakis, T. Katsila, A 3-miRNA signature enables risk stratification in glioblastoma multiforme patients with different clinical outcomes, Curr. Oncol. 29 (2022) 4315–4331, https://doi.org/10.3390/curroncol29060345.
[36] R.T. Powell, A. Olar, S. Narang, G. Rao, E. Sulman, G.N. Fuller, A. Rao, 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. Inf. 8 (2017), https://doi.org/10.4103/jpi.jpi_43_16.
[37] M. Nalisnik, D.A. Gutman, J. Kong, L.A.D. Cooper, An interactive learning framework for scalable classification of pathology images, in: 2015 IEEE International Conference on Big Data (Big Data), IEEE, 2015, pp. 928–935, https://doi.org/10.1109/BigData.2015.7363841.
[38] N. Nayak, H. Chang, A. Borowsky, P. Spellman, B. Parvin, Classification of tumor histopathology via sparse feature learning, in: Proc IEEE Int Symp Biomed Imaging 2013, 2013, https://doi.org/10.1109/ISBI.2013.6556499.
[39] S. Wen, T.M. Kurc, Y. Gao, T. Zhao, J.H. Saltz, W. Zhu, A methodology for texture feature-based quality assessment in nucleus segmentation of histopathology image, J. Pathol. Inf. 8 (2017) 38, https://doi.org/10.4103/jpi.jpi_43_17.
[40] Y. Chang, F. He, J. Wang, S. Chen, J. Li, J. Liu, Y. Yu, L. Su, A. Ma, C. Allen, Y. Lin, S. Sun, B. Liu, J. Javier Otero, D. Chung, H. Fu, Z. Li, D. Xu, Q. Ma, Define and visualize pathological architectures of human tissues from spatially resolved transcriptomics using deep learning, Comput. Struct. Biotechnol. J. 20 (2022) 4600–4617, https://doi.org/10.1016/j.csbj.2022.08.029.
[41] A.Z. Shirazi, E. Fornaciari, N.S. Bagherian, L.M. Ebert, B. Koszyca, G.A. Gomez, DeepSurvNet: deep survival convolutional network for brain cancer survival rate classification based on histopathological images, Med. Biol. Eng. Comput. 58 (2020) 1031–1045, https://doi.org/10.1007/s11517-020-02147-3.
[42] Y. Liu, Y. Jia, C. Hou, N. Li, N. Zhang, X. Yan, L. Yang, Y. Guo, H. Chen, J. Li, Y. Hao, J. Liu, Pathological prognosis classification of patients with neuroblastoma using computational pathology analysis, Comput. Biol. Med. 149 (2022) 105980, https://doi.org/10.1016/j.compbiomed.2022.105980.
[43] B. Lessmann, T.W. Nattkemper, V.H. Hans, A. Degenhard, A method for linking computed image features to histological semantics in neuropathology, J. Biomed. Inf. 40 (2007) 631–641, https://doi.org/10.1016/j.jbi.2007.06.007.
[44] O. Attallah, MB-AI-His: histopathological diagnosis of pediatric medulloblastoma and its subtypes via AI, Diagnostics 11 (2021) 359, https://doi.org/10.3390/diagnostics11020359.
[45] K. Faust, Q. Xie, D. Han, K. Goyle, Z. Volynskaya, U. Djuric, P. Diamandis, Visualizing histopathologic deep learning classification and anomaly detection using nonlinear feature space dimensionality reduction, BMC Bioinf. 19 (2018) 173, https://doi.org/10.1186/s12859-018-2184-4.
[46] C. Han, H. Yao, B. Zhao, Z. Li, Z. Shi, L. Wu, X. Chen, J. Qu, K. Zhao, R. Lan, C. Liang, X. Pan, Z. Liu, Meta multi-task nuclei segmentation with fewer training samples, Med. Image Anal. 80 (2022), https://doi.org/10.1016/j.media.2022.102481.
[47] O. Attallah, CoMB-deep: composite deep learning-based pipeline for classifying childhood medulloblastoma and its classes, Front. Neuroinf. 15 (2021), https://doi.org/10.3389/fninf.2021.663592.
[48] V.M. Ravi, P. Will, J. Kueckelhaus, N. Sun, K. Joseph, H. Salié, L. Vollmer, U. Kuliesiute, J. von Ehr, J.K. Benotmane, N. Neidert, M. Follo, F. Scherer, J. M. Goeldner, S.P. Behringer, P. Franco, M. Khiat, J. Zhang, U.G. Hofmann, C. Fung, F.L. Ricklefs, K. Lamszus, M. Boerries, M. Ku, J. Beck, R. Sankowski, M. Schwabenland, M. Prinz, U. Schüller, S. Killmer, B. Bengsch, A.K. Walch, D. Delev, O. Schnell, D.H. Heiland, Spatially resolved multi-omics deciphers bidirectional tumor-host interdependence in glioblastoma, Cancer Cell 40 (2022) 639–655.e13, https://doi.org/10.1016/j.ccell.2022.05.009.
