

DOI: https://doi.org/10.48009/1_iis_2026_115

Cost-Efficient glioma classification using machine learning with optimized molecular biomarker selection

Gia Bao Nguyen, *Robert Morris University*, gxnst757@mail.rmu.edu

John C. Stewart, *Robert Morris University*, stewartj@rmu.edu

G. Alan Davis, *Robert Morris University*, davis@rmu.edu

Abstract

Determining whether a brain tumor is a high-grade or low-grade glioma has a profound impact on clinical decision-making, patient prognosis, and the overall treatment strategy. High-grade gliomas are aggressive, fast-growing, and typically require intensive intervention. By contrast, low-grade gliomas often follow a slower progression and may be managed more conservatively. Currently, the most challenging part of glioma classification using machine learning is not only achieving high predictive accuracy, but also addressing the high cost of molecular profiling, particularly in healthcare systems that rely on single-gene sequencing. While molecular biomarkers play an important role in glioma classification, this highlights the need for an effective biomarker selection framework to reduce the cost of molecular profiling without compromising predictive performance. In this study, we examine the effectiveness of different feature selection methods in identifying an optimal biomarker set for training glioma classification models. The results indicate that a model trained on only the five most informative genes can achieve comparable accuracy to a model trained on a set of 20 genes. By reducing the number of biomarkers needed from 20 to 5, the cost of molecular profiling may be reduced by approximately fourfold.

Keywords: Artificial Intelligence, Machine Learning, Classification, Health Informatics, Glioma Classification

Introduction

Gliomas represent the most common class of malignant tumors affecting the central nervous system. These tumors originate in glial cells, which perform essential supportive functions for neuronal activity, including regulation of inflammatory processes, metabolic waste removal, and maintaining overall neural homeostasis. Because glial cells are distributed throughout the brain and spinal cord, gliomas may arise in a wide range of bodily locations and may exhibit substantial differences in their biological behavior and clinical presentation. Gliomas are commonly classified according to their degree of malignancy, ranging from relatively slow-growing, lower-grade gliomas (LGG; World Health Organization grades II–III), to highly aggressive and rapidly progressing glioblastoma (GBM; World Health Organization grade IV). This distinction is clinically meaningful, as lower-grade gliomas are generally associated with more favorable prognoses and better survival rates than GBM (Sun et al., 2024).

Despite the recent rise of AI Agents and Large Language Models (LLMs), traditional machine learning approaches are still widely applied for glioma classification because of their explainability (Vershina et

al., 2025). In practice, clinicians rarely rely solely on machine learning to make final decisions; instead, they integrate the results of different diagnosis methods such as biopsy and neuroimaging (Christodoulou et al., 2026). However, machine learning still plays an important role in supporting clinical reasoning and accelerating the classification process.

Most traditional machine learning approaches rely on mutation biomarkers as key predictive features (Bhatele & Bhadauria, 2022; Taşçı et al., 2022b). According to DeWitt et al. (2017), these biomarkers can be identified through two primary methods: single-gene sequencing and next-generation sequencing (NGS) panel. The cost of single-gene sequencing is \$420, whereas a next-generation sequencing (NGS) panel costs around \$1,800. In healthcare systems with well-established molecular profiling infrastructure, NGS panels are often preferred because they provide a more cost-effective approach when multiple genes must be evaluated simultaneously (Loong et al., 2023). However, the clinical adoption of NGS remains uneven across regions. In regions with limited clinical adoption, such as Asia-Pacific (APAC), clinicians still rely on single-gene sequencing because NGS panel is not widely adopted (Loong et al., 2023). In such circumstances, the cost of molecular profiling increases linearly with the number of genes being tested. For example, assessing 20 genes individually could cost approximately \$8,400, compared with \$1,800 for a single NGS panel (Loong et al., 2023). From a machine learning perspective, this creates a practical challenge. Although larger biomarker panels may provide additional predictive information, they also increase molecular profiling costs in settings where sequential single-gene testing remains common. Existing glioma classification studies frequently incorporate approximately 20 molecular biomarkers to maximize predictive performance; however, the trade-off between predictive accuracy and the number of biomarkers required for testing has received limited attention (Sánchez-Marqués et al., 2024; Taşçı et al., 2022b). These considerations suggest the need for a cost-conscious biomarker selection strategy and motivate the investigation of whether a substantially reduced biomarker set can preserve predictive performance. Specially, the research question addressed in the current study was the following . . .

RQ1. How should machine learning-based glioma classification systems be designed to optimize the trade-off between predictive performance and molecular profiling cost under biomarker constraints?

