



Review article

Machine learning applications to neuroimaging for glioma detection and classification: An artificial intelligence augmented systematic review



Quinlan D. Buchlak^{a,*}, Nazanin Esmaili^{a,b}, Jean-Christophe Leveque^c, Christine Bennett^a, Farrokh Farrokhi^c, Massimo Piccardi^b

^a School of Medicine, The University of Notre Dame Australia, Sydney, NSW, Australia

^b Faculty of Engineering and IT, University of Technology Sydney, Ultimo, NSW, Australia

^c Neuroscience Institute, Virginia Mason Medical Center, Seattle, WA, USA

ARTICLE INFO

Article history:

Received 24 November 2020

Accepted 30 April 2021

Keywords:

Machine learning
Multimodal neuroimaging
Neurosurgery
Image processing
Radiomics
Glioblastoma
Brain tumor classification
Glioma
Glioma grading
Deep learning
Convolutional neural networks
T1-MR image
T2-MR image
FLAIR

ABSTRACT

Glioma is the most common primary intraparenchymal tumor of the brain and the 5-year survival rate of high-grade glioma is poor. Magnetic resonance imaging (MRI) is essential for detecting, characterizing and monitoring brain tumors but definitive diagnosis still relies on surgical pathology. Machine learning has been applied to the analysis of MRI data in glioma research and has the potential to change clinical practice and improve patient outcomes. This systematic review synthesizes and analyzes the current state of machine learning applications to glioma MRI data and explores the use of machine learning for systematic review automation. Various datapoints were extracted from the 153 studies that met inclusion criteria and analyzed. Natural language processing (NLP) analysis involved keyword extraction, topic modeling and document classification. Machine learning has been applied to tumor grading and diagnosis, tumor segmentation, non-invasive genomic biomarker identification, detection of progression and patient survival prediction. Model performance was generally strong (AUC = 0.87 ± 0.09 ; sensitivity = 0.87 ± 0.10 ; specificity = 0.86 ± 0.10 ; precision = 0.88 ± 0.11). Convolutional neural network, support vector machine and random forest algorithms were top performers. Deep learning document classifiers yielded acceptable performance (mean 5-fold cross-validation AUC = 0.71). Machine learning tools and data resources were synthesized and summarized to facilitate future research. Machine learning has been widely applied to the processing of MRI data in glioma research and has demonstrated substantial utility. NLP and transfer learning resources enabled the successful development of a replicable method for automating the systematic review article screening process, which has potential for shortening the time from discovery to clinical application in medicine.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

Gliomas are tumors arising from glial cells in the central nervous system and are the most common primary intraparenchymal tumors of the brain [42,109,224]. The World Health Organization classifies them into four grades based on their pathological and genetic properties [146,147]. Grades I and II are considered low-grade gliomas (LGG), while grades III and IV are considered high-grade (HGG) [11]. Gliomas include astrocytomas, ependymomas, oligoastrocytomas and oligodendrogliomas. Glioblastoma is a grade IV astrocytoma and is the most aggressive primary brain tumor, with a limited median survival after diagnosis of just

12–15 months [76,104,237]. The incidence of LGG is approximately 1/100,000 per year [97]. The 5-year survival rate for LGG (WHO grade II) ranges between 42 and 92% [191]. Five-year survival rates for grade III gliomas range between 27% and 52%, while grade IV gliomas are associated with five-year survival rates of 5% [164]. Glioblastoma survival at 10 years is estimated at 0.71% [219]. Low-grade gliomas are associated with epilepsy, degraded cognition and poorer quality of life [1,117] and it appears that more than half of patients with LGG are likely to experience recurrence after treatment [195,222]. Differentiation between different classes of glioma is necessary for risk stratification and treatment planning. The determination of tumor grade currently relies on surgical biopsy, an invasive procedure, making continued development of safer, less-invasive diagnostics and more effective glioma treatment methods a worthy pursuit.

* Corresponding author at: The University of Notre Dame Australia, 160 Oxford St, Sydney 2015, Australia.

E-mail address: quinlan.buchlak1@my.nd.edu.au (Q.D. Buchlak).

Magnetic resonance imaging (MRI) is an essential tool used for detecting, characterizing and monitoring brain tumors and typically utilizes various sequence types [4,10] which highlight distinct and complementary brain tissue characteristics including the tumor core, tissue necrosis and surrounding edema [119]. MRI sequences include T1-weighted imaging (T1), contrast-enhanced T1-weighted imaging (T1CE), T2-weighted imaging (T2), diffusion-weighted imaging (DWI) and fluid attenuated inversion-recovery (FLAIR). T1 typically shows bright glioma borders, while FLAIR and T2 show brighter internal tumor regions [247]. Gadolinium contrast enhancement on MRI correlates with neovascularity and tumor cell infiltration in high grade gliomas [223] and is often regarded as a sign of malignancy [94] although not all malignant gliomas demonstrate contrast enhancement [197]. Glioblastoma has been shown to extend beyond the area of enhancing tissue on MRI into areas that may have been deemed normal or oedematous by clinicians [79,234]. This extension beyond enhancing tissue can result in incomplete surgical resection and tumor recurrence [25].

Accurate tumor diagnosis is necessary for effective determination of a patient's prognosis [237]. Molecular and genetic profiling of gliomas is linked with better treatment outcomes [224] but often necessitates invasive brain biopsy, which carries with it surgical risks. Additionally, single biopsy specimens do not account for spatial variation within tumors. The segmentation of brain tumors using MRI has an important role in early diagnosis and treatment planning [238]. Tumors can manifest in different locations, sizes and shapes, and demonstrate substantial structural diversity, adding to the complexity of image analysis and classification [83]. Tumor borders are often diffuse and their appearance complex and heterogeneous [229]. Automated or semi-automated techniques for image analysis have the potential to improve efficiency and accuracy [27,171] above that achieved by human radiologists alone. Treatment of HGG involves surgical resection and debulking, combined with adjuvant radiotherapy and chemotherapy [35,90,192,205]. Higher volume resection appears to be associated with improved survival [192], but may adversely impact postoperative functional status [78]. For tumors that return, recurrence or progression of tumor often occurs 6–8 months after initial treatment [6]. Machine learning, which is a branch of artificial intelligence that enables effective prediction and classification [38,69], has been used to detect progression and differentiate it from post-treatment radiotherapy effects [205].

Radiomics describes the detailed quantification of tumor phenotypes using neuroimaging metrics and the development of image-based biomarkers to facilitate clinical decision support and targeted treatment [4,73]. Machine learning underpins radiomics [228] and modern approaches to patient safety improvement [40]. The field of neuroimaging is increasingly exploiting the capabilities of machine learning for clinically useful classification, achieving individualised patient outcome prediction and semiautomated tumor delineation and diagnosis. Machine learning algorithms applied in this domain include random forest (RF), logistic regression (LR), support vector machines (SVM), neural networks (NN) and deep learning convolutional neural networks (CNN). CNNs are designed to extract feature maps from input images and perform classification functions using these feature maps [99]. They are considered the most suitable machine learning method for medical image processing [128]. Deep learning algorithms have demonstrated impressive performance in image pattern recognition and classification [212]. Machine learning algorithms excel at learning complex non-linear relationships in large, high-dimensional datasets. The relative characteristics and capabilities of machine learning algorithms have been previously described [36,38,95,128]. Applications of machine learning methods within the field of radiomics include detection of genetic

mutations [7], effectively dealing with missing data [177], artifact reduction [103] and noise removal [10,64]. Automated tumor segmentation involves training algorithms to determine whether discrete image voxels belong to one of numerous possible tissue classes including gray matter, white matter, cerebrospinal fluid, skull, soft tissue and pathological tissue. Machine learning methods coupled with MRI data have the potential to supersede invasive tumor classification methods requiring surgical biopsy [49,179]. Machine learning holds great potential for improving prognosis and diagnostic accuracy [162]. Building effective technological tools leveraging the capabilities of machine learning may improve patient safety in earlier detection of recurrence and resource efficiency by allowing elimination of surgical biopsies. Machine learning applications to radiology are abundant and the number of bioinformatics and machine learning analysis tools is increasing rapidly [73]. The state of the science with regard to machine learning applications to glioma MRI research has not recently been comprehensively synthesized and analysed.

The time between a discovery being published in the academic literature and these insights being translated into clinical practice can be many years [20]. Systematic reviews facilitate knowledge synthesis, dissemination and translation but themselves may require up to a year to complete [56]. Furthermore, they are resource intensive. A systematic review involving the retrieval of 10,000 citations may cost upwards of USD \$300,000 [127]. It has been estimated that most systematic reviews cost at least USD \$140,000 [153]. Major academic institutions each appear to publish an average of 132 systematic reviews per year, amounting to a total yearly cost in excess of USD \$18.5 million per institution [153]. This investment may result in little benefit if the review fails to provide a useful, complete or up-to-date synthesis of the available evidence [60]. It has been estimated that 44% of meta-analyses produced in the pharmaceutical industry are not successfully published [153]. Automation tools can be used to make this process more efficient [24], potentially facilitating the ability to produce a moderately sized systematic review in just two weeks [56]. The application of machine learning techniques to the automation of systematic reviews has so far been limited and calls have been made to deploy machine learning to improve the efficiency of the process [153]. Natural language processing (NLP) is based on machine learning technology and is used to convert unstructured language data into a quantitative representation that can be analysed by computers [38,151,180]. NLP has been designed to capture semantics and context and uses vector-based representations of words as inputs into machine learning algorithms. This technique has been applied to the analysis of unstructured medical records, operative reports and investigation data and has been shown to be accurate, efficient and valid [139,231]. NLP can be applied to develop diagnostic systems, process clinical notes to identify themes and generate data for predictive modelling, develop question answering systems, translate languages, automatically describe images and automate text summarisation [41,196,236]. NLP has the potential to at least partially automate the systematic review process [38]. A semi-automated system leveraging modern transfer learning technologies [84] may improve the efficiency and speed of systematic reviews and reduce the time between discovery and clinical application in medicine.

The aims of this study were three-fold: to (1) systematically review machine learning methods and tools applied to glioma MRI data, (2) analyse article characteristics and impact, and (3) explore the application of NLP methods to the corpus. The study was designed to answer the following research questions: How has machine learning been applied to process clinical MRI data in glioma research? What level of impact have papers in this field achieved and where have they been produced? Is it possible to

successfully deploy NLP to partially automate the systematic review process?

2. Method

Our method was guided by the standards of the Institute of Medicine [158] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [155]. A prospective systematic review protocol was developed and approved by senior authors. A comprehensive search strategy was developed [190] and applied to the following databases in accordance with recommendations [34]: PubMed, ScienceDirect, Embase, Ovid, Ebsco, Google Scholar and Scopus. Relevant studies were identified up to May 2020. Risk of bias was assessed using PROBAST (prediction model risk of bias assessment tool) [157,226].

2.1. Study identification

The search was conducted between April and May 2020. The search strategy employed comprehensive combinations of domain-specific and methodological search terms, both keywords and MeSH terms (Table 1). The following search query was inserted into PubMed and generated 375 results: [(machine learning OR artificial intelligence OR computer vision) AND (neuroimaging OR MRI OR magnetic resonance imaging) AND (glioma OR glioblastoma OR astrocytoma OR oligoastrocytoma OR oligodendroglioma OR ependymoma)]. We hand-searched both referenced and citing articles for each included study to identify additional relevant papers [186]. The reference lists of review articles related to the focus of this review were mined.

2.2. Inclusion and exclusion criteria

Inclusion criteria were: original research published in a peer-reviewed journal; involved the use of one or more sequences of MRI neuroimaging data (e.g., T1, T2, T1CE, FLAIR); included adult

patients diagnosed with high- or low-grade glioma; published in English; and involved the application of machine learning techniques for outcome classification or prediction. Included articles were limited to those published within the last 10 years to focus on methodologies applied within the era of deep learning [149]. A machine learning study was defined here as one that employed an algorithm to predict or classify one or more outcomes and involved the use of one or more of the following performance metrics: accuracy, area under the receiver operating characteristics curve (AUC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score or Dice similarity coefficient (DSC). Exclusion criteria were: review article; described machine learning methods but not a specific clinical application; book or book chapter; conference paper or abstract; non-medical application; classified outcomes not related to glioma; included paediatric patients; and did not involve the processing of MRI data. If a study involved the use of multivariate logistic regression, but not as a classifier, it was excluded. Studies that relied primarily on Cox regression and fuzzy methods were excluded. Articles were selected for inclusion by one researcher (QDB) and selections were verified by a second researcher (NE). Articles classified as appropriately meeting inclusion and exclusion criteria independently by both researchers were included in the analysis, with disagreements resolved by discussion.

2.3. Data collection and extraction

Numerous datapoints were collected and coded for each included study. Data recorded included: abstract, purpose, outcome predicted, study design, the number of institutions involved, machine learning algorithms applied and the top performing algorithm, sample size, MRI sequences analysed, tumor types, software tools used by researchers, number and types of features extracted, modeling methods employed, performance assessment and validation methods applied, PROBAST bias assessment and whether or not the study presented underlying equations, involved data augmentation or compared multiple algorithms. Lead and senior authors of each included paper were recorded, along with the lead author's institution and country. When a study developed multiple models, the details and performance statistics of the top performing model were recorded for analysis. Citation numbers were gathered from Google Scholar for every included study over a two-day period to limit variation. Impact factors were collected from journal websites. Scimago Journal and Country Rank (www.scimagojr.com) was used to determine journal rank score and h-index. Mean citations per year was calculated by dividing the total number of citations by years since publication, excluding the incomplete year of 2020.

2.4. Analysis

Quantitative and qualitative analyses were conducted. Data extracted from each article were used to calculate descriptives and build predictive models. Analyses were conducted using custom Python scripts and the scipy [105], scikitlearn [169], nltk [29], gensim [183,184], keras [54] and simpletransformers [178] packages.

2.4.1. Natural language processing and machine learning modeling
Abstracts were used to feed an NLP analysis that consisted of three phases: (1) keyword identification, (2) topic modeling using the Latent Dirichlet Allocation (LDA) approach [32], and (3) training and testing document classifiers (Fig. 1). For keyword identification, text was tokenized, tokens were converted to lower case, numeric characters were dropped, English stop words were removed, the remaining tokens were lemmatized

Table 1
MeSH terms used in the search strategy.

	Pathology	Methodology
MeSH Terms	• Glioma	• Machine Learning
	• Glioblastoma	• Artificial Intelligence
	• Astrocytoma	• Computational Intelligence
	• Oligodendroglioma	• Neuroimaging
	• Ependymoma	• Magnetic resonance imaging
	• Brain Neoplasms	• Deep Learning
	• Diffuse Intrinsic Pontine Glioma	• Supervised Machine Learning
	• Glioma, Subependymal	• Support Vector Machine
	• Ganglioglioma	• Unsupervised Machine Learning
	• Gliosarcoma	• Neural Networks, Computer
Keywords	• Medulloblastoma	• Neural network
		• Convolutional neural network
		• K nearest neighbors
		• Decision tree
		• Random forest
		• Naïve bayes
		• Classification
		• Prediction
		• Image processing

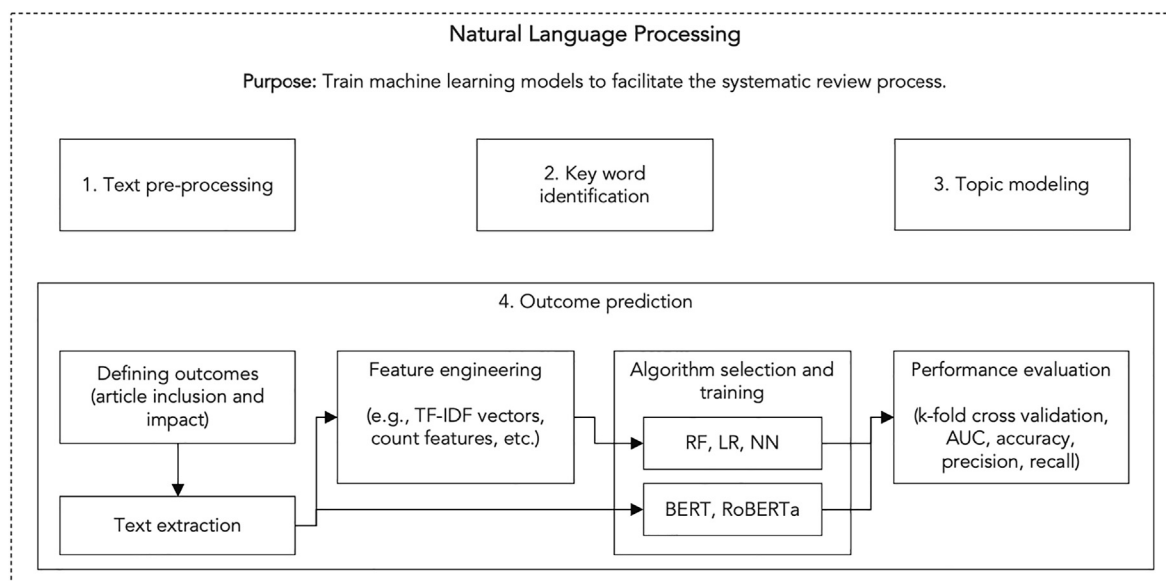


Fig. 1. Overview of the natural language processing analysis applied.

and the most frequently occurring meaningful words were counted. Topic modeling involved the same pre-processing across all included abstracts, with the addition of token vectorization and filtering to remove common and rare words. Tokens that appeared in <15 abstracts and those that appeared in >50% of abstracts were filtered out. A corpus was created, term frequency-inverse document frequency (TF-IDF) calculations were conducted and LDA models were developed to classify up to 100 topics. LDA model performance was assessed with the coherence metric.

