

Research Progress of Gliomas in Machine Learning

Yameng Wu [†], Yu Guo [†], Jun Ma , Yu Sa , Qifeng Li and Ning Zhang ^{*}

Tianjin Key Lab of BME Measurement, Department of Biomedical Engineering, Tianjin University, Tianjin 300000, China; wuyameng@tju.edu.cn (Y.W.); guoyu@tju.edu.cn (Y.G.); majun1@tju.edu.cn (J.M.); sayu@tju.edu.cn (Y.S.); qfli@tju.edu.cn (Q.L.)

^{*} Correspondence: zhni@tju.edu.cn; Tel.: +86-22-27892060

[†] These authors have contributed equally to this work.

Abstract: In the field of gliomas research, the broad availability of genetic and image information originated by computer technologies and the booming of biomedical publications has led to the advent of the big-data era. Machine learning methods were applied as possible approaches to speed up the data mining processes. In this article, we reviewed the present situation and future orientations of machine learning application in gliomas within the context of workflows to integrate analysis for precision cancer care. Publicly available tools or algorithms for key machine learning technologies in the literature mining for glioma clinical research were reviewed and compared. Further, the existing solutions of machine learning methods and their limitations in glioma prediction and diagnostics, such as overfitting and class imbalanced, were critically analyzed.

Keywords: gliomas; machine learning; prediction; radiomics; gene expression



Citation: Wu, Y.; Guo, Y.; Ma, J.; Sa, Y.; Li, Q.; Zhang, N. Research Progress of Gliomas in Machine Learning. *Cells* **2021**, *10*, 3169. <https://doi.org/10.3390/cells10113169>

Academic Editor: Anyuan Guo

Received: 4 September 2021

Accepted: 5 November 2021

Published: 15 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. The Background of Glioma and Machine Learning

Glioma is a common primary central nervous system (CNS) tumor which originates from glial cells [1]. According to the classification scheme of the 2016 World Health Organization (WHO) Classification of Tumors of the Central Nervous System, adult diffuse gliomas consist of astrocytoma, oligodendroglioma, and glioblastoma. Gliomas can be divided into I to IV grades, in which stage I and II are low-grade gliomas, and stage III and IV are high-grade gliomas [2]. Grade IV gliomas are also known as glioblastoma multiforme (GBM), which is the most lethal brain cancer, and patients usually survive less than one year [3]. In the new 2016 WHO classification, each diagnostic category of diffuse glioma is defined by combining genotype and phenotype, which indicates the importance of molecular biomarkers, such as isocitrate dehydrogenase (IDH), mutational status in glioma has been recognized. With the improvement of computer computing speed and the development of computing methods, the Next Generation Sequencing (NGS) and imaging technology are developing vigorously. Genomics [4] and imaging information play an important role in helping scientists and physicians to understand the pathophysiological mechanisms in diagnoses and prognoses, and in choosing treatment plans. It is worth noting that the rapid development of technology has produced an enormous wealth of data, which can be used to infer the qualitative and quantitative relationship between DNA, RNA, proteins, and other cell molecules by mathematical and statistical tools. One of the main challenges of bioinformatics using NGS and image information is how to effectively transform medical big data into available knowledge. A key difficulty is that it takes a long time to process the large amount of data, which makes the traditional statistics-based algorithm less effective. To solve this problem, both industry and scientific research turn to machine learning methods for help.

Term machine learning refers to a series of processes that involves creating and evaluating algorithms that contribute to pattern recognition, classification, and prediction

based on models generated from data. Because of its potential applications, machine learning methods are widely used in various fields, from pattern recognition [5], computer vision [6], spacecraft engineering [7], finance [8], entertainment [9], computational biology [10] to biomedical [11] and medical applications [12]. It is helping researchers process high-throughput data and image data more effectively and accurately.

Machine learning is a branch of artificial intelligence, which can imitate the computer algorithms of human intelligence. It is the computational algorithms that can automatically study experiences from data [13]. Machine learning algorithms build the model based on sample data, known as “training data”, to make predictions or decisions without being explicitly programmed to do so. It integrates and absorbs the ideas of different disciplines, such as artificial intelligence, probability and statistics, computer science, information theory, psychology, cybernetics, and philosophy [14,15]. The relationship between biology and machine learning has a long and complex history [16,17]. The early machine learning technology is called perceptron, which simulates the neurons of the human brain, and an artificial neural network (ANN) is produced [18]. Further development of artificial neural network structures, such as adaptive resonance theory (ART) [19] and neocognitron [20], are inspired by the human visual nervous system.

1.2. The Connections of Machine Learning and Glioma Research

Automated pattern recognition through machine learning is essential due to the enormity and complexity of biomedical data; manual analysis is both inefficient and untenable. Machine learning will help realize a future of knowing gliomas by unlocking the potential of large biomedical and patient datasets. It is known that machine learning could tackle medical tasks. Figure 1 shows the timeline of the combination of glioma and machine learning. Early uses of machine learning have shown promise to predict glioma biomarkers from MRI texture features [21], discover new biomarkers [22], classify tumor type [23], and fusion heterogeneous data from metabolic and molecular datasets [24]. Machine learning can assist biomedical scientists and medical professionals by identifying and summarizing meaningful patterns from gliomas’ large datasets [25].

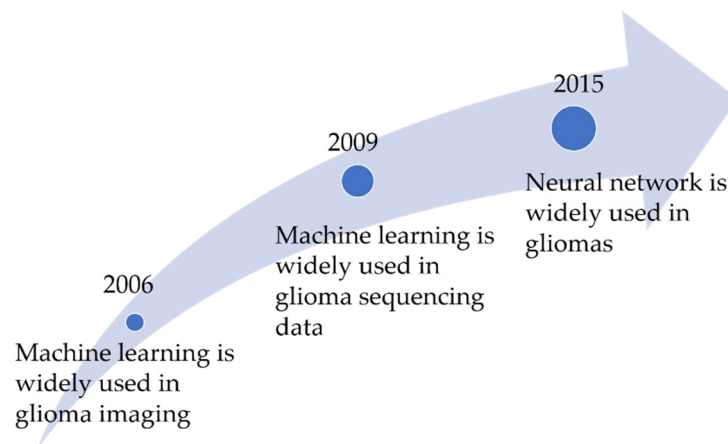


Figure 1. A timeline of the development of the combination of gliomas and machine learning.

Recognizing the increasing importance of understanding the pathogenesis of glioma, and the increasing reliance on machine learning for prediction, we believe that it is of great significance to review the published research on the prediction and prognosis of glioma of machine learning applications. The purpose of this review is to summarize over the past 20 years machine learning applications used in glioma. The purpose of this review is to statistic the literature on gliomas in machine learning over the past 20 years and summarize some typical research in recent years.

1.3. The Structure of This Article

This paper is divided into four sections. Section 1 presents the introduction of background in gliomas and machine learning methods and Section 2 gives an overview on the characteristics of machine learning and a literature review on the application of machine learning in gliomas. Section 3 analyzed several publications on machine learning approaches applied in gliomas over the past few years and further grouped them into three categories: the expression prediction of biomarkers, grade prediction, and prognosis of gliomas. The challenges with a few recommendations for future directions were presented in Section 4 and the last section is an overall discussion.

2. The Machine Learning Algorithms

2.1. Machine Learning Techniques

The machine learning algorithm is a branch of artificial intelligence research that builds and examines algorithms to boost pattern recognition, classification, and prediction. It uses various statistical, probability, and optimization tools to “learn” patterns from the large, noisy, or complex data sets, and then uses the prior training to classify new data, to identify novel patterns, or predict new trends [26]. Machine learning methods are similar to the methods that human beings usually use for learning; however it can draw a lot of energy from statistics and probability, fundamentally, it has more powerful functions because machine learning can carry out reasoning or decision-making, which cannot be achieved by using traditional statistical algorithms [27].

