Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/31/20230173

Deep learning-based glioma grading and feature visualization analysis

Yunmeng Chai[1,4,†], Yuehan Wang[2,†], Wenqi Xue[3,†]

[1]Faculty of Environment and Life, Beijing University of Technology, Beijing, 100124, China
[2]School of Computer Science and Engineering, Tianjin University of Technology, Tianjin, 300382, China
[3]School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou, Zhejiang, 310018, China

[4]chaiyunmeng@emails.bjut.edu.cn
[†]All the authors contributed equally and their names were listed in alphabetical order.

Abstract. Gliomas can be separated into high- and low-grade gliomas according to the classification method developed by the World Health Organization (WHO). Glioma classification is significantly related to prognosis, and accurate glioma classification is very important. This study aims to evaluate and verify the analytical performance of different models based on deep learning for glioma grading. Firstly, the glioma grading clinical and mutation feature data sets were included. The training cohort included 20 genes with the most common mutations and 2 clinical features in the Cancer Genome Atlas Low Grade Glioma (TCGA-LGG) and TCGA- Glioblastoma Multiforme (GBM) glioma projects. Then, four pre-trained models are used to extract deep learning features from the data. Preprocessing is used to reduce redundancy and select the most predictive value. To assess the performance, indexes, including the area under the data working curve (AUC) and the accuracy prediction value, are leveraged. Finally, the prediction performance of the test queue is compared to determine the optimal classification model.

Keywords: neural network, deep learning, glioma grading.

1. Introduction

Gliomas, namely heterogeneous tumors, are thought to originate from glial stem cells or neuroblasts, and their morphology is similar to the type of glial cells found in the normal brain. Based on the World Health Organization (WHO) standards, tumors could be graded (grade 2, grade 3 and grade 4) based on the longitudinal axis. Grade 2 is that the cancer cells look abnormal, grow slowly but can invade normal tissues; grade 3 refers to the fact that cancer cells do not look like normal cells, and the number of these cancer cells increases rapidly, which is called anaplastic carcinoma. Grade 4 means that cancer cells do not look like normal cells and grow rapidly. Glioma grade classification is significantly correlated with prognosis [1,2]. Some studies have reported that the genetic differences or prognostic differences between gliomas at grade 2 and grade 3 are not clear. Therefore, they are referred to “low-grade gliomas”, and grade 4 gliomas refer to “high-grade gliomas” [3,4].

DOI: 10.54254/2755-2721/31/20230173

The grading of glioma can guide clinical decision-making. The treatment plan varies with different gradings and is also related to other factors such as the prognosis and quality of life. Therefore, accurate glioma grading is very important. Accurate diagnosis and grading of gliomas require histological examination of tumor specimens. However, although histopathology is considered to be the gold standard for the diagnosis and grading of gliomas, the inherent lag characteristics of postoperative histopathology limit its availability in making initial treatment decisions. At the same time, due to the spatial heterogeneity of tumors, preoperative biopsy may have the risk of limited tissue or sampling errors, and may have insufficient grading [5]. Therefore, a precise and noninvasive grading approach is essential for the treatment and prognosis of glioma patients.

Deep neural systems have advanced quickly in the last few years in the area of medical picture segmentation, classification, and detection, and have been widely used. Many artificial intelligence models, specifically deep neural network (DNN), can provide diagnostic accuracy close to that of doctors [6]. Deep learning requires a large amount of data sets during training to achieve fullness.

2. Method

2.1. Dataset

The Glioma Grading Clinical and Mutation Features Dataset, sourced from The Cancer Genome Atlas (TCGA) project is leveraged as the dataset [7]. The dataset includes the 20 most frequently mutated genes and 2 clinical features from the TCGA-LGG and TCGA-GBM glioma projects. In this dataset, the number of samples with a categorical variable value of low-grade glioma (LGG) is 487 and the number of samples with a categorical variable value of high-grade glioma (HGG) is 352. The number of samples with low grade glioma (LGG) accounts for 58% of the total number of samples and this dataset is a more balanced dataset.