Subsequently, and separate from topic modeling, a document classification analysis was conducted. Multiple document classifiers were trained to explore the development of a system to facilitate automated article screening and selection and to predict article impact. The primary classification outcome was article inclusion in this systematic review. Secondary outcomes measured impact, which included the number of citations per year and binary variables that were created using this data (e.g., >10 citations per year). Regression models were developed to predict number of citations and classification models were developed to classify binary impact outcome variables. Feature engineering involved developing TF-IDF vectors and text-based features for each abstract. TF-IDF vectors represented the relative importance of a term within a document and within the entire corpus. Text-based features included word count, character count, average word length (word density), punctuation count, uppercase count, title case word count, and counts of nouns, verbs, adjectives, adverbs and pronouns. Applied algorithms for this document classification task included RF, LR, NN and pre-trained deep neural networks. A transfer learning approach for text classification was designed employing the Bidirectional Encoder Representations from Transformers (BERT) [63] and RoBERTa [142] pre-trained language models. These models enabled the use of raw, unprocessed text as input rather than engineered features. A sliding window was applied to manage the maximum sequence length limitation. Model classification performance was assessed using 5-fold cross validation, with accuracy, area under the receiver operating characteristic curve (AUC), precision and recall metrics.

3. Results

3.1. Study selection

The search resulted in the retrieval of 15,001 records. We assessed 536 full text articles; 337 were excluded, which left 153 studies for analysis (Fig. 2). Data elements were extracted and the purpose, methodology, model performance and risk of bias were summarized for each included study.

3.2. Machine learning application domains

Diagnosis or classification of tumor grade was the primary focus of 41 studies (Table 2), while 35 studies developed models for tumor detection and automated tumor segmentation on MRI (Table 3) and 10 focused on detecting and predicting tumor progression (Table 4). The majority of included studies (48; Table 5) focused on developing machine learning models to noninvasively classify glioma genetic and molecular characteristics. Seventeen studies developed models for glioma survival prediction (Table 6).

3.3. Summarizing methods applied and descriptive statistics

Descriptive statistics are presented in Table 7. Most studies (73%) compared the performance of multiple machine learning models when predicting a selected outcome. Authors either varied the algorithm applied or the inputs to a single algorithm or both, while recording model performance metrics. Almost all studies were retrospective and involved the application of supervised machine learning methods. One study employed a prospective design, one employed unsupervised methods and one employed semi-supervised methods. The most common algorithms that were reported by authors as demonstrating the strongest performance were SVM ($n = 46$), CNN ($n = 42$), RF ($n = 28$), NN ($n = 10$) and LR ($n = 10$). A noted advantage of applying CNNs was that they removed the need for feature engineering and selection, which can be a complex process required by other algorithms (e.g., SVM and RF).

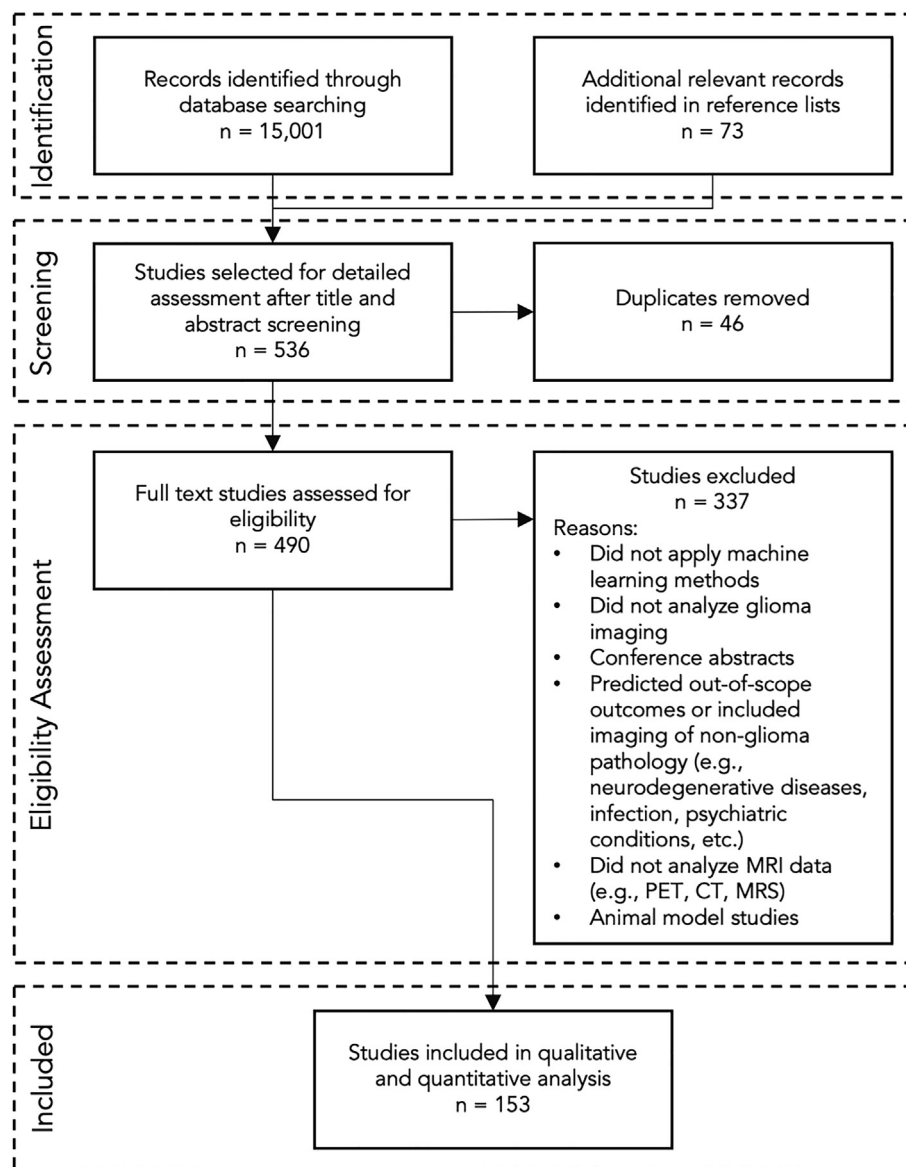


Fig. 2. Overview of study screening and selection.

Most studies (78%) described a robust model validation process, which either involved some form of cross-validation (e.g., 3-, 5-, 10-fold, etc.) or a train-test methodology with additional external validation. Performance evaluation processes ranged from: (1) simply splitting the dataset into train and test sets; through to (2) running k-fold cross-validation; and up to (3) running k-fold cross-validation, then validating the algorithm on an additional unseen holdout segment of the dataset and then further assessing the performance of the trained algorithm on external datasets (e.g., Brain Tumor Segmentation Challenge [BraTS] challenge datasets). A simple two-dataset train-test split paradigm was not considered to be a sufficiently robust model validation process. Training datasets were also referred to as “discovery” or “development” datasets. The terms referring to hold-out datasets, “test” and “validation”, were used interchangeably.

Top performing algorithms were plotted by year (Fig. 3). RF, CNN and SVM algorithms were highlighted by authors as being

the most effective classifiers. The number of papers applying these algorithms to MRI glioma data appeared to be increasing exponentially until 2019, when the number of publications dropped. Most machine learning research involving glioma MRI data has been published within the last six years. When plotting the country of each lead author by year, it was clear that most glioma MRI machine learning studies produced since 2017 were from the USA, China and South Korea. The most prevalent institutions from each of these countries were presented (Fig. 4).

Many studies employed data augmentation methods ($n = 31$). These methods included the synthetic minority oversampling technique (SMOTE) [111,145,166,218,246] and random oversampling of examples (ROSE) [109,111,166]. Numerous studies augmented their MRI datasets with images that were manipulated in some way. Image manipulation techniques included horizontal and vertical flipping [7,23,45,75,82,95,215], rotating [7,23,45,75,93,171,215], shifting [7,95,137], zooming [45,95], adding noise

Table 2

Overview of included studies that focus on tumor grading and diagnosis, summarizing top-performing machine learning algorithms, specific outcomes classified, validation methods and performance metrics. Note: Risk of bias is based on assessment with the PROBAST tool [157,226]. ACC, accuracy. AUC, area under the receiver operating characteristics curve. CNN, convolutional neural network. CV, cross-validation. DNN, deep neural network. GBM, gradient boosting machine. HGG, high grade glioma. LGG, low grade glioma. PCNSL, primary central nervous system lymphoma. PPV, positive predictive value. SEN, sensitivity. SPC, specificity. SVM, support vector machine. *Testing or validation (not training) metrics are reported.

Study	Top performing algorithm	Application / outcomes classified	Validation methods	Performance measures*	Risk of bias
Tumor grading and diagnosis					
Alcaide-Leon et al. (2017) [9]	SVM	Glioma v PCNSL	10-fold CV x100	AUC = 0.88	+
Alis et al. (2020) [11]	Neural network	LGG v HGG	Train-test	ACC = 88.3; AUC = 0.92; SEN = 0.88; SPC = 0.89	+
Anaraki et al. (2019) [18]	CNN	Glioma grade	Train-test	ACC = 93.1; SEN = 0.97; SPC = 0.98; PPV = 0.93	+
Cho et al. (2018) [52]	Random forest	LGG v HGG	5-fold CV	ACC = 89.0; AUC = 0.92; SEN = 0.94; SPC = 0.73	+
Citak-Er et al. (2018) [55]	SVM	LGG v HGG	10-fold CV	ACC = 93.0; SEN = 0.96; SPC = 0.87	+
Hsieha et al. (2017) [87]	Logistic regression	LGG v HGG	LOOCV	ACC = 93.0; AUC = 0.94; SEN = 0.97; NPV = 0.99	+
Hu et al. (2015) [89]	SVM	LGG v HGG	Train-testLOOCV	ACC = 81.8; SEN = 1.00; SPC = 0.69; PPV = 0.69	+
Inano et al. (2014) [94]	Clustering & SVM	LGG v HGG	5-fold CV	ACC = 80.4; AUC = 0.91; SEN = 0.85; SPC = 0.75	+
Ismael et al. (2020) [95]	CNN	Tumor type	Train-test	ACC = 99.0; SEN = 0.99; PPV = 0.99; F1 = 0.99	+
Hemanth et al. (2019) [85]	CNN	Tumor type	Train-test	ACC = 96.4; SEN = 0.92; SPC = 0.97	+
Jeong et al. (2019) [101]	Random forest	LGG v HGG	LOOCV	ACC = 96.0; AUC = 0.94	–
Khawaldeh et al. (2019) [110]	CNN	Glioma grade	Train-test-validation	ACC = 91.2; SEN = 0.92; PPV = 0.92; F1 = 0.92	+
Kim et al. (2018) [116]	Logistic regression	Glioblastoma v PCNSL	Train-test10-fold CV	ACC = 94.7; AUC = 0.96; SEN = 0.97; SPC = 0.93	+
Koley et al. (2016) [119]	Random forest	Tumor type	Train-test3 & 10-fold CV	ACC = 97.2; SEN = 0.97; PPV = 0.98; F1 = 0.97	+
Kunimatsu et al. (2019) [124]	SVM	Glioblastoma v PCNSL	LOOCV	ACC = 80.0; AUC = 0.99	+
KV & Rajendran (2019) [125]	CNN	Glioma grade	Train-test	SEN = 0.93; SPC = 0.98; PPV = 0.95; F1 = 0.94	+
Mohsen et al. (2017) [156]	DNN	Tumor type	7-fold CV	ACC = 97.0; AUC = 0.98; SEN = 0.97; PPV = 0.97; F1 = 0.97	+
Nakamoto et al. (2019) [159]	SVM	Glioma grade	Train-testLOOCV	ACC = 86.6; AUC = 0.93; SEN = 0.90; SPC = 0.80	+
Nazir et al. (2015) [160]	Neural network	Glioblastoma classification	Train-test-validation	ACC = 94.0	+
Park et al. (2019) [166]	GBM, random forest, elastic net	Glioma grade	Train-test; CV; External validation	ACC = 79.4; AUC = 0.85; SEN = 0.93; SPC = 0.70	+
Qian et al. (2019) [175]	SVM	Glioblastoma v metastases	5-fold CV	ACC = 82.7; AUC = 0.90; SEN = 0.80; SPC = 0.87; PPV = 0.90; NPV = 0.73	+
Sajjad et al. (2019) [188]	CNN	Segmentation and glioma grading	Train-testCV	ACC = 94.6; SEN = 0.88; SPC = 0.96	–
Salçin et al. (2019) [189]	CNN	Tumor type (meningioma, glioma, pituitary)	Train-test	ACC = 91.7; SEN = 0.88; SPC = 1.00; PPV = 1.00; F1 = 0.94	?
Sengupta et al. (2019) [200]	SVM	Glioma grade	12-fold CV	AUC GrIIvGrIII = 0.98, GrIIvGrIV = 0.92; SEN GrIIvGrIII = 0.92, GrIIvGrIV = 1.00; SPC GrIIvGrIII = 1.00, GrIIvGrIV = 0.83	?
Subashini et al. (2016) [207]	SVM	LGG v HGG	Train-test	ACC = 91.0	+
Suh et al. (2019) [209]	Random forest	Glioblastoma v PCNSL	10-fold CV	AUC = 0.92	+
Sun et al. (2019) [210]	Neural network	LGG v HGG	10-fold CV	ACC = 94.4; AUC = 0.97	+
Svolos et al. (2013) [211]	SVM	LGG v HGG	Train-test10-fold CV with 200 repetitions	ACC = 98.0; SEN = 0.51–0.94; SPC = 0.71–1.00	+
Takahashi et al. (2019) [213]	Logistic regression, SVM	LGG v HGG	Train-test5-fold CVLOOCV	ACC = 91.0; AUC = 0.93	+
Tian et al. (2018) [218]	SVM	LGG v HGG; Grade III v Grade IV	10-fold CV, 100 repetitions	ACC LGGvHGG = 96.8, GrIIvGrIV = 98.1; AUC LGGvHGG = 0.99, GrIIvGrIV = 0.99; SEN LGGvHGG = 0.96, GrIIvGrIV = 0.99; SPC LGGvHGG = 0.97, GrIIvGrIV = 0.97	+
Vamvakas et al. (2019) [221]	SVM	LGG v HGG	LOOCV	ACC = 95.5; AUC = 0.96; SEN = 0.95; SPC = 0.96	+
Wiestler et al. (2016) [225]	Random forest	Glioma grade	5-fold CV	ACC = 91.8; AUC = 0.94; SEN = 0.89; SPC = 1.00; PPV = 1.00; NPV = 0.77	+
Wong et al. (2018) [227]	CNN	No tumor v LGG v HGG	Train-test5-fold CV	ACC = 82.0; F1 = 0.80	+

Table 2 (continued)

Study	Top performing algorithm	Application / outcomes classified	Validation methods	Performance measures*	Risk of bias
Wu et al. (2018) [230]	Linear regression	LGG v HGG	10-fold CV	ACC = 91.3; AUC = 0.96; SEN = 0.94; SPC = 0.87	+
Xiao et al. (2018) [232]	Naïve Bayes, SVM	Glioblastoma v PCNSL	10-fold CV	ACC = 88.0; AUC = 0.90; SEN = 0.92; SPC = 0.91	+
Yang et al. (2015) [235]	Random forest	Glioblastoma subtype; survival	3-fold CV	AUC classical = 0.72, mesenchymal = 0.70, neural = 0.75, proneural = 0.82, 12mo survival = 0.69	+
Yang et al. (2018) [239]	CNN multilayer perceptron network & CNN	LGG v HGG	5-fold CV	ACC = 86.7; AUC = 0.94	+
Yun et al. (2019) [241]		Glioblastoma v PCNSL	Train-test-validation10-fold CV	AUC = 0.95; SEN = 0.93; SPC = 0.82	+
Zacharaki et al. (2011) [242]	KNN	Tumor type & grade	LOOCV	ACC mets v glioma = 96.9%; high-grade from low-grade = 94.5%	+
Zhang et al. (2017) [246]	SVM; KNN	LGG v HGG; WHO II, III, IV	LOOCV	ACC LGG v HGG = 94.5, WHO Gr = 96.1; AUC LGG v HGG = 94.5, WHO Gr = 97.1	+
Zollner et al. (2010) [253]	SVM	LGG v HGG	Train-test10-fold CV	ACC = 87.0	+

[239], brightening [95], shearing [45,188], adjusting resolution by down- and up-sampling [150] and including images of other tumor types [171]. One study directly copied samples to balance classes in the training set [165]. Most studies employing data augmentation found that it improved model performance, while some did not [82].