Review of Literature

The use of molecular biomarkers as input variables, particularly genomic data, offers substantial advantages for both glioma classification and for understanding the biological mechanisms underlying chemoresistance. Levenson et al. (2019) identified three primary methodological approaches for detecting biomarkers associated with chemoresistance in gliomas: magnetic resonance imaging (MRI), metabolomics, and genomics. While MRI provides valuable structural and anatomical information, it does not directly capture molecular-level alterations and is, therefore, limited in its ability to identify specific biological pathways involved in therapeutic resistance (Shuhendler et al., 2015). In contrast, genomic approaches enable detailed characterization of the molecular abnormalities present in gliomas, including mutations in key regulatory genes such as IDH, and PTEN (Li & Sun, 2025; Ferris et al., 2017). These gene mutations have been shown to be closely associated with tumor behavior, treatment response, and patient prognosis (Li & Sun, 2025). This perspective is consistent with the approach adopted in the present study, in which genomic data are used not only as input variables for the glioma classification model, but also to examine the association of specific genetic mutations on glioma risk and progression; analyses that cannot be achieved by using imaging data alone (Levenson et al., 2019).

Recent multi-omics research further supports the prognostic value of genomic biomarkers. Li and Sun (2025), for example, conducted a large-scale integrative analysis combining genomic, DNA methylation, and transcriptomic data to construct multi-layer molecular interaction networks for glioma. Their findings identified IDH1 and PTEN as the most influential genes among 178 measured genomic variables in determining glioma prognosis. Moreover, the study demonstrated that the activity states of these genes provide more informative prognostic indicators than traditional histological classification methods alone (Li & Sun, 2025).

The application of molecular biomarkers within machine learning frameworks for glioma classification has also been extensively explored. Taşçı et al. (2022b), for example, developed an ensemble-based model that incorporated hierarchical voting-based feature selection for glioma grading using the TCGA dataset. These authors' approach achieved a classification accuracy of 87.61%, indicating strong predictive performance. However, the authors use 20 molecular biomarkers as predictive features, which may impose a significant cost burden on patients in healthcare systems where NGS panels are not available, as discussed earlier. To address this limitation, our study applies a range of feature-selection techniques to identify informative biomarkers and evaluates their glioma classification effectiveness by training and comparing machine learning models using the resulting biomarker subsets. This selection strategy enables clinicians to identify the optimal set of five most informative biomarkers for each traditional machine learning approach, which reduces molecular profiling cost by fourfold when conducting single-gene sequencing.

Methodology

In the current study, three feature selection methods were applied to identify the five most informative molecular biomarkers on the TCGA glioma dataset:

- **Chi-Square Statistical Analysis (Filter):** Evaluates the association between each molecular biomarker and glioma class (see Table 1). The Chi-Square test was performed under the premise that all 4 assumptions were satisfied (Wang & Zhou, 2021).
- **Feature Importance of Random Forest (Embedded):** Estimates the importance of each biomarker based on its contribution to reducing impurity across decision trees within a Random Forest model (see Table 2). Biomarkers with higher importance scores are considered more influential in predicting glioma subtypes (Rogers & Gunn, 2006).
- **Recursive Feature Elimination with Logistic Regression (Wrapped):** Iteratively removes the least important biomarkers based on the coefficients of a Logistic Regression model (see Table 3). The process continues until the desired number of biomarkers is retained, resulting in a subset that maximizes predictive performance (Han et al., 2021).

The resulting biomarker subset of each method was then used to train five different traditional machine learning models: Logistic Regression, Random Forest, XGBoost, Support Vector Machine (SVM), and Naive Bayes. The specific objective of these models was to classify glioma cases into two clinically meaningful categories: lower-grade glioma (LGG), and glioblastoma (GBM). The dataset used consisted of 839 patient records and 24 features for input, of which 20 were frequently mutated genetic biomarkers. All biomarkers were encoded as binary categorical features, where 1 indicates a mutated status and 0 indicates a non-mutated status. Among the 839 patients, 487 were diagnosed with low-grade glioma (LGG) and 352 with glioblastoma (GBM). To evaluate the performance of machine learning models, we employed a k-fold cross-validation technique with both k=5 and k=10, and examined the average accuracy and F1-Score. For the 10-fold cross-validation setting, out-of-fold predictions from each test fold were collected and aggregated into a single prediction vector for each model. These aggregated predictions were used to

perform McNemar's test for statistical comparison between models. Finally, the model achieving the highest average accuracy across the three feature selection methods was selected and compared with the best-performing baseline model trained on all 20 biomarkers. Model development and analysis were conducted in Python within a Jupyter Notebook environment.

