2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA)

Research on Glioma Grading Auxiliary Diagnosis Method Based on Feature Selection and Classifier Fusion

Zunpeng Yu
China Academy of Industrial Internet
Beijing, China
yuzunpeng@china-aii.com

Jiehao Chen
China Academy of Industrial Internet
Beijing, China
* Corresponding author: chenjiehao@china-aii.com

***Abstract*—This paper investigates the feasibility of brain glioma grading based on gene expression data, proposes a modelling method for assisted diagnostic decision-making for brain glioma grading, and constructs an assisted grading diagnostic decision-making model, which assists doctors and other healthcare professionals in clinical diagnosis and treatment decisions. Three feature selection algorithms—RFE (Recursive Feature Elimination), Boruta, and SHAP (Shapley Additive Explanations)—were employed to identify important genes related to glioma grading, addressing the high-dimensionality issue of gene expression data. Subsequently, ten traditional machine learning classifiers, including SVM (Support Vector Machine), RF (Random Forest), and KNN (K-Nearest Neighbors), are used to build the grading classification decision models. Finally, the biological functions associated with the selected feature genes are analyzed, revealing biological processes related to glioma grade. The identified feature genes were then translated into medical knowledge.**

Keywords- Gene Expression Data; Machine Learning; Feature Selection; Data Mining; Knowledge Discovery

I. INTRODUCTION

With the continuous development and maturity of artificial intelligence technology, its integration with the medical field has become increasingly close. In recent years, intelligent medical data analysis has become a research hotspot [1, 2]. In the theoretical research on decision management and knowledge services based on medical big data, Ye et al. [3] studied the process of "data-information-knowledge" in big data knowledge management, pointing out that knowledge management based on big data can be used to assist decision-making and management services. Ma et al. [4] expanded the theories and methods of knowledge management and knowledge services in the information resource management field to smart healthcare, conducting a foundational and theoretical exploration of knowledge management based on big health data in the context of smart health. Wu et al. [5] introduced the practical applications of data-driven smart healthcare services, which can use patient data for knowledge discovery and precision medicine management, helping experts identify disease-related information and knowledge. The clinical healthcare data can be used to provide personalized medical information services for patients, medical experts, and other users. The design and development of machine learning models can extract information from medical data, discover disease-related medical knowledge, and identify biomarkers related to diseases for auxiliary diagnosis and prediction of brain diseases. It has important theoretical and practical significance for improving the efficiency of clinical doctors in diagnosing diseases.

The accurate classification of gliomas grade is not only helpful for understanding the malignancy and prognosis of the disease but also plays a crucial role in formulating subsequent treatment plans. Different grades of gliomas may require different treatment approaches or dosages, enabling more precise treatment and achieving better outcomes. Additionally, the identified feature genes can be applied in clinical diagnosis as important biomarkers for future grading and diagnosis of patients, providing a basis for doctors' decision-making. These genes can also undergo further research as potential drug development and treatment targets. However, current auxiliary diagnostic methods are primarily based on medical imaging data [6, 7]. In contrast, genetic data can offer a deeper understanding of the disease's pathogenesis, providing doctors with more in-depth medical information and knowledge.

II. FRAMEWORK FOR DIAGNOSTIC DECISION SUPPORT METHODS

Medical data, especially genomic data, are characterized by high dimensionality and information redundancy. One of the challenges doctors face in the diagnostic and treatment decision-making process is how to select relevant information from a vast amount of data related to the disease. Feature extraction and redundancy removal from high-dimensional medical genomic data are crucial for improving the robustness of classification models and reducing model and computational complexity. This is also a key issue and goal in developing decision-support models. Deep learning models are characterized by complex parameters, poor interpretability, large amounts of data required, and high computational resource requirements [8, 9]. This paper tries to build a framework for assisted diagnostic modeling using traditional, easy-to-understand machine learning algorithms. Moreover, different feature extraction methods and classification algorithms have their own advantages, and they need to be combined and compared to select the most effective comprehensive decision-making model.