Efficiency improvement was the focus of multiple studies [85,106]. One study developed a deep learning model that used FLAIR, T1CE, and T2 scans, which achieved performance comparable to models trained on these three sequences plus the T1 sequence, removing the need for T1 data [248]. Another study implemented a deep learning model that required less input data to train than other published CNNs; it required MRI sequences from just 10 patient cases to achieve adequate performance [251]. Machine learning was applied to improve image resolution to reduce required scan time [114] and gadolinium contrast agent dose [75]. Some studies successfully trained deep learning algorithms to perform both tumor segmentation and diagnostic classification [154]. Genetic algorithms have been used to automate elements of the machine learning pipeline and model development in a systematic way [206,240].

3.4. Algorithm performance metrics

Numerous measures were used to assess the performance of machine learning models. DSC measures the similarity between two datasets and is widely used in the validation of image segmentation algorithms [171]. Other applied performance measures included sensitivity, specificity, accuracy and AUC. Aggregated performance statistics across all included articles and within each of the application domains are presented (Table 8). Mean algorithm performance outcomes were high, with a mean AUC, accuracy, precision and recall across all studies of 0.87, 0.86, 0.88 and 0.87, respectively. The most commonly reported performance statistics across all studies were accuracy (n = 122), AUC (n = 119), recall/sensitivity (n = 110) and specificity (n = 86). DSC was the metric of choice in the tumor segmentation literature. In all other subdomains, AUC was the metric of choice. Aggregated statistics provide useful benchmarks for future machine learning practitioners assessing the relative performance of their models.

3.5. Natural language processing results

NLP was used for keyword identification, topic modeling and document classification. The top 20 meaningful keywords across all included articles were identified (Table 9). Topic modeling with an LDA and TF-IDF approach showed that coherence was maximised at five topics (Fig. 5). The five topics identified (Table 10) appeared to align with the research subdomain groupings.

3.5.1. Training document classifiers to automate systematic review article screening

A total of 475 abstracts were included in the analysis designed to train a classifier to predict study inclusion; 153 abstracts from included articles and 322 from articles that passed the systematic review screening process but did not fulfil inclusion criteria. RF, LR and NN models trained on TF-IDF data demonstrated good classification performance (AUC = 0.79–0.80), while models trained on count data demonstrated poor performance (AUC = 0.50–0.67; Fig. 6). The document classifier based on BERT yielded good results (mean 5-fold CV AUC = 0.71), with the added advantage of removing the need for data preprocessing. RoBERTa, however, demonstrated unsatisfactory classification performance. NLP was also applied to predict study impact using the text of included abstracts (n = 153). It was, however, not possible to successfully train classification or regression models to predict impact features using any of the above methods.

3.6. Tools used for image processing and machine learning

The tools used by authors for image processing and machine learning were recorded and synthesized (Table 11). A wide array of open source and commercial tools were used in this body of research. The BraTS datasets are an open source repository of deidentified multi-institutional radiological data. They have been used frequently in machine learning image processing competitions to optimise algorithms. Many studies (19%) used the BraTS datasets. Numerous researchers custom-coded their own software for image processing and analysis [94].

Table 3

Overview of included studies that focused on tumor detection and segmentation, summarizing top-performing machine learning algorithms, specific outcomes classified, validation methods and performance metrics. Note: Risk of bias is based on assessment with the PROBAST tool [157,226]. ACC, accuracy. AUC, area under the receiver operating characteristics curve. CNN, convolutional neural network. CV, cross-validation. DSC, Dice similarity coefficient. HGG, high grade glioma. LGG, low grade glioma. LIPC, local independent projection-based classification. PPV, positive predictive value. SEN, sensitivity. SPC, specificity. SVM, support vector machine. *Testing or validation (not training) metrics are reported.

Study	Top performing algorithm	Application / outcomes classified	Validation methods	Performance measures*	Risk of bias
Tumor detection and segmentation					
AlBadawy et al. (2018) [8]	CNN	Tumor segmentation	Train-test10-fold CV	DSC = 0.72	+
Aliotta et al. (2019) [10]	Neural network	Tumor segmentationDTI reconstruction	Train-testLOOCV	AUC = 0.95	+
Amin et al. (2017) [15]	SVM	Cancerous v non-cancerous tissue	5, 10, 15, 20, 25, 30-fold CV	ACC = 97.1; AUC = 0.98; SEN = 0.92; SPC = 0.98	+
Amin et al. (2018) [14]	Random forest	Tumor segmentation	Train-test5-fold CV	ACC = 92.0; AUC = 0.93; SEN = 0.99; SPC = 0.95; PPV = 0.95; NPV = 0.98	+
Amin et al. (2018) [16]	CNN	Tumor segmentation and classification	Train-testExternal validation	ACC = 98.7; SEN = 0.97; SPC = 0.98; PPV = 0.99; DSC = 0.98	+
Amin et al. (2019) [13]	KNN	Tumor detection	10-fold CVExternal validation	ACC = 100; AUC = 1.00; SEN = 1.00; SPC = 1.00	+
Amin et al. (2019) [17]	CNN	Tumor segmentation	Train-testExternal validation	ACC = 99–100; SEN = 0.97–1.00; SPC = 1.00; PPV = 0.97–1.00; DSC = 0.98–1.00	+
Bauer et al. (2013) [26]	SVM	Tumor segmentation	Not reported	DSC = 0.41–0.72	?
Bonte et al. (2018) [33]	Random forest	Tumor segmentation	Train-test	DSC LGG: 40.9% core & 68.4% total region; HGG: 75.0% core & 80.1% total region.	+
Cui et al. (2018) [61]	CNN	Tumor segmentation	10-fold CV	SEN complete = 0.87, core = 0.84, enhancing = 0.76; SPC complete = 0.91, core = 0.77, enhancing = 0.87; DSC complete = 0.90, core = 0.81, enhancing = 0.81	+
El-Dahshan et al. (2010) [66]	KNN	Normal / abnormal MRI	Train	ACC = 98.0; SEN = 1.00; SPC = 0.90	–
Goetz et al. (2015) [74]	Neural network	Tumorous v healthy tissue	Train-testLOOCV	ACC = 86.0; SEN = 0.78; PPV = 0.93; DSC = 0.86	?
Gupta et al. (2017) [80]	SVM	Tumorous v non-tumorous tissue	Train-test10-fold CV	ACC = 98.1; SEN = 1.00; SPC = 0.96; PPV = 0.97	+
Havaei et al. (2017) [82]	CNN	Tumor segmentation	Train-test-validation	SEN: complete 0.87, core 0.79, enhancing 0.80; SPC: complete 0.89, core 0.79, enhancing 0.68	+
Hoseini et al. (2019) [86]	CNN	Tumor segmentation	Train-test	ACC = 91.1	?
Huang et al. (2014) [92]	k-means, LIPC	Tumor segmentation	Train-test	DSC: complete 0.84, core 0.69, enhancing 0.59	+
Huang et al. (2020) [93]	CNN	Tumor diagnosis	Train-test-validation	ACC = 96.9	+
Javed et al. (2013) [100]	SVM	Tumor detection	Train	ACC = 1.00	?
Kayalibay et al. (2017) [108]	CNN	Tumor segmentation	Train-test-validationCV	SEN whole = 0.92, core = 0.87, enhanced = 0.68; PPVwhole = 0.90, core = 0.81, enhanced = 0.57; DSCwhole = 0.90, core = 0.79, enhanced = 0.44	+
Kamnitsas et al. (2017) [106]	CNN	Tumor segmentation	Train-test5-fold CV	SEN: complete 0.81, core 0.72, enhancing 0.73; PPV: complete 0.92, core 0.84, enhancing 0.76; DSC: complete 0.90, core 0.75, enhancing 0.72	+
Kickingereder et al. (2019) [112]	Neural network	Tumor detection and segmentation	Train-test-validation5-fold CV	DSC enhancing = 0.89–0.91, non-enhancing = 0.93	+
Ma et al. (2018) [150]	Random forest	Tumor segmentation	Ensemble method	SEN = 0.85; SPC = 0.89; DSC = 0.86	+
Menze et al. (2014) [152]	Logistic regression;Markov models;Random forest; Neural network	Tumor segmentation	Train-testCV	DSC = 0.88	+
Mlynarski et al. (2019) [154]	DNN	Tumor segmentation	5-fold CV	DSC = 0.81	+

Table 3 (continued)

Study	Top performing algorithm	Application / outcomes classified	Validation methods	Performance measures*	Risk of bias
Pereira et al. (2016) [171]	CNN	Tumor segmentation	Train-test-validation	Combined whole tumor: SEN = 0.86; PPV = 0.85; DSC = 0.78	+
Perkuhn et al. (2018) [172]	CNN	Tumor segmentation	Train-test	ACC = 100; DSC whole = 0.86, enhancing = 0.78, necrosis = 0.62	+
Saouli et al. (2018) [194]	CNN	Tumor segmentation	Not reported	SEN whole tumor = 0.82; SPC whole tumor = 0.74; DSC whole tumor = 0.89	+
Seetha et al. (2018) [198]	CNN	Tumor detection	Train-test	ACC = 97.5	–
Shusharina et al. (2020) [202]	CNN	Tumor segmentation	6-fold CV	DSC = 0.79	+
Tang et al. (2020) [215]	CNN	Tumor segmentation	Train-test 4-fold CV	SEN = 0.88; PPV = 0.81; DSC = 0.84	+
Thaha et al. (2019) [217]	CNN	Tumor segmentation	Train-test	ACC = 92.0; SEN = 0.90; PPV = 0.87	+
Yang et al. (2019) [238]	CNN	Tumor segmentation	5-fold CV	SEN: complete 0.96, core 0.92, enhancing 0.83; PPV: complete 0.83, core 0.70, enhancing 0.92; DSC: complete 0.89, core 0.80, enhancing 0.87	+
Zhao et al. (2016) [247]	CNN	Tumor segmentation	Train-test	ACC = 81.0	+
Zhao et al. (2018) [248]	CNN and Markov random fields	Tumor segmentation	Train-test	PPV: complete 0.89, core 0.76, enhancing 0.64; DSC: complete 0.87, core 0.83, enhancing 0.78	+
Zhuge et al. (2017) [251]	Neural networks	Tumor segmentation	Train-test	SEN = 0.85; DSC = 0.83	+

Table 4

Overview of included studies that focused on detecting tumor progression and recurrence, summarizing top-performing machine learning algorithms, specific outcomes classified, validation methods and performance metrics. Note: Risk of bias is based on assessment with the PROBAST tool [157,226]. ACC, accuracy. AUC, area under the receiver operating characteristics curve. CNN, convolutional neural network. CV, cross-validation. LOOCV, leave one out cross validation. PPV, positive predictive value. SEN, sensitivity. SPC, specificity. SVM, support vector machine. *Testing or validation (not training) metrics are reported.

Study	Top performing algorithm	Application / outcomes classified	Validation methods	Performance measures*	Risk of bias
Detecting tumor progression and recurrence					
Akbari et al. (2016) [6]	SVM	Infiltration and tumor recurrence	LOOCV External validation	ACC = 91.3; AUC = 0.84; SEN = 0.91; SPC = 0.94	+
Costabile et al. (2019) [59]	Logistic regression	Tumor recurrence: Normal tissue, abnormal tissue, tumor	Train	AUC: radiotherapy = 0.75; No radiotherapy = 0.84; All combined = 0.60	+
Gaw et al. (2019) [72]	Neural network	GBM cell density	LOOCV	r = 0.84 (predicted v actual)	?
Hu et al. (2019) [90]	Bayesian framework transfer learning	Identification of tumor invasion	LOOCV	r = 0.88 (predicted v actual tumor invasion)	+
Hu et al. (2011) [91]	SVM	Radiation necrosis v recurrent tumor	Train-test	AUC = 0.94; SEN = 0.90; SPC = 0.94	+
Ismail et al. (2018) [96]	SVM	Tumor progression	Train-test	ACC = 90.2	+
Jang et al. (2018) [99]	CNN	Progression v pseudoprogression	Train-test 10-fold CV	AUC = 0.83; SEN = 0.74; PPV = 0.74; F1 = 0.74	+
Qian et al. (2016) [174]	SVM	Progression v pseudoprogression	10-fold CV	ACC = 86.7; AUC = 0.92	+
Rathore et al. (2018) [181]	SVM	Tumor recurrence	Train-test 10-fold CV	ACC = 89.5; AUC = 0.91; SEN = 0.97; SPC = 0.77	+
Sengupta et al. (2018) [199]	SVM	Edema v nonenhancing tumor	9-fold CV	AUC = 0.97; SEN = 0.93; SPC = 0.93	–

Table 5

Overview of included studies that focused on detecting molecular or genomic characteristics, summarizing top-performing machine learning algorithms, specific outcomes classified, validation methods and performance metrics. Note: Risk of bias is based on assessment with the PROBAST tool [157,226]. ACC, accuracy. AUC, area under the receiver operating characteristics curve. CNN, convolutional neural network. CV, cross-validation. DNN, deep neural network. DSC, Dice similarity coefficient. HGG, high grade glioma. IDH, Isocitrate dehydrogenase. LASSO, Least Absolute Shrinkage and Selection Operator. LGG, low grade glioma. LOOCV, leave one out cross validation. MCC, Matthews correlation coefficient. MGMT, O (6)-Methylguanine-DNA methyltransferase. MNMC, Multi-label Nonlinear Matrix Completion. PPV, positive predictive value. SEN, sensitivity. SPC, specificity. SVM, support vector machine. *Testing or validation (not training) metrics are reported.

Study	Top performing algorithm	Application / outcomes classified	Validation methods	Performance measures*	Risk of bias
Detecting molecular or genomic characteristics					
Akbari et al. (2018) [5]	SVM ensemble	EGFRvIII	10-fold CV External validation	ACC = 87.0; AUC = 0.86; SEN = 0.79; SPC = 0.90	+
Akkus et al. (2017) [7]	CNN	1p/19q	Train-test-validation	ACC = 87.7; SEN = 0.93; SPC = 0.82	+
Alis et al. (2020) [12]	Random forest	IDH1	Train-test-10-fold CV	ACC = 86.9; AUC = 0.93; SEN = 0.60; SPC = 1.00	+
Arita et al. (2018) [19]	LASSO	IDH; TERT	Train-test	ACC = 87.0	+
Bangalore Yogananda et al. (2020) [23]	CNN	IDH	3-fold CV	ACC T2 only = 97.14, T2 T1CE FLAIR = 97.12; T2 only = 0.98, T2 T1CE FLAIR = 0.99; T2 only = 0.97, T2 T1CE FLAIR = 0.98	+
Bisdas et al. (2018) [31]	SVM	IDH; grade II and III glioma	LOOCV	ACC Grade = 78.1, IDH = 83.8%; SEN Grade = 0.77, IDH = 0.96; SPC Grade = 0.79, IDH = 0.55; AUC Grade = 0.79, IDH = 0.87	+
Chang et al. (2018) [47]	CNN	IDH1; 1p/19q; MGMT	5-fold CV	ACC = 83–94	+
Chang et al. (2018) [45]	CNN	IDH	Train-test-validation	ACC = 82.1; AUC = 0.95; SEN = 0.79; SPC = 0.87	+
Chen et al. (2018) [49]	MNMC	IDH1 & MGMT	10-fold CV	ACC: MGMT = 75.6, IDH = 87.0; AUC: MGMT = 0.79, IDH = 0.89; SEN: MGMT = 0.79, IDH = 0.92; SPC: MGMT = 0.68, IDH = 0.85	?
De Looze et al. (2018) [145]	Random forest	IDHGrII/III, GrII/IV, GrIII/IV	5-fold CV	AUC IDH 0.88, GrII/III = 0.98, GrII/IV = 1.00, GrIII/IV = 0.97; SEN IDH 0.81, GrII/III = 0.82, GrII/IV = 1.00, GrIII/IV = 0.83; SPC IDH 0.77, GrII/III = 0.94, GrII/IV = 1.00, GrIII/IV = 0.97	+
Eichinger et al. (2017) [65]	Neural network	IDH	Train-test	ACC = 95.0; AUC = 0.95	+
Fellah et al. (2013) [70]	Random forest	Glioma grade; 1p/19q	Train-test	ACC = 61.0	?
Han et al. (2018) [81]	Random forest	1p/19q	Train-test	ACC = 70.0; AUC = 0.76; SEN = 0.68; SPC = 0.71	+
Hu et al. (2017) [88]	Decision tree	EGFR, PDGFRA, PTEN, CDKN2A, RB1, TP53	LOOCV	ACC EGFR BAT = 75, ENH = 71.9; PDGFRA = 77.1; PTEN BAT = 68.8, ENH = 43.8; CDKN2A BAT = 87.5, ENH = 59.4; RB1 BAT = 87.5, ENH = 43.8; TP53 BAT = 37.5, ENH = 37.5; AUC EGFR BAT = 0.94, ENH = 1.00; PDGFRA = 0.89; PTEN BAT = 0.90, ENH = 1.00; CDKN2A BAT = 0.95, ENH = 0.94; RB1 BAT = 0.95, ENH = 1.00; TP53 BAT = 1.00, ENH = 1.00; AUC = 0.94	+
Jakola et al. (2018) [98]	Logistic regression	IDH; 1p/19q	Train	AUC = 0.94	+
Kanas et al. (2017) [107]	KNN	MGMT	10-fold CV x500	ACC = 73.6; SEN = 0.85; SPC = 0.66	+
Kesler et al. (2019) [109]	Random forest	IDH	LOOCV	ACC = 89.0; AUC = 0.95; SEN = 0.90; SPC = 0.89	+
Kickingereder et al. (2016) [111]	Logistic regression & gradient boosting	MES, RTK, MGMT, EGFR, PDGFRA, MDM4, CDK4, PTEN, NF1, RB1	10-fold CV	ACC EGFR = 63.0, RTKII = 61.0; AUC = 0.63–0.69	+
Kim et al. (2020) [115]	Logistic regression	IDH	Train-test-10-fold CV	AUC IDH = 0.75, grade = 0.82; SEN IDH = 0.54, grade = 0.71; SPC IDH = 0.87, grade = 0.82	+
Kocak et al. (2020) [118]	Neural network	1p/19q	10-fold CV	ACC = 83.8; AUC = 0.89	+
Korfiatis et al. (2016) [120]	SVM	MGMT	5-fold CV	AUC = 0.85; SEN = 0.80; SPC = 0.81	+
Korfiatis et al. (2017) [121]	CNN	MGMT	Train-testCV	ACC = 94.9; SEN = 0.97; PPV = 0.97; F1 = 0.97	+
Lee et al. (2020) [130]	Logistic regression	IDH; 1p/19q	Train	ACC = 83.6; AUC = 0.81; SEN = 0.90; SPC = 0.74	+
Lee et al. (2019) [129]	KNN & AdaBoost	IDH	Train-test	ACC = 83.4	+