Supervised learning, unsupervised learning, and intensive learning are the three main aspects of machine learning. The typical examples of supervised learning are support vector machine [28], random forests [29], decision tree [30], etc., and the commonly used unsupervised learning algorithm is clustering [31]. It mainly uses supervised learning [32] and unsupervised learning [33] to deal with tasks in biology and medicine. Supervised learning predicts labels or classes on future data based on past data that includes labels/classes. Unsupervised learning identifies structure, usually clusters, amongst unlabeled data.

In the research of gliomas and other tumors, when using machine learning technology to analyze omics data, the preprocessing methods are used to reduce the dimension of features. After this procedure, the input features needed by the machine learning algorithm are identified. Then the machine learning algorithm is used to build and test the prediction model. After that, the model based on input features can predict the output of tumor samples. Figure 2 shows this process briefly.

Collecting and integrating data from various channels is the first step in the machine learning process. In the scene of supervised learning, we also need to label the data. Data preprocessing mainly includes data normalization and data whitening. Machine learning can achieve the following goals: classification, regression, clustering, anomaly detection, and so on. The test set prepared in data preprocessing is used to test the model. After the evaluation of the model, parameter tuning will optimize the training model.

2.2. Machine Learning Methods and Gliomas

The occurrence and development of gliomas are usually associated with genetic abnormalities which can be revealed by gene expression data and imaging information and another supplementary way at the molecular level. The diagnosis and treatment of gliomas entail genomics and clinical medical imaging data. Novel computational methods such as machine learning have promoted the automatic extraction of important tumor markers for clinical treatment planning and post-treatment monitoring. For example, studies have shown that in some experiments of cancer survival prediction, the machine learning algorithm model performs better than the traditional regression model [34]. Machine learning methods were being extensively used in medical practice ranging from detection and prediction of tumors with CT [35] and MR images [23] to the identification of malignant tumors through proteomics [36] and genomics [37]. From the aforementioned cases, ma-

chine learning methods are particularly suitable for biological and medical tasks, especially those that rely on complex proteomic or genomic data.

The machine learning techniques including decision tree and Naive Bayes have been widely used in the diagnosis and treatment of gliomas for decades [38,39]. In recent years, the amount of glioma research involving machine learning methods has increased dramatically. Owing to the open-access policies of many journals and the steady increase of scientific publications, the published papers are widely available.

PubMed Central (Available online: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>, accessed on 8 November 2021) is the U.S. National Library of Medicine's web-based journal literature search system, which currently contains over 28 million citations from Medline, life science journals, and online books (PubMed). The number of publications each year, as retrieved in PubMed, has surpassed one million since 2011. Figure 3 shows the quantity and trend that we found in PubMed Central by using the keywords shown in Table S1 by year (from 2002 up to 2021). The correlation of the unique papers was evaluated by reading the titles and abstracts and recognizing papers that used identifiable machine learning methods, as well as molecular, clinical, histological, physiological, or epidemiological data in carrying out a glioma prognosis or prediction. The chosen papers covered decision trees, Naive Bayes, random forests, minimum Redundancy Maximum Relevance Feature Selection (mRMR), k-nearest neighbors (k-NN), Convolutional Neural Network (CNN), Support Vector Machine (SVM), and other algorithms. No restrictions were implemented in the resulted hits except the exclusion of papers published before 2000. Based on the recent PubMed results concerning the subject of machine learning and gliomas more than 171 articles have been published thus far. The overwhelming majority of these publications employed one or more machine learning algorithms and merging data from the heterogeneous data source for the identification of gliomas as well as for the treatment and prognosis of gliomas.

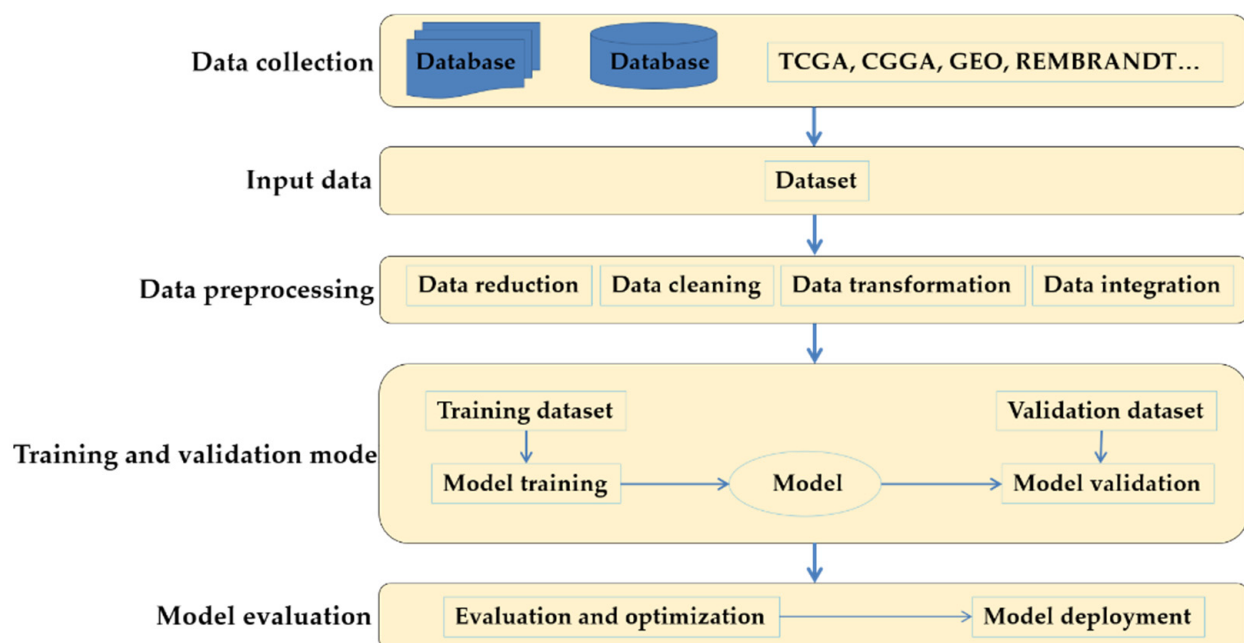


Figure 2. A typical machine learning process.

As mentioned above, the number of publications presented in Figure 3 refers to the accurate numbers retrieved from the databases without any modification. It can be observed from the picture that the enthusiasm of researchers in this particular field has increased exponentially since entering the 21st century, and in the past three years, research heat has increased the fastest.

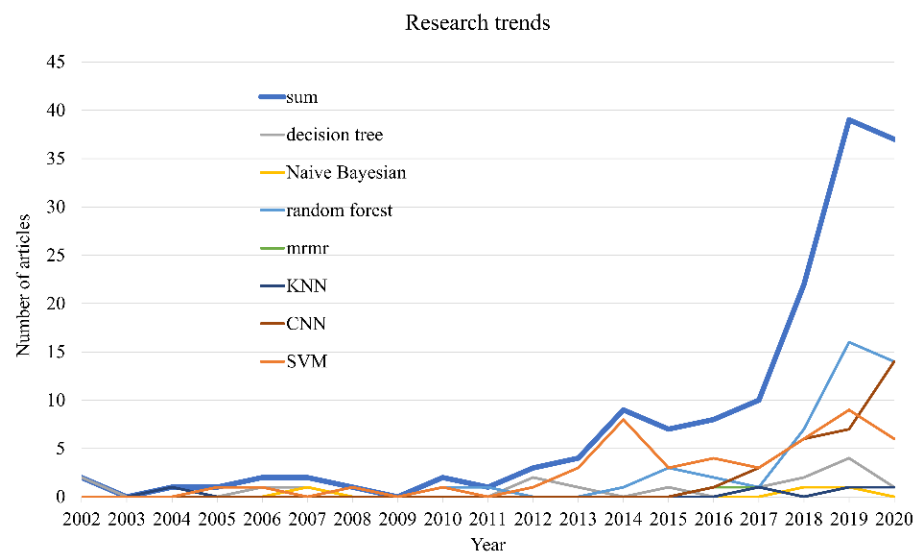


Figure 3. A line chart showing the exponential growth in published papers using machine learning algorithms to solve glioma tasks. The data were collected using a variety of key word searches through PubMed. In this figure, two axes have been plotted. The *y*-axis represents the number for publications related to “glioma” and “machine learning methods”. The *x*-axis represents the publication year. Each line represents the cumulative total of papers published over a year period. The earliest papers appeared in the early 2000s.

