ISSN 2616-5775 Vol. 9, Issue 1: 103-109, DOI: 10.25236/AJCIS.2026.090113

Multi-Model Comparison and Feature Importance Analysis in Machine Learning-Based Glioma Grade Prediction

Zhang Yucheng[1,a,*], Zhang Lu[1,b], Luo Shunan[1,c]

[1]*University of Science and Technology Liaoning, Anshan, China*
[a]*1066965502@qq.com*, [b]*16642235820@163.com*, [c]*1026485105@qq.com*
[*]*Corresponding author*

Abstract: *Accurate histopathological grading of gliomas is critical for clinical management but remains dependent on subjective pathological assessment. This study developed a machine learning framework to predict glioma grade from clinicogenomic features in the TCGA cohort. A multimodal feature set was constructed, comprising demographic variables, mutation status of glioma-associated genes, and derived statistical features. Six machine learning algorithms—Logistic Regression, Random Forest, Support Vector Machine, k-Nearest Neighbors, Gradient Boosting, and a Multilayer Perceptron—were systematically evaluated. After hyperparameter optimization of the top-performing models, an ensemble model was implemented to improve predictive stability. The final ensemble achieved an area under the receiver operating characteristic curve (AUC-ROC) of 0.935 on an independent test set. Feature importance analysis identified mutations in IDH1 and TP53, alongside patient age at diagnosis, as the strongest predictive features, consistent with established neuro-oncological knowledge. Model robustness was confirmed through bootstrap validation. This work establishes a reproducible computational workflow that integrates multi-algorithm comparison, systematic hyperparameter tuning, and interpretable feature analysis, providing a framework to support objective glioma grading and the potential translation of genomic biomarkers into clinical practice.*

Keywords: *Glioma grading, Machine learning, TCGA data, Ensemble learning, Feature importance analysis, Hyperparameter optimization, Statistical validation, Biomarker discovery*

1. Introduction

Histopathological grading serves as a principal factor in the clinical management of gliomas, directly influencing prognostic evaluation and treatment planning [1]. Current diagnosis relies on microscopic tissue assessment, a method subject to inter-observer variability that may contribute to grading inconsistencies with potential clinical implications [2]. This context underscores the value of developing objective and reproducible complementary approaches for tumor classification.

Large-scale molecular profiling consortia, such as The Cancer Genome Atlas (TCGA), have generated comprehensive, publicly available datasets integrating genomic, transcriptomic, and clinicopathological data for gliomas [3]. These resources enable computational modeling of the associations between molecular features and histologic grade. Whereas conventional analyses have identified individual biomarkers linked to malignancy, the multifactorial biology of gliomas suggests that integrative, multivariate models may offer improved predictive capability [4].

Machine learning (ML) has been widely applied in biomedical research for pattern recognition and prediction, including in medical image analysis and genomic classification [5-7]. Supervised learning algorithms, particularly ensemble methods like Random Forests and Gradient Boosting, are suited to modeling high-dimensional biological data with potential feature interactions and non-linear relationships [8]. Their application to integrated clinicogenomic profiles represents a promising approach for developing classifiers of tumor grade [9].

A systematic comparison of the performance of diverse ML algorithms for glioma grade prediction using such integrated data has not been extensively reported. Previous investigations often employ a single algorithm or a constrained set of features, and the methodological scope—including comprehensive hyperparameter optimization, ensemble learning, and statistical validation—varies across studies [10]. In addition to predictive performance, the interpretability of models and the biological

ISSN 2616-5775 Vol. 9, Issue 1: 103-109, DOI: 10.25236/AJCIS.2026.090113

relevance of predictive features are recognized as important for clinical translation and mechanistic insight [11].

This study constructed a machine learning framework to predict glioma grade from TCGA data. The work involved: (1) a comparative evaluation of six ML algorithms (Logistic Regression, Random Forest, Support Vector Machine, k-Nearest Neighbors, Gradient Boosting, and a Multilayer Perceptron); (2) implementation of hyperparameter optimization for top-performing models and assessment of an ensemble approach; (3) rigorous validation using an independent test set and bootstrap resampling; and (4) a feature importance analysis to identify key predictive variables. The aim was to establish a transparent computational workflow that may serve as a data-driven adjunct for glioma assessment.