This study mainly explores the relationship between gene expression data and different grades of gliomas. Due to the high-dimensional and redundant nature of genomic data, the study aims to effectively address the high-dimensional issue of

979-8-3315-0976-7/25/$31.00 ©2025 IEEE

gene expression data and remove irrelevant, redundant information related to the decision-making process. This will enhance the classification ability of the algorithm model and transform genetic features into meaningful physiological processes related to glioma grades, which can be understood by doctors as medical knowledge. Additionally, through screening and identification of signature genes, the focus can be placed on testing these key genes, which helps save doctors' time for testing and evaluation, reduces examination costs for patients, and so on.

The model construction process mainly includes data preprocessing, feature selection, and classifier construction. The model framework is shown in Figure 1.

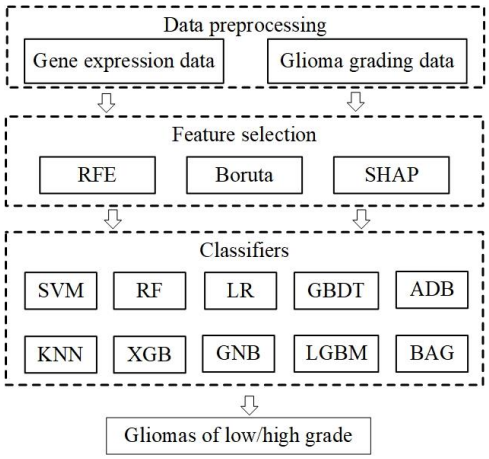

Figure 1. Glioma Grade Classification and Diagnosis Method Based on Gene Expression Data.

The gene expression data in this study was obtained from the raw data of microarray sequencing, and the gene expression data needs to be pre-processed. The fluorescence signal intensity contained in the fluorescence image was analyzed to form the intermediate data, which was further cleaned and screened, as well as other preprocessing to obtain the expression amount of each gene in each sample to create the gene expression matrix. Three different feature selection algorithms, RFE (Recursive Feature Elimination) [10], Boruta [11], and SHAP (SHapley Additive exPlanations) [12], were selected on the basis of the gene expression data matrix and were used to select feature genes from tens of thousands of genes for glioma grading diagnosis. Since the three feature selection algorithms need to be combined with specific classifiers during feature selection, RFE and Boruta were chosen to be combined with Random Forest, while SHAP was combined with XGBoost. Finally, based on the selected feature genes, ten different classifiers were chosen for glioma grade classification: Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), as well as ensemble learning classifiers including Light Gradient Boosting Machine (LGBM), XGBoost (XGB), AdaBoost (ADB), Gradient Boosting Decision Tree (GBDT), and Bagging (BAG). These classifiers were used to make classification decisions for glioma grading and output the decision results. During the model construction process, cross-validation and grid search were performed on the training data after feature filtering to optimize the parameters of some models, aiming to find the best parameter combination for model training.

III. Experimental Design

A. Dataset Collection

The gene expression data collected in this study come from public datasets in two open databases, TCGA and GEO. The TCGA database classifies gliomas into high-grade gliomas (GBM, Grade IV) and low-grade gliomas (LGG, Grade II and III). Since GBM and LGG gene expression data are obtained using different sequencing technologies, and due to technical and experimental factors, there are challenges in merging the datasets. Additionally, there are significant differences between sequencing chips from different manufacturers. So, in this paper, only sequencing data from similar chip platforms from the same manufacturer were downloaded to provide reliable results and minimize experimental and system biases.

Gene expression data were matched with clinical information based on patients' coding information. Based on the clinical information of each patient, datasets and samples with unclear pathological grading of brain tumors were excluded, while datasets and samples with clear pathological grading were retained. In the end, 15 datasets were obtained, and duplicate samples within these datasets were removed. A total of 1,902 samples were included, with 1,244 samples from high-grade gliomas and 658 samples from low-grade gliomas. Detailed demographic and clinical information for each dataset subset is shown in Table I.