3. A Survey of Machine Learning Applications in Gliomas

MR images and histopathology are the most important indicators in the determination of whether it is gliomas or other abnormalities. Machine learning methods have good performance at pattern recognition of images. Gene mutations are essential in the determination of tumor molecular grade. Uncovering more genes related with gliomas may help us to know the pathway of gliomas. Machine learning methods are good at processing high-dimension gene expression profiles. Therefore, the problems involving gliomas radiograph, gene expression, and histopathology often tend to be tackled by machine learning methods.

Most of the glioma biomarkers were tested by gene detection, gene methylation detection, and an immunohistochemistry (IHC) test. The diagnosis of glioma grades often requires not only pathology but also molecular biomarkers. Traditional machine learning methods have greater advantages in computing speed, computing scale, and cost-savings in mining omics data. Hence, less complicated methods such as decision tree, Naïve Bayes, SVM and KNN are better choices.

Treatment decisions are based on histopathological diagnosis and grading. To understand more accurately about the tumor, IDH mutation, 1p19q deletion, and MGMT status tests are recommended in general. In some cases, TERT, EGFR, BRAF V600E and H3K27 tests are also needed. Some other biomarkers are regarded as the common glioma protein targets, such as Ki67, S100 and GFAP, which may not be solid. Their roles in gliomas are still under investigation and controversies have been observed in experiments [40].

Based on our analysis of the most recent studies on gliomas machine learning models, we categorized the current publications according to their functionality into three main aspects: (i) predicting the expression of biomarkers; (ii) cancer grade prediction; and (iii) the survivability risk prediction. The first case means the application of machine learning methods in predicting the status of IDH, 1p19q, MGMT, and other biomarkers in gliomas. The second case reviewed the computing methods for predicting glioma grades, and the third case tries to predict outcomes after the diagnosis of gliomas, such as life expectancy, survivability, progression, and tumor drug sensitivity.

In this perspective, we outline a vision for how machine learning can be applied to make critical advances in gliomas. We focus on those three areas in the next three subsections. Though there are some areas which have a relatively small number of studies, we do not discuss them in this review (Table 1). We reviewed several chosen papers published in recent years in those three aspects, which are typical in solving glioma clinical issues, data processing, and model building, by presenting the adopted techniques and the proposed model configuration. Table 2 depicts some of the publications presented in this review.

Table 1. Some related articles of other research directions were not discussed in detail in our review.

Category	Reference	Machine Learning Methods	Data Type
feature selection	Zöllner et al. [41]	SVM	dynamic susceptibility contrast MRI
	Sun et al. [42]	L1-SVM + multi-layer perceptron	radiomics
	Abusamra [43]	SVM, KNN, RF and eight other feature selection methods	gene expression data
automatic segmentation	Wu et al. [44]	SVM	T2 weighted MRI
	Chen et al. [45]	multiscale 3D convolutional neural network	MRI
recurrence	X. Gao et al. [46]	SVM	Pre- and post-contrast T1WI and T2 FLAIR
	Rathore [47]	PCA	MRI

Table 2. Publications discussed in this review.

Category	Reference	Machine Learning Algorithm	Training Data	Year	Aims
Biomarkers prediction	Hsu, J.BK. et al. [48]	random forests	gene expression profile	2019	Identify gene biomarkers
	J. Haubold et al. [49]	linear SVM, random forest	multiparametric 18F-FET PET-MRI and MR Fingerprinting magnetic resonance imaging (MRI), positron emission tomography (PET), and computed tomography(CT)	2020	Identify ATRX, IDH1, and 1p19q status
	Y. Matsui et al. [50]	Neural network		2020	Identify IDH1, and 1p19q status with multimodal data
Grades classification	C. Lu et al. [51]	SVM and ensemble learning approaches	multimodal MR radiomics	2018	grades classification
	A. Sengupta et al. [52]	random forests and SVM	Conventional MRI images and 3D T1 perfusion MRI data	2019	feature selection before classification
	B. Niu et al. [53]	random forests and Complement Naive Bayes	gene expression data	2020	imbalanced data problem
Prognosis prediction	P. Mobadersany et al. [54]	convolutional neural network	pathology images and genomics	2018	predict glioma outcomes
	X. Gong et al. [55]	LASSO	Transcriptomic data	2021	develop a signature associated with the tumor immune investigated the relationship between tumor shape and prognosis
	N. Czarnek et al. [4]	Khachiyan and Cox proportional hazards	Axial preoperative fluid-attenuated inversion recovery (FLAIR) and post contrast T1 (T1 + C) images	2017	

3.1. The Expression Prediction of Biomarkers in Gliomas

The findings of pathological results are the premise of rational treatment. Presently, the malignant and benign tumors and molecular subtypes are determined by pathological examination during surgery or biopsy [56]. Then, the following immunohistochemistry (IHC) test determines the molecular biomarkers of tumor tissues at the microscopic level. These pathologic biomarkers, typical proteins, genes, and other biomarkers, are useful indicators for diagnosis, prognosis, or treatment response [57]. However, obtaining such information for gliomas requires invasive approaches. Surgical decision-making could be difficult and time-consuming for many patients. Those patients who are not eligible for surgery or seek non-surgical treatment may have limited treatment options without patho-

logical guidance. Therefore, presurgical glioma status and the expression of biomarkers are valued and preferred with non-invasive approaches. In recent years, many studies have reported that it is possible to predict the expression of biomarkers in gliomas from pre-operative medical images by using machine learning [58].

This section aimed to discuss the research using machine learning methods on the prediction of the expression of biomarkers in gliomas. Based on our survey, we here present the most relevant and recent publications that proposed the use of machine learning techniques for biomarker expression predicting. A work using random forests regression algorithms to predict the glioma pathologic biomarkers and tumor grades on gene expression profile is commented on by [48]. It is a typical study among the other published works in the glioma detection field using gene expression profile. In their work, random forests frequently used machine learning models were employed to identify potential biomarkers related to glioma survival. Only utilizing histological information in studying various types of gliomas is restricted. Previous studies have shown that gene expression profiling provides an objective method to classify tumors. Machine learning algorithms are good at mining patterns from large-scale gene expression profiles. The authors claimed that their machine learning-based approach can identify 104 genes which can be used as core genes related to patient survival. Ten genes can potentially serve as indicators to classify patients with gliomas into different risk groups and could be used to estimate the prognosis of patients with gliomas.

In clinic, the detection of glioma biomarkers is always through gene detection or immunohistochemistry test, while some research tends to predict the status of these biomarkers by medical images. MRI has proven to be indispensable for brain tumor imaging and most research focused on conventional MR sequences for evaluation. Except Ki67, GFAP and S100, other glioma biomarkers, such as ATRX, IDH1 and 1p/19q statues can also be identified by medical imaging. In Haubold et al. research [49], the authors enhanced the imaging platform for quantitative and radiomic analysis by introducing MR fingerprinting as an additional MR sequence. In the study, linear SVM and random forests were used for prediction of the mutational status of ATRX, IDH1, and 1p19q of patients with cerebral gliomas, which were based on data from multiparametric ¹⁸F-FET PET-MRI. A training set was used for 3-folds cross-validation training with 20 repeats, each of which was trained on a different subset of the samples to provide for a robust and diverse committee of linear SVM and random forest, while pooling across the repeats of the predictions on the testing folds. Compared with the research which use conventional MRI [59], this study used more advanced MR fingerprinting as an additional MR sequence and it resulted in a better performance. The publication showed that MRI can perform well in the prediction of expression of biomarkers in gliomas.