2.2. Models

2.2.1. Extreme Gradient Boosting (XGBoost). It is a decision tree algorithm with gradient boosting that is effective. The model is greatly enhanced over the original Gradient Boosting Decision Tree (GBDT). As a forward addition model, the algorithm's central principle is Boosting, which uses specific techniques to combine several weak learners into one strong learner. Each tree's result is the difference value between the ground truth and the prediction of all the preceding trees. The final result is then derived by summing up all the outcomes, increasing the model's overall efficacy. XGBoost is made up of multiple Classification and Regression Tree (CART) trees, so it can handle problems such as classification and regression [8].

2.2.2. Adaboost. It is an iterative approach, and its main principle is to learn various weak classifiers on the training data. By integrating these classifiers, a strong classifier could be constructed to produce more accurate prediction. This model is designed for altering the data distribution, weighing each sample according on whether or not it was correctly identified each time during training, together with the precision of the latest performance. The updated data set with the amended weights is given to the lower classifier for training. Afterwards, strong classifier could be created by integrating the results of each training session's classifiers. Applying the Adaboost allows some unnecessary input of the training data to be excluded and placed on top of the key training data.

2.2.3. Gradient Boosting. Boosting-based models contains several techniques inspired by gradient descent. By leveraging the negative gradient information from current classifiers, weak models could be sequentially trained. By cumulatively integrating these trained weak classifier, a strong classifier could be bulit. The Multiple Additive Regression Tree (MART) is another name for the Gradient Boosting technique, which uses a decision tree as the weak classifier. The decision tree used in GBDT is typically a CART.

DOI: 10.54254/2755-2721/31/20230173

2.2.4. Random forest. The random forest classification model is formed by combining multiple decision tree classification models. The fundamental idea is to draw a subset of samples leveraging bootstrap. Then a decision tree model could be built based on each sample subset separately. Finally, each record is voted according to each decision tree model's output to get the final classification. Random forest classification has good prediction accuracy. With the growing number of decision trees, this model does not produce the problem of overfitting. In addition, random forest classification has a better tolerance for noise and outliers.

2.2.5. Logistic regression. This model belongs to a branch of linear regression approach, leveraging the regression idea to solve the classification problem algorithm, usually used for binary classification problems. The basic idea of it to use the linear function value obtained from the logistic regression model brought from the sigmoid function for transformation, and then the value obtained from the transformation is compared with the threshold, and the attribute markers are finally obtained. The linear regression model predicts the target variable with a linearly weighted combination of characteristic attributes, and the weight of each attribute reflects the importance of the attribute to the prediction result. Logistic regression has the advantages of simplicity and ease of use, fast computation, and high interpretability. However, logistic regression models are very sensitive to outliers and are prone to overfitting or underfitting problems.

2.2.6. Neural network. A deep neural network is a multi-layer unsupervised neural network. Deep neural networks use supervised learning to train each layer, where a layer takes the output of its previous layer for learning, and then using supervised learning at each layer to fine-tune each layer by adding a classifier for classification [9]. After feature mapping through each layer, the features of the original sample space are mapped to another feature space. Deep neural networks have high accuracy and are more capable of learning. In addition, deep neural network models are less susceptible to noise interference? However, deep neural networks have the disadvantage that they cannot reflect the features extracted from each layer and the results are not interpretable [10].

3. Result and discussion

3.1. Experimental setup

This work sets up a deep neural network with an input layer containing 50 nodes, choosing tanh as the activation function, containing a hidden layer containing 20 nodes, choosing relu as the activation function, and an output layer containing 1 node, choosing sigmoid as the activation function, while choosing the optimiser function as adam and the loss function set to binary crossentropy neural network.

The experimental results show that as the training processed, the overall training loss tends to decrease, and the validating loss tends to be flat, and the neural network may have overfitting problems. To reduce overfitting, a Dropout layer is added with a dropout rate at 0.25.

3.2. Visualization analysis

Visualizing the data through the heat map of correlation coefficients, it could be observed that the characteristic relationships with relatively large linear correlations are as follows: Grade (glioma category) has a linear correlation with age at diagnosis of 0.53, Grade (glioma category) has a linear correlation with IDH1 (isocitrate dehydrogenase) of -0.71. age at diagnosis has a linear correlation with IDH1 (isocitrate dehydrogenase) of -0.57, ATRX and TP53 have a linear correlation of 0.55. diagnosis had a linear correlation of -0.57 with IDH1 (isocitrate dehydrogenase) and ATRX and TP53 had a linear correlation of 0.55. Grade (glioma category) had the highest linear correlation of -0.71 with IDH1, so results show that IDH1 may become an important variable in classifying gliomas.