TABLE I. Demographic and Clinical Details of Each Dataset Subset

Data set	Number of GBM samples	Number of LGG samples	Age(years) median (min~max)	Gender (M/F)
TCGA-GBM[13]	524	--	59 10~89	320/204
GSE15824[14]	15	13	45 13~70	23/5
GSE16011[15]	159	109	51.88 14.38~81.18	180/88
GSE19728[15]	5	10	64 18~80	12/3
GSE21354[16]	--	9	--	--
GSE24072[17]	24	7	65 30~76	18/13
GSE30336[18]	--	52	43 18~81	34/18
GSE36245[19]	46	--	18.5 0~48(Two samples with unknown age)	23/23
GSE4271[20]	56	21	45 22~82	52/25
GSE4290[21]	77	76	--	--
GSE4412[22]	50	24	42.5 18~82	28/46

GSE45921[23]	4	16	--	--
GSE53733[24]	70	--	--	--
GSE61374[25]	--	137	41 21~80	84/53
GSE68848[26]	134	184	--	--
GSE7696[27]	80	--	52.25 26.4~70.3	59/21
Total	1244	658		

B. Data Preprocessing

Gene expression data extraction, data filtering, and dataset merging were performed on raw data to obtain the final gene feature matrix data for all samples based on bioinformatics methods. In the data matrix, each column represents a gene, each row represents a sample, and the values indicate the expression levels of the genes, reflecting the gene expression intensity. In this section, a total of 1,902 gliomas patient samples were obtained, with each sample containing expression data for 12,458 genes, corresponding to a 12,458-dimensional feature space.

C. Experimental Setup

As mentioned above, we used three feature selection methods, including RFE and Boruta combined with Random Forest and SHAP combined with XGBoost, to select gene features. Then, we applied 10 different classifiers for grading classification decisions. A 5-fold cross-validation was used to perform a grid search for optimizing the classifier's hyperparameters. The experimental process is as follows:

First, the dataset was randomly divided into a training set (80%) and a test set (20%). Feature selection was then performed on the training set, selecting the top 100 feature genes each time. Cross-validation was conducted on the training set with only the selected feature genes to find the optimal hyperparameter combination for each classifier. The model parameters optimization settings are shown in Table II.

The classifier was trained on the training set based on the optimal hyperparameter combination and feature gene expression data. The goal of parameter selection is to improve the classifier's capability. Finally, the test set was used to evaluate the model's performance.

To comprehensively assess the stability of the overall framework, we repeated the above experiment 20 times for each model, with each run involving a random division of the training and test sets. The overall performance of the model was calculated as the average of the 20 runs.

TABLE II. HYPERPARAMETER SETTINGS FOR GRID SEARCH OPTIMIZATION OF EACH CLASSIFIER

Classifiers	Parameters for grid search
ADB	n_estimators：20，50，70，90，100，150，200 learning rate：0.1，0.3，0.5，0.7，0.9
BAG	n_estimators：20，50，70，90，100，150，200
GBDT	n_estimators：20，50，70，90，100，150，200 learning rate：0.1，0.3，0.5，0.7，0.9
LGBM	n_estimators：20，50，70，90，100，150，200 learning rate：0.1，0.3，0.5，0.7，0.9
LR	C：0.001，0.01，0.1，1，10，100
RF	n_estimators：20，50，70，90，100，150，200 criterion：gini，entropy
SVM	n_estimators：20，50，70，90，100，150，200 criterion：gini，entropy
XGB	n_estimators：20，50，70，90，100，150，200 learning rate：0.1，0.3，0.5，0.7，0.9

IV. MODEL EVALUATION METRICS

This paper used sensitivity (Sen), specificity (Spe), accuracy (Acc), precision (Pre), and area under the curve (AUC) metrics to comprehensively evaluate the model's performance.

V. EXPERIMENTAL RESULTS AND ANALYSIS

Based on three different gene feature selection algorithms, ten commonly used machine learning classifiers were used to construct decision models for auxiliary grading classification of gliomas. Twenty random trials were conducted for each model, and the average results from all twenty random experiments on the test sets were taken as the performance metrics for each model. Tables III, IV, and V show the performance of decision models that combine three feature selection methods (RFE, Boruta, and SHAP) with ten different classifiers. The relevant metrics are calculated based on high-grade gliomas as the benchmark.