Another interesting article also published in 2020 [50] proposed a neural network model that used multimodal data including MRI, positron emission tomography (PET), and computed tomography (CT) for the prediction of mutations in the IDH gene and the codeletion of chromosome arms 1p and 19q (1p19q). We should highlight an important aspect of this work regarding using multi-modality data to directly predict the groups of molecular expressions. Residual networks were employed in this study to extract features followed by training a committee of 217 models using leave-one-out cross-validation. It was found that the deep neural network could accurately predict IDH mutation and 1p19q codeletion when the MRI, PET, and CT data were combined. The authors claimed that the model was trained only according to low-grade gliomas data, which could be useful when other researchers applied the method on glioblastomas. This procedure of establishing different distinct molecular subtypes could be useful in distinguishing low-grade gliomas (LGG) and glioblastomas. However, the accuracy is a little lower than just using MRI as the results of using imaging fusion technology do not show advantages in the accuracy of biomarker identifying.

3.2. Glioma Grades Classification

Treatment options and responses differ from diagnosis [60]. The significance of the grading system is to mark the likely growth rate of the tumor and the possibility of tumor spreading in the brain, which can be used to predict curative effects and carry out the treatment plan. The grade and classification of brain tumors are formulated by the WHO. Glioma staging covers different types of tumors, many of which have significant differences in biological characteristics, prognosis, and treatment.

According to the latest WHO guidelines for the diagnosis of gliomas, the diagnosis of gliomas should include not only pathological indicators, but also the status of molecular biomarkers. Some researchers are now exploring whether machine learning can simplify the process of glioma grading. A work that studied multimodal MR radiomics that can stratify gliomas' molecular subtypes is proposed in [51]. They suggested a three-level machine learning model composed of four binary classifiers to stratify five molecular subtypes of gliomas. They exploited multimodal MR radiomics to provide a reliable alternative to determine the pathology and molecular subtypes of gliomas. The approach proposed by the authors is comprised of two parts, the classification model-building part, and the performance evaluation part. The classification model building phase is itself comprised of six SVM and three ensemble learning approaches. The algorithm SVM is ideal for glioma prediction, which is easy to control the complexity of decision rules and the frequency of errors, as well as overfitting is unlikely to occur. The decision tree is another popular choice, as it results in a more human-friendly structure that can provide an understanding of how the system makes a choice. This work employed advanced multimodal MR radiomics to construct more comprehensive functional and metabolic radiomics in the characterization of gliomas. The authors claim that their model can effectively stratify five molecular subtypes to benefit the diagnosis and monitoring of gliomas. If gliomas molecular grades can be classified only through medical images, there will be less burden of patients and more efficiency of the hospital process.

Some of the researchers are concerned about the effect of supervised feature selection. Feature selection is necessary as the data usually contain many irrelevant, redundant, and noisy expressions. Effective data engineering can avoid the “garbage in, garbage out” consequence in machine learning problems. For example, Sengupta et al. [52] proposed an optimized SVM classifier to handle the problems of glioma grading using T1 perfusion parameters and volume of tumor components. The authors applied random forests to obtain optimal features for building an SVM classifier, which provided better grading results than the above result. The classifier was evaluated using 12-fold cross-validation and can achieve satisfactory classifications with an error of 3.7% for grade II vs. III, 5.26% for grade III vs. IV, and 9.43% for Grade II vs. III vs. IV. After effective data engineering, the performance was better than the last research. A potential limitation in this study was the absence of Grade I patients. The model could likely be further optimized by eliminating data imbalance between groups. In this research, most of the methods require manual tumor delineation, which is one of the limitations of this kind of experiment.

In the scenario of glioma prediction, it is common to occur data imbalance. Some research solved this issue and reached a more reliable result. The advantages of the Naive Bayes algorithm is not only easy to understand and can efficiently train, but also can be founded based on statistical modeling, solving the problem of data imbalance between groups was also effective. Niu et al. [53] employed five machine learning methods in a study aiming to predict the glioma stages based on the selected key genes. A total number of 527 gene expression data of brain tissue samples of Homo sapiens downloaded from the GEO (Available online: <http://www.ncbi.nlm.nih.gov/geo/>, accessed on 8 November 2021) database were considered in this study. A specific classification model was followed with the employment of two algorithms, namely random forests and Complement Naive Bayes. As a result, the prediction accuracy by using random forests was 97.1% for Grade I-II, and the prediction accuracy by using Complement Naive Bayes was 72.8% for Grade III-VI; any bias could be avoided when building the most

effective model of imbalanced data. In their study, the authors selected 19 genes between Grade II glioma through bioinformatics method and protein-protein interaction network followed by the random forests. They filtered 21 genes between Grade III glioma and Grade II through bioinformatics method and protein-protein interaction network followed by the Complement Naive Bayes algorithm. The authors combined the more accurate gene expression data in order to yield a comprehensive insight into the molecules involved in the pathogenesis of gliomas.

3.3. The Prognosis Prediction of Gliomas

The incidence rate, cancer recurrence, and cancer survivability are three important cancer prognosis predictive facets. The first scenario predicts the possibility of developing cancer before the disease occurs. The second scenario tries to predict the likelihood of redeveloping cancer and the third scenario predicts the outcomes after the disease diagnosis, such as life expectancy, survivability, progress, and tumor drug sensitivity [61].

The term survival refers to the time interval from the beginning of a patient to the occurrence of an event, such as the period of the beginning and end of a recovery, or the time from the cancer diagnosis until death [62]. In medical research, survival analysis is often used to evaluate data from time-to-event. Survival analysis is a different field concerned with predicting the time until a medical condition occurs. From the perspective of machine learning, survival analysis is a ranking problem in which data points are ranked on their survival times rather than predicting the actual survival times [63]. Oncologists often face the difficult tasks of predicting the prognosis and survival of patients with refractory malignant tumors [64]. Their assessments in these cases are based on clinical experience and comprehensive knowledge of patients. However, physicians usually tend to overestimate the survival risk of patients with advanced cancer, as such predictions are largely unreliable, inaccurate, and generally more optimistic. Survivability prediction is related to several predictive facets that consist of genetic factors, size, as well as grade and stage of the tumor. When physicians have a good understanding of the prognosis of patients, patients are likely to receive more accurate treatment [65]. Necessity is apparent for physicians to have the ability for formulating a correct estimation of survival among patients with advanced and incurable cancers in the medical decision-making process, which involved data analysis, classification, and prediction.

Some research merged multi omics data of gliomas and tried to mine more information for glioma prognosis. An interesting article published in 2018 [54] proposed a convolutional neural network-based model with traditional survival models from pathology images. It can tolerate noisy inputs as well as be used for both regression and classification, which is suitable for the prediction of prognosis. The authors advocate the idea that the treatment planning for gliomas is dependent on many factors, including patient age and grade, which have been limited by considerable intra- and inter-observer variability. The article integrated the microscopic images of tissue biopsies and genomic biomarkers into a single prediction framework. They provided a comprehensive insight into the molecules involved in the pathogenesis of glioma. The authors claimed that their approach surpasses the prognostic accuracy of human experts using the current clinical standard for classifying brain tumors. They presented an innovative approach for objective, accurate, and integrated predictions of patient outcomes. Further analysis is still required to combine pathology images with rich genomic and clinical annotations to clarify the mechanisms underlying glioma tumor genesis and development.

A work by [55] developed a signature associated with the tumor immune microenvironment using machine learning. The study showed the improvement in predicting the survival rates of glioma patients by using LASSO Cox regression algorithm. Based on biological knowledge, the authors investigated the immunogenomic landscape of glioma followed by developing an immune-relevant prognostic 15-gene signature for glioma patients. The dataset performed through this study consists of transcriptomic and clinical data found in the Chinese Glioma Genome Atlas (CGGA) (Available online: <http://www.cgga.org.cn/>,

accessed on 8 November 2021) databases [66], as well as of immune cells derived from the TISIDB database. Based on the results of this study, the results showed that glioma patients in the high-infiltration group have worse overall survival than patients in the low-infiltration group. They also claimed that their 15-immune-relevant-gene signature model showed effectiveness and breadth in predicting prognosis in glioma. However, existing limitations of the current article are related to the fact that the impact of other variables related to prognosis (such as age, genetic alterations, WHO grade, and treatments) is not considered, which may have led to misprediction results. Furthermore, the authors made clear that the function of the genes in the signature was a lack of further basic experiments for validation. The immune-related gene signature may help in identifying pathways associated with glioma and potential immune targets for treatment in the future.