[49] A. Yonekura, H. Kawanaka, V.B.S. Prasath, B.J. Aronow, H. Takase, Automatic disease stage classification of glioblastoma multiforme histopathological images using deep convolutional neural network, Biomed Eng Lett 8 (2018) 321–327, https://doi.org/10.1007/s13534-018-0077-0.
[50] R. Lasfar, G. Tóth, The difference of model robustness assessment using cross-validation and bootstrap methods, J. Chemom. 38 (2024) e3530, https://doi.org/10.1002/CEM.3530.
[51] L. Pei, K.A. Jones, Z.A. Shboul, J.Y. Chen, K.M. Iftekharuddin, Deep neural network analysis of pathology images with integrated molecular data for enhanced glioma classification and grading, Front. Oncol. 11 (2021), https://doi.org/10.3389/fonc.2021.668694.
[52] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, P.A. Heng, DCAN: deep contour-aware networks for object instance segmentation from histology images, Med. Image Anal. 36 (2017) 135–146, https://doi.org/10.1016/j.media.2016.11.004.
[53] A.Z. Shirazi, M.D. McDonnell, E. Fornaciari, N.S. Bagherian, K.G. Scheer, M. S. Samuel, M. Yaghoobi, R.J. Ormsby, S. Poonnoose, D.J. Tumes, G.A. Gomez, 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 (2021) 337–350, https://doi.org/10.1038/s41416-021-01394-x.
[54] R.J. Chen, M.Y. Lu, D.F.K. Williamson, T.Y. Chen, J. Lipkova, Z. Noor, M. Shaban, M. Shady, M. Williams, B. Joo, F. Mahmood, Pan-cancer integrative histology-genomic analysis via multimodal deep learning, Cancer Cell 40 (2022) 865–878.e6, https://doi.org/10.1016/j.ccell.2022.07.004.
[55] A.A. Bidgoli, S. Rahnamayan, T. Dehkharghanian, A. Riasatian, S. Kalra, M. Zaveri, C.J.V. Campbell, A. Parwani, L. Pantanowitz, H.R. Tizhoosh, Evolutionary deep feature selection for compact representation of gigapixel images in digital pathology, Artif. Intell. Med. 132 (2022) 102368, https://doi.org/10.1016/j.artmed.2022.102368.
[56] X. Li, Q. Tang, J. Yu, Y. Wang, Z. Shi, Microvascularity detection and quantification in glioma: a novel deep-learning-based framework, Lab. Invest. 99 (2019) 1515–1526, https://doi.org/10.1038/s41374-019-0272-3.
[57] N. Bergmann, C. Delbridge, J. Gempt, A. Feuchtinger, A. Walch, L. Schirmer, W. Bunk, T. Aschenbrenner, F. Liesche-Starnecker, J. Schlegel, The intratumoral heterogeneity reflects the intertumoral subtypes of glioblastoma multiforme: a regional immunohistochemistry analysis, Front. Oncol. 10 (2020), https://doi.org/10.3389/fonc.2020.00494.
[58] F. Liesche-Starnecker, K. Mayer, F. Kofler, S. Baur, F. Schmidt-Graf, J. Kempter, G. Prokop, N. Pfarr, W. Wei, J. Gempt, S.E. Combs, C. Zimmer, B. Meyer, B. Wiestler, J. Schlegel, Immunohistochemically characterized intratumoral heterogeneity is a prognostic marker in human glioblastoma, Cancers 12 (2020) 2964, https://doi.org/10.3390/cancers12102964.
[59] L. Garofano, S. Migliozzi, Y.T. Oh, F. D'Angelo, R.D. Najac, A. Ko, B. Frangaj, F. P. Caruso, K. Yu, J. Yuan, W. Zhao, A. Luisa Di Stefano, F. Bielle, T. Jiang, P. Sims, M.L. Suvà, F. Tang, X.D. Su, M. Ceccarelli, M. Sanson, A. Lasorella, A. Iavarone, Pathway-based classification of glioblastoma uncovers a mitochondrial subtype with therapeutic vulnerabilities, Nat. Can. (Ott.) 2 (2021) 141, https://doi.org/10.1038/S43018-020-00159-4.