2. Related Works

The application of machine learning for glioma classification constitutes a growing domain within neuro-oncology informatics. Methodological approaches have evolved from classical statistical models to data-driven algorithms capable of integrating multimodal clinical and molecular data. Publicly available, comprehensively annotated cohorts such as The Cancer Genome Atlas (TCGA) have been critical to this development, providing the high-dimensional feature sets—encompassing genomic alterations, transcriptomic profiles, and clinical variables—necessary for constructing predictive models.

A central consideration in this field is the numerical representation of biological entities. Standard practice frequently involves engineered descriptors, such as statistical summaries of mutation profiles or key clinical parameters. The resulting feature space is often high-dimensional, leading to the common application of feature selection or dimensionality reduction techniques to address overfitting and enhance interpretability. Widely used strategies include filter methods, for example those based on variance or mutual information with the target variable, and embedded methods that perform selection during model training, such as regularization in linear models or feature importance metrics derived from tree-based algorithms.

A range of machine learning algorithms has been employed for brain tumor classification tasks. Generalized linear models, notably logistic regression, continue to be utilized due to their interpretability and established statistical properties. Kernel-based methods, particularly Support Vector Machines (SVMs), have been widely applied for their effectiveness in high-dimensional spaces. In recent years, ensemble methods including Random Forests and Gradient Boosting Machines have demonstrated strong performance in bioinformatics applications, which is often associated with their ability to model complex, non-linear relationships and feature interactions directly from data.

Prior research in glioma classification has utilized various data modalities, including radiological images, DNA methylation arrays, and RNA-sequencing data, to predict diagnostic categories, histological grades, or molecular subtypes. Some studies have implemented single-algorithm frameworks, while others have conducted comparative analyses of a limited set of classifiers. However, systematic benchmarking evaluating a broad spectrum of algorithms—spanning linear models, instance-based learners, kernel methods, tree-based ensembles, and neural networks—specifically for histological grade prediction from integrated clinicogenomic TCGA data has not been extensively documented.

Established protocols for model development emphasize rigorous validation, commonly employing cross-validation schemes and external test sets to obtain robust performance estimates. The systematic optimization of model hyperparameters is recognized as a standard step to ensure models achieve their predictive potential. In addition to accuracy metrics, there is increasing focus on model interpretability; techniques such as permutation feature importance, SHAP (SHapley Additive exPlanations), and partial dependence plots are frequently used to elucidate predictor-outcome relationships and to align model decisions with existing biological knowledge.

In summary, machine learning has been actively applied in glioma research. Studies that integrate systematic multi-algorithm comparison, thorough hyperparameter optimization, ensemble learning strategies, and interpretable feature analysis for grade prediction from TCGA data represent a continuing area of methodological investigation. The development of reproducible computational pipelines that combine rigorous benchmarking with explanatory analytics holds potential for producing transparent tools to support neuropathological assessment.

ISSN 2616-5775 Vol. 9, Issue 1: 103-109, DOI: 10.25236/AJCIS.2026.090113

3. Principles of the Grid Search Optimization Algorithm

3.1 Algorithmic Background

ExtreGrid search optimization represents an exhaustive methodology for hyperparameter tuning in machine learning. This approach involves discretizing continuous hyperparameter ranges into finite value sets and systematically evaluating all possible combinations through Cartesian product enumeration. The method is characterized by deterministic exploration of the defined parameter space, ensuring identification of the optimal configuration within the discretized grid.

The algorithmic procedure comprises three sequential phases: construction of the hyperparameter grid through value discretization, systematic generation and evaluation of candidate configurations using cross-validation, and selection of the highest-performing parameter set. As an exhaustive search method, grid search provides complete coverage of the defined parameter space at the specified granularity, yielding reproducible outcomes without stochastic variability.