Table 5 (continued)

Study	Top performing algorithm	Application / outcomes classified	Validation methods	Performance measures*	Risk of bias
Li et al. (2018) [134]	Logistic regression	EGFR	Train-test	ACC = 90.0; AUC = 0.95	+
Li et al. (2018) [135]	Random forest	MGMT	Train-test	ACC = 80.0; AUC = 0.88; SEN = 0.70; SPC = 0.86	+
Li et al. (2018) [133]	SVM	ARTX	Train-test External validation	ACC = 76.9; AUC = 0.73; SEN = 0.57; SPC = 0.86	+
Li et al. (2019) [132]	SVM	PTEN	Train-test	AUC = 0.79	+
Li et al. (2018) [233]	SVM	p53	Train-test 10-fold CV	ACC = 70.7; AUC = 0.76; SEN = 0.62; SPC = 0.85	+
Li et al. (2017) [136]	DNN	IDH	LOOCV	ACC = 92.4; AUC = 0.95; SEN = 0.94; SPC = 0.87; PPV = 0.95; NPV = 0.84; MCC = 0.80	+
Liang et al. (2018) [137]	DNN	IDH	5-fold CV External validation	ACC = 84.6; AUC = 0.86; SEN = 0.79; SPC = 0.88	+
Liu et al. (2018) [140]	CNN, SVM	H3K27M; segmentation	5-fold CV	ACC = 96.5; AUC = 0.95; SEN = 0.97; DSC = 0.77	?
Liu et al. (2019) [144]	Logistic regression	IDH; 1p/19q	Train	AUC IDH = 0.85, IDH + codeletion = 0.81, IDHwildtype = 0.888	+
Lu et al. (2018) [148]	SVM	IDH; 1p/19q; LGG v HGG	Train-test 5-fold CV	LGG v HGG ACC = 87.7, SEN = 0.83, SPC = 0.91; HGG IDH ACC = 88.9, SEN = 0.88, SPC = 1.00; LGG IDH ACC = 91.7, SEN = 0.86, SPC = 1.00; LGG 1p/19q ACC = 80.0, SEN = 1.00, SPC = 0.67	+
Pan et al. (2019) [165]	Random forest	H3K27M	Train-test	ACC = 84.4; AUC = 0.83	+
Park et al. (2018) [167]	LASSO	IDH; 1p/19q	Train-test 5, 10-fold CV	AUC = 0.78	+
Park et al. (2018) [168]	LASSO	IDH; 1p/19q	Train-test 5, 10-fold CV	AUC = 0.81	+
Rathore et al. (2018) [182]	K-means clustering, SVM	EGFR	Train-test cluster x1000 10-fold CV	3 clusters: rim-enhancing, irregular and solid. EGFR ACC = 80.2	+
Ren et al. (2019) [185]	SVM	IDH; ARTX	10-fold CV	IDH ACC = 94.7, AUC = 0.93, SEN = 1.00, SPC = 0.86, PPV = 0.92, NPV = 1.00; ARTX ACC = 91.7, AUC = 0.93, SEN = 0.95, SPC = 0.88, PPV = 0.90, NPV = 0.94	+
Shofty et al. (2018) [201]	Ensemble bagged trees	IDH; 1p/19q	5-fold CV	ACC = 87.0; AUC = 0.87; SEN = 0.92; SPC = 0.83	+
Su et al. (2020) [206]	TPOT autoML	H3K27M	Train-test 10-fold CV	ACC = 80.0; AUC = 0.90; SEN = 0.64; SPC = 0.93; PPV = 0.91	+
Tan et al. (2019) [214]	SVM	IDH	Train-test 5-fold CV	AUC = 0.90	+
Wu et al. (2019) [228]	Random forest	IDH	Train-test x2010-fold CV x5	ACC = 87.6; AUC = 0.93	+
Yu et al. (2017) [240]	SVM	IDH	LOOCV External validation	ACC = 83.0; AUC = 0.79; SEN = 0.88; SPC = 0.67; PPV = 0.91; NPV = 0.57; MCC = 0.51	+
Zhang et al. (2017) [244]	Random forest	IDH	Train-test	ACC = 89.0; AUC = 0.92	+
Zhang et al. (2018) [245]	SVM	IDH; TP53	Train-test 10-fold CV	ACC IDH = 79.2, TP53 = 86.9; SEN IDH = 0.85, TP53 = 0.85; SPC IDH = 0.70, TP53 = 0.86	+
Zhou et al. (2019) [249]	Random forest	IDH; 1p/19q	Train-test External validation	Overall ACC = 78.2; AUC IDH = 0.92, 1p/19q = 0.72	+
Zinn et al. (2018) [252]	Extreme gradient boosting machine	POSTN	Train-test LOOCV	AUC = 0.92; SEN = 0.93; SPC = 0.92	+

Table 6

Overview of included studies that focused on survival prediction, summarizing top-performing machine learning algorithms, specific outcomes classified, validation methods and performance metrics. Note: Risk of bias is based on assessment with the PROBAST tool [157,226]. ACC, accuracy. AUC, area under the receiver operating characteristics curve. CNN, convolutional neural network. CV, cross-validation. DNN, deep neural network. GBM, gradient boosting machine. HGG, high grade glioma. LOOCV, leave one out cross validation. OS, overall survival. PFS, progression free survival. PPV, positive predictive value. SEN, sensitivity. SPC, specificity. SVM, support vector machine. *Testing or validation (not training) metrics are reported.

Study	Top performing algorithm	Application / outcomes classified	Validation methods	Performance measures*	Risk of bias
Survival prediction					
Bae et al. (2018) [21]	Random survival forest	OS, PFS	Train-test10-fold CV	AUC PFS = 0.74, OS = 0.78	+
Chaddad et al. (2018) [43]	Random forest	OS (320d), PFS (213d)	Train-test	AUC OS = 0.74, PFS = 0.84	+
Chang et al. (2016) [46]	Random forest	OS (289d), PFS (147d)	Train-testLOOCV	ACC OS = 77.0, PFS = 65.0	+
Choi et al. (2020) [53]	Random survival forest	OS	External validation	AUC = 0.71	+
Emblem et al. (2014) [67]	SVM	Survival	Train-test10-fold CV	AUC: 1-year 0.77, 2-year 0.76, 3-year 0.81, 4-year 0.82; SEN: 1-year 0.79, 2-year 0.81, 3-year 0.87, 4-year 0.88; SPC: 1-year 0.79, 2-year 0.71, 3-year 0.75, 4-year 0.80	–
Emblem et al. (2015) [68]	SVM	Survival	Train-test10-fold CV	AUC: 6-month 0.78, 1-year 0.70, 2-year 0.70, 3-year 0.80	?
Kim et al. (2019) [113]	K-means	Glioblastoma clustering	Clustering	3 clusters; 1 associated with survival	+
Lao et al. (2017) [126]	CNN	OS	Train-test	AUC = 0.71	+
Liao et al. (2019) [138]	GBM	Short (<12mo) v long survival	Train-test	ACC = 81.0; AUC = 0.81	+
Liu et al. (2017) [143]	SVM	Short (<12mo) v long survival	10-fold CV	ACC = 78.0; AUC = 0.81; SEN = 0.78; SPC = 0.79	+
Liu et al. (2019) [141]	SVM	Survival time	LOOCV	ACC = 86.8; SEN = 0.85; SPC = 0.88	+
Peeken et al. (2019) [170]	Random survival forest	OS (12mo), PFS (4.3mo)	Train-test	AUC = 0.80	+
Sanghani et al. (2018) [193]	SVM	Short (<10mo), medium (10–15mo) and long survival (>15mo)	10-fold CV	ACC 2-class = 98.7, 3-class = 88.95	+
Suchorska et al. (2019) [208]	Random survival forest	High v low risk	Train-test	AUC = 0.85	+
Upadhaya et al. (2015) [220]	SVM	Short v long survival (threshold = 14.8mo)	LOOCV	ACC = 90.0; SEN = 0.85; SPC = 0.95	+
Zacharaki et al. (2012) [243]	J48 decision tree	Short- and long-term survivors (18mo)	L3OCV	ACC = 85.1; AUC = 0.84	+
Zhou et al. (2017) [250]	SVM	Short- v long-term survival	LOOCV	ACC = 86.0; SEN = 0.86; PPV = 0.94	+
Other applications					
Gong et al. (2018) [75]	CNN	Image enhancement	Train-test	Image enhanced, reduced contrast dose	+
Kim et al. (2018) [114]	CNN	Resolution enhancement	Train-test	Resolution enhanced	+

Table 7
Descriptive statistics for publications that met inclusion criteria.

Variable	Mean (±SD)	Median
Years since publication	2.43 (±1.98)	2.00
Sample size (n patients included)	198.56 (±429.37)	123.00
Number of contributing institutions	1.83 (±3.95)	1.00
Citations	67.61 (±213.59)	23.00
Mean citations per year	22.18 (±59.22)	9.57
Journal impact factor	4.27 (±3.45)	3.65
SCImago Journal Rank	1.59 (±1.68)	1.41
Journal H index (SCImago)	113.72 (±66.56)	111.00

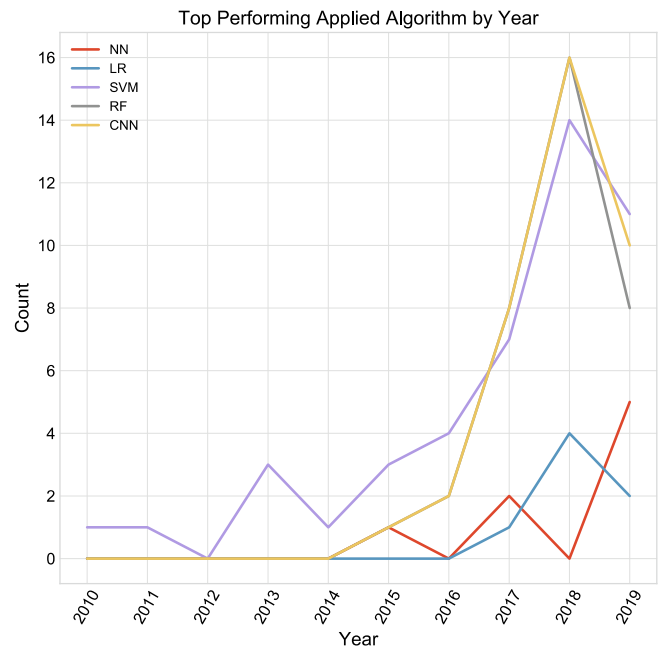


Fig. 3. A summary of the top performing applied algorithms by year. Algorithms reported by authors to demonstrate superior performance when compared to other algorithms and those applied in > 5 papers are shown.

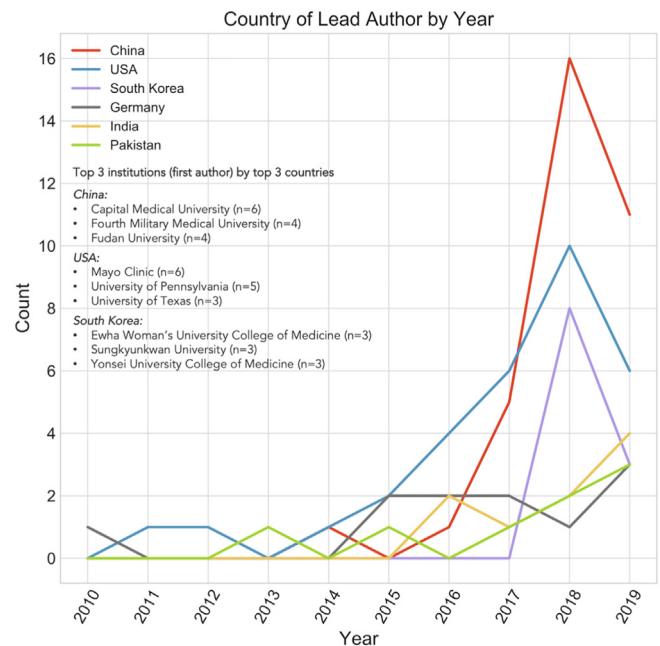


Fig. 4. Country of lead author by year. Countries that produced > 5 publications in the included set of articles overall are shown.

4. Discussion

This study was designed to systematically review machine learning methods applied to glioma MRI data, analyse article characteristics, and explore the application of NLP methods to the systematic review process. This review comprehensively mapped and synthesized the ecosystem of research that has applied machine learning techniques to MRI data for image processing and outcome prediction in glioma patients over the last decade. Primary application domains identified were (1) tumor grading and diagnosis, (2) tumor segmentation, (3) non-invasive genomic biomarker identification, (4) detection of tumor progression and (5) survival prediction. Machine learning methods appeared to demonstrate substantial impact and utility within each of these research domains and the performance of developed models was generally strong. Of particular note was the substantial corpus of studies that focused on the non-invasive identification of tumor genetic and molecular biomarkers. This body of work has the potential to substantially change clinical practice and may eliminate the need for brain biopsy and associated investigations, which are important clinical decision making issues that come up commonly in practice. Machine learning models typically demonstrated high levels of performance and the synthesized metrics presented may provide useful benchmarks for future researchers. It has been suggested that machine learning models are capable of more accurate prognosis prediction than histopathological categorization alone [243]. The top 3 algorithms highlighted as being most effective by authors were random forest, convolutional neural network and support vector machine algorithms. It was common practice for machine learning studies to train and compare multiple models, varying inputs or the algorithms applied or both for a given output feature or set of outputs. The application of transfer learning, using pre-trained deep neural networks appeared to show particular promise for improved diagnostics and segmentation [93]. It was evident that radiomics has made substantial contributions to preoperative glioma diagnosis and treatment planning [221]. Applications of machine learning to MRI glioma data were heterogeneous, which highlights the utility and broad applicability of machine learning technologies. In addition to the domains mentioned above, specific and notable applications included image enhancement resulting in reduced gadolinium contrast dosage [75] and image correction and improvement [74,114]. An inventory of tools used by researchers, both open-source and commercial, was developed and presented to facilitate future research. The number of studies applying machine learning methods to MRI glioma data appeared to be increasing exponentially until a drop in the number of publications in 2019. This drop may be a consequence of a shifting global market for machine learning capabilities, shifts in research priorities or long peer-review and publication processing times. Machine learning capabilities applied to MRI data show substantial potential to deliver effective clinical decision support and improved care in the management of glioma patients [4].

Multiple supervised machine learning classifiers were successfully trained to automate the systematic review screening process using NLP methods. Classification models developed using transfer learning methods, leveraging BERT, yielded acceptable text classification performance results. This kind of approach may be used to more efficiently generate future systematic reviews by partially automating article screening. Because the present review was based on publicly available data and open-source technologies, it can be easily replicated. Furthermore, an updated future review on the same topic can be efficiently produced by (1) collecting the abstracts of included articles, (2) training a classifier with these abstracts, (3) conducting an updated comprehensive literature

Table 8
Aggregated performance metrics across included studies and divided by research subdomain. AUC, area under the receiver operating characteristics curve. DSC, Dice similarity coefficient. NPV, negative predictive value. PPV, positive predictive value. SD, standard deviation. Note: na denotes ≤ 3 studies. Some studies presented multiple high performing models predicting different outcomes (e.g., multiple genetic biomarkers), so n-models in the metrics columns in some cases exceeds n-studies in the research subdomain column.