Many researchers have taken an extraordinary effort to develop the prognosis prediction model of gliomas, and proposed a survival prediction model of this highly malignant tumor based on MRI radiation characteristics, imaging features from fluorodeoxyglucose-positron emission tomography (FDG-PET), combined with genetic and clinical risk factors. Because of its non-invasiveness and easy access, this field has received more and more attention. In [4], the authors investigated the relationship between tumor shape, quantified using algorithmic analysis of magnetic resonance images, and survival. This study uncovered the relationship between tumor 3D shape and prognosis, which cannot be found only by human flesh eyes. In this study, the researchers gathered each patient's Fluid Attenuated Inversion Recovery (FLAIR) abnormality and manually delineated enhancing tumor. They implemented a set of features that capture the intricacies of the two and three dimensional shapes and that are independent of the imaging equipment and acquisition parameters. The results showed that a 3D complexity measure bounding ellipsoid volume ratio (BEVR) was strongly prognostic of survival. Lower values of BEVR are associated with poorer survival and indicate a higher level of irregularity of the tumor, which might be associated with a more rapid tumor growth. In addition, three enhancing-tumor based shape features were clinical independent factors. The proposed analysis can be used to help physicians and caregivers customize treatment based on better survival estimates for patients with GBMs. There was also a limitation in this research which is the limited sample size of 68 patients. For machine learning tasks with a small number of samples, a better solution is to apply cross-validation for model selection and preventing overfitting [67].

4. Future Challenges

Several challenges must be addressed before the adaptation of machine learning in oncology and specifically, gliomas. Based on the research in the scope of this review, we summarized several main issues listed in Table 3, including data, algorithm, and application. Details are presented in the following sub-sections.

Table 3. Challenges in future research.

Category	Challenges
Data aspect	lack of annotated data data quality and integrity data class imbalance
Model aspect	Overfitting Clever hans lack of comparing with different models generalizability of models reproducibility of model
Clinical application aspect	Physicians' knowledge limitations

4.1. The Challenges in Data

One of the most common issues seen among the studies surveyed in this review was the lack of attention paid to the objective dataset. It is considerable to recognize

that several substantive challenges for machine learning glioma analysis at the data level include the (i) lack of annotated data [68]; (ii) ensure data quality and integrity; as well as (iii) data imbalance.

In the proper training and convergence of machine learning techniques process, the tremendous volume of high-quality and well-annotated data is necessary; however, the multi-institutional nature of most clinical trials for gliomas limit the sharing of patient data between institutions, so the multitude of heterogeneous datasets are difficult to aggregate. Even though large cohorts across many institutions can be aggregated, annotation is a time-consuming process and requires a high degree of expertise. Considering that manual annotations are often costly, the future development of customized semi-automated annotation tools and iterative re-annotation strategies may provide a promising solution by relying on machine learning algorithms to provide initial ground-truth estimates, which are then refined by human experts [69]. In genomics, batch effects [70] often occurs when merging data from different institutions and different batches of experiments. Mean-centering, standardization, ratio-based and EJLR (Extended Johnson-Li-Rabinovic) method [71] are widely used to solve this problem.

As for the second problem, for large datasets data entry and data verification are of essential importance. Further verification or spot checking of data integrity is also a valuable exercise, and implement by a knowledgeable expert. However, in most machine learning papers, the methods employed to ensure the quality and integrity of data are rarely discussed. In terms of data imbalance [72], it is a common issue but has little attention when building and analyzing cancer prediction models. This problem makes the classifier focus on learning most of the data classes so it can poorly classify the samples which belong to the minor class. The paper mentioned above [53] involved in this problem gave a solution. It is feasible for imbalanced data on the high dimensional level to mitigate negative effects through SMOTE [73], adaptive boosting [74].

4.2. The Challenges in Algorithm

There are also many tough challenges in the algorithm aspect. For a limited sample size, almost any models are prone to overfitting, which results in an artificially inflated algorithm accuracy [75,76] Although most of the research on cancer prediction models centered on improving the predictability and learning of good representations, they ignored the problem of overfitting due to limited samples as referred by [76]. For example, small samples and large dimensions were particularly big problems for microarray studies, which often have tens of thousands of genes (i.e., features), but only hundreds of samples [74]. In addition, feature selection is a helpful method to prevent overfitting, as shown in previous study [49]. The sample-per-feature ratio is too small to be highly susceptible to overtraining. The problem with overfitting models is that with more and more test cases available, the robustness of models cannot be guaranteed. In overfitting, in addition to striving for large, heterogeneous datasets, L1 regularization is also effective [77], and batch normalization [78] is giving good results in other research. Generally, 5-fold or 10-fold cross-validation is enough to verify most of any learning algorithm [79,80]. Cross-validation is internal validation which is critical to create a robust model that can consistently deal with novel data. In addition to the standard practice of internal validation, it is particularly beneficial to use external data sources for validation testing [81]. The external validation set must be large enough to ensure reproducibility, as well as to help minimize generated objective bias.

The performances of different multiple predictor models is another challenge in algorithms. While sophisticated and advanced neural network algorithms always work well in large heterogeneous datasets, in biomedical tasks complex models are not always the best tools for the task. Overfitting may happen easily on a complex model. Sometimes conventional algorithms, such as support vector machine k-nearest neighbors, and Naïve Bayes can be more popular than sophisticated algorithms, hence their explanatory ability. The assessment of different predictors is a significant step in determining the proper

algorithm for the study. Any newly-mentioned model should be compared with the existing models. In our previously discussed research, [53] used random forests and Complement Naive Bayes for comparing, [82] and used random survival forest, SVM, and Cox proportional hazards models to predict survival. However, testing more than one machine learning algorithm is not carried out in most of the papers which we reviewed.

Generalizability of machine learning algorithms is another important issue that should be highlighted here. Most complex machine learning algorithms do not perform well in a small cohort, especially for the large, standardized datasets from multiple institutions with clinical, neuroimaging, and neuropathologic data of gliomas, but up to now most gliomas research focused on relatively small patient populations. Large and heterogeneous populations of gliomas that cover diverse patient populations are needed to fully realize the power of machine learning for the diagnosis and treatment of gliomas. The performance of the model in security and privacy also affects the generalizability, such as Clever Hans. At present, it is much easier to find adversarial examples than to design a model that can defend the adversarial examples. However, there are still some methods to defend adversarial perturbation, such as FGSM (fast gradient sign method) [83] and defensive distillation [84].

The reproducibility of machine learning algorithms is one of the main challenges to transform theoretical research into clinical practice [85]. Machine learning methods, as a computational experiment process like any experiments, are essentially depending on a hypothesis. It follows the defined procedures and needs to verify. Providing the datasets used for training and testing and detailed methodological documentation is of paramount importance. Researchers using open databases such as the Cancer Genome Atlas (TCGA) should at least explain the query used to download the experimental dataset in the supplementary material to permit others to verify and reproduce the results. Parameters of model training should also be stated in detail including how the sets were partitioned. In general, the results of a good machine learning experiment should be as reproducible as other standard laboratory protocols.

4.3. The Challenges in Clinical Application

Additional challenges relate to the deployment of machine learning applications in a clinical setting. In early studies of the application of machine learning in gliomas, there is no prospective research to confirm that machine learning can bring benefits to patients' prognosis. Currently, physicians receive very little training in computer/data science and most computer scientists are not familiar with the complexity of clinical patient management [73]. In addition, image analysis faced many challenges, including the lack of standardization of image acquisition, poor reproducibility, complex quantitative features, and so on [86]. Therefore, comparing results among different institutions may be a challenge, hence limiting the clinical applications.

Another issue is how to understand and trust the information generated by machine learning "black box" [87]. Different from other disciplines, clinical practices can be with a very low fault tolerance rate. The quality of patient life is dependent on the decision maker, which means decisions should be made with a high degree of confidence. Giving meaning to the black box can also form the basis for future work in medical imaging, radiomics, and machine learning.