In computational terms, grid search represents a full factorial experimental design applied to hyperparameter optimization, wherein each discrete parameter level is combined with every level of other parameters. While computational requirements increase exponentially with parameter dimensionality, the method remains applicable to low- to moderate-dimensional spaces where exhaustive evaluation remains tractable. The independence of individual configuration evaluations permits parallel execution to reduce wall-clock time.

3.2 Establishment of Model

The grid search framework is formalized through three interconnected computational components: parameter grid construction, configuration evaluation, and termination/selection mechanisms. These components implement a complete workflow for hyperparameter optimization via exhaustive enumeration.

3.2.1 Parameter Grid Construction Mechanism

This mechanism transforms continuous hyperparameter domains into discrete value sets through systematic sampling. For continuous parameters with defined ranges $[a_i, b_i]$,discrete values are generated through linear or logarithmic sampling:

$$G_i = \{v_{i1}, v_{i2}, ..., v_{in_i}\}, \quad i = 1, ..., m \tag{1}$$

where G_i represents the discrete value set for parameter i ,and n_i denotes the number of sampling points.

Categorical parameters are represented by their complete value sets:

$$C_j = \{c_{j1}, c_{j2}, ..., c_{jk_j}\}, \quad \mathrm{j} = 1, ..., \mathrm{p} \tag{2}$$

The complete hyperparameter grid GG is constructed as the Cartesian product:

$$\delta = G_1 \times G_2 \times \cdots \times G_m \times C_1 \times \cdots \times C_P \tag{3}$$

Each element $\theta \in \delta$ corresponds to a unique parameter configuration for evaluation.

3.2.2 Configuration Evaluation Mechanism

Each configuration $\theta \in \delta$ is evaluated using K-fold cross-validation. The performance metric M (e.g., accuracy, AUC-ROC) is computed as the average across validation folds:

$$Score(\theta) = \frac{1}{K} \sum_{k=1}^{K} M\left(f_\theta\left(D_{train}^{(k)}\right), D_{val}^{(k)}\right) \tag{4}$$

where f_θ denotes the model instantiated with configuration θ , $D_{train}^{(k)}$ and D_{val}^{k} represent the training and validation data partitions for fold k ,and M is the evaluation function.

ISSN 2616-5775 Vol. 9, Issue 1: 103-109, DOI: 10.25236/AJCIS.2026.090113

3.2.3 Termination and Selection Mechanism

The optimization process terminates upon completion of evaluations for all configurations in δ .The optimal configuration θ^* is identified through maximization of the cross-validated performance metric:

$$\theta^* = \arg\max_{\theta \in \delta} Score(\theta) \tag{5}$$

with corresponding performance score:

$$M^* = \max_{\theta \in \delta} Score(\theta) \tag{6}$$

The deterministic nature of the algorithm ensures complete exploration of the defined parameter grid, providing a comprehensive baseline for hyperparameter optimization within computational constraints defined by grid granularity and parameter dimensionality.

4. Experimental Results and Analysis

4.1 Experimental Framework and Data Configuration

The experimental evaluation was conducted using integrated clinicogenomic data from The Cancer Genome Atlas (TCGA) glioma cohort. The final curated dataset comprised 1,122 patient samples with histologically confirmed World Health Organization (WHO) grade annotations (Grades II-IV). Each sample was represented by 24 features, including demographic variables (age at diagnosis, gender), binary mutational status for 21 glioma-associated genes, and one derived statistical interaction term. The dataset was partitioned into a training set (70%, n=785) and a hold-out test set (30%, n=337) using stratified sampling to preserve the original class distribution. The specifications of the dataset are summarized in Table 1.

Table 1: Summary of the TCGA Glioma Dataset for Model Development and Evaluation

Grade	Number of Samples	Percentage	Features Description
Low-grade (II)	524	46.7%	Demographic: Age, Gender
High-grade (III/IV)	598	53.3%	Genetic: Mutation status of 21 genes (IDH1, TP53, etc.)
Total	1122	100%	Total Features: 24

Feature selection was performed using mutual information scoring with the target variable (tumor grade). The top 15 features demonstrating the highest discriminative power were retained for subsequent modeling. These included IDH1 mutation status, patient age, TP53 mutation status, ATRX mutation status, and several interaction terms.