The Chi-Square statistical analysis was conducted under two hypotheses. The null hypothesis (H_0) assumes no association between the biomarker and glioma grade, while the alternative hypothesis (H_1) assumes an association. The null hypothesis was rejected when the p-value was less than 0.05, which indicates a statistically significant association. For each biomarker, a 2x2 contingency table was constructed between mutation status (0/1) and glioma grade (LGG and GBM). Afterward, the biomarkers were ranked in descending order based on their Chi-Square statistics (χ^2). According to Preacher (2001), the Chi-Square statistic (χ^2) is calculated as:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where O_{ij} and E_{ij} denote the observed and expected frequencies, respectively, for the cell corresponding to the i^{th} biomarker status (mutated or non-mutated) and the j^{th} glioma grade category (LGG or GBM). Higher Chi-square values indicate stronger associations between a biomarker and glioma grade.

Table 1. Chi-Square Statistical Analysis (Top 5 features based on Chi-Square)

Feature	Chi-Square (χ^2)	p-value
IDH1	417.860789	7.129739e-93
PTEN	111.130416	5.540342e-26
ATRX	81.602246	1.664249e-19
CIC	75.456953	3.734608e-18
EGFR	47.514868	5.458835e-12

The Chi-Square test showed that the five most informative biomarkers were IDH1, PTEN, ATRX, CIC, and EGFR. In addition, Feature Importance was estimated using a Random Forest Classifier trained with 100 decision trees. Importance scores were computed based on the mean decrease in Gini impurity aggregated across all trees. According to Louppe et al. (2013), the feature importance (FI) for a biomarker (f) is defined as:

$$FI(f) = \sum \frac{N_{node}}{N_{total}} \times \Delta Gini$$

where N_{node} is the number of samples at a node, N_{total} is the total number of samples, and $\Delta Gini$ denotes the decrease in Gini impurity resulting from a split using biomarker (f). Higher feature importance indicates a strong contribution of the biomarker to model predictions.

Table 2. Feature Importance of Random Forest (Top 5 features based on Importance Score)

Feature	Importance Score (FI)
IDH1	0.461578
PTEN	0.084544
CIC	0.067658
ATRX	0.066360
IDH2	0.045343

Based on Feature Importance of Random Forest, the five most informative biomarkers were IDH1, PTEN, CIC, ATRX, and IDH2. The only difference between the Random Forest and Chi-square results was the inclusion of IDH2.

Recursive Feature Elimination (RFE) with Logistic Regression selected informative biomarkers by recursively removing the least important features based on the magnitude of their regression coefficients, until the optimal subset of five biomarkers is obtained. At first, a Logistic Regression model was trained using all 20 biomarkers, and the coefficient vector $\beta = [\beta_1, \beta_2, \dots, \beta_p]$ was estimated. Feature importance was defined as $I_j = |\beta_j|$, where β_j represents the coefficient of the j^{th} biomarker. According to Zhang et al. (2008), this iterative process can be formulated as:

$$S_{t+1} = S_t \setminus \left\{ \arg \min_{j \in S_t} |\beta_j| \right\}$$

where S_t denotes the feature subset at iteration t . Similar to the feature importance of Random Forest, a higher absolute coefficient indicates a strong contribution of biomarker to the model's performance.

Table 3. Recursive Feature Elimination with Logistic Regression (Top 5 features based on Absolute Coef)

Feature	Absolute Coef ($ \beta $)
IDH1	4.487869
IDH2	3.287480
NOTCH1	2.894272
TP53	0.966881
PIK3R1	0.958309

Table 3 lists the five most informative biomarkers for RFE. Those biomarkers were IDH1, IDH2, NOTCH1, TP53 and PIK3R1.

After conducting feature selection, we obtained three distinct sets of molecular biomarkers. Each biomarker set was used to train five traditional machine learning models: Logistic Regression, Random Forest, XGBoost, Support Vector Machine (SVM), and Naive Bayes. Model performance was evaluated using stratified k-fold cross-validation. Given the relatively small sample size ($n = 839$), a simple train-test split was not preferred, as it leads to high variance in performance estimates depending on how the data are partitioned (Sivakumar et al., 2024). Specifically, instead of splitting the TCGA dataset D into a single training and testing set, it was partitioned into k disjoint folds $D_1 \cup D_2 \cup \dots \cup D_k$ where $D_i \cap D_j = \emptyset$ for $i \neq j$ and i, j denote the fold indices (Berrar, 2019). Stratified cross-validation ensured that the class distribution within each fold D_j approximates the overall distribution in the dataset D . For each fold t :