DOI: 10.54254/2755-2721/31/20230173

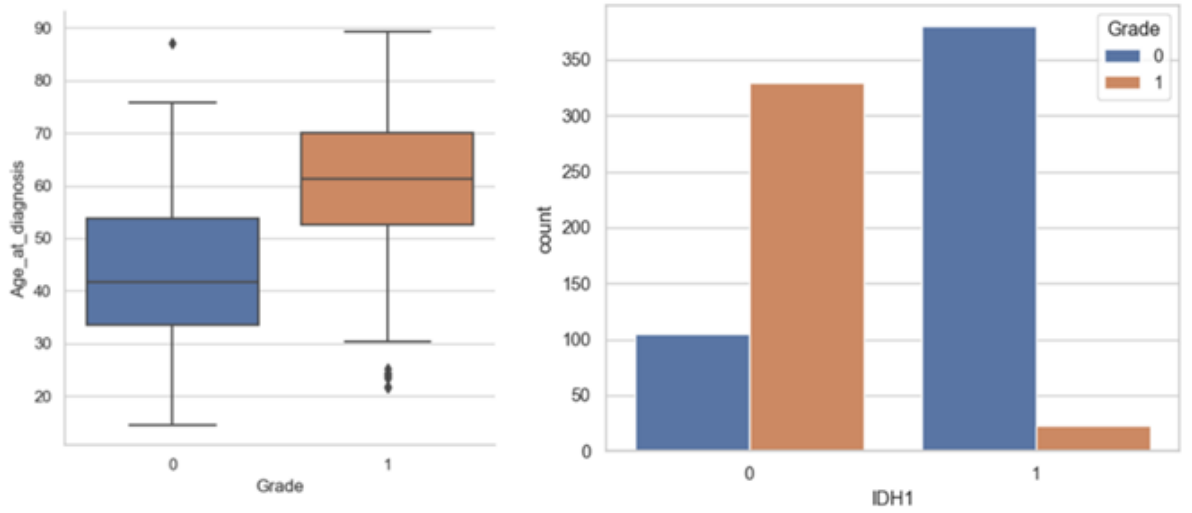

Figure 1. Relationship between Age_at_diagnosis and glioma type (left); isocitrate dehydrogenase and glioma type (right) (Picture credit: Original).

Figure 1 shows that patients diagnosed with low-grade glioma (LGG) are in their 30s and 40s, while patients diagnosed with high-grade glioma (HGG) are in their 60s and 65s. The likelihood of having a high-grade glioma increases with age. In Figure 1 the presence of mutations in IDH1 is significantly associated with the type of glioma, and mutations in IDH1 may lead to an increased risk of LGG.