As shown in Table III, for the decision models based on the RFE feature selection algorithm, the model built using the GNB classifier has a lower accuracy (83.88%), while the accuracy of other models is around 88%. The BAG classifier model achieves the highest classification accuracy, reaching 88.96%, and the GBDT classifier model has the highest AUC value at 95.34%. It can be observed that the specificity metric for all models is lower compared to other metrics, indicating that the models tend to misclassify low-grade gliomas as high-grade gliomas in the classification decision process.

TABLE III. PERFORMANCE OF MODELS BASED ON THE RFE FEATURE SELECTION ALGORITHM

Classifiers	Acc	AUC	Pre	Sen	Spe
ADB	0.8818	0.9458	0.8962	0.9255	0.8010
BAG	**0.8896**	0.9462	0.9043	0.9290	0.8176
GBDT	0.8875	**0.9534**	0.8996	**0.9314**	0.8070
GNB	0.8388	0.8981	**0.9133**	0.8316	**0.8535**
KNN	0.8803	0.9287	0.8921	0.9286	0.7919
LGBM	0.8871	0.9454	0.9024	0.9276	0.8134
LR	0.8783	0.9420	0.8961	0.9199	0.8017
RF	0.8895	0.9507	0.9023	0.9313	0.8127
SVM	0.8886	0.9383	0.9022	0.9300	0.8127
XGB	0.8848	0.9440	0.8980	0.9287	0.8043

As shown in Table IV, for the Boruta feature selection algorithm, the classification model based on the RF classifier exhibits the best overall performance, with the highest accuracy, AUC, precision, and sensitivity, achieving 89.36%, 0.9547,

90.47%, and 93.53%, respectively. The ADB classifier model has the highest specificity at 94.77%.

TABLE IV. PERFORMANCE OF MODELS BASED ON THE BORUTA FEATURE SELECTION ALGORITHM

Classifiers	Acc	AUC	Pre	Sen	Spe
ADB	0.8787	0.9477	0.8406	0.8227	**0.9477**
BAG	0.8904	0.9501	0.9027	0.9324	0.8131
GBDT	0.8824	0.9515	0.8961	0.9272	0.8000
GNB	0.8192	0.9010	0.9088	0.8029	0.8500
KNN	0.8797	0.9312	0.8912	0.9288	0.7897
LGBM	0.8902	0.9519	0.9031	0.9314	0.8146
LR	0.8728	0.9455	0.8932	0.9140	0.7974
RF	**0.8936**	**0.9547**	**0.9047**	**0.9352**	0.8172
SVM	0.8877	0.9439	0.9041	0.9260	0.8179
XGB	0.8878	0.9482	0.8994	0.9324	0.8064

As shown in Table V, for the SHAP feature selection algorithm, the model using RF as the classifier exhibits the best performance, with the highest accuracy, AUC, precision, and sensitivity among all models, achieving 89.92%, 0.9582, 91.01%, and 93.80%, respectively. The SVM model performs second best, outperforming several other ensemble learning classifiers.

TABLE V. PERFORMANCE OF MODELS BASED ON THE SHAP FEATURE SELECTION ALGORITHM

Classifiers	Acc	AUC	Pre	Sen	Spe
ADB	0.8822	0.9490	0.9021	0.9191	0.8149
BAG	0.8899	0.9506	0.9025	0.9318	0.8132
GBDT	0.8896	0.9537	0.9030	0.9306	0.8149
GNB	0.7917	0.8864	0.9051	0.7598	**0.8522**
KNN	0.8843	0.9333	0.8975	0.9284	0.8034
LGBM	0.8887	0.9462	0.9064	0.9249	0.8224
LR	0.8815	0.9499	0.9001	0.9204	0.8104
RF	**0.8992**	**0.9582**	**0.9101**	**0.9380**	0.8281
SVM	0.8919	0.9444	0.9064	0.9302	0.8218
XGB	0.8843	0.9451	0.8999	0.9256	0.8093

From the above summary, it can be concluded that in the gliomas grading auxiliary diagnosis decision models based on three feature selection algorithms, all models, except for the GNB classifier, achieved good classification performance. Among all the auxiliary diagnosis decision models combining feature selection algorithms and classifiers, the SHAP-RF combination achieved the best performance. This model effectively distinguishes between high-grade and low-grade glioma patients, addressing the issue of glioma grading diagnosis, and can further assist clinical doctors in diagnostic and treatment management decisions.