- The training set was defined as: $D^{train}_t = D \setminus D_t$
- The testing set was defined as: $D^{test}_t = D_t$

The model was then trained as:

$$f^{(t)} = \mathcal{A}(D^{train}_t)$$

where $f^{(t)}$ denotes the model trained at fold t , and $\mathcal{A}$ denotes the learning algorithm. For each sample $x_i \in D^{test}_t$, the predicted label $\hat{y}$ is obtained as:

$$\hat{y}_i = f^{(t)}(x_i)$$

After that, the out-of-fold (OOF) prediction vector was constructed by aggregating predictions from all test folds. Each sample is predicted exactly once, as it appears in the test set of only one-fold. Each fold produces a set of predictions corresponding to the samples in its test partition. These partial predictions are then concatenated to form the final OOF prediction vector. The OOF prediction vectors from two models, the best performing model across the three feature selection methods (Model A) and the best performing baseline model (Model B), were used to perform McNemar's test. According to Lachenbruch (2014), McNemar's test evaluates the difference between two classifiers by comparing their paired predictions on each sample. The test statistic is computed based on the number of discordant predictions, where model A is correct and model B is incorrect, and vice versa. In the current study, we analyzed model performance on both $k=5$ and $k=10$. The out-of-fold (OOF) predictions were constructed only for the 10-fold cross-validation setting.

Results

The baseline models were trained on all 20 molecular biomarkers. To evaluate their performances, k -fold cross-validation was employed with both $k=5$ and $k=10$.

Table 4. Performance Comparison of Baseline Models Trained on 20 Biomarkers

Model	Accuracy (K=5)	Accuracy (K=10)	F1 (K=5)	F1 (K=10)
Logistic Regression	0.871250	0.871285	0.872051	0.871950
SVM	0.866481	0.865333	0.867279	0.865991
Random Forest	0.859338	0.862937	0.860126	0.863589
Naive Bayes	0.729398	0.740175	0.725199	0.736007
XGBoost	0.856972	0.864128	0.857816	0.864856

Among the baseline models, Logistic Regression demonstrated the highest performance for glioma classification, achieving an accuracy of approximately 87.12%. SVM, Random Forest, and XGBoost also achieved competitive accuracy ranging from 85% to 86%. The models exhibited stable performance across both 5-fold and 10-fold cross-validation, which indicated a strong generalizability. However, these models relied on a complete set of 20 biomarkers. This characteristic makes the models not cost-efficient, if samples are obtained using single-gene sequencing.

The second evaluation focused on models trained on the set of five biomarkers obtained from the Chi-Square Statistical Analysis, which included IDH1, PTEN, ATRX, CIC, and EGFR.

Table 5. Performance Comparison of Models Trained on IDH1, PTEN, ATRX, CIC, and EGFR

Model	Accuracy (K =5)	Accuracy (K=10)	F1 (K =5)	F1 (K =10)
Logistic Regression	0.853372	0.853385	0.854239	0.871950
SVM	0.852174	0.853371	0.853054	0.865991
Random Forest	0.853372	0.854575	0.854246	0.863589
Naive Bayes	0.854555	0.854561	0.855422	0.736007
XGBoost	0.852181	0.858147	0.853043	0.864856

Models trained on IDH1, PTEN, ATRX, CIC, and EGFR demonstrated high stability, with all models achieving an accuracy of approximately 85% under both 5-fold and 10-fold cross-validation. The Naive Bayes model benefited the most from feature selection using the Chi-Square test, achieving nearly a 12% improvement in accuracy compared to its baseline. However, its F1-score under 10-fold cross-validation remained low with only about 73%, which indicated that the model was having difficulty balancing precision and recall. Overall, while none of these models could achieve a better performance than the baseline Logistic Regression trained on a full set of 20 biomarkers, feature selection using the Chi-Square Test should be considered a trade-off between predictive accuracy and the cost of gen profiling, rather than an optimal approach for glioma classification.

The third evaluation focused on models trained on the set of five biomarkers identified by Feature Importance of Random Forest. These biomarkers included IDH1, PTEN, CIC, ATRX, and IDH2.

Table 6. Performance Comparison of Models Trained on IDH1, PTEN, CIC, ATRX, and IDH2

Model	Accuracy (K =5)	Accuracy (K=10)	F1 (K =5)	F1 (K =10)
Logistic Regression	0.867679	0.867714	0.868490	0.868428
SVM	0.864100	0.865318	0.864860	0.865962
Random Forest	0.865291	0.866509	0.866051	0.867150
Naive Bayes	0.864107	0.866538	0.864905	0.867241
XGBoost	0.865291	0.867699	0.866033	0.868327

Models trained on IDH1, PTEN, CIC, ATRX, and IDH2 demonstrated a slight improvement over those based on Chi-Square feature selection, achieving an accuracy of approximately 86%. This suggested that replacing EGFR with IDH2 could improve predictive performance. Additionally, feature selection based on Random Forest's Feature Importance also led to a significant improvement in the F1-score of the Naive Bayes model (from 73% to 86% under 10-fold cross-validation). However, similar to Chi-Square-based models, none of these models could outperform the baseline Logistics Regression with an accuracy of 87.12%.