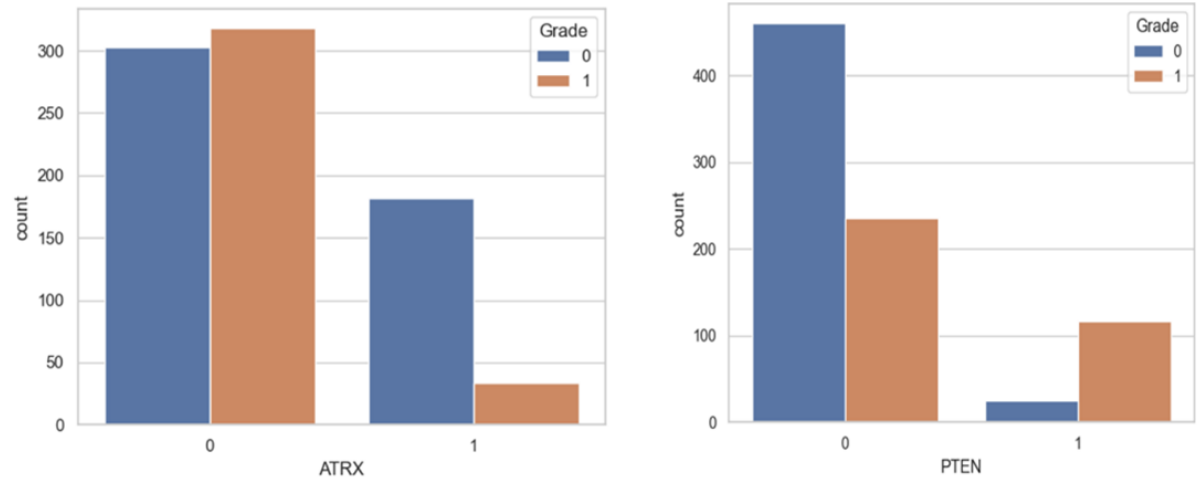

Figure 2. Relationship between chromatin remodelling protein (ATRX) and glioma type (left); homologous phosphatase-tensin (PTEN) and glioma type (right) (Picture credit: Original).

Figure 2 demonstrates that mutations in ATRX were obviously greater in patients with HGG than in those with LGG. There are strong correlations between PTEN and the type of glioma. A significantly greater proportion of patients with HGG had PTEN mutations than those with LGG.

DOI: 10.54254/2755-2721/31/20230173

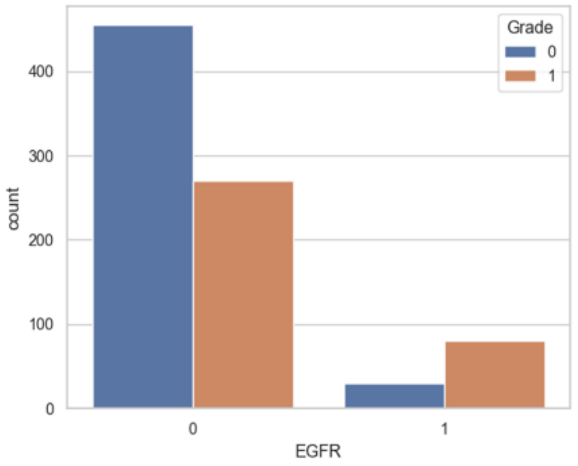

Figure 3. Relationship between epidermal growth factor receptor and glioma type (Picture credit: Original).

There was a significant association between the presence of EGFR mutations and the type of glioma suffered as demonstrated in Figure 3. The proportion of patients with EGFR mutations was greater in patients with HGG than in patients with LGG.

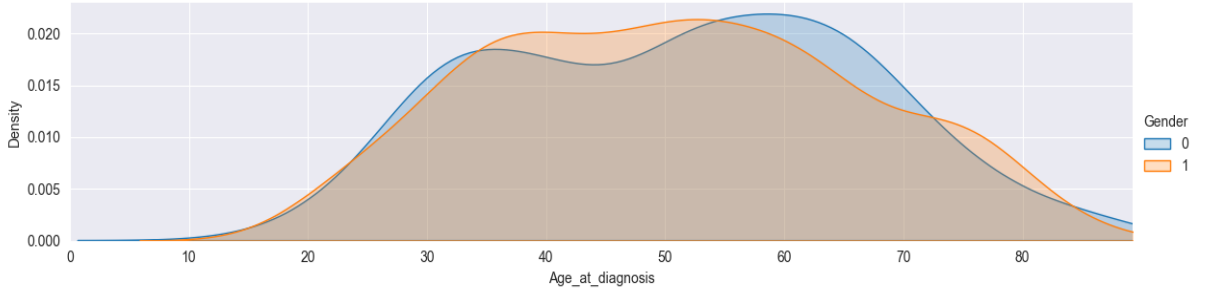

Figure 4. Relationship between Age at diagnosis and the patients' gender (Picture credit: Original).

As can be seen from Figure 4, more patients are female around the age of 35-55, more patients are male around the age of 55-73 and more patients are female than male around the age of 73-85.

Overall, glioma patients are concentrated between the ages of 30-40 and 50-65 years; more patients are male; more patients are diagnosed with LGG, and more patients are diagnosed with glioma without mutations in IDH1 and ATRX.