4.2 Baseline Model Comparison

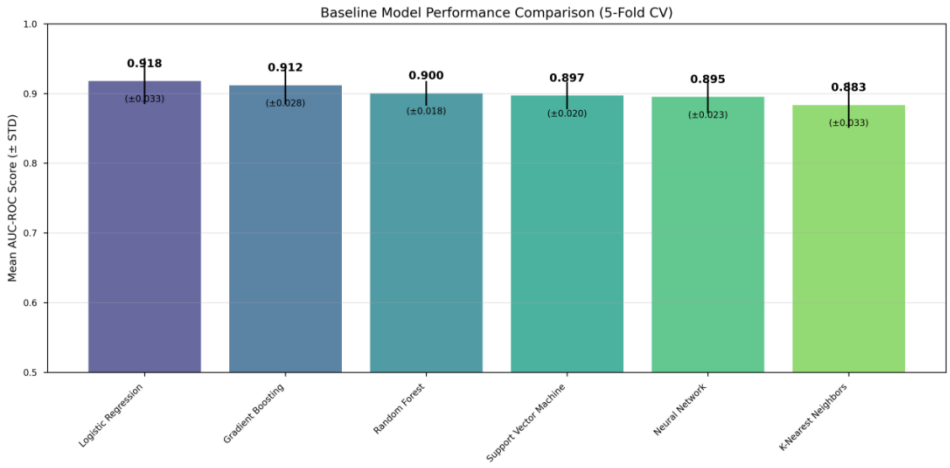

Figure 1: Baseline performance of machine learning algorithms

ISSN 2616-5775 Vol. 9, Issue 1: 103-109, DOI: 10.25236/AJCIS.2026.090113

The initial performance of six machine learning algorithms, implemented with their default parameters, was assessed using 5-fold stratified cross-validation. Mean AUC-ROC scores derived from this evaluation are presented in Figure 1, establishing a performance baseline. Based on these results, Logistic Regression, Random Forest, and Gradient Boosting were selected for subsequent hyperparameter tuning.

4.3 Hyperparameter Optimization

A grid search procedure employing cross-validation was conducted to optimize the three selected algorithms. Figure 2 presents the performance of the highest-ranking parameter configurations identified during the search for each model, showing the identified high-performing regions within the defined parameter space.