The fourth evaluation focused on models trained on IDH1, IDH2, NOTCH1, TP53 and PIK3R1. These biomarkers were selected using Recursive Feature Elimination (RFE) with Logistic Regression.

Table 7. Performance Comparison of Models Trained on IDH1, IDH2, NOTCH1, TP53 and PIK3R1

Model	Accuracy (K=5)	Accuracy (K=10)	F1 (K=5)	F1 (K=10)
Logistic Regression	0.868869	0.868904	0.869687	0.869618
SVM	0.872441	0.872476	0.873227	0.873149
Random Forest	0.872441	0.871285	0.873227	0.871997
Naive Bayes	0.486270	0.486317	0.380751	0.380052
XGBoost	0.868869	0.868904	0.869687	0.869619

The use of Recursive Feature Elimination (RFE) as the biomarker selection method substantially improved the performance of most glioma classifiers, with the exception of Naive Bayes. Among the evaluated models, Support Vector Machine (SVM) and Random Forest exhibited the most notable improvements, both achieving an accuracy of around 87.24% under 5-fold cross-validation. Under 10-fold cross-validation, SVM achieved a slightly higher average accuracy (87.24% > 87.12%) and F1-Score (87.31% > 87.19%) than Random Forest. Therefore, the SVM trained on IDH1, IDH2, NOTCH1, TP53 and PIK3R1 was selected as the best-performing model across the three feature selection methods. The only limitation was observed in the Naive Bayes model. In contrast to the other classifiers, applying RFE-based feature selection reduced its classification accuracy to only 48.62%. This suggests that the feature subset selected by RFE may not be well-suited to the probabilistic assumptions underlying the Naive Bayes algorithm.