5. Conclusions

This article reviewed the most recent machine learning-based gliomas application models. The considered papers were published during the last 2 decades and used gene expression datasets, as well as medical imaging cohorts, for cancer susceptibility, grade, and survivability risk. This review presented some commonly-used architectures, datasets, and the accuracies of each suggested model. Analyzing the considered papers indicated that machine learning methods can serve as filters, predictors, and classification methods.

It is also obvious that machine learning methods generally improve the performance or predictive accuracy of most scenarios, especially when compared to conventional medical or expert-based systems. Overall, we believe that if the quality of studies continues to improve, it is likely that the use of machine learning classifiers will become much more commonplace in many clinical and hospital settings.

This study has summarized the most recent approaches and their related machine learning architectures. We also highlighted some critical points that have to be considered when building a machine learning-based prediction model, such as reproducibility and data quality. More powerful machine learning-based approaches can be suggested in the future by choosing different model parameters or combining two or more of the presented approaches.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/cells10113169/s1>, Table S1: We retrieved in PubMed Central by using the keywords in this table. (When retrieving, we use not only the abbreviations of the algorithms but also their full names, and only abbreviations are reflected in this table.).

Author Contributions: N.Z. conceived and supervised the project. Y.G. and Y.W. were responsible for the design, data preprocessing, computational analyses, and drafted the manuscript with revisions provided. J.M., Y.S., Q.L., N.Z. and Y.W. participated in the design of the study and performed the computational analysis. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work was supported by grants from Key Projects in the Science & Technology Pillar Program of Tianjin (20YFZCSN00530) and Independent Innovation Foundation of Tianjin University (2020XYF-0110, 2020XY-0058).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Northcott, P.A.; Robinson, G.W.; Kratz, C.P.; Mabbott, D.J.; Pomeroy, S.L.; Clifford, S.C.; Rutkowski, S.; Ellison, D.W.; Malkin, D.; Taylor, M.D. Medulloblastoma. *Nat. Rev. Dis. Prim.* **2019**, *5*, 1–20. [[CrossRef](#)] [[PubMed](#)]
2. Louis, D.N.; Perry, A.; Reifenberger, G.; von Deimling, A.; Figarella-Branger, D.; Cavenee, W.K.; Ohgaki, H.; Wiestler, O.D.; Kleihues, P.; Ellison, D.W. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A summary. *Acta Neuropathol.* **2016**, *131*, 803–820. [[CrossRef](#)] [[PubMed](#)]
3. Stupp, R.; Mason, W.P.; van den Bent, M.J.; Weller, M.; Fisher, B.; Taphoorn, M.J.B.; Belanger, K.; Brandes, A.A.; Marosi, C.; Bogdahn, U.; et al. Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *N. Engl. J. Med.* **2005**, *352*, 987–996. [[CrossRef](#)]
4. Czarnek, N.; Clark, K.; Peters, K.B.; Mazurowski, M.A. Algorithmic three-dimensional analysis of tumor shape in MRI improves prognosis of survival in glioblastoma: A multi-institutional study. *J. Neurooncol.* **2017**, *132*, 55–62. [[CrossRef](#)]
5. Bulgarevich, D.S.; Tsukamoto, S.; Kasuya, T.; Demura, M.; Watanabe, M. Pattern recognition with machine learning on optical microscopy images of typical metallurgical microstructures. *Sci. Rep.* **2018**, *8*, 2078. [[CrossRef](#)] [[PubMed](#)]
6. Gillies, A. *Machine Learning Procedures for Generating Image Domain Feature Detectors*; The University of Michigan: Ann Arbor, MI, USA, 1985.
7. Zhang, X.; Qin, W.; Zhuang, H.; Guo, Y.; Yang, P.; Han, H.; Liu, Y.; Chen, B.; Chen, Z. A real-time recognition method for telemetry status of spacecraft in-orbit based on neural network pattern recognition. In Proceedings of the 2018 3rd International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Huhhot, China, 14–16 September 2018; Volume 2018, pp. 361–363. [[CrossRef](#)]
8. Poongodi, M.; Sharma, A.; Vijayakumar, V.; Bhardwaj, V.; Sharma, A.P.; Iqbal, R.; Kumar, R. Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system. *Comput. Electr. Eng.* **2020**, *81*, 106527. [[CrossRef](#)]
9. Xu, X.; Wang, J.; Peng, H.; Wu, R. Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Comput. Human Behav.* **2019**, *98*, 166–173. [[CrossRef](#)]
10. Sun, S.; Dong, B.; Zou, Q. Revisiting genome-wide association studies from statistical modelling to machine learning. *Brief. Bioinform.* **2021**, *22*, bbaa263. [[CrossRef](#)]