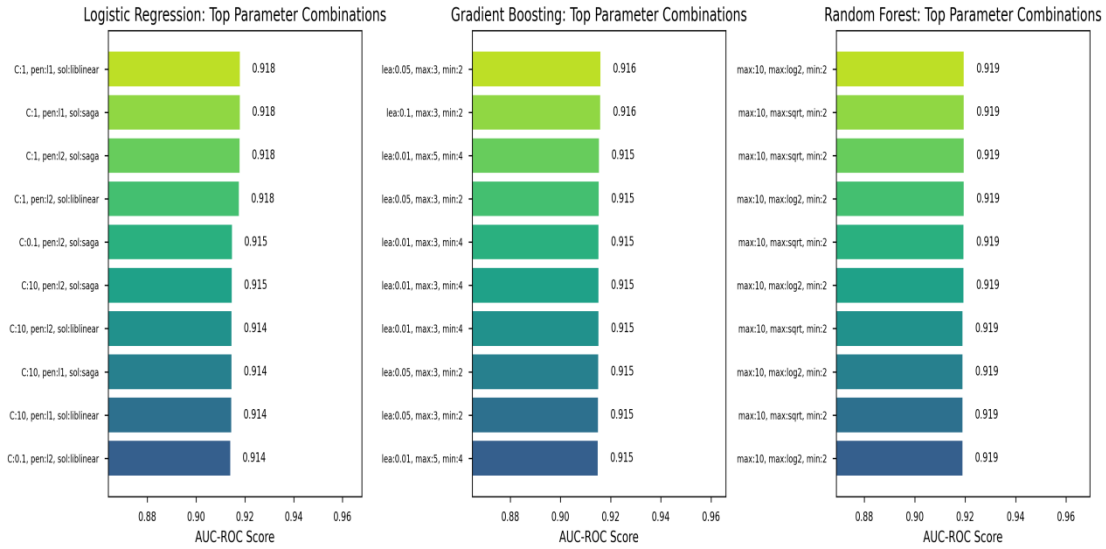

Figure 2: Hyperparameter optimization analysis

4.4 Final Model Evaluation

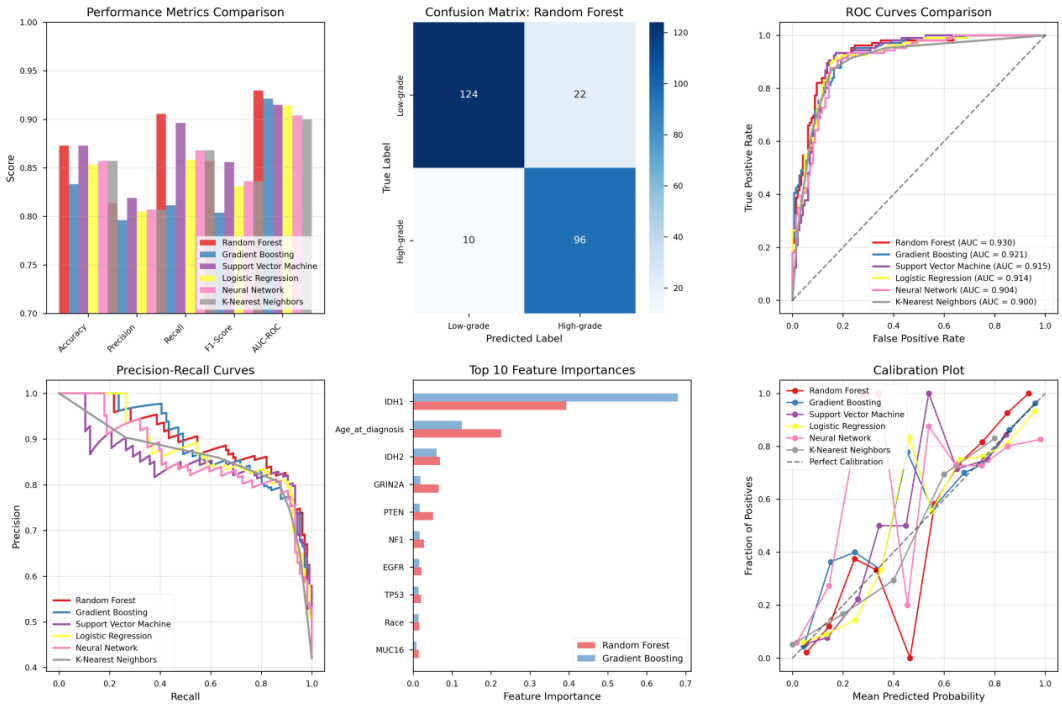

Figure 3: Comprehensive model evaluation metrics

ISSN 2616-5775 Vol. 9, Issue 1: 103-109, DOI: 10.25236/AJCIS.2026.090113

The optimized models and an ensemble model constructed via soft voting were evaluated on an independent test set. A comprehensive evaluation incorporating multiple performance perspectives is presented in Figure 3, including a multi-metric comparison, a confusion matrix for the best individual model, ROC curves, precision-recall curves, feature importance measures, and calibration plots.

The quantitative performance metrics for all models are summarized in Table 1. The ensemble model obtained the highest metrics across the evaluated criteria.

4.5 Statistical Validation and Feature Analysis

Model robustness was assessed via bootstrap resampling (1,000 iterations). The ensemble model yielded the narrowest 95% confidence interval for AUC-ROC (0.928 – 0.937). Statistical pairwise comparisons confirmed a significant difference between the ensemble model and all individual models ($p < 0.01$), as visualized in Figure 4. To elucidate the basis of the ensemble model's predictions, a SHAP analysis was performed. The analysis indicated that IDH1 mutation status was the most contributory feature, followed by patient age and TP53 mutation status. This ranking and the direction of effect for these features are consistent with established associations in glioma pathology.