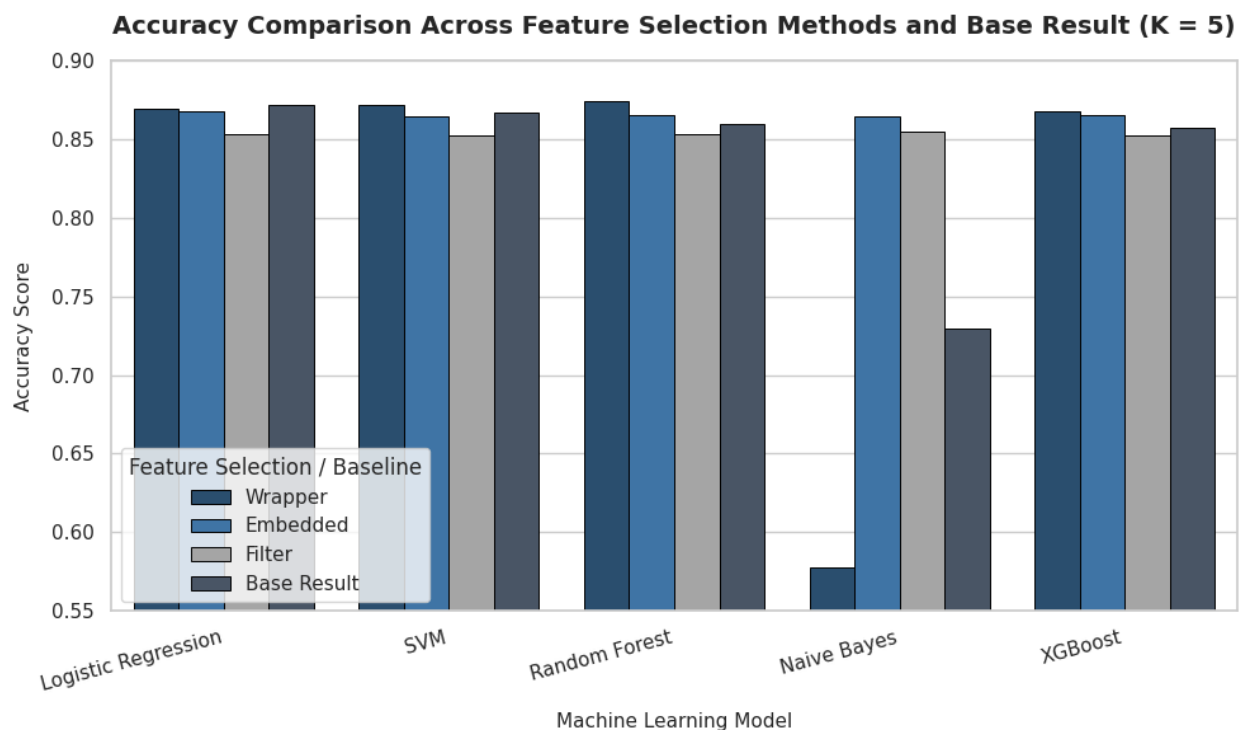


Figure 1. Accuracy Comparison Across Feature Selection Methods and Base Result (K = 5)

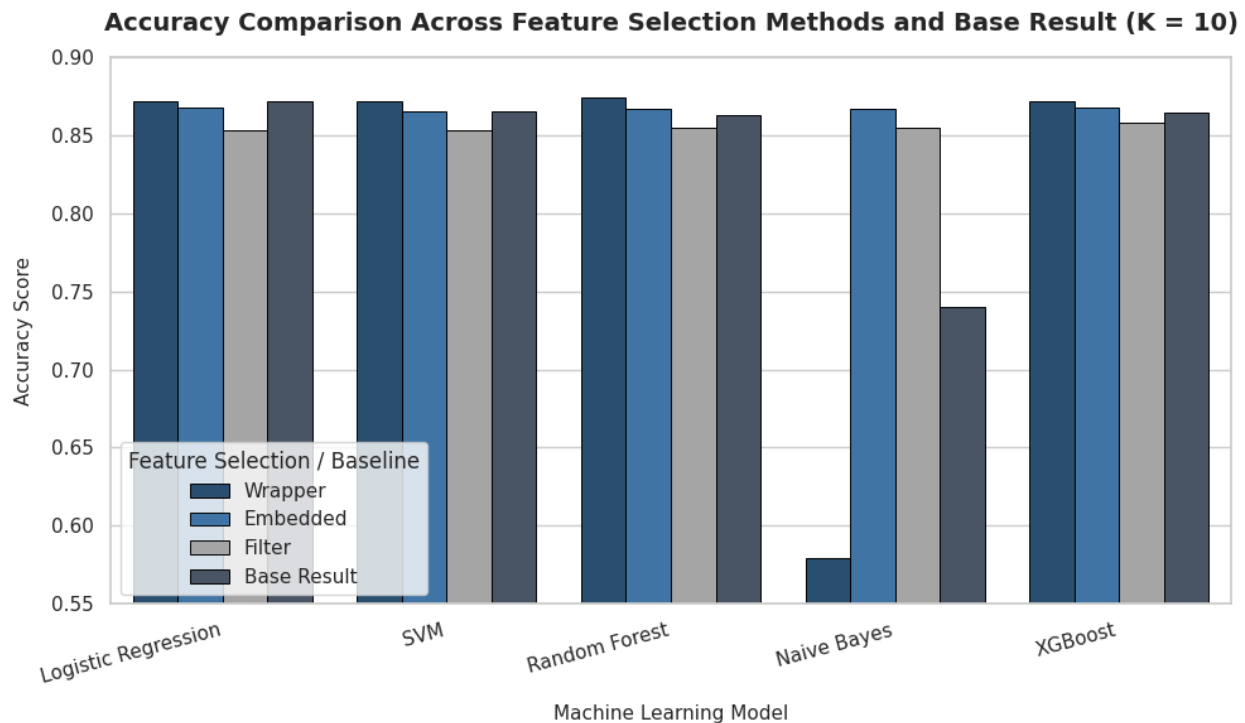


Figure 2. Accuracy Comparison Across Feature Selection Methods and Base Result (K = 10)

In terms of addressing the research question (RQ1) proposed in the current study, the SVM model (Model A) trained on IDH1, IDH2, NOTCH1, TP53, and PIK3R1 not only outperformed other models across the three feature selection methods but also outperformed the baseline Logistic Regression model (Model B). This finding supported the primary research question of this study: a carefully selected subset of only five biomarkers could achieve, and even exceed, the predictive performance of models trained on the full set of 20 biomarkers. To determine whether the observed improvement was statistically significant, McNemar's test was performed using the out-of-fold predictions obtained from the 10-fold cross-validation.