Research subdomain (n studies)	Performance metrics (mean $\pm$ SD, n models)							
	AUC	Accuracy	Recall (Sensitivity)	Specificity	Precision (PPV)	NPV	F1	DSC
All studies (n = 153)	0.87 $\pm$ 0.09 (n = 119)	0.86 $\pm$ 0.12 (n = 122)	0.87 $\pm$ 0.10 (n = 110)	0.86 $\pm$ 0.10 (n = 86)	0.88 $\pm$ 0.11 (n = 38)	0.85 $\pm$ 0.15 (n = 9)	0.92 $\pm$ 0.09 (n = 10)	0.81 $\pm$ 0.11 (n = 39)
Diagnosis (n = 43)	0.91 $\pm$ 0.08 (n = 30)	0.91 $\pm$ 0.06 (n = 15)	0.93 $\pm$ 0.05 (n = 29)	0.89 $\pm$ 0.10 (n = 24)	0.93 $\pm$ 0.09 (n = 10)	0.83 $\pm$ 0.11 (n = 3)	0.93 $\pm$ 0.06 (n = 7)	na
Segmentation (n = 35)	0.97 $\pm$ 0.03 (n = 4)	0.95 $\pm$ 0.06 (n = 15)	0.87 $\pm$ 0.09 (n = 29)	0.90 $\pm$ 0.10 (n = 14)	0.85 $\pm$ 0.12 (n = 20)	na	na	0.81 $\pm$ 0.11 (n = 37)
Progression (n = 8)	0.86 $\pm$ 0.13 (n = 7)	0.89 $\pm$ 0.02 (n = 4)	0.89 $\pm$ 0.09 (n = 5)	0.90 $\pm$ 0.08 (n = 4)	na	na	na	na
Genomics (n = 48)	0.88 $\pm$ 0.09 (n = 58)	0.80 $\pm$ 0.14 (n = 56)	0.83 $\pm$ 0.12 (n = 38)	0.84 $\pm$ 0.11 (n = 36)	0.93 $\pm$ 0.03 (n = 5)	0.84 $\pm$ 0.19 (n = 4)	na	na
Survival (n = 26)	0.78 $\pm$ 0.05 (n = 19)	0.84 $\pm$ 0.09 (n = 10)	0.84 $\pm$ 0.04 (n = 8)	0.81 $\pm$ 0.08 (n = 7)	na	na	na	na

Table 9
Top 20 keywords identified across all abstracts of included studies.

Identified keyword	Frequency
'feature'	447
'tumor'	368
'image'	291
'patient'	277
'glioma'	250
'model'	238
'mri'	236
'using'	224
'imaging'	180
'brain'	175
'accuracy'	158
'segmentation'	149
'data'	139
'grade'	135
'learning'	132
'classification'	128
'status'	127
'performance'	127
'machine'	126
'training'	121

Table 10
Topic modeling results based on an LDA and TF-IDF approach.

Topic number	Identified topic keywords and corresponding weights
1	0.017***grading + 0.016***grade + 0.015***tissue + 0.014***segmentation + 0.013***set + 0.012***information + 0.010***novel + 0.010***treatment + 0.009***large + 0.009***prediction
2	0.020***mutation + 0.016***gbm + 0.016***status + 0.012***idh + 0.012***multiparametric + 0.012***clinical + 0.011***auc + 0.010***parameter + 0.010***histogram + 0.010***prediction
3	0.018***radiomics + 0.017***survival + 0.015***gbm + 0.014***svm + 0.013***brain + 0.013***detection + 0.012***classification + 0.011***time + 0.011***proposed + 0.011***mr
4	0.019***gbm + 0.014***radiomics + 0.014***cohort + 0.014***idh + 0.012***auc + 0.012***three + 0.012***n + 0.011***conventional + 0.010***enhancing + 0.010***classifier
5	0.016***segmentation + 0.014***radiomic + 0.013***brain + 0.012***grade + 0.011***status + 0.011***network + 0.010***idh + 0.010***classification + 0.010***training + 0.010***classifier

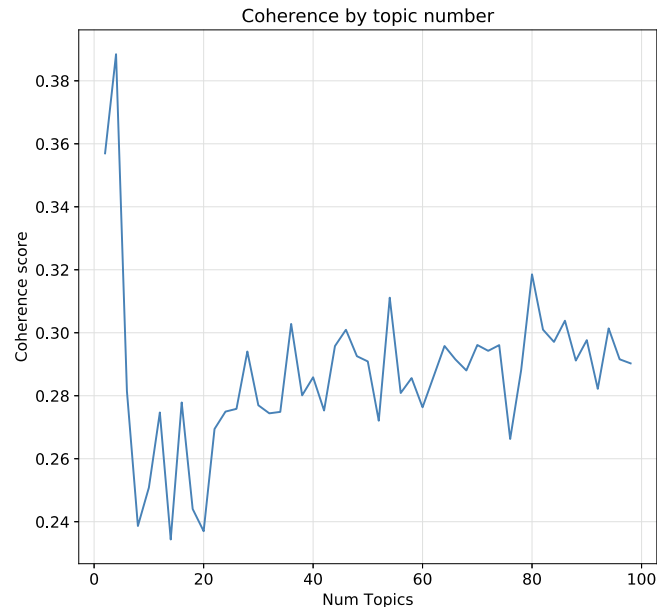


Fig. 5. Plotting coherence by topic number during topic modeling.

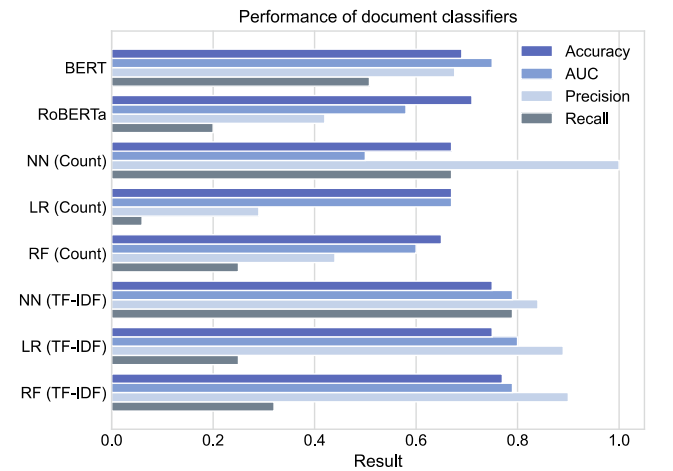


Fig. 6. Five-fold cross validation classification performance statistics for each of the models trained to predict study inclusion.

search, and (4) running the new abstracts collected in the search through the trained model to yield a classification score for each. In fact, this approach may be applied to any systematic review, potentially substantially reducing the time required to produce

Table 11

Image processing and machine learning tools and datasets used by the authors of included studies.

Function	Tools
Image processing, feature extraction and tumor segmentation	• 3D Slicer (www.slicer.org) • AnalyzeDirect software (version 11.0, Mayo Clinic, Rochester, MN, USA) • Brainlab • Brainstat (GE Healthcare) • Brain Suite • Brain Tumor Image Analysis (BraTumIA) • BrainWeb [58] • COVALIC (Comparison and Validation of Image Computing) toolkit (github.com/InsightSoftwareConsortium/covalic) • FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL) • Functool (GE, Milwaukee, WI, USA) • Insight Segmentation and Registration Toolkit (ITK) (www.itk.org) • ITK-SNAP • LIFEx (www.lifexsoft.org) [161] • MITK (www.mitk.org; German Cancer Research Center) • MN152 (Montreal Neurological Institute (MNI)) brain atlas • Monaco (www.elekta.com) • MRlcron software (The NeuroImaging Tools and Resources Collaboratory, www.mccauslandcenter.sc.edu/mricron) • NordicICE (Nordic NeuroLab, Bergen, Norway) • Olea Sphere v2.3, Olea Medical, La Ciotat, France • Pyradiomics (http://www.radiomics.io/pyradiomics.html) [77] • R (R Foundation for Statistical Computing, Vienna, Austria) [216] ◦ RNiftyReg ◦ Advanced Normalization Tools (ANTs) repository (ANTsR). • SLIC 3D [3] • Statistical Parametric Mapping package (SPM; www.fil.ion.ucl.ac.uk/spm/) • SUSAN [204] • TumorSim [173] • TexRAD (www.texrad.com)
Machine learning	• CAFE [102] • Concatenated random forest modeling (github.com/ntustison/BraTS2013) • Deeplearning4j (DL4J) • Matlab ◦ SSIM function ◦ Statistics and Machine-Learning Toolbox ◦ Classification learner tool ◦ LIBSVM [44] • MXNet deep learning framework [50] • Python (www.python.org) ◦ Tensorflow [2] ◦ Scikitlearn [169] ◦ Orange [62] ◦ Theano [28] ◦ NetworkX ◦ DEAP [71] • R (R Foundation for Statistical Computing, Vienna, Austria) [216] ◦ Boruta ◦ mRMR ◦ e1071 ◦ pROC ◦ caret [123] ◦ ROSE ◦ DMwR ◦ mlr [30] ◦ randomForestSRC • SPSS • TPOT (https://github.com/rhiever/tpot) [163] • Wndchrm (Weighted Neighbour Distance using Compound Hierarchy of Algorithms Representing Morphology) • Weka (https://www.cs.waikato.ac.nz/ml/weka/)
Datasets and data sources	• BraTS datasets [152] • Chinese Genome Atlas (CGGA) database • MRIs from Nanfang Hospital, Guangzhou, China, and General Hospital, Tianjing Medical University, China [51] • National Center for Biotechnology Information's Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) • Radiopaedia • The Cancer Imaging Archive (TCIA) [57] • The Cancer Genome Atlas (TCGA) database (http://cancergenome.nih.gov)
CNN architectures	• AlexNet [122] • CNN_S [48] • DeepMedic [106] • DFFM (github.com/tfnfyd110/DFFM) • GoogLeNet [212] • InputCascadeCNN [82] • ResNet [84] • U-net [187] • VGG-16 and VGG-19 [203]

updated reviews and perhaps closing the translational gap that presently exists between published research and clinical practice.

CNNs appeared to afford a collection of advantages over more traditional machine learning algorithms. Deep learning has demonstrated superior performance [93,176]. CNNs appeared to be best applied for successfully conducting segmentation and detection. CNNs do not require a separate feature extraction and selection process [95] and may, therefore, be more amenable to replication and verification. Variations in feature extraction and selection methods in classical machine learning studies inhibit reproducibility [131]. Despite the availability of open-source deep learning software and the demonstrated high classification performance of CNNs, the application of more classical machine learning algorithms (e.g., SVM and RF) in recent years is still prevalent, perhaps because traditional machine learning approaches demonstrate superior performance on smaller datasets when compared with deep learning [22], MRI data is relatively scarce and expensive to acquire and glioma cases are rare. Data augmentation has commonly been applied to overcome data shortages and appears to have a significant impact on the performance of deep learning models [171]. Using multiple different techniques to create an augmented dataset appears to generate optimal results [95], although methods deploying CNNs successfully to smaller datasets have been developed [251] and warrant further investigation. Genetic algorithms have been used to optimize the architecture of CNNs and this approach appears to facilitate the efficient development of high performing deep learning classifiers [18]. The disadvantages of CNNs include the requirement for more time, energy, hardware and expertise to train, validate and apply. Transfer learning [93] can facilitate the process of efficiently developing a CNN-based system.

Various model validation methods were applied. The most prevalent was some form of cross-validation. External validation is useful in developing insight into the generalizability of a model and was also widely applied. One study showed that the tumor segmentation performance of a CNN trained and tested on data from the same institution degraded substantially when tested on data from a different institution [8]. A caveat for consideration here with regard to external validation is that organizational specificity may be of value to clinicians. Machine learning models trained on a given organization's sample are tuned to that organization's patient population and may generate useful results for that population despite the fact that they may not generalize. It has been suggested that machine learning researchers may consider striking a balance between generalizability and specificity [39]. When referring to a second hold-out performance evaluation dataset, the terms "validation" and "test" were used interchangeably. Studies referred to (1) "training" and "test" sets, (2) "training" and "validation" sets, and (3) "training", "testing" and "validation" sets. Ismael et al. 2020 [95] described a distinction between patient-level and image-level validation methods. The patient-level method results in validation datasets that contain data that is completely separate and unseen by the model. Because MRI scans consist of multiple layered images, the image-level splitting method results in validation datasets that contain some images from patients included in the training set. In this case, the validation and testing sets are not truly unseen by the model, potentially resulting in less optimal validation and overly optimistic performance evaluation. Both methods were employed in this body of literature.

There was substantial variability in the nature of the performance metrics reported. Some authors reported only model accuracy (well recognised as a flawed metric) while others included a

full suite of metrics including AUC, sensitivity, specificity, NPV, and PPV, among others. Variability in the reporting of performance metrics may limit insight into the utility and safety of a model, particularly when the nature of the error is of consequence to patient safety and clinical outcomes. Studies would do well to report multiple metrics including accuracy, AUC, sensitivity (i.e., recall), specificity and precision as a minimum.

Efficiency improvement was a notable theme in this body of literature. As algorithms and hardware develop over time, it is evident that efficiency improvement opportunities continually present themselves. The tumor segmentation task and CNN algorithms are particularly computationally intensive [95]. Algorithm training time efficiency gains appear to have been a recurring theme over the last decade. It has been clearly shown that CNNs demonstrate excellent image segmentation and classification performance, so more recent studies have focused on achieving comparable performance with less data and in less time. Systems that can generate accurate and clinically useful results using minimal MRI sequences represent an opportunity to improve healthcare efficiency [33]. There are opportunities for future researchers to make a contribution by continuing to explore methods for improving efficiency.

It was evident that the existence of open-source de-identified clinical datasets has stimulated this field of research. The BraTS datasets in particular, consisting of labelled glioma radiology data, appear to have been a valuable resource for many researchers [13,17,33,74,171,194,238,248]. These datasets remain online and accessible as a benchmarking resource. While the dataset has grown over time, sample sizes are still somewhat limited. Other accessible datasets of note included The Cancer Imaging Archive and The Cancer Genome Atlas. The continued development of larger, deidentified, publicly available neuroimaging datasets would be of substantial utility to researchers pursuing patient care improvements by developing machine learning driven software tools.

A clinical decision support system may be designed and developed to synthesize and comprehensively deliver the functionality explored in this body of literature in clinical practice, resulting in tangible benefits to patients. Such a system may adhere to the following constraints. It may employ (1) computationally efficient and high performing CNNs that are interpretable, (2) minimal standard preoperative MRI sequence inputs that necessitate minimal or no contrast administration, (3) effective automatic tumor segmentation, (4) classification of a full suite of clinically important genetic and molecular markers, (5) prediction of tumor progression, and (6) prediction of patient survival. A system incorporating these elements of functionality may become an indispensable tool for clinicians and patients, facilitating more informed clinical decision making and treatment optimization. It appears that some researchers are successfully pursuing the development of an automated MRI-based brain tumor diagnostic system to facilitate clinical practice [95]. Incorporating the full set of functions outlined above within a single system may represent a worthy challenge. Tangible clinical change will only be achieved if viable and readily usable clinical systems that leverage trained models are developed and implemented in clinical practice [37].

4.1. Limitations and future research

Multiple future research opportunities were identified. Almost no research has investigated the application of reinforcement learning or generative adversarial networks within this problem domain and few studies employed unsupervised learning. There

may be an opportunity to make a contribution by exploring the applicability of these techniques. An advantage of unsupervised methods is that they do not require expensive and time-consuming case labelling by experts. There is a risk, though, that the derived categorisation, may not align with existing grading taxonomies and may be difficult to interpret; therefore, weakly-supervised approaches may offer a viable trade-off. Different algorithms appear to demonstrate performance variability on different components of the tumor identification and segmentation task [152]. An optimal machine learning approach may involve the use of multiple algorithms in an ensemble that leverages strengths and mitigates weaknesses on different components of the classification and segmentation tasks. Feature importance was rarely discussed or visualized in the studies included. Improving algorithm interpretability remains an opportunity for development for applied machine learning researchers.

This review incorporated a relatively large sample size which may have underpinned the observed performance of the NLP classification models developed. Most systematic reviews, however, tend to include less articles and the optimisation of deep language models to achieve successful classification with smaller sample sizes would be beneficial to the research community. When considering the automation of the systematic review article screening process, the application of domain-specific, tuned deep language models like Clinical-BERT or bioBERT may result in stronger classification results. This is an opportunity for future researchers. To limit scope, this review excluded conference papers. However, numerous leading machine learning MRI processing approaches have been published only as conference papers, which are more common in the computer science community. We point interested readers, in particular, to the MICCAI BraTS and BrainLes papers [22].