[60] M. Martinez-Lage, T.M. Lynch, Y. Bi, C. Cocito, G.P. Way, S. Pal, J. Haller, R.E. Yan, A. Ziober, A. Nguyen, M. Kandpal, D.M. O'Rourke, J.P. Greenfield, C.S. Greene, R. V. Davuluri, N. Dahmane, Immune landscapes associated with different glioblastoma molecular subtypes, Acta Neuropathol Commun 7 (2019) 1–12, https://doi.org/10.1186/S40478-019-0803-6/FIGURES/8.
[61] J. Kang, Z. Ullah, J. Gwak, MRI-based brain tumor classification using ensemble of deep features and machine learning classifiers, Sensors 21 (2021) 1–21, https://doi.org/10.3390/s21062222.
[62] P. Saha, M.S. Sadi, M.M. Islam, EMCNet: automated COVID-19 diagnosis from X-ray images using convolutional neural network and ensemble of machine learning classifiers, Inform. Med. Unlocked 22 (2021) 100505, https://doi.org/10.1016/J.IMU.2020.100505.
[63] E. Upschulte, S. Harmeling, K. Amunts, T. Dickscheid, Contour proposal networks for biomedical instance segmentation, Med. Image Anal. 77 (2022) 102371, https://doi.org/10.1016/J.MEDIA.2022.102371.
[64] J.P. Thakkar, T.A. Dolecek, C. Horbinski, Q.T. Ostrom, D.D. Lightner, J. S. Barnholtz-Sloan, J.L. Villano, Epidemiologic and molecular prognostic review of glioblastoma, Cancer Epidemiol. Biomarkers Prev. 23 (2014) 1985–1996, https://doi.org/10.1158/1055-9965.EPI-14-0275/68016/AM/EPIDEMIOLOGIC-AND-MOLECULAR-PROGNOSTIC-REVIEW-OF.
[65] G. Reifenberger, V.P. Collins, Pathology and molecular genetics of astrocytic gliomas, J. Mol. Med. (Berl.) 82 (2004) 656–670, https://doi.org/10.1007/s00109-004-0564-x.
[66] O.S. Al-Kadi, A multiresolution clinical decision support system based on fractal model design for classification of histological brain tumours, Comput. Med. Imag. Graph. 41 (2015) 67–79, https://doi.org/10.1016/j.compmedimag.2014.05.013.
[67] H. Xu, L. Liu, X. Lei, M. Mandal, C. Lu, An unsupervised method for histological image segmentation based on tissue cluster level graph cut, Comput. Med. Imag. Graph. 93 (2021), https://doi.org/10.1016/j.compmedimag.2021.101974.
[68] E.I. Papageorgiou, P.P. Spyridonos, D.T. Glotsos, C.D. Stylios, P. Ravazoula, G. N. Nikiforidis, P.P. Groumpos, Brain tumor characterization using the soft computing technique of fuzzy cognitive maps, Applied Soft Computing Journal 8 (2008) 820–828, https://doi.org/10.1016/j.asoc.2007.06.006.
[69] C. Cong, S. Liu, A. Di Ieva, M. Pagnucco, S. Berkovsky, Y. Song, Colour adaptive generative networks for stain normalisation of histopathology images, Med. Image Anal. 102580 (2022), https://doi.org/10.1016/j.media.2022.102580.
[70] J. Peng, Z. Luo, CS-Net: instance-aware cellular segmentation with hierarchical dimension-decomposed convolutions and slice-attentive learning, Knowl. Base Syst. 232 (2021), https://doi.org/10.1016/j.knosys.2021.107485.
[71] M. Komosiński, K. Komosiński, K. Krawiec, Evolutionary weighting of image features for diagnosing of CNS tumors, Artif. Intell. Med. (2000), https://doi.org/10.1016/S0933-3657(99)00048-2.
[72] E. Gangoso, B. Southgate, L. Bradley, S. Rus, F. Galvez-Cancino, N. McGivern, E. Güç, C.A. Kapourani, A. Byron, K.M. Ferguson, N. Alfazema, G. Morrison, V. Grant, C. Blin, I.F. Sou, M.A. Marques-Torrejon, L. Conde, S. Parrinello, J. Herrero, S. Beck, S. Brandner, P.M. Brennan, P. Bertone, J.W. Pollard, S. A. Quezada, D. Sproul, M.C. Frame, A. Serrels, S.M. Pollard, Glioblastomas acquire myeloid-affiliated transcriptional programs via epigenetic immunoediting to elicit immune evasion, Cell 184 (2021) 2454–2470.e26, https://doi.org/10.1016/j.cell.2021.03.023.
[73] S.E.A. Raza, L. Cheung, M. Shaban, S. Graham, D. Epstein, S. Pelengaris, M. Khan, N.M. Rajpoot, Micro-Net: a unified model for segmentation of various objects in microscopy images, Med. Image Anal. 52 (2019) 160–173, https://doi.org/10.1016/j.media.2018.12.003.