Table 8. Contingency Table for McNemar's Test Comparing Model A and Model B

Model	Model B (Correct)	Model B (Incorrect)
Model A (Correct)	728	4
Model A (Incorrect)	3	104

McNemar's test showed no statistically significant difference between the two models (test statistic = 3.0, $p = 1.0$). This means there is insufficient evidence to conclude that the improvement is not due to random variation. However, within the scope of the current dataset ($n = 839$), the five-biomarker model demonstrated slightly better empirical performance than the model trained on 20 biomarkers, suggesting that reducing the cost of molecular profiling may be achieved without compromising predictive performance.

Conclusions and Discussion

The current study demonstrated that glioma classifiers using only five molecular biomarkers may still achieve competitive predictive performance compared with models trained on a set of 20 biomarkers.

Among the evaluated feature selection methods, Recursive Feature Elimination (RFE) combined with Support Vector Machine (SVM) produced the best overall performance, outperforming the strongest baseline model while requiring only one-quarter of the original biomarkers. From a clinical perspective, IDH1, IDH2, NOTCH1, TP53, and PIK3R1 should be prioritized for biomarker assessment in settings where access to comprehensive molecular profiling is limited. These findings suggest that careful biomarker selection may substantially reduce the cost of molecular profiling using single-gene sequencing without compromising diagnostic accuracy.

In terms of limitations, it should be noted that the conclusions of this study are based on a single cohort of 839 samples and should therefore be interpreted within the context of this dataset. In addition, the models did not include demographic variables such as age, gender, and race, which may provide additional predictive value in real-world clinical settings, especially where genomic testing is limited. Further, the TCGA dataset contains only 839 samples, which is relatively small for assessing the generalizability of model performance; however, this limitation was partially addressed through the use of 5-fold and 10-fold cross-validation. The limited sample size also reduces the sensitivity of the McNemar's test, making it more difficult to detect small differences between models (Lachenbruch, 2014). As a result, there was insufficient evidence to reject the null hypothesis. In addition, the absence of some clinically relevant biomarkers associated with high-grade glioma, such as MET and TERT, represents another limitation of the dataset (Brennan et al., 2013).

Validation on independent and larger cohorts is also required to assess the generalizability of the proposed approach. Particularly, future studies should investigate the generalizability of the selected biomarker panels across different populations and sequencing platforms. Extending this work to multi-class glioma classification and integrating additional omics data may further enhance predictive performance and clinical applicability.

Although most machine learning models in the current study demonstrated promising performance with accuracies of approximately 85%, relying solely on machine learning predictions for clinical decision-making is not recommended. The most appropriate approach is to integrate machine learning outputs with established diagnostic methods, such as biopsy, MRI imaging, and clinical expertise.

References

- Berrar, D. (2019). *Cross-validation*. In S. Ranganathan et al. (Eds.), *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>
- Bhatele, K. R., & Bhadauria, S. S. (2022). Machine learning application in glioma classification: Review and comparison analysis. *Archives of Computational Methods in Engineering*, 29(1), 247–274. <https://doi.org/10.1007/s11831-021-09572-z>
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197–227. <https://doi.org/10.1007/s11749-016-0481-7>
- Blumenthal, D. T., Artzi, M., Liberman, G., Bokstein, F., Aizenstein, O., & Ben Bashat, D. (2017). Differentiation of high-grade glioma components using support vector machine. *American Journal of Neuroradiology*, 38(5), 908–914. <https://doi.org/10.3174/ajnr.A5127>