3.3. Performance comparison

Comparing the performance of each model as shown in Figure 1, it can be found that each value of the deep neural network model is higher than the corresponding value of the other models. Therefore, the deep neural network is more effective than the other models. the precision of LGG class of deep neural network reaches 0.91, the f1 value of LGG class reaches 0.90, the accuracy reaches 0.89, and the auc value reaches 0.89, which indicates that the deep neural network network has a good accuracy and model generalization ability. Secondly, just considering the model accuracy and AUC value, Adaboost model and xgboost model also have good performance. The model accuracy and auc value of adaboost model is 0.87, and the model accuracy and auc value of xgboost model is 0.86. The model accuracy of logistic regression model is 0.87, but its auc value is only 0.84, which is not as good as the model generalization ability of dnn model, xgboost model and adaboost model. It is worth noting that the accuracy of the random forest model is significantly lower than that of the other models, only 0.80, and the predictive effect of this modality is not good (Table 1).

DOI: 10.54254/2755-2721/31/20230173

Table 1. Performance comparison.

Methods	Pre -0	Pre -1	Rec -0	Rec -1	F1 -0	F1 -1	ACC	AUC
XGboost	0.88	0.83	0.86	0.86	0.87	0.84	0.86	0.86
Adaboost	0.88	0.83	0.89	0.85	0.88	0.85	0.87	0.87
Gradientboost	0.86	0.83	0.86	0.83	0.86	0.83	0.85	0.84
RandomForest	0.79	0.81	0.86	0.73	0.83	0.77	0.80	0.84
LogisticRegression	0.90	0.85	0.87	0.88	0.88	0.86	0.87	0.84
DNN	0.91	0.87	0.89	0.89	0.90	0.88	0.89	0.89

The change in loss values and accuracy of the deep neural network model is demonstrated in Figure 5. It can be seen that the training set loss values and the test and loss values show an overall trend of convergence.

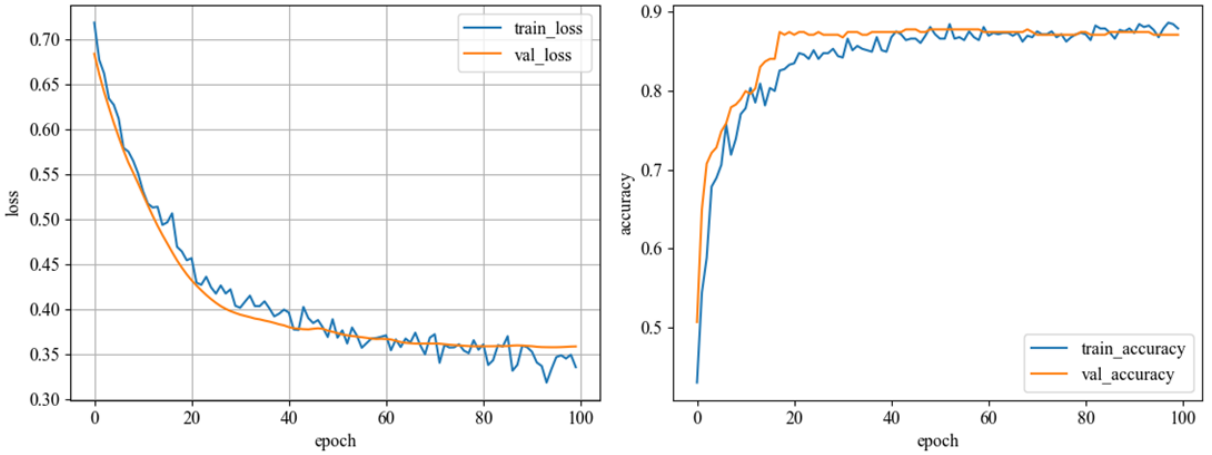

Figure 5. Loss curve and accuracy (Picture credit: Original).

This work used logistic regression feature screening and feature screening based on decision tree information gain to screen 22 attributes, and screened 12 attributes in order of importance, respectively. The attributes screened by the two methods were approximately the same, with a few discrepancies. Both feature screening methods showed that IDH1 and age at diagnosis were the attributes with the highest importance and were significantly more important than the other attributes. Apart from IDH1 and age at confirmation, other attributes with higher importance were IDH2, CIC and ATRX (Table 2).