[74] M.U. Oner, J. Chen, E. Revkov, A. James, S.Y. Heng, A.N. Kaya, J.J.S. Alvarez, A. Takano, X.M. Cheng, T.K.H. Lim, D.S.W. Tan, W. Zhai, A.J. Skanderup, W. K. Sung, H.K. Lee, Obtaining spatially resolved tumor purity maps using deep multiple instance learning in a pan-cancer study, Patterns 3 (2022), https://doi.org/10.1016/j.patter.2021.100399.

[75] A.A. Mughal, L. Zhang, A. Fayzullin, A. Server, Y. Li, Y. Wu, R. Glass, T. Meling, I. A. Langmoen, T.B. Leergaard, E.O. Vik-Mo, Patterns of invasive growth in malignant gliomas—the Hippocampus emerges as an invasion-spared brain region, Neoplasia 20 (2018) 643–656, https://doi.org/10.1016/j.neo.2018.04.001.

[76] S.K. McBrayer, J.R. Mayers, G.J. DiNatale, D.D. Shi, J. Khanal, A.A. Chakraborty, K.A. Sarosiek, K.J. Briggs, A.K. Robbins, T. Sewastianik, S.J. Shareef, B. A. Olenchock, S.J. Parker, K. Tateishi, J.B. Spinelli, M. Islam, M.C. Haigis, R. E. Looper, K.L. Ligon, B.E. Bernstein, R.D. Carrasco, D.P. Cahill, J.M. Asara, C. M. Metallo, N.H. Yennawar, M.G. Vander Heiden, W.G. Kaelin, Transaminase inhibition by 2-hydroxyglutarate impairs glutamate biosynthesis and redox homeostasis in glioma, Cell 175 (2018) 101–116.e25, https://doi.org/10.1016/j.cell.2018.08.038.