- Brennan, C. W., Verhaak, R. G. W., McKenna, A., Campos, B., Nouchmeh, H., Salama, S. R., Zheng, S., Chakravarty, D., Sanborn, J. Z., Berman, S. H., Beroukhi, R., Bernard, B., Wu, C. J., Genovese, G., Shmulevich, I., Barnholtz-Sloan, J., Zou, L., Vegesna, R., Shukla, S. A., Ciriello, G., ... TCGA Research Network. (2013). *The somatic genomic landscape of glioblastoma*. *Cell*, 155(2), 462–477. <https://doi.org/10.1016/j.cell.2013.09.034>
- Christodoulou, R. C., Papageorgiou, P. S., Eller, D., Vassiliou, E., Papageorgiou, S. G., Solomou, E. E., & Georgiou, M. F. (2026). Multimodal artificial intelligence in glioma management: Integrating neuroimaging and hematologic biomarkers for precision oncology. *Frontiers in Oncology*, 16, Article 1812518. <https://doi.org/10.3389/fonc.2026.1812518>
- DeWitt, J. C., Jordan, J. T., Frosch, M. P., Samore, W. R., Iafrate, A. J., Louis, D. N., & Lennerz, J. K. (2017). Cost-effectiveness of IDH testing in diffuse gliomas according to the 2016 WHO classification of tumors of the central nervous system recommendations. *Neuro-Oncology*, 19(12), 1640–1650. <https://doi.org/10.1093/neuonc/now120>
- El Naqa, I., & Murphy, M. J. (2015). What is machine learning? In I. El Naqa & M. J. Murphy (Eds.), *Machine learning in radiation oncology: Theory and applications* (pp. 3–11). Springer International Publishing. <https://doi.org/10.1007/978-3-319-18305-3>
- Ferris, S. P., Hofmann, J. W., Solomon, D. A., & Perry, A. (2017). Characterization of gliomas: From morphology to molecules. *Virchows Archiv*, 471(2), 257–269. <https://doi.org/10.1007/s00428-017-2181-4>
- Han, Y., Huang, L., & Zhou, F. (2021). A dynamic recursive feature elimination framework (dRFE) to further refine a set of OMIC biomarkers. *Bioinformatics*, 37(15), 2183–2189. <https://doi.org/10.1093/bioinformatics/btab055>
- Heugenhauer, J., Galijasevic, M., Mangesius, S., Goebel, G., Buchroithner, J., Erhart, F., Pichler, J., Widhalm, G., Stockhammer, G., Iglseder, S., Freyschlag, C. F., Oberndorfer, S., Bordihn, K., von Campe, G., Czech, T., Surböck, B., Urbanic Purkart, T., Marosi, C., Felzmann, T., & Nowosielski, M. (2022). MRI response assessment in glioblastoma patients treated with dendritic-cell-based immunotherapy. *Cancers*, 14(6), 1579. <https://doi.org/10.3390/cancers14061579>
- Jackson, R. J., Fuller, G. N., Abi-Said, D., Lang, F. F., Gokaslan, Z. L., Shi, W., Wildrick, D. M., & Sawaya, R. (2001). Limitations of stereotactic biopsy in the initial management of gliomas. *Neuro-Oncology*, 3(3), 193–200. <https://doi.org/10.1093/neuonc/3.3.193>
- Jozsa, F., Gaier, C., Ma, Y., Kitchen, N., McEvoy, A., Miserocchi, A., Samandouras, G., Sethi, H., Thorne, L., Hill, C., & Darie, L. (2023). Safety and efficacy of brain biopsy: Results from a single institution retrospective cohort study. *Brain & Spine*, 3, 101763. <https://doi.org/10.1016/j.bas.2023.101763>
- Lachenbruch, P. A. (2014). *McNemar test*. In N. Balakrishnan, T. Colton, B. Everitt, W. Piegorisch, F. Ruggeri, & J. L. Teugels (Eds.), *Wiley StatsRef: Statistics Reference Online*. Wiley. <https://doi.org/10.1002/9781118445112.stat04876>

- Levenson, C. W., Morgan, T. J., Twigg, P. D., Jr., Logan, T. M., & Schepkin, V. D. (2019). Use of MRI, metabolomic, and genomic biomarkers to identify mechanisms of chemoresistance in glioma. *Cancer Drug Resistance*, 2(3), 862–876. <https://doi.org/10.20517/cdr.2019.18>
- Li, Y., & Sun, H. (2025). Multi-omics analysis identifies novel genes involved in glioma prognosis. *Scientific Reports*, 15, 5806. <https://doi.org/10.1038/s41598-025-90658-0>
- Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A review of machine learning interpretability methods. *Entropy*, 23(1), 18. <https://doi.org/10.3390/e23010018>
- Loong, H. H., et al. (2023). Recommendations for the use of next-generation sequencing in patients with metastatic cancer in the Asia-Pacific region: A report from the APODDC working group. *ESMO Open*, 8(4), Article 101586. <https://doi.org/10.1016/j.esmoop.2023.101586>
- Louppe, G., Wehenkel, L., Suter, A., & Geurts, P. (2013). *Understanding variable importances in forests of randomized trees*. In *Advances in neural information processing systems* (Vol. 26).
- Ostrom, Q. T., Gittleman, H., Stetson, L., Virk, S. M., & Barnholtz-Sloan, J. S. (2015). Epidemiology of gliomas. In J. Raizer & A. Parsa (Eds.), *Current understanding and treatment of gliomas* (Vol. 163, pp. 1–14). Springer. https://doi.org/10.1007/978-3-319-12048-5_1
- Preacher, K. J. (2001). *Calculation for the chi-square test: An interactive calculation tool for chi-square tests of goodness of fit and independence*. Ohio State University. <http://quantpsy.org/chisq/chisq.htm>
- Rizwan, M., Shabbir, A., Javed, A. R., Shabbir, M., Baker, T., & Al-Jumeily, D. O