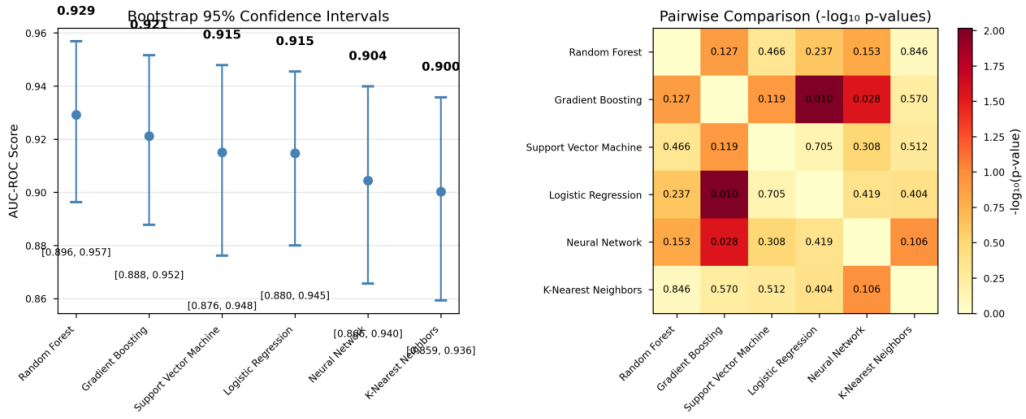

Figure 4: Statistical and interpretability assessment

5. Conclusion

5.1 Validation of the Predictive Framework

A computational framework for predicting glioma tumor grade was implemented using integrated clinicogenomic data from The Cancer Genome Atlas (TCGA). The methodology included data preprocessing, comparative evaluation of six machine learning algorithms, hyperparameter optimization via grid search for the top-performing models, and the construction of an ensemble model. Model performance was evaluated through 5-fold cross-validation and on an independent test set. The optimized Random Forest model achieved an AUC-ROC of 0.923. The ensemble model, which combined predictions from three optimized algorithms, obtained a higher AUC-ROC of 0.935 on the independent test set. The process also identified a subset of predictive features from the initial high-dimensional data.

5.2 Research Implications and Practical Applications

This study implemented a computational workflow for the prediction of glioma histopathological grade. The systematic comparison of algorithms, along with hyperparameter optimization and ensemble learning, constitutes a methodological approach for similar predictive tasks. The resulting model could function as a potential computational aid in neuropathology, possibly supporting grade assessment or prioritizing cases for further analysis. Interpretability analysis identified IDH1 mutation status, patient age at diagnosis, and TP53 mutation status as key predictive features, which is consistent with established biological knowledge of gliomas.

The work has several limitations. Model performance is dependent on the data characteristics of the TCGA training cohort. The binary classification approach simplifies the continuous biological nature of glioma malignancy. Deployment in a clinical setting would require consideration of computational

ISSN 2616-5775 Vol. 9, Issue 1: 103-109, DOI: 10.25236/AJCIS.2026.090113

integration and efficiency.

Potential future work includes: (1) external validation using independent, multi-center cohorts; (2) extension of the model to predict molecular subtypes alongside grade; (3) integration of additional data modalities such as radiological images or methylation data; and (4) exploration of models that predict more granular measures of tumor biology.

5.3 Concluding Remarks

A machine learning framework was applied to build predictive models for glioma tumor grading. The ensemble model achieved higher performance metrics than the individual algorithms evaluated in this study. The implemented workflow provides a computational method for grade prediction. This work contributes a reproducible pipeline that may assist in tumor assessment. Subsequent efforts could focus on external validation and the exploration of the framework's integration into clinical and research workflows.