11. Zhou, G.D.; Zhang, J.; Su, J.; Shen, D.; Tan, C.L. Recognizing names in biomedical texts: A machine learning approach. *Bioinformatics* **2004**, *20*, 1178–1190. [[CrossRef](#)]
12. Razzak, M.I.; Naz, S.; Zaib, A. Deep learning for medical image processing: Overview, challenges and the future. *Lect. Notes Comput. Vis. Biomech.* **2018**, *26*, 323–350. [[CrossRef](#)]
13. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [[CrossRef](#)]
14. Pentakalos, O. *Introduction to Machine Learning*; CRC Press: Boca Raton, FL, USA, 2020; pp. 262–290. ISBN 1420011782.
15. Bharadwaj; Prakash, K.B.; Kanagachidambaresan, G.R. Pattern Recognition and Machine Learning. In *Programming with TensorFlow*; Prakash, K.B., Kanagachidambaresan, G.R., Eds.; EAI/Springer Innovations in Communication and Computing; Springer: Cham, Berlin/Heidelberg, Germany, 2021; pp. 105–144. ISBN 1493938436. [[CrossRef](#)]
16. He, W.; Jia, C.; Zou, Q. 4mCPred: Machine learning methods for DNA N 4 -methylcytosine sites prediction. *Bioinformatics* **2019**, *35*, 593–601. [[CrossRef](#)]
17. Meng, C.; Zhang, J.; Ye, X.; Guo, F.; Zou, Q. Review and comparative analysis of machine learning-based phage virion protein identification methods. *Biochim. Biophys. Acta-Proteins Proteom.* **2020**, *1868*, 140406. [[CrossRef](#)] [[PubMed](#)]
18. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386–408. [[CrossRef](#)]
19. Carpenter, G.A.; Grossberg, S. The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network. *Computer* **1988**, *21*, 77–88. [[CrossRef](#)]
20. Fukushima, K.; Miyake, S.; Ito, T. Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition. In *IEEE Transactions on Systems, Man and Cybernetics*; Springer: Berlin/Heidelberg, Germany, 1983; Volume SMC-13, pp. 826–834.
21. Korfiatis, P.; Kline, T.L.; Coufalova, L.; Lachance, D.H.; Parney, I.F.; Carter, R.E.; Buckner, J.C.; Erickson, B.J. MRI texture features as biomarkers to predict MGMT methylation status in glioblastomas. *Med. Phys.* **2016**, *43*, 2835–2844. [[CrossRef](#)] [[PubMed](#)]
22. Wu, S.; Li, J.; Cao, M.; Yang, J.; Li, Y.X.; Li, Y.Y. A novel integrated gene coexpression analysis approach reveals a prognostic three-transcription-factor signature for glioma molecular subtypes. *BMC Syst. Biol.* **2016**, *10*, 71. [[CrossRef](#)] [[PubMed](#)]
23. Zacharaki, E.I.; Wang, S.; Chawla, S.; Yoo, D.S.; Wolf, R.; Melhem, E.R.; Davatzikos, C. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magn. Reson. Med.* **2009**, *62*, 1609–1618. [[CrossRef](#)] [[PubMed](#)]
24. Metsis, V.; Huang, H.; Andronesi, O.C.; Makedon, F.; Tzika, A. Heterogeneous data fusion for brain tumor classification. *Oncol. Rep.* **2012**, *28*, 1413–1416. [[CrossRef](#)]
25. Rajkomar, A.; Dean, J.; Kohane, I. Machine Learning in Medicine. *N. Engl. J. Med.* **2019**, *380*, 1347–1358. [[CrossRef](#)]
26. Kaelbling, L.P.; Littman, M.L.; Moore, A.W. Reinforcement Learning: A Survey. *J. Artif. Intell. Res.* **1996**, *4*, 237–285. [[CrossRef](#)]
27. Clarke, M.R.B.; Duda, R.O.; Hart, P.E. Pattern Classification and Scene Analysis. *J. R. Stat. Soc. Ser. A* **1974**, *137*, 442. [[CrossRef](#)]
28. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
29. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
30. He, Z.; Wu, Z.; Xu, G.; Liu, Y.; Zou, Q. Decision Tree for Sequences. *IEEE Trans. Knowl. Data Eng.* **2021**, phrase indicating stage of publication (submitted; accepted; in press). [[CrossRef](#)]
31. Singh, A. Foundations of Machine Learning. *SSRN Electron. J.* **2019**. [[CrossRef](#)]
32. Zhu, X. Computer Semi-Supervised Learning Literature Survey. *Tech rep, computer sciences, University of Wisconsin-Madison* **2005**, *2*, 1–59.
33. Hofmann, T. Unsupervised learning by probabilistic Latent Semantic Analysis. *Mach. Learn.* **2001**, *42*, 177–196. [[CrossRef](#)]
34. Spechler, S.J. Barrett esophagus and risk of esophageal cancer: A clinical review. *JAMA-J. Am. Med. Assoc.* **2013**, *310*, 627–636. [[CrossRef](#)]
35. Li, W.; Jia, F.; Hu, Q. Automatic Segmentation of Liver Tumor in CT Images with Deep Convolutional Neural Networks. *J. Comput. Commun.* **2015**, *3*, 146–151. [[CrossRef](#)]
36. Wang, J.X.; Zhang, B.; Yu, J.K.; Liu, J.; Yang, M.Q.; Zheng, S. Application of serum protein fingerprinting coupled with artificial neural network model in diagnosis of hepatocellular carcinoma. *Chin. Med. J.* **2005**, *118*, 1278–1286.
37. Li, M.; Guo, Y.; Feng, Y.M.; Zhang, N. Identification of triple-negative breast cancer genes and a novel high-risk breast cancer prediction model development based on ppi data and support vector machines. *Front. Genet.* **2019**, *10*, 180. [[CrossRef](#)] [[PubMed](#)]
38. Mousavi, H.; Rao, G.; Rao, A.K.; Monga, V. Automated discrimination of lower and higher grade gliomas based on histopathological image analysis. *J. Pathol. Inform.* **2015**, *6*, 15. [[CrossRef](#)]
39. Ertosun, M.G.; Rubin, D.L. Automated Grading of Gliomas using Deep Learning in Digital Pathology Images: A modular approach with ensemble of convolutional neural networks. *AMIA Annu. Symp. Proc.* **2015**, *2015*, 1899–1908. [[PubMed](#)]
40. van Bodegraven, E.J.; van Asperen, J.V.; Robe, P.A.J.; Hol, E.M. Importance of GFAP isoform-specific analyses in astrocytoma. *Glia* **2019**, *67*, 1417–1433. [[CrossRef](#)] [[PubMed](#)]
41. Zöllner, F.G.; Emblem, K.E.; Schad, L.R. SVM-based glioma grading: Optimization by feature reduction analysis. *Z. Med. Phys.* **2012**, *22*, 205–214. [[CrossRef](#)] [[PubMed](#)]
42. Harari, M.B.; Parola, H.R.; Hartwell, C.J.; Riegelman, A. Literature searches in systematic reviews and meta-analyses: A review, evaluation, and recommendations. *J. Vocat. Behav.* **2020**, *118*, 103377. [[CrossRef](#)]
43. Abusamra, H. A comparative study of feature selection and classification methods for gene expression data of glioma. *Procedia Comput. Sci.* **2013**, *23*, 5–14. [[CrossRef](#)]