5. Conclusion

Machine learning has been widely applied to, and has demonstrated substantial utility in, the processing of MRI data in glioma research. Many high performing machine learning models have been published, predicting a broad array of clinically useful outcomes. This study highlights the potential of machine learning methods for the development of personalised clinical neurosurgical care. These techniques may be deployed to improve treatment precision and prolong patient survival. The identification and extraction of biological tumor information using non-invasive image analysis, in particular, holds great promise for treatment advancement. Furthermore, the development and public release of deep transformer-based language models appears to have significant potential to facilitate efficient academic research synthesis and translation to clinical practice, perhaps substantially shortening the time from discovery to application in medicine.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Aaronson NK, Taphoorn MJB, Heimans JJ, Postma TJ, Gundy CM, Beute GN, et al. Compromised health-related quality of life in patients with low-grade glioma. *J Clin Oncol* 2011;29:4430–5.
- [2] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M (2016) Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv Prepr arXiv160304467*.
- [3] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell* 2012;34:2274–82.
- [4] Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5:1–9.
- [5] Akbari H, Bakas S, Pisapia JM, Nasrallah MP, Rozycki M, Martinez-Lage M, et al. In vivo evaluation of EGFRvIII mutation in primary glioblastoma patients via complex multiparametric MRI signature. *Neuro Oncol* 2018;20:1068–79.
- [6] Akbari H, Macyszyn L, Da X, Bilello M, Wolf RL, Martinez-Lage M, et al. Imaging surrogates of infiltration obtained via multiparametric imaging pattern analysis predict subsequent location of recurrence of glioblastoma. *Neurosurgery* 2016;78:572–80.
- [7] Akkus Z, Ali I, Sedlár J, Agrawal JP, Parney IF, Giannini C, et al. Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence. *J Digit Imaging* 2017;30:469–76.
- [8] AlBadawy EA, Saha A, Mazurowski MA. Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing. *Med Phys* 2018;45:1150–8.
- [9] Alcaide-Leon P, Dufort P, Geraldo AF, Alshafai L, Maralani PJ, Spears J, et al. Differentiation of enhancing glioma and primary central nervous system lymphoma by texture-based machine learning. *Am J Neuroradiol* 2017;38:1145–50.
- [10] Aliotta E, Nourzadeh H, Sanders J, Muller D, Ennis DB. Highly accelerated, model-free diffusion tensor MRI reconstruction using neural networks. *Med Phys* 2019;46:1581–91.
- [11] Alis D, Bagcilar O, Senli YD, Isler C, Yergin M, Kocer N, et al. The diagnostic value of quantitative texture analysis of conventional MRI sequences using artificial neural networks in grading gliomas. *Clin Radiol* 2020.
- [12] Alis D, Bagcilar O, Senli YD, Yergin M, Isler C, Kocer N, et al. Machine learning-based quantitative texture analysis of conventional MRI combined with ADC maps for assessment of IDH1 mutation in high-grade gliomas. *Jpn J Radiol* 2020;38:135–43.
- [13] Amin J, Sharif M, Raza M, Saba T, Anjum MA. Brain tumor detection using statistical and machine learning method. *Comput Methods Programs Biomed* 2019;177:69–79.
- [14] Amin J, Sharif M, Raza M, Yasmin M. Detection of brain tumor based on features fusion and machine learning. *J Ambient Intell Humaniz Comput* 2018:1–17.
- [15] Amin J, Sharif M, Yasmin M, Fernandes SL. A distinctive approach in brain tumor detection and classification using MRI. *Pattern Recognit Lett* 2017.
- [16] Amin J, Sharif M, Yasmin M, Fernandes SL. Big data analysis for brain tumor detection: Deep convolutional neural networks. *Futur Gener Comput Syst* 2018;87:290–7.
- [17] Amin J, Sharif M, Yasmin M, Saba T, Anjum MA, Fernandes SL. A new approach for brain tumor segmentation and classification based on score level fusion using transfer learning. *J Med Syst* 2019;43:326.
- [18] Anaraki AK, Ayati M, Kazemi F. Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. *Biocybern Biomed Eng* 2019;39:63–74.
- [19] Arita H, Kinoshita M, Kawaguchi A, Takahashi M, Narita Y, Terakawa Y, et al. Lesion location implemented magnetic resonance imaging radiomics for predicting IDH and TERT promoter mutations in grade II/III gliomas. *Sci Rep* 2018;8:1–10.
- [20] Avorn J, Fischer M. 'Bench to behavior': translating comparative effectiveness research into improved clinical practice. *Health Aff* 2010;29:1891–900.
- [21] Bae S, Choi YS, Ahn SS, Chang JH, Kang S-G, Kim EH, et al. Radiomic MRI phenotyping of glioblastoma: improving survival prediction. *Radiology* 2018;289:797–806.
- [22] Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, Shinohara RT, Berger C, Ha SM, Rozycki M. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge; 2018. *arXiv Prepr arXiv181102629*.
- [23] Bangalore Yogananda CG, Shah BR, Vejdani-Jahromi M, Nalawade SS, Murugesan GK, Yu FF, et al. A novel fully automated MRI-based deep-learning method for classification of IDH mutation status in brain gliomas. *Neuro Oncol* 2020;22:402–11.
- [24] Bannach-Brown A, Clark J. Systematic Review & Meta-Analysis: Automation tools to help your review. In: 16th Australian Conference on Personality & Individual Differences: Advances and Challenges in Personality and Individual Differences-Theories and Applications, 2018.
- [25] Barajas Jr RF, Hodgson JG, Chang JS, Vandenberg SR, Yeh R-F, Parsa AT, et al. Glioblastoma multiforme regional genetic and cellular expression patterns: influence on anatomic and physiologic MR imaging. *Radiology* 2010;254:564–76.
- [26] Bauer S, Lu H, May CP, Nolte L, Büchler P, Reyes M. Integrated segmentation of brain tumor images for radiotherapy and neurosurgery. *Int J Imaging Syst Technol* 2013;23:59–63.
- [27] Bauer S, Wiest R, Nolte L-P, Reyes M. A survey of MRI-based medical image analysis for brain tumor studies. *Phys Med Biol* 2013;58:R97.
- [28] Bergstra J, Breuleux O, Bastien F, Lamblin P, Pascanu R, Desjardins G, et al. Theano: A CPU and GPU math compiler in Python. In: *In: Proc. 9th Python in Science Conf. p. 3–10*.

- [29] Bird S, Klein E, Loper E. Natural language processing with Python: analyzing text with the natural language toolkit. "O'Reilly Media, Inc."; 2009.
- [30] Bischl B, Lang M, Kothhoff L, Schiffner J, Richter J, Studerus E, et al. mlr: Machine Learning in R. *J Mach Learn Res* 2016;17:5938–42.
- [31] Bisdas S, Shen H, Thust S, Katsaros V, Stranjalis G, Boskos C, et al. Texture analysis-and support vector machine-assisted diffusional kurtosis imaging may allow in vivo gliomas grading and IDH-mutation status prediction: a preliminary study. *Sci Rep* 2018;8:1–9.
- [32] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Mach Learn Res* 2003;3:993–1022.
- [33] Bonte S, Goethals I, Van Holen R. Machine learning based brain tumour segmentation on limited data using local texture and abnormality. *Comput Biol Med* 2018;98:39–47.
- [34] Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Syst Rev* 2017;6:245.
- [35] Brandsma D, Stalpers L, Taal W, Sminia P, van den Bent MJ. Clinical features, mechanisms, and management of pseudoprogression in malignant gliomas. *Lancet Oncol* 2008;9:453–61.
- [36] Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
- [37] Buchlak QD, Esmaili N, Leveque J-C, Bennett C, Piccardi M, Farrokhi F. Ethical thinking machines in surgery and the requirement for clinical leadership. *Am J Surg* 2020.
- [38] Buchlak QD, Esmaili N, Leveque J-C, Farrokhi F, Bennett C, Piccardi M, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. *Neurosurg Rev* 2019;1–19. <https://doi.org/10.1007/s10143-019-01163-8>.
- [39] Buchlak QD, Yanamadala V, Leveque J-C, Edwards A, Nold K, Sethi R. The Seattle spine score: Predicting 30-day complication risk in adult spinal deformity surgery. *J Clin Neurosci* 2017. <https://doi.org/10.1016/j.jocn.2017.06.012>.
- [40] Buchlak QD, Yanamadala V, Leveque J-C, Sethi R. Complication avoidance with pre-operative screening: insights from the Seattle spine team. *Curr Rev Musculoskelet Med* 2016;9. <https://doi.org/10.1007/s12178-016-9351-x>.
- [41] Cai T, Giannopoulos AA, Yu S, Kelil T, Ripley B, Kumamaru KK, et al. Natural language processing technologies in radiology research and clinical applications. *Radiographics* 2016;36:176–91.
- [42] Cha S. Update on brain tumor imaging: from anatomy to physiology. *Am J Neuroradiol* 2006;27:475–87.
- [43] Chaddad A, Sabri S, Niazi T, Abdulkarim B. Prediction of survival with multi-scale radiomic analysis in glioblastoma patients. *Med Biol Eng Comput* 2018;56:2287–300.
- [44] Chang C-C, Lin C-J. LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol* 2011;2:1–27.
- [45] Chang K, Bai HX, Zhou H, Su C, Bi WL, Agbodza E, et al. Residual convolutional neural network for the determination of IDH status in low-and high-grade gliomas from MR imaging. *Clin Cancer Res* 2018;24:1073–81.
- [46] Chang K, Zhang B, Guo X, Zong M, Rahman R, Sanchez D, et al. Multimodal imaging patterns predict survival in recurrent glioblastoma patients treated with bevacizumab. *Neuro Oncol* 2016;18:1680–7.
- [47] Chang P, Grinband J, Weinberg BD, Bardis M, Khy M, Cadena G, et al. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *Am J Neuroradiol* 2018;39:1201–7.
- [48] Chatfield K, Simonyan K, Vedaldi A, Zisserman A. Return of the devil in the details: Delving deep into convolutional nets; 2014 arXiv Prepr arXiv14053531.
- [49] Chen L, Zhang H, Lu J, Thung K, Aibaidula A, Liu L, et al. Multi-label nonlinear matrix completion with transductive multi-task feature selection for joint MGMT and IDH1 status prediction of patient with high-grade gliomas. *IEEE Trans Med Imaging* 2018;37:1775–87.
- [50] Chen T, Li M, Li Y, Lin M, Wang N, Wang M, Xiao T, Xu B, Zhang C, Zhang Z (2015) Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems. arXiv Prepr arXiv151201274
- [51] Cheng J. Brain tumor dataset; 2017.
- [52] Cho H, Lee S, Kim J, Park H. Classification of the glioma grading using radiomics analysis. *PeerJ* 2018;6:e5982.
- [53] Choi YS, Ahn SS, Chang JH, Kang S-G, Kim EH, Kim SH, et al. Machine learning and radiomic phenotyping of lower grade gliomas: improving survival prediction. *Eur Radiol* 2020;1–9.
- [54] Chollet F. Keras: The python deep learning library. *Astrophys Source Code Libr*. 2018.
- [55] Citak-Er F, Firat Z, Kovanlikaya I, Ture U, Ozturk-Isik E. Machine-learning in grading of gliomas based on multi-parametric magnetic resonance imaging at 3T. *Comput Biol Med* 2018;99:154–60.
- [56] Clark J, Glasziou P, Del Mar C, Bannach-Brown A, Stehlik P, Scott AM. A full systematic review was completed in 2 weeks using automation tools: a case study. *J Clin Epidemiol* 2020;121:81–90.
- [57] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging* 2013;26:1045–57.
- [58] Cocosco CA, Kollokian V, Kwan R-K-S, Pike GB, Evans AC. Brainweb: Online interface to a 3D MRI simulated brain database. *NeuroImage*. Citeseer. 1997.
- [59] Costabile JD, Thompson JA, Alaswad E, Ormond DR. Biopsy Confirmed Glioma Recurrence Predicted by Multi-Modal Neuroimaging Metrics. *J Clin Med* 2019;8:1287.
- [60] Créquit P, Trinquart L, Yavchitz A, Ravaud P. Wasted research when systematic reviews fail to provide a complete and up-to-date evidence synthesis: the example of lung cancer. *BMC Med* 2016;14:8.
- [61] Cui S, Mao L, Jiang J, Liu C, Xiong S (2018) Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded neural network. *J Healthc Eng* 2018
- [62] Demšar J, Curk T, Erjavec A, Gorup Č, Hočevar T, Milutinovič M, et al. Orange: data mining toolbox in Python. *J Mach Learn Res* 2013;14:2349–53.
- [63] Devlin J, Chang M-W, Lee K, Toutanova K Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv Prepr; 2018. arXiv181004805
- [64] Dong W, Wang P, Yin W, Shi G, Wu F, Lu X. Denoising prior driven deep neural network for image restoration. *IEEE Trans Pattern Anal Mach Intell* 2018;41:2305–18.
- [65] Eichinger P, Alberts E, Delbridge C, Trebesch S, Valentinič A, Bette S, et al. Diffusion tensor image features predict IDH genotype in newly diagnosed WHO grade II/III gliomas. *Sci Rep* 2017. <https://doi.org/10.1038/s41598-017-13679-4>.
- [66] El-Dahshan E-S-A, Hosny T, Salem A-B-M. Hybrid intelligent techniques for MRI brain images classification. *Digit Signal Process* 2010;20:433–41.
- [67] Emblem KE, Due-Tønnessen P, Hald JK, Bjørnerud A, Pinho MC, Scheie D, et al. Machine learning in preoperative glioma MRI: Survival associations by perfusion-based support vector machine outperforms traditional MRI. *J Magn Reson imaging* 2014;40:47–54.
- [68] Emblem KE, Pinho MC, Zöllner FG, Due-Tønnessen P, Hald JK, Schad LR, et al. A generic support vector machine model for preoperative glioma survival associations. *Radiology* 2015;275:228–34.
- [69] Farrokhi F, Buchlak QD, Sikora M, Esmaili N, Marsans M, McLeod P, et al. Investigating risk factors and predicting complications in deep brain stimulation surgery with machine learning algorithms. *World Neurosurg*. 2019.
- [70] Fellah S, Caudal D, De Paula AM, Dory-Lautrec P, Figarella-Branger D, Chinot O, et al. Multimodal MR imaging (diffusion, perfusion, and spectroscopy): is it possible to distinguish oligodendroglial tumor grade and 1p/19q codeletion in the pretherapeutic diagnosis? *Am J Neuroradiol* 2013;34:1326–33.
- [71] Fortin F-A, De Rainville F-M, Gardner M-A, Parizeau M, Gagné C. DEAP: Evolutionary algorithms made easy. *J Mach Learn Res* 2012;13:2171–5.
- [72] Gaw N, Hawkins-Daarud A, Hu LS, Yoon H, Wang L, Xu Y, et al. Integration of machine learning and mechanistic models accurately predicts variation in cell density of glioblastoma using multiparametric MRI. *Sci Rep* 2019;9:1–9.
- [73] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016;278:563–77.
- [74] Goetz M, Weber C, Binczyk F, Polanska J, Tarnawski R, Bobek-Billewicz B, et al. DALSA: domain adaptation for supervised learning from sparsely annotated MR images. *IEEE Trans Med Imaging* 2015;35:184–96.
- [75] Gong E, Pauly JM, Wintermark M, Zaharchuk G. Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. *J Magn Reson Imaging* 2018;48:330–40.
- [76] Grabowski MM, Recinos PF, Nowacki AS, Schroeder JL, Angelov L, Barnett GH, et al. Residual tumor volume versus extent of resection: predictors of survival after surgery for glioblastoma. *J Neurosurg* 2014;121:1115–23.
- [77] Van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017;77:e104–7.
- [78] Gulati S, Jakola AS, Nerland US, Weber C, Solheim O. The risk of getting worse: surgically acquired deficits, perioperative complications, and functional outcomes after primary resection of glioblastoma. *World Neurosurg* 2011;76:572–9.
- [79] Guo J, Yao C, Chen H, Zhuang D, Tang W, Ren G, et al. The relationship between Cho/NAA and glioma metabolism: implementation for margin delineation of cerebral gliomas. *Acta Neurochir (Wien)* 2012;154:1361–70.
- [80] Gupta N, Khanna P. A non-invasive and adaptive CAD system to detect brain tumor from T2-weighted MRIs using customized Otsu's thresholding with prominent features and supervised learning. *Signal Process Image Commun* 2017;59:18–26.
- [81] Han Y, Xie Z, Zang Y, Zhang S, Gu D, Zhou M, et al. Non-invasive genotype prediction of chromosome 1p/19q co-deletion by development and validation of an MRI-based radiomics signature in lower-grade gliomas. *J Neurooncol* 2018;140:297–306.
- [82] Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al. Brain tumor segmentation with deep neural networks. *Med Image Anal* 2017;35:18–31.
- [83] Havaei M, Guizard N, Larochelle H, Jodoin P-M. Deep learning trends for focal brain pathology segmentation in MRI. In: *Machine learning for health informatics*. Springer; 2016. p. 125–48.
- [84] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. p. 770–8.