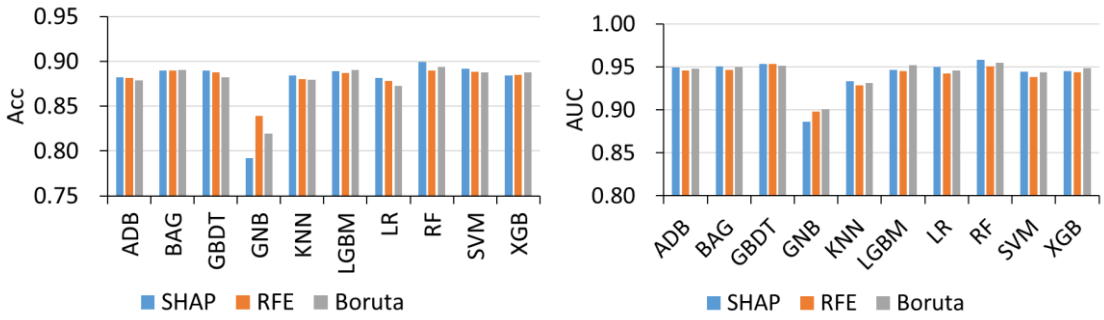

Figure 2. Comparison of Key Performance Metrics for Models Built by Different Classifiers Using Various Feature Selection Algorithms.

In addition, the impact of different feature selection algorithms on model performance was compared. As shown in Figure 2, the performance metrics of the SHAP feature selection method generally outperform those of the other two methods in the classifier models.

VI. MEDICAL KNOWLEDGE DISCOVERY BASED ON FEATURE GENES

This study used three feature selection algorithms for gene feature selection related to glioma grading. Each algorithm was run 20 times, and the top 100 genes selected in each run were then statistically analyzed. The results from the three feature selection algorithms were integrated to identify disease risk genes and discover medical knowledge.

The selected feature genes from the three feature selection algorithms were summarized. By combining the strengths of the three algorithms, the frequency with which each gene was selected across the 20 random trials of all three algorithms was calculated. The more frequently a gene was selected, the more strongly it is related to the glioma grade. Figure 3 displays the feature genes that were selected more than 20 times, sorted by the number of selections in descending order. Integrating the results from the three feature selection methods helps reduce the number of false positive feature genes, filtering out genes that may not truly be related to glioma grade. These genes can serve as biomarkers for clinical diagnosis and therapeutic targets. Annotating and interpreting these genes can reveal the underlying mechanisms of glioma grade development, transforming this knowledge into medical insights for glioma grading, enhancing the understanding of glioma's causes and progression, and assisting doctors in diagnostic and treatment decisions.

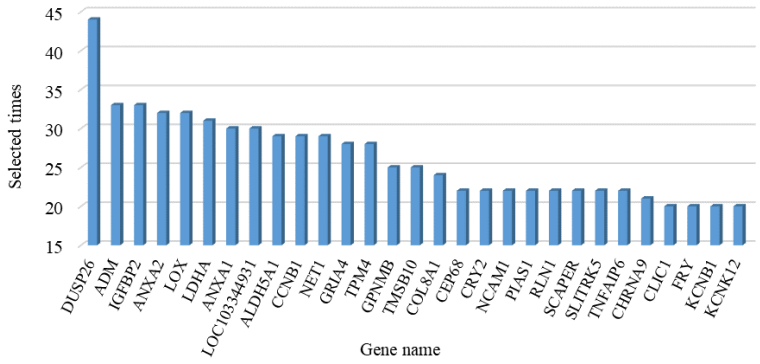

Figure 3. Sorting of Feature Genes by Selection Frequency Across All Model Experiments.

In addition, this study also used literature mining methods to conduct a deeper exploration and interpretation of the top seven feature genes with the highest selection frequency, which are the genes most related to glioma grade. It was found that these feature genes have already been proven to be associated with tumor occurrence or progression, indicating that the glioma grade-related feature genes selected by the model have high reliability and can serve as a basis for clinical diagnosis and decision-making. Moreover, the identified feature genes may include genes with unknown functions and those that have not been previously studied. Further clinical observation or experimental research can be conducted to analyze their disease-related functions.