[77] V. Brindha, P. Jayashree, P. Karthik, P. Manikandan, Tumor grading model employing geometric analysis of histopathological images with characteristic nuclei dictionary, Comput. Biol. Med. 149 (2022) 106008, https://doi.org/10.1016/j.compbiomed.2022.106008.

[78] C. Zhong, J. Han, A. Borowsky, B. Parvin, Y. Wang, H. Chang, When machine vision meets histology: a comparative evaluation of model architecture for classification of histology sections, Med. Image Anal. 35 (2017) 530–543, https://doi.org/10.1016/j.media.2016.08.010.

[79] I. Alzoubi, G. Bao, R. Zhang, C. Loh, Y. Zheng, S. Cherepanoff, G. Gracie, M. Lee, M. Kuligowski, K.L. Alexander, M.E. Buckland, X. Wang, M.B. Graeber, An open-source AI framework for the analysis of single cells in whole-slide images with a note on CD276 in glioblastoma, Cancers 14 (2022) 3441, https://doi.org/10.3390/cancers14143441.

[80] K. Fatima, A. Arooj, H. Majeed, A new texture and shape based technique for improving meningioma classification, Microsc. Res. Tech. 77 (2014) 862–873, https://doi.org/10.1002/jemt.22409.

[81] L. Hou, D. Samaras, T.M. Kurc, Y. Gao, J.E. Davis, J.H. Saltz, Patch-based convolutional neural network for whole slide tissue image classification, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2016, pp. 2424–2433, https://doi.org/10.1109/CVPR.2016.266.

[82] L. Jin, F. Shi, Q. Chun, H. Chen, Y. Ma, S. Wu, N.U.F. Hameed, C. Mei, J. Lu, J. Zhang, A. Aibaidula, D. Shen, J. Wu, Artificial intelligence neuropathologist for glioma classification using deep learning on hematoxylin and eosin stained slide images and molecular markers, Neuro Oncol. 23 (2021) 44–52, https://doi.org/10.1093/neuonc/noaa163.

[83] M. Nalisnik, M. Amgad, S. Lee, S.H. Halani, J.E. Velazquez Vega, D.J. Brat, D. A. Gutman, L.A.D. Cooper, Interactive phenotyping of large-scale histology imaging data with HistomicsML, Sci. Rep. 7 (2017) 14588, https://doi.org/10.1038/s41598-017-15092-3.

[84] F. Orzan, F. Pagani, M. Cominelli, L. Triggiani, S. Calza, F. De Bacco, D. Medicina, P. Balzarini, P.P. Panciani, R. Liserre, M. Buglione, M.M. Fontanella, E. Medico, R. Galli, C. Isella, C. Boccaccio, P.L. Poliani, A Simplified Integrated Molecular and Immunohistochemistry-Based Algorithm Allows High Accuracy Prediction of Glioblastoma Transcriptional Subtypes, vol. 100, Laboratory Investigation, 2020, pp. 1330–1344, https://doi.org/10.1038/s41374-020-0437-0.

[85] H. Su, F. Xing, X. Kong, Y. Xie, S. Zhang, L. Yang, Robust cell detection and segmentation in histopathological images using sparse reconstruction and stacked denoising autoencoders, Med Image Comput Comput Assist Interv (2015), https://doi.org/10.1007/978-3-319-24574-4_46.

[86] T.H. Vu, H.S. Mousavi, V. Monga, G. Rao, U.K.A. Rao, Histopathological image classification using discriminative feature-oriented dictionary learning, IEEE Trans. Med. Imag. 35 (2016) 738–751, https://doi.org/10.1109/TMI.2015.2493530.

[87] F. Xing, Y. Xie, L. Yang, An automatic learning-based framework for robust nucleus segmentation, IEEE Trans. Med. Imag. 35 (2015) 550–566, https://doi.org/10.1109/TMI.2015.2481436.