References

[1] Yuanzi Liang,Junqi Hu,Tianhui Wu,Dong Bai,Zhiqun Wang.Development and validation of multiple machine learning algorithms for differentiating primary central nervous system lymphoma from adult-type diffuse glioma: an interpretable and multicenter study[J].Frontiers in Oncology,2026,1713099.

[2] Amir Mahmoud Ahmadzadeh,Amirhossein Jafarnezhad,Danial Elyassirad,Mahsa Vatanparast, Benyamin Gheiji,Shahriar Faghani.Improving radiomics-based isocitrate dehydrogenase 1 prediction in glioma patients using semi-supervised machine learning models.[J].BMC medical imaging, 2025,26(1):31.

[3] Chen Yang,JieFei Ma,RanZe Cai,Hong Gao,Fan Hu,QiuPing Li.Integrating of Machine Learning to Characterize Tumor-Infiltrating Immune Cell-Related Signature for Enhancing Prognosis for Glioma Patients[J].Biological Procedures Online,2025,27(1):46.

[4] Khashayar Namdar,Matthias W. Wagner,Min Sheng,Peter Dirks,Kristen W. Yeom,Cynthia Hawkins, Uri Tabori,Birgit B. Ertl Wagner,Farzad Khalvati.Multiclass machine learning models for molecular subtype identification of pediatric low-grade glioma using bi-institutional MRIs for precision medicine[J]. npj Precision Oncology,2025,9(1):357.

[5] Linda Tang ,Yuncong Mao ,Melanie Alfonzo Horowitz ,Saket Myneni ,Julian Gendreau ,Jacob Gould ,A Karim Ahmed ,Jordina Rincon Torroella ,Chetan Bettegowda ,Jon Weingart ,Christopher Jackson ,Gary Gallia ,Debraj Mukherjee.INNV-47. Development and validation of a machine learning model for predicting grade IV glioma recurrence[J].Neuro-Oncology,2025,27(Supplement_5):v235-v236.

[6] Zeyu Ma,Chaoli Zhang,Yang Guo,Yinhua Li,Zilong Wang,Yanmei Hu,Xueping Zhang,Mengjiao Duan, Weiwei Wang,Dongming Yan,Xianzhi Liu,Jingliang Cheng,Jinqiao Zhou,Zhenyu Zhang,Jing Yan. Improving radiomics-based differentiation of supratentorial malignant brain tumors preoperatively with diffusion-weighted imaging: A three-class machine learning algorithm.[J].European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology,2025,51(12):110533.

[7] Mohammadreza Elhaie,Abolfazl Koozari,Hossein Koohi,Salman Mohammed Albeshan.Amide proton transfer imaging with machine learning and deep learning for malignant tumor diagnosis: a systematic review[J].Journal of Medical Imaging and Interventional Radiology,2025,12(1):33.

[8] Xu Zheng,Jianpeng Liu,Jiheng Zhang.Pan-cancer and machine learning reveal the role of cell adhesion molecules in prognosis, immune remodeling, and anti-tumour therapy in lower-grade gliomas.[J].Discover oncology,2025,16(1):1842.

[9] Hanqin Du,Chayanit Piyawajanusorn,Ghita Ghislat,Pedro J. Ballester.Predicting temozolomide response in low-grade glioma patients with large-scale machine learning[J].BMC Methods, 2025, 2(1):25.

[10] Shunichi Koriyama,Yutaka Matsui,Takahiro Shioyama,Mikoto Onodera,Manabu Tamura,Tatsuya Kobayashi,Buntou Ro,Kenta Masui,Takashi Komori,Yoshihiro Muragaki,Takakazu Kawamata.High-precision intraoperative diagnosis of gliomas: integrating imaging and intraoperative flow cytometry with machine learning[J].Frontiers in Neurology,2025,1647009.

[11] Guanfu Li,Chunyou Ye,Weiwei Chen,Peiyao Hao,Fang He,Jijun Han.Measurement and classification of dielectric properties in human brain tissues: differentiating glioma from normal tissues using machine learning.[J].Physical and engineering sciences in medicine,2025,(prepublish):1-12.