VII. Conclusion

This study addresses the high-dimensional and information-redundant nature of gene expression data by selecting three feature selection methods for feature gene screening and dimensionality reduction. Based on the selected feature genes, decision models for glioma grading and diagnosis were constructed using ten machine-learning classifiers, and the performance of each model was evaluated. The results show that most of the constructed models have superior performance and can effectively assist in gliomas grading decision-making. Feature genes related to glioma grade were identified, which can serve as biomarkers for diagnostic decisions and therapeutic targets. Additionally, functional annotation and interpretation of the selected feature genes were conducted to mine relevant medical information, transforming it into medical knowledge, thereby enhancing doctors' understanding and assisting them in clinical decision-making.

Acknowledgment

The authors sincerely acknowledge the patients who provided research samples, their families, and the public databases: TCGA, GEO.

References

[1] S. K. Zhou, C. Davatzikos, J. S. Duncan, B. V. Ginneken, A. Madabhushi, J. L. Prince et al., "A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises," Proceedings of the IEEE, vol. 109, no. 5, pp. 820-838, 2021.

[2] P. Mehta, H. U. Ahmed, M. Emberton, S. Punwani, S. Ourselin, "Computer-aided diagnosis of prostate cancer using multiparametric MRI and clinical features: A patient-level classification framework," Medical Image Analysis, vol. 73, p. 102153, 2021.

[3] Y. P. Ye, H. T. Chen, H. Chen, "Knowledge Management Process, Technical Tools, Models and Strategies in the Period of the Big Data," Library and Information Service, vol. 63, no. 5, pp. 5-13, 2019.

[4] F. C. Ma, L. Q. Zhou, "Knowledge Management and Services for Smart Health. Journal of Library Science in China," Journal of Library Science in China, vol. 44, no. 05, pp. 4-19, 2018.

[5] D. Wu, Z. J. Liu, "Research on the Application and the Trend of Intelligent Information Services from Big Data Perspective," Journal of Information Resources Management, vol. 8, no. 2, pp. 28-39+75, 2018.

[6] P. X. Yu, D. W. Wang, R. G. Zhang, M. Deng, H. Y. Yang, L. J. Wu et al., "Spatial resolution enhancement using deep learning improves chest disease diagnosis based on thick slice CT," Npj Digital Medicine, vol. 7, no. 1, p. 335, 2024.

[7] J. D. Zhang, Z. W. Shi, Z. H. Li, Z. Y. Liu, D. G. Shen, "A robust and efficient AI assistant for breast tumor segmentation from DCE-MRI via a spatial-temporal framework," Patterns (N Y), vol. 4, no. 9, p. 100826, 2023.

[8] S. A. Harmon, T. H. Sanford, S. Xu, E. B. Turkbey, H. Roth, Z. Y. Xu et al., "Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets," Nature Communications, vol. 11, no. 1, p. 4080, Aug 14 2020.

[9] H. F. Wang, C. M. Zhang, L. Y. Shi, S. Y. Ng, H. Yan, K. C. Yeung et al., "An intelligent composite model incorporating global / regional X-rays and clinical parameters to predict progressive adolescent idiopathic scoliosis curvatures and facilitate population screening," eBioMedicine, vol. 95, p. 104768, 2023.

[10] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," (in English), Mach Learn, vol. 46, no. 1-3, pp. 389-422, 2002.

[11] W. R. R. Miron B. Kursa, "Feature Selection with the Boruta Package," J Stat Softw, vol. 36, no. 11, pp. 1-13, 2010.

[12] S. M. Lundberg. S. I. Lee., "A Unified Approach to Interpreting Model Predictions," 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA., pp. 1-10, 2017.

[13] C. W. Brennan, R. G. W. Verhaak, A. McKenna, B. Campos, H. Noushmehr, S. R. Salama et al., "The somatic genomic landscape of glioblastoma," Cell, vol. 155, no. 2, pp. 462-77, Oct 10 2013.