Table 2. Feature importance from various models.

Attribute	Importance	Attribute	Importance
IDH1	3.353	Age_at_diagnosis	0.368
Age_at_diagnosis	2.267	IDH1	0.297
IDH2	1.812	ATRX	0.044
NOTCH1	1.230	CIC	0.043
GRIN2A	1.161	PTEN	0.042
CIC	1.143	IDH2	0.026
CSMD3	0.835	TP53	0.024
ATRX	0.687	Gender	0.023
EGFR	0.023	EGFR	0.023
NF1	0.017	MUC16	0.017
TP53	0.015	RB1	0.015
RB1	0.014	NF1	0.014

DOI: 10.54254/2755-2721/31/20230173

The model accuracy and auc values before and after feature screening are shown in Table 3 below. From the data in the table, it can be found that feature screening has almost no positive effect on the model performance improvement, and even the model performance will be decreased. After feature screening, the auc value of the logistic regression model increases from 0.84 to 0.87, and the generalization ability of the model is improved. After the feature screening using logistic regression, the model accuracy and auc value of the deep neural network model have improved and reached 0.90 respectively, which is the best result among all the models.

Table 3. Result comparison with machine learning models with DNN.

	No feature screening		Logistic regression feature screening		Decision tree feature screening	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
XGboost	0.86	0.86	0.87	0.87	0.86	0.86
Adaboost	0.87	0.87	0.86	0.86	0.86	0.86
Gradientboost	0.85	0.84	0.84	0.84	0.83	0.82
RandomForest	0.80	0.84	0.80	0.79	0.81	0.81
LogisticRegression	0.87	0.84	0.87	0.87	0.87	0.87
DNN	0.89	0.89	0.90	0.90	0.89	0.89

4. Conclusion

The performance of different models was compared in this project, and the DNN model with logistic regression feature screening was finally adopted for the glioma grading study. The accuracy of the grading study was effectively improved by preprocessing. By using multiple integrated learning models, the model with the highest accuracy was better selected and trained. The accuracy of the glioma grading study was verified by using DNN models with publicly available datasets. The test results show that the DNN model algorithm for logistic regression feature screening can accurately perform grading studies of gliomas.

References

[1] Miklja, Z., Pasternak, A., Stallard, S., Nicolaides, T., Kline-Nunnally, C., Cole, et al. (2019). Molecular profiling and targeted therapy in pediatric gliomas: review and consensus recommendations. Neuro-oncology, 21(8), 968-980.

[2] Youssef, G., & Miller, J. J. (2020). Lower grade gliomas. Current neurology and neuroscience reports, 20, 1-9.

[3] Komori, T. (2022). Grading of adult diffuse gliomas according to the 2021 WHO Classification of Tumors of the Central Nervous System. Laboratory Investigation, 102(2), 126-133.

[4] Mair, M. J., Geurts, M., van den Bent, M. J., & Berghoff, A. S. (2021). A basic review on systemic treatment options in WHO grade II-III gliomas. Cancer treatment reviews, 92, 102124.

[5] Zhuge, Y., Ning, H., Mathen, P., Cheng, J. Y., Krauze, A. V., Camphausen, K., & Miller, R. W. (2020). Automated glioma grading on conventional MRI images using deep convolutional neural networks. Medical physics, 47(7), 3044-3053.

[6] Montavon, G., Samek, W., & Müller, K. R. (2018). Methods for interpreting and understanding deep neural networks. Digital signal processing, 73, 1-15.

[7] Wang, Z., Jensen, M. A., & Zenklusen, J. C. (2016). A practical guide to the cancer genome atlas (TCGA). Statistical Genomics: Methods and Protocols, 111-141.

[8] Azmi, S. S., & Baliga, S. (2020). An overview of boosting decision tree algorithms utilizing AdaBoost and XGBoost boosting strategies. Int. Res. J. Eng. Technol, 7(5), 6867-6870.

[9] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

[10] Mathew, A., Amudha, P., & Sivakumari, S. (2021). Deep learning techniques: an overview. Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020, 599-608.