- [85] Hemanth DJ, Anitha J, Naaji A, Geman O, Popescu DE. A modified deep convolutional neural network for abnormal brain image classification. *IEEE Access* 2018;7:4275–83.
- [86] Hoseini F, Shahbahrami A, Bayat P. AdaptAhead optimization algorithm for learning deep CNN applied to MRI segmentation. *J Digit Imaging* 2019;32:105–15.
- [87] Hsieh K-L-C, Chen C-Y, Lo C-M. Quantitative glioma grading using transformed gray-scale invariant textures of MRI. *Comput Biol Med* 2017;83:102–8.
- [88] Hu LS, Ning S, Eschbacher JM, Baxter LC, Gaw N, Ranjbar S, et al. Radiogenomics to characterize regional genetic heterogeneity in glioblastoma. *Neuro Oncol* 2017;19:128–37.
- [89] Hu LS, Ning S, Eschbacher JM, Gaw N, Dueck AC, Smith KA, et al. Multiparametric MRI and texture analysis to visualize spatial histologic heterogeneity and tumor extent in glioblastoma. *PLoS ONE* 2015;10.
- [90] Hu LS, Yoon H, Eschbacher JM, Baxter LC, Dueck AC, Nespodzany A, et al. Accurate patient-specific machine learning models of glioblastoma invasion using transfer learning. *Am J Neuroradiol* 2019;40:418–25.
- [91] Hu X, Wong KK, Young GS, Guo L, Wong ST. Support vector machine multiparametric MRI identification of pseudoprogression from tumor recurrence in patients with resected glioblastoma. *J Magn Reson Imaging* 2011;33:296–305.
- [92] Huang M, Yang W, Wu Y, Jiang J, Chen W, Feng Q. Brain tumor segmentation based on local independent projection-based classification. *IEEE Trans Biomed Eng* 2014;61:2633–45.
- [93] Huang Z, Zhao Y, Li X, Zhao X, Liu Y, Song G, et al. Application of innovative image processing methods and AdaBound-SE-DenseNet to optimize the diagnosis performance of meningiomas and gliomas. *Biomed Signal Process Control* 2020;59:101926.
- [94] Inano R, Oishi N, Kunieda T, Arakawa Y, Yamao Y, Shibata S, et al. Voxel-based clustered imaging by multiparameter diffusion tensor images for glioma grading. *NeuroImage Clin* 2014;5:396–407.
- [95] Ismael SAA, Mohammed A, Hefny H. An enhanced deep learning approach for brain cancer MRI images classification using residual networks. *Artif Intell Med* 2020;102:101779.
- [96] Ismail M, Hill V, Statsevych V, Huang R, Prasanna P, Correa R, et al. Shape features of the lesion habitat to differentiate brain tumor progression from pseudoprogression on routine multiparametric MRI: a multisite study. *Am J Neuroradiol* 2018;39:2187–93.
- [97] Jakola AS, Myrnes KS, Kloster R, Torp SH, Lindal S, Unsgård G, et al. Comparison of a strategy favoring early surgical resection vs a strategy favoring watchful waiting in low-grade gliomas. *JAMA* 2012;308:1881–8.
- [98] Jakola AS, Zhang Y-H, Skjalsvik AJ, Solheim O, Bø HK, Berntsen EM, et al. Quantitative texture analysis in the prediction of IDH status in low-grade gliomas. *Clin Neurol Neurosurg* 2018;164:114–20.
- [99] Jang B-S, Jeon SH, Kim IH, Kim IA. Prediction of pseudoprogression versus progression using machine learning algorithm in glioblastoma. *Sci Rep* 2018;8:1–9.
- [100] Javed U, Riaz MM, Ghafoor A, Cheema TA. MRI brain classification using texture features, fuzzy weighting and support vector machine. *Prog Electromagn Res* 2013;53:73–88.
- [101] Jeong J, Wang L, Ji B, Lei Y, Ali A, Liu T, et al. Machine-learning based classification of glioblastoma using delta-radiomic features derived from dynamic susceptibility contrast enhanced magnetic resonance images: Introduction. *Quant Imaging Med Surg* 2019;9:1201.
- [102] Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Caffe: Convolutional architecture for fast feature embedding. In: *Proceedings of the 22nd ACM international conference on Multimedia*. p. 675–8.
- [103] Jin KH, McCann MT, Froustey E, Unser M. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans Image Process* 2017;26:4509–22.
- [104] Johnson DR, O'Neill BP. Glioblastoma survival in the United States before and during the temozolomide era. *J Neurooncol* 2012;107:359–64.
- [105] Jones E, Oliphant T, Peterson P (2014) *{SciPy}: Open source scientific tools for {Python}*
- [106] Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 2017;36:61–78.
- [107] Kanas VG, Zacharaki EI, Thomas GA, Zinn PO, Megalooikonomou V, Colen RR. Learning MRI-based classification models for MGMT methylation status prediction in glioblastoma. *Comput Methods Programs Biomed* 2017;140:249–57.
- [108] Kayalibay B, Jensen G, van der Smagt P (2017) CNN-based segmentation of medical imaging data. *arXiv Prepr arXiv170103056*
- [109] Kesler SR, Harrison RA, Petersen ML, Rao V, Dyson H, Alfaro-Munoz K, et al. Pre-surgical connectome features predict IDH status in diffuse gliomas. *Oncotarget* 2019;10:6484.
- [110] Khawaldeh S, Pervaiz U, Rafiq A, Alkhalwaldeh RS. Noninvasive grading of glioma tumor using magnetic resonance imaging with convolutional neural networks. *Appl Sci* 2018;8:27.
- [111] Kickingreder P, Bonekamp D, Nowosielski M, Kratz A, Sill M, Burth S, et al. Radiogenomics of glioblastoma: machine learning-based classification of molecular characteristics by using multiparametric and multiregional MR imaging features. *Radiology* 2016;281:907–18.
- [112] Kickingreder P, Isensee F, Tursunova I, Petersen J, Neuberger U, Bonekamp D, et al. Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. *Lancet Oncol* 2019;20:728–40.
- [113] Kim BS, Kim ST, Kim JH, Seol HJ, Nam D-H, Shin HJ, et al. Apparent Diffusion Coefficient as a Predictive Biomarker for Survival in Patients with Treatment-Naive Glioblastoma Using Quantitative Multiparametric Magnetic Resonance Profiling. *World Neurosurg* 2019;122:e812–20.
- [114] Kim KH, Do W, Park S. Improving resolution of MR images with an adversarial network incorporating images with different contrast. *Med Phys* 2018;45:3120–31.
- [115] Kim M, Jung SY, Park JE, Jo Y, Park SY, Nam SJ, et al. Diffusion-and perfusion-weighted MRI radiomics model may predict isocitrate dehydrogenase (IDH) mutation and tumor aggressiveness in diffuse lower grade glioma. *Eur Radiol* 2020;30:2142–51.
- [116] Kim Y, Cho H, Kim ST, Park H, Nam D, Kong D-S. Radiomics features to distinguish glioblastoma from primary central nervous system lymphoma on multi-parametric MRI. *Neuroradiology* 2018;60:1297–305.
- [117] Klein M, Engelberts NHJ, van der Ploeg HM, Kasteleijn-Nolst Trenité DGA, Aaronson NK, Taphoorn MJB, et al. Epilepsy in low-grade gliomas: The impact on cognitive function and quality of life. *Ann Neurol Off J Am Neurol Assoc Child Neurol Soc* 2003;54:514–20.
- [118] Kocak B, Durmaz ES, Ates E, Sel I, Gunes ST, Kaya OK, et al. Radiogenomics of lower-grade gliomas: machine learning-based MRI texture analysis for predicting 1p/19q codeletion status. *Eur Radiol* 2020;30:877–86.
- [119] Koley S, Sadhu AK, Mitra P, Chakraborty B, Chakraborty C. Delineation and diagnosis of brain tumors from post contrast T1-weighted MR images using rough granular computing and random forest. *Appl Soft Comput* 2016;41:453–65.
- [120] Korfiatis P, Kline TL, Coufalova L, Lachance DH, Parney IF, Carter RE, et al. MRI texture features as biomarkers to predict MGMT methylation status in glioblastomas. *Med Phys* 2016;43:2835–44.
- [121] Korfiatis P, Kline TL, Lachance DH, Parney IF, Buckner JC, Erickson BJ. Residual deep convolutional neural network predicts MGMT methylation status. *J Digit Imaging* 2017;30:622–8.
- [122] Krizhevsky A, Sutskever I, Hinton GE (2012) *Imagenet classification with deep convolutional neural networks*. In: *Advances in neural information processing systems*. pp 1097–1105
- [123] Kuhn M. Building predictive models in R using the caret package. *J Stat Softw* 2008;28:1–26.
- [124] Kunitatsu A, Kunitatsu N, Yasaka K, Akai H, Kamiya K, Watadani T, et al. Machine learning-based texture analysis of contrast-enhanced MR imaging to differentiate between glioblastoma and primary central nervous system lymphoma. *Magn Reson Med Sci* 2019;18:44.
- [125] Am KV, Rajendran VR. Glioma Tumor Grade Identification Using Artificial Intelligent Techniques. *J Med Syst* 2019;43:113.
- [126] Lao J, Chen Y, Li Z-C, Li Q, Zhang J, Liu J, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci Rep* 2017;7:1–8.
- [127] Lau J. Systematic review automation thematic series; 2019.
- [128] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436.
- [129] Lee MH, Kim J, Kim S-T, Shin H-M, You H-J, Choi JW, et al. Prediction of IDH1 Mutation Status in Glioblastoma Using Machine Learning Technique Based on Quantitative Radiomic Data. *World Neurosurg* 2019;125:e688–96.
- [130] Lee MK, Park JE, Jo Y, Park SY, Kim SJ, Kim HS. Advanced imaging parameters improve the prediction of diffuse lower-grade gliomas subtype, IDH mutant with no 1p19q codeletion: added value to the T2/FLAIR mismatch sign. *Eur Radiol* 2020;30:844–54.
- [131] Li Q, Bai H, Chen Y, Sun Q, Liu L, Zhou S, et al. A fully-automatic multiparametric radiomics model: towards reproducible and prognostic imaging signature for prediction of overall survival in glioblastoma multiforme. *Sci Rep* 2017;7:1–9.
- [132] Li Y, Liang Y, Sun Z, Xu K, Fan X, Li S, et al. Radiogenomic analysis of PTEN mutation in glioblastoma using preoperative multi-parametric magnetic resonance imaging. *Neuroradiology* 2019;61:1229–37.
- [133] Li Y, Liu X, Qian Z, Sun Z, Xu K, Wang K, et al. Genotype prediction of ATRX mutation in lower-grade gliomas using an MRI radiomics signature. *Eur Radiol* 2018. <https://doi.org/10.1007/s00330-017-5267-0>.
- [134] Li Y, Liu X, Xu K, Qian Z, Wang K, Fan X, et al. MRI features can predict EGFR expression in lower grade gliomas: a voxel-based radiomic analysis. *Eur Radiol* 2018;28:356–62.
- [135] Li Z-C, Bai H, Sun Q, Li Q, Liu L, Zou Y, et al. Multiregional radiomics features from multiparametric MRI for prediction of MGMT methylation status in glioblastoma multiforme: a multicentre study. *Eur Radiol* 2018;28:3640–50.
- [136] Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. *Sci Rep* 2017;7:1–11.
- [137] Liang S, Zhang R, Liang D, Song T, Ai T, Xia C, et al. Multimodal 3D DenseNet for IDH genotype prediction in gliomas. *Genes (Basel)* 2018;9:382.
- [138] Liao X, Cai B, Tian B, Luo Y, Song W, Li Y. Machine-learning based radiogenomics analysis of MRI features and metagenes in glioblastoma multiforme patients with different survival time. *J Cell Mol Med* 2019;23:4375–85.
- [139] Lilley EJ, Lindvall C, Lillemoe KD, Tulskey JA, Wiener DC, Cooper Z. Measuring processes of care in palliative surgery: a novel approach using natural language processing. *Ann Surg* 2018;267:823–5.

- [140] Liu J, Chen F, Pan C, Zhu M, Zhang X, Zhang L, et al. A cascaded deep convolutional neural network for joint segmentation and genotype prediction of brainstem gliomas. *IEEE Trans Biomed Eng* 2018;65:1943–52.
- [141] Liu L, Zhang H, Wu J, Yu Z, Chen X, Rekik I, et al. Overall survival time prediction for high-grade glioma patients based on large-scale brain functional networks. *Brain Imaging Behav* 2019;13:1333–51.
- [142] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V., Roberta: A robustly optimized bert pretraining approach;2019. arXiv Prepr arXiv190711692.
- [143] Liu Y, Xu X, Yin L, Zhang X, Li L, Lu H. Relationship between glioblastoma heterogeneity and survival time: an MR imaging texture analysis. *Am J Neuroradiol* 2017;38:1695–701.
- [144] Liu Z, Zhang T, Jiang H, Xu W, Zhang J. Conventional MR-based preoperative nomograms for prediction of IDH/1p19q subtype in low-grade glioma. *Acad Radiol* 2019;26:1062–70.
- [145] De Looze C, Beausang A, Cryan J, Loftus T, Buckley PG, Farrell M, et al. Machine learning: a useful radiological adjunct in determination of a newly diagnosed glioma's grade and IDH status. *J Neurooncol* 2018;139:491–9.
- [146] Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, et al. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* 2007;114:97–109.
- [147] Louis DN, Perry A, Reifenberger G, Von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol* 2016;131:803–20.
- [148] Lu C-F, Hsu F-T, Hsieh K-L-C, Kao Y-C-J, Cheng S-J, Hsu J-B-K, et al. Machine learning-based radiomics for molecular subtyping of gliomas. *Clin Cancer Res* 2018;24:4429–36.
- [149] Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys* 2019;29:102–27.
- [150] Ma C, Luo G, Wang K. Concatenated and connected random forests with multiscale patch driven active contour model for automated brain tumor segmentation of MR images. *IEEE Trans Med Imaging* 2018;37:1943–54.
- [151] Manning C, Surdeanu M, Bauer J, Finkel J, Bethard S, McClosky D. The Stanford CoreNLP natural language processing toolkit. In: In: Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations. p. 55–60.
- [152] Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging* 2014;34:1993–2024.
- [153] Michelson M, Reuter K. The significant cost of systematic reviews and meta-analyses: A call for greater involvement of machine learning to assess the promise of clinical trials. *Contemp Clin Trials Commun* 2019;16:100443.
- [154] Mlynarski P, Delingette H, Criminisi A, Ayache N. Deep learning with mixed supervision for brain tumor segmentation. *J Med Imaging* 2019;6:34002.
- [155] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264–9.
- [156] Mohsen H, El-Dahshan E-S-A, El-Horbaty E-S-M, Salem A-B-M. Classification using deep learning neural networks for brain tumors. *Futur Comput Informatics J* 2018;3:68–71.
- [157] Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med* 2019;170:W1–W33.
- [158] Morton S, Berg A, Levit L, Eden J. Finding what works in health care: standards for systematic reviews. National Academies Press; 2011.
- [159] Nakamoto T, Takahashi W, Haga A, Takahashi S, Kiryu N, Nawa K, et al. Prediction of malignant glioma grades using contrast-enhanced T1-weighted and T2-weighted magnetic resonance images based on a radiomic analysis. *Sci Rep* 2019;9:1–12.
- [160] Nazir M, Wahid F, Ali Khan S. A simple and intelligent approach for brain MRI classification. *J Intell Fuzzy Syst* 2015;28:1127–35.
- [161] Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res* 2018;78:4786–9.
- [162] Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med* 2016;375:1216.
- [163] Olson RS, Urbanowicz RJ, Andrews PC, Lavender NA, Moore JH., Automating biomedical data science through tree-based pipeline optimization. In: European Conference on the Applications of Evolutionary Computation. Springer; 2016. pp 123–137.
- [164] Ostrom QT, Gittleman H, Fulop J, Liu M, Blanda R, Kromer C, et al. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2008–2012. *Neuro Oncol* 2015;17:iv1–iv62.
- [165] Pan C, Liu J, Tang J, Chen X, Chen F, Wu Y, et al. A machine learning-based prediction model of H3K27M mutations in brainstem gliomas using conventional MRI and clinical features. *Radiother Oncol* 2019;130:172–9.
- [166] Park YW, Choi YS, Ahn SS, Chang JH, Kim SH, Lee S-K. Radiomics MRI phenotyping with machine learning to predict the grade of lower-grade gliomas: a study focused on nonenhancing tumors. *Korean J Radiol* 2019;20:1381–9.
- [167] Park YW, Han K, Ahn SS, Bae S, Choi YS, Chang JH, et al. Prediction of IDH1-mutation and 1p/19q-codeletion status using preoperative MR imaging phenotypes in lower grade gliomas. *Am J Neuroradiol* 2018;39:37–42.
- [168] Park YW, Han K, Ahn SS, Choi YS, Chang JH, Kim SH, et al. Whole-tumor histogram and texture analyses of DTI for evaluation of IDH1-mutation and 1p/19q-codeletion status in World Health Organization grade II gliomas. *Am J Neuroradiol* 2018;39:693–8.
- [169] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011;12:2825–30.
- [170] Peeken JC, Goldberg T, Pyka T, Bernhofer M, Wiestler B, Kessel KA, et al. Combining multimodal imaging and treatment features improves machine learning-based prognostic assessment in patients with glioblastoma multiforme. *Cancer Med* 2019;8:128–36.
- [171] Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans Med Imaging* 2016;35:1240–51.
- [172] Perkuhn M, Stavrinou P, Thiele F, Shakirin G, Mohan M, Garmpis D, et al. Clinical evaluation of a multiparametric deep learning model for glioblastoma segmentation using heterogeneous magnetic resonance imaging data from clinical routine. *Invest Radiol* 2018;53:647–54.
- [173] Prastawa M, Bullitt E, Gerig G. Simulation of brain tumors in MR images for evaluation of segmentation efficacy. *Med Image Anal* 2009;13:297–311.
- [174] Qian X, Tan H, Zhang J, Zhao W, Chan MD, Zhou X. Stratification of pseudoprogression and true progression of glioblastoma multiforme based on longitudinal diffusion tensor imaging without segmentation. *Med Phys* 2016;43:5889–902.
- [175] Qian Z, Li Y, Wang Y, Li L, Li R, Wang K, et al. Differentiation of glioblastoma from solitary brain metastases using radiomic machine-learning classifiers. *Cancer Lett* 2019;451:128–35.
- [176] Qiu Y, Yan S, Gundreddy RR, Wang Y, Cheng S, Liu H, et al. A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology. *J Xray Sci Technol* 2017;25:751–63.
- [177] Quan TM, Nguyen-Duc T, Jeong W-K. Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. *IEEE Trans Med Imaging* 2018;37:1488–97.
- [178] Rajapakse T., Simple Transformers. simpletransformers.ai; 2020.
- [179] Ranjith G, Parvathy R, Vikas V, Chandrasekharan K, Nair S. Machine learning methods for the classification of gliomas: initial results using features extracted from MR spectroscopy. *Neuroradiol J* 2015;28:106–11.
- [180] Raschka S, Mirjalili V. Python machine learning. Packt Publishing Ltd.; 2017.
- [181] Rathore S, Akbari H, Doshi J, Shukla G, Rozycki M, Bilello M, et al. Radiomic signature of infiltration in peritumoral edema predicts subsequent recurrence in glioblastoma: implications for personalized radiotherapy planning. *J Med Imaging* 2018;5:21219.
- [182] Rathore S, Akbari H, Rozycki M, Abdullah KG, Nasrallah MP, Binder ZA, et al. Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. *Sci Rep* 2018;8:1–12.
- [183] Rehurek R., Scalability of Semantic Analysis in Natural Language Processing; 2011.
- [184] Rehurek R, Sojka P., Software framework for topic modelling with large corpora. In: In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. Citeseer; 2010.
- [185] Ren Y, Zhang X, Rui W, Pang H, Qiu T, Wang J, et al. Noninvasive prediction of IDH1 mutation and ATRX expression loss in low-grade gliomas using multiparametric MR radiomic features. *J Magn Reson Imaging* 2019;49:808–17.
- [186] Richards D. Handsearching still a valuable element of the systematic review. *Evid Based Dent* 2008;9:85.
- [187] Ronneberger O, Fischer P, Brox T., U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention; 2015, Springer, pp 234–241.
- [188] Sajjad M, Khan S, Muhammad K, Wu W, Ullah A, Baik SW. Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *J Comput Sci* 2019;30:174–82.
- [189] Salçin K. Detection and classification of brain tumours from MRI images using faster R-CNN. *Teh Glas* 2019;13:337–42.
- [190] Sampson M, McGowan J, Tetzlaff J, Cogo E, Moher D. No consensus exists on search reporting methods for systematic reviews. *J Clin Epidemiol* 2008;61:748–54.
- [191] Sanai N, Berger MS. Glioma extent of resection and its impact on patient outcome. *Neurosurgery* 2008;62:753–66.
- [192] Sanai N, Polley M-Y, McDermott MW, Parsa AT, Berger MS. An extent of resection threshold for newly diagnosed glioblastomas. *J Neurosurg* 2011;115:3–8.
- [193] Sanghani P, Ang BT, King NKK, Ren H. Overall survival prediction in glioblastoma multiforme patients from volumetric, shape and texture features using machine learning. *Surg Oncol* 2018;27:709–14.
- [194] Saouli R, Akil M, Kachouri R. Fully automatic brain tumor segmentation using end-to-end incremental deep neural networks in MRI images. *Comput Methods Programs Biomed* 2018;166:39–49.
- [195] Schomas DA, Laack NNI, Rao RD, Meyer FB, Shaw EG, O'Neill BP, et al. Intracranial low-grade gliomas in adults: 30-year experience with long-term follow-up at Mayo Clinic. *Neuro Oncol* 2009;11:437–45.
- [196] Schwartz JT, Gao M, Geng EA, Mody KS, Mikhail CM, Cho SK. Applications of machine learning using electronic medical records in spine surgery. *Neurospine* 2019;16:643.