[14] J. A. Khan, T. K. Mandal, T. K. Das, Y. Singh, B. Pillai, and S. Maiti, "Magnetite (Fe3O4) nanocrystals affect the expression of genes involved in the TGF-beta signalling pathway," Molecular BioSystems, vol. 7, no. 5, pp. 1481-6, May 2011.

[15] L. A. M. Gravendeel, M. C. M. Kouwenhoven, O. Gevaert, J. J. Rooi, A. P. Stubbs, J. E. Duijm et al., "Intrinsic Gene Expression Profiles of Gliomas Are a Better Predictor of Survival than Histology," Cancer Research, vol. 69, no. 23, pp. 9065-9072, Dec 1 2009.

[16] Z. Y. Liu, Z. Q. Yao, C. Li, Y. C. Lu, C. F. Gao, "Gene expression profiling in human high-grade astrocytomas," International Journal of Genomics, vol. 2011, p. 245137, 2011.

[17] J. L. Garcia, J. Couceiro, J. A. G. Moreta, J. M. G. Valero, A. S. Briz, V. Sauzeau et al., "Expression of VAV1 in the tumour microenvironment of glioblastoma multiforme," J Neuro-Oncol, vol. 110, no. 1, pp. 69-77, 2012, doi: 10.1007/s11060-012-0936-y.

[18] S. Turcan, D. Rohle, A. Goenka, L. A. Walsh, F. Fang, E. Yilmaz et al., "IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype," Nature, vol. 483, no. 7390, pp. 479-483, Mar 22 2012.

[19] D. Sturm, H. Witt, V. Hovestadt, D. A. K. Quang, D. T. W. Jones, C. Konermann et al., "Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma," Cancer Cell, vol. 22, no. 4, pp. 425-437, 2012 .

[20] H. S. Phillips, S. Kharbanda, R. H. Chen, W. F. Forrest, R. H. Soriano, T. D. Wu et al., "Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis," Cancer Cell, vol. 9, no. 3, pp. 157-173, 2006.

[21] L. X. Sun, A. M. Hui, Q. Su, A. Vortmeyer, Y. Kotliarov, S. Pastorino et al., "Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain," Cancer Cell, vol. 9, no. 4, pp. 287-300, 2006.

[22] W. A. Freije, F. E. C. Vargas, Z. X. Fang, S. Horvath, T. Cloughesy, L. M. Liau et al., "Gene expression profiling of gliomas strongly predicts survival," Cancer Research, vol. 64, no. 18, pp. 6503-6510, 2004.

[23] J. X. Zhou, T. Xu, Y. Yan, R. Qin, H. X. Wang, X. P. Zhang et al., "MicroRNA-326 Functions as a Tumor Suppressor in Glioma by Targeting the Nin One Binding Protein (NOB1)," Plos One, vol. 8, no. 7, 2013.

[24] G. Reifenberger, R. G. Weber, V. Riehmer, K. Kaulich, E. Willscher, H. Wirth et al., "Molecular characterization of long-term survivors of glioblastoma using genome-and transcriptome-wide profiling," International Journal of Cancer, vol. 135, no. 8, pp. 1822-1831, 2014.

[25] M. Weller, R. G. Weber, E. Willscher, V. Riehmer, B. Hentschel, M. Kreuz et al., "Molecular classification of diffuse cerebral WHO grade II/III gliomas using genome- and transcriptome-wide profiling improves stratification of prognostically distinct patient groups," Acta Neuropathologica, vol. 129, no. 5, pp. 679-693, 2015.

[26] S. Madhavan, J. C. Zenklusen, Y. Kotliarov, H. Sahni, H. A. Fine, and K. Buetow, "Rembrandt: Helping Personalized Medicine Become a Reality through Integrative Translational Research," Molecular Cancer Research, vol. 7, no. 2, pp. 157-167, 2009.

[27] A. Murat, E. Migliavacca, T. Gorlia, W. L. Lambiv, T. Shay, M. F. Hamou et al., "Stem cell-related "Self-Renewal" signature and high epidermal growth factor receptor expression associated with resistance to concomitant chemoradiotherapy in glioblastoma," Journal of Clinical Oncology, vol. 26, no. 18, pp. 3015-3024, 2008.