44. Wu, Y.; Zhao, Z.; Wu, W.; Lin, Y.; Wang, M. Automatic glioma segmentation based on adaptive superpixel. *BMC Med. Imaging* **2019**, *19*, 73. [[CrossRef](#)]
45. Chen, W.; Liu, B.; Peng, S.; Sun, J.; Qiao, X. Computer-Aided Grading of Gliomas Combining Automatic Segmentation and Radiomics. *Int. J. Biomed. Imaging* **2018**, 2018. [[CrossRef](#)]
46. Gao, X.Y.; Wang, Y.D.; Wu, S.M.; Rui, W.T.; Ma, D.N.; Duan, Y.; Zhang, A.N.; Yao, Z.W.; Yang, G.; Yu, Y.P. Differentiation of treatment-related effects from glioma recurrence using machine learning classifiers based upon pre-and post-contrast T1WI and T2 FLAIR subtraction features: A two-center study. *Cancer Manag. Res.* **2020**, *12*, 3191–3201. [[CrossRef](#)]
47. Rathore, S.; Akbari, H.; Doshi, J. Radiomic signature of infiltration in peritumoral edema predicts subsequent recurrence in glioblastoma: Implications for personalized radiotherapy planning. *J. Med. Imaging* **2018**, *5*, 1. [[CrossRef](#)] [[PubMed](#)]
48. Hsu, J.B.K.; Chang, T.H.; Lee, G.A.; Lee, T.Y.; Chen, C.Y. Identification of potential biomarkers related to glioma survival by gene expression profile analysis. *BMC Med. Genom.* **2019**, *11*, 34. [[CrossRef](#)] [[PubMed](#)]
49. Haubold, J.; Demircioglu, A.; Gratz, M.; Glas, M.; Wrede, K.; Sure, U.; Antoch, G.; Keyvani, K.; Nittka, M.; Kannengiesser, S.; et al. Non-invasive tumor decoding and phenotyping of cerebral gliomas utilizing multiparametric 18F-FET PET-MRI and MR Fingerprinting. *Eur. J. Nucl. Med. Mol. Imaging* **2020**, *47*, 1435–1445. [[CrossRef](#)]
50. Matsui, Y.; Maruyama, T.; Nitta, M.; Saito, T.; Tsuzuki, S.; Tamura, M.; Kusuda, K.; Fukuya, Y.; Asano, H.; Kawamata, T.; et al. Prediction of lower-grade glioma molecular subtypes using deep learning. *J. Neurooncol.* **2020**, *146*, 321–327. [[CrossRef](#)]
51. Lu, C.F.; Hsu, F.T.; Hsieh, K.L.C.; Kao, Y.C.J.; Cheng, S.J.; Hsu, J.B.K.; Tsai, P.H.; Chen, R.J.; Huang, C.C.; Yen, Y.; et al. Machine learning-based radiomics for molecular subtyping of gliomas. *Clin. Cancer Res.* **2018**, *24*, 4429–4436. [[CrossRef](#)] [[PubMed](#)]
52. Sengupta, A.; Ramaniharan, A.K.; Gupta, R.K.; 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*, 1295–1306. [[CrossRef](#)]
53. Niu, B.; Liang, C.; Lu, Y.; Zhao, M.; Chen, Q.; Zhang, Y.; Zheng, L.; Chou, K.C. Glioma stages prediction based on machine learning algorithm combined with protein-protein interaction networks. *Genomics* **2020**, *112*, 837–847. [[CrossRef](#)]
54. Mobadersany, P.; Yousefi, S.; Amgad, M.; Gutman, D.A.; Barnholtz-Sloan, J.S.; Velázquez Vega, J.E.; Brat, D.J.; Cooper, L.A.D. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E2970–E2979. [[CrossRef](#)]
55. Gong, X.; Liu, L.; Xiong, J.; Li, X.; Xu, J.; Xiao, Y.; Li, J.; Luo, X.; Mao, D.; Liu, L. Construction of a Prognostic Gene Signature Associated with Immune Infiltration in Glioma: A Comprehensive Analysis Based on the CGGA. *J. Oncol.* **2021**, 2021. [[CrossRef](#)]
56. Jackson, R.J.; Fuller, G.N.; Abi-Said, D.; Lang, F.F.; Gokaslan, Z.L.; Shi, W.M.; Wildrick, D.M.; Sawaya, R. Limitations of stereotactic biopsy in the initial management of gliomas. *Neuro-Oncol.* **2001**, *3*, 193–200. [[CrossRef](#)] [[PubMed](#)]
57. Kristensen, B.W.; Priesterbach-Ackley, L.P.; Petersen, J.K.; Wesseling, P. Molecular pathology of tumors of the central nervous system. *Ann. Oncol.* **2019**, *30*, 1265–1278. [[CrossRef](#)] [[PubMed](#)]
58. Akkus, Z.; Ali, I.; Sedlář, J.; Agrawal, J.P.; Parney, I.F.; Giannini, C.; Erickson, B.J. Predicting Deletion of Chromosomal Arms 1p/19q in Low-Grade Gliomas from MR Images Using Machine Intelligence. *J. Digit. Imaging* **2017**, *30*, 469–476. [[CrossRef](#)] [[PubMed](#)]
59. Gao, M.; Huang, S.; Pan, X.; Liao, X.; Yang, R.; Liu, J. Machine Learning-Based Radiomics Predicting Tumor Grades and Expression of Multiple Pathologic Biomarkers in Gliomas. *Front. Oncol.* **2020**, *10*, 1676. [[CrossRef](#)] [[PubMed](#)]
60. Ahmed, R.; Oborski, M.J.; Hwang, M.; Lieberman, F.S.; Mountz, J.M. Malignant gliomas: Current perspectives in diagnosis, treatment, and early response assessment using advanced quantitative imaging methods. *Cancer Manag. Res.* **2014**, *6*, 149–170. [[CrossRef](#)]
61. Bashiri, A.; Ghazisaeedi, M.; Safdari, R.; Shahmoradi, L.; Ehtesham, H. Improving the prediction of survival in cancer patients by using machine learning techniques: Experience of gene expression data: A narrative review. *Iran. J. Public Health* **2017**, *46*, 165–172. [[PubMed](#)]
62. Somi, M.H.; Ahmadzadeh, R.; Farhang, S.; Mirinejhad, S.K.; Jazayeri, E.; Sadeghi, M.; Naghashi, S. Evaluation of treatment and survival rates in patients with esophageal cancer referred to Imam Khomeini Hospital, Tabriz, Iran. *Govaresh* **2012**, *17*, 33–38.
63. Raykar, V.C.; Steck, H.; Krishnapuram, B.; Dehing-Oberije, C.; Lambin, P. On ranking in survival analysis: Bounds on the concordance index. In Proceedings of the Advances in Neural Information Processing Systems 20-2007 Conference, Vancouver, BC, Canada, 3 December 2007; pp. 1209–1216.
64. Zhao, S.; Lin, Y.; Xu, W.; Jiang, W.; Zhai, Z.; Wang, P.; Yu, W.; Li, Z.; Gong, L.; Peng, Y.; et al. Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1 α . *Science* **2009**, *324*, 261–265. [[CrossRef](#)]
65. Krishnan, M.; Temel, J.S.; Wright, A.A.; Bernacki, R.; Selvaggi, K.; Balboni, T. Predicting life expectancy in patients with advanced incurable cancer: A review. *J. Support. Oncol.* **2013**, *11*, 68–74. [[CrossRef](#)]
66. Zhao, Z.; Zhang, K.N.; Wang, Q.; Li, G.; Zeng, F.; Zhang, Y.; Wu, F.; Chai, R.; Wang, Z.; Zhang, C.; et al. Chinese Glioma Genome Atlas (CGGA): A Comprehensive Resource with Functional Genomic Data from Chinese Glioma Patients. *Genom. Proteom. Bioinforma.* **2021**, *19*, 1–12. [[CrossRef](#)]
67. Schaffer, C. Technical Note: Selecting a Classification Method by Cross-Validation. *Mach. Learn.* **1993**, *13*, 135–143. [[CrossRef](#)]
68. Choy, G.; Khalilzadeh, O.; Michalski, M.; Do, S.; Samir, A.E.; Pianykh, O.S.; Geis, J.R.; Pandharipande, P.V.; Brink, J.A.; Dreyer, K.J. Current applications and future impact of machine learning in radiology. *Radiology* **2018**, *288*, 318–328. [[CrossRef](#)] [[PubMed](#)]
69. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]

70. Luo, J.; Schumacher, M.; Scherer, A.; Sanoudou, D.; Megherbi, D.; Davison, T.; Shi, T.; Tong, W.; Shi, L.; Hong, H.; et al. A comparison of batch effect removal methods for enhancement of prediction performance using MAQC-II microarray gene expression data. *Pharm. J.* **2010**, *10*, 278–291. [[CrossRef](#)] [[PubMed](#)]
71. Johnson, W.E.; Li, C.; Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **2007**, *8*, 118–127. [[CrossRef](#)]
72. Krawczyk, B. Learning from imbalanced data: Open challenges and future directions. *Prog. Artif. Intell.* **2016**, *5*, 221–232. [[CrossRef](#)]
73. Danaee, P.; Ghaeini, R.; Hendrix, D.A. A deep learning approach for cancer detection and relevant gene identification. *Pac. Symp. Biocomput.* **2017**, 219–229. [[CrossRef](#)]
74. Margineantu, D.; Dietterich, T.G. Pruning Adaptive Boosting. In Proceedings of the Fourteenth International Conference on Machine Learning, San Francisco, CA, USA, 8–12 July 1997; Citeseer: Vancouver, BC, Canada, 1997; pp. 211–218.
75. Hawkins, D.M. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12. [[CrossRef](#)]
76. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
77. Moore, R.C.; DeNero, J. L1 and L2 Regularization for Multiclass Hinge Loss Models. In Proceedings of the Symposium on Machine Learning in Speech and Language Processing, Bellevue, WA, USA, 27 June 2011; 2009; Volume 3.
78. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, PMLR, Lille, France, 7–8 July 2015; Volume 1, pp. 448–456.
79. Browne, M.W. Cross-validation methods. *J. Math. Psychol.* **2000**, *44*, 108–132. [[CrossRef](#)]
80. Cawley, G.C.; Talbot, N.L.C. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079–2107.
81. Ng, A.Y. Preventing “overfitting” of cross-validation data. In Proceedings of the ICML, San Francisco, CA, USA, 8–12 July 1997; Morgan Kaufmann: Burlington, MA, USA, 1997; Volume 97, pp. 245–253.
82. Chen, H.; Li, C.; Zheng, L.; Lu, W.; Li, Y.; Wei, Q. A machine learning-based survival prediction model of high grade glioma by integration of clinical and dose-volume histogram parameters. *Cancer Med.* **2021**, *10*, 2774–2786. [[CrossRef](#)] [[PubMed](#)]
83. Goodfellow, I.J.; Shlens, J.; Szegedy, C. Explaining and harnessing adversarial examples. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015-Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015.
84. Papernot, N.; McDaniel, P.; Wu, X.; Jha, S.; Swami, A. Distillation as a Defense to Adversarial Perturbations Against Deep Neural Networks. In Proceedings of the 2016 IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, 22–26 May 2016; pp. 582–597. [[CrossRef](#)]
85. McDermott, M.B.A.; Wang, S.; Marinsek, N.; Ranganath, R.; Foschini, L.; Ghassemi, M. Reproducibility in machine learning for health research: Still a ways to go. *Sci. Transl. Med.* **2021**, *13*. phrase indicating stage of publication (submitted; accepted; in press). [[CrossRef](#)] [[PubMed](#)]
86. Rizzo, S.; Botta, F.; Raimondi, S.; Origgi, D.; Fanciullo, C.; Morganti, A.G.; Bellomi, M. Radiomics: The facts and the challenges of image analysis. *Eur. Radiol. Exp.* **2018**, *2*, 36. [[CrossRef](#)]
87. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why should i trust you?” Explaining the predictions of any classifier. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM 2016, New York, NY, USA, 13–17 August 2016; pp. 1135–1144. [[CrossRef](#)]