- [197] Scott JN, Brasher PMA, Sevcik RJ, Rewcastle NB, Forsyth PA. How often are nonenhancing supratentorial gliomas malignant? A population study. *Neurology* 2002;59:947–9.
- [198] Seetha J, Raja SS. Brain tumor classification using convolutional neural networks. *Biomed Pharmacol J* 2018;11:1457.
- [199] Sengupta A, Agarwal S, Gupta PK, Ahlawat S, Patir R, Gupta RK, et al. On differentiation between vasogenic edema and non-enhancing tumor in high-grade glioma patients using a support vector machine classifier based upon pre and post-surgery MRI images. *Eur J Radiol* 2018;106:199–208.
- [200] Sengupta A, Ramaniharan AK, Gupta RK, Agarwal S, Singh A. Glioma grading using a machine-learning framework based on optimized features obtained from T1 perfusion MRI and volumes of tumor components. *J Magn Reson Imaging* 2019;50:1295–306.
- [201] Shofty B, Artzi M, Ben BD, Liberman G, Haim O, Kashanian A, et al. MRI radiomics analysis of molecular alterations in low-grade gliomas. *Int J Comput Assist Radiol Surg* 2018;13:563–71.
- [202] Shusharina N, Söderberg J, Edmunds D, Löfman F, Shih H, Bortfeld T. Automated delineation of the clinical target volume using anatomically constrained 3D expansion of the gross tumor volume. *Radiother Oncol* 2020;146:37–43.
- [203] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition; 2014. arXiv Prepr arXiv14091556.
- [204] Smith SM, Brady JM. SUSAN—a new approach to low level image processing. *Int J Comput Vis* 1997;23:45–78.
- [205] Stupp R, Mason WP, Van Den Bent MJ, Weller M, Fisher B, Taphoorn MJB, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* 2005;352:987–96.
- [206] Su X, Chen N, Sun H, Liu Y, Yang X, Wang W, et al. Automated machine learning based on radiomics features predicts H3 K27M mutation in midline gliomas of the brain. *Neuro Oncol* 2020;22:393–401.
- [207] Subashini MM, Sahoo SK, Sunil V, Easwaran S. A non-invasive methodology for the grade identification of astrocytoma using image processing and artificial intelligence techniques. *Expert Syst Appl* 2016;43:186–96.
- [208] Suchorska B, Schüller U, Biczok A, Lenski M, Albert NL, Giese A, et al. Contrast enhancement is a prognostic factor in IDH1/2 mutant, but not in wild-type WHO grade II/III glioma as confirmed by machine learning. *Eur J Cancer* 2019;107:15–27.
- [209] Suh HB, Choi YS, Bae S, Ahn SS, Chang JH, Kang S-G, et al. Primary central nervous system lymphoma and atypical glioblastoma: differentiation using radiomics approach. *Eur Radiol* 2018;28:3832–9.
- [210] Sun P, Wang D, Mok VC, Shi L. Comparison of Feature Selection Methods and Machine Learning Classifiers for Radiomics Analysis in Glioma Grading. *IEEE Access* 2019;7:102010–20.
- [211] Svolos P, Tsolaki E, Kapsalaki E, Theodorou K, Fountas K, Fezoulidis I, et al. Investigating brain tumor differentiation with diffusion and perfusion metrics at 3T MRI using pattern recognition techniques. *Magn Reson Imaging* 2013;31:1567–77.
- [212] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. p. 1–9.
- [213] Takahashi S, Takahashi W, Tanaka S, Haga A, Nakamoto T, Suzuki Y, et al. Radiomics Analysis for Glioma Malignancy Evaluation Using Diffusion Kurtosis and Tensor Imaging. *Int J Radiat Oncol Biol Phys* 2019;105:784–91.
- [214] Tan Y, Zhang S, Wei J, Dong D, Wang X, Yang G, et al. A radiomics nomogram may improve the prediction of IDH genotype for astrocytoma before surgery. *Eur Radiol* 2019;29:3325–37.
- [215] Tang F, Liang S, Zhong T, Huang X, Deng X, Zhang Y, et al. Postoperative glioma segmentation in CT image using deep feature fusion model guided by multi-sequence MRIs. *Eur Radiol* 2020;30:823–32.
- [216] Team RC. R: A language and environment for statistical computing; 2013.
- [217] Thaha MM, Kumar KPM, Murugan BS, Dhanasekaran S, Vijayakarthick P, Selvi AS. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. *J Med Syst* 2019;43:294.
- [218] Tian Q, Yan L, Zhang X, Zhang X, Hu Y, Han Y, et al. Radiomics strategy for glioma grading using texture features from multiparametric MRI. *J Magn Reson Imaging* 2018;48:1518–28.
- [219] Tykocki T, Eltayeb M. Ten-year survival in glioblastoma. A systematic review. *J Clin Neurosci* 2018;54:7–13.
- [220] Upadhyaya T, Morvan Y, Stindel E, Le Reste P-J, Hatt M. A framework for multimodal imaging-based prognostic model building: preliminary study on multimodal MRI in glioblastoma multiforme. *IRBM* 2015;36:345–50.
- [221] Vamvakas A, Williams SC, Theodorou K, Kapsalaki E, Fountas K, Kappas C, et al. Imaging biomarker analysis of advanced multiparametric MRI for glioma grading. *Phys Medica* 2019;60:188–98.
- [222] Veneti S, Huse JT. The evolving molecular genetics of low-grade glioma. *Adv Anat Pathol* 2015;22:94.
- [223] Watanabe M, Tanaka R, Takeda N. Magnetic resonance imaging and histopathology of cerebral gliomas. *Neuroradiology* 1992;34:463–9.
- [224] Weller M, Van Den Bent M, Tonn JC, Stupp R, Preusser M, Cohen-Jonathan-Moyal E, et al. European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. *Lancet Oncol* 2017;18:e315–29.
- [225] Wiestler B, Kluge A, Lukas M, Gempt J, Ringel F, Schlegel J, et al. Multiparametric MRI-based differentiation of WHO grade II/III glioma and WHO grade IV glioblastoma. *Sci Rep* 2016;6:1–6.
- [226] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019;170:51–8.
- [227] Wong KCL, Syeda-Mahmood T, Moradi M. Building medical image classifiers with very limited data using segmentation networks. *Med Image Anal* 2018;49:105–16.
- [228] Wu S, Meng J, Yu Q, Li P, Fu S. Radiomics-based machine learning methods for isocitrate dehydrogenase genotype prediction of diffuse gliomas. *J Cancer Res Clin Oncol* 2019;145:543–50.
- [229] Wu Y-P, Lin Y-S, Wu W-G, Yang C, Gu J-Q, Bai Y, Wang M-Y (2017) Semiautomatic segmentation of glioma on mobile devices. *J Healthc Eng*; 2017.
- [230] Wu Y, Liu B, Wu W, Lin Y, Yang C, Wang M. Grading glioma by radiomics with feature selection based on mutual information. *J Ambient Intell Humaniz Comput* 2018;9:1671–82.
- [231] Wyles CC, Tibbo ME, Fu S, Wang Y, Sohn S, Kremers WK, et al. Use of natural language processing algorithms to identify common data elements in operative notes for total hip arthroplasty. *JBJS* 2019;101:1931–8.
- [232] Xiao D-D, Yan P-F, Wang Y-X, Osman MS, Zhao H-Y. Glioblastoma and primary central nervous system lymphoma: preoperative differentiation by using MRI-based 3D texture analysis. *Clin Neurol Neurosurg* 2018;173:84–90.
- [233] Yiming Li, Zenghui Qian, Kaibin Xu, Kai Wang, Xing Fan, Shaowu Li, Tao Jiang, Xing Liu, Yinyan Wang (2018) MRI features predict p53 status in lower-grade gliomas via a machine-learning approach. *NeuroImage Clin*. doi: 10.1016/j.nicl.2017.10.030.
- [234] Yamahara T, Numa Y, Oishi T, Kawaguchi T, Seno T, Asai A, et al. Morphological and flow cytometric analysis of cell infiltration in glioblastoma: a comparison of autopsy brain and neuroimaging. *Brain Tumor Pathol* 2010;27:81–7.
- [235] Yang D, Rao G, Martinez J, Veeraraghavan A, Rao A. Evaluation of tumor-derived MRI-texture features for discrimination of molecular subtypes and prediction of 12-month survival status in glioblastoma. *Med Phys* 2015;42:6725–35.
- [236] Yang H, Luo L, Chueng LP, Ling D, Chin F. Deep learning and its applications to natural language processing. In: *Deep learning: Fundamentals, theory and applications*. Springer; 2019. p. 89–109.
- [237] Yang I, Aghi MK. New advances that enable identification of glioblastoma recurrence. *Nat Rev Clin Oncol* 2009;6:648.
- [238] Yang T, Song J, Li L. A deep learning model integrating SK-TPCNN and random forests for brain tumor segmentation in MRI. *Biocybern Biomed Eng* 2019;39:613–23.
- [239] Yang Y, Yan L-F, Zhang X, Han Y, Nan H-Y, Hu Y-C, et al. Glioma grading on conventional MR images: a deep learning study with transfer learning. *Front Neurosci* 2018;12:804.
- [240] Yu J, Shi Z, Lian Y, Li Z, Liu T, Gao Y, et al. Noninvasive IDH1 mutation estimation based on a quantitative radiomics approach for grade II glioma. *Eur Radiol* 2017;27:3509–22.
- [241] Yun J, Park JE, Lee H, Ham S, Kim N, Kim HS. Radiomic features and multilayer perceptron network classifier: a robust MRI classification strategy for distinguishing glioblastoma from primary central nervous system lymphoma. *Sci Rep* 2019;9:1–10.
- [242] Zacharaki EI, Kanas VG, Davatzikos C. Investigating machine learning techniques for MRI-based classification of brain neoplasms. *Int J Comput Assist Radiol Surg* 2011;6:821–8.
- [243] Zacharaki EI, Morita N, Bhatt P, O'Rourke DM, Melhem ER, Davatzikos C. Survival analysis of patients with high-grade gliomas based on data mining of imaging variables. *Am J Neuroradiol* 2012;33:1065–71.
- [244] Zhang B, Chang K, Ramkissoon S, Tanguturi S, Bi WL, Reardon DA, et al. Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas. *Neuro Oncol* 2017;19:109–17.
- [245] Zhang X, Tian Q, Wang L, Liu Y, Li B, Liang Z, et al. Radiomics strategy for molecular subtype stratification of lower-grade glioma: detecting IDH and TP53 mutations based on multimodal MRI. *J Magn Reson Imaging* 2018;48:916–26.
- [246] Zhang X, Yan L-F, Hu Y-C, Li G, Yang Y, Han Y, et al. Optimizing a machine learning based glioma grading system using multi-parametric MRI histogram and texture features. *Oncotarget* 2017;8:47816.
- [247] Zhao L, Jia K. Multiscale CNNs for brain tumor segmentation and diagnosis. *Comput Math Methods Med*; 2016.
- [248] Zhao X, Wu Y, Song G, Li Z, Zhang Y, Fan Y. A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Med Image Anal* 2018;43:98–111.
- [249] Zhou H, Chang K, Bai HX, Xiao B, Su C, Bi WL, et al. Machine learning reveals multimodal MRI patterns predictive of isocitrate dehydrogenase and 1p/19q status in diffuse low-and high-grade gliomas. *J Neurooncol* 2019;142:299–307.
- [250] Zhou M, Chaudhury B, Hall LO, Goldof DB, Gillies RJ, Gatenby RA. Identifying spatial imaging biomarkers of glioblastoma multiforme for survival group prediction. *J Magn Reson Imaging* 2017;46:115–23.

- [251] Zhuge Y, Krauze AV, Ning H, Cheng JY, Arora BC, Camphausen K, et al. Brain tumor segmentation using holistically nested neural networks in MRI images. *Med Phys* 2017;44:5234–43.
- [252] Zinn PO, Singh SK, Kotrotsou A, Hassan I, Thomas G, Luedi MM, et al. A coclinical radiogenomic validation study: conserved magnetic resonance radiomic appearance of periostin-expressing glioblastoma in patients and xenograft models. *Clin cancer Res* 2018;24:6288–99.
- [253] Zöllner FG, Emblem KE, Schad LR. Support vector machines in DSC-based glioma imaging: Suggestions for optimal characterization. *Magn Reson Med* 2010;64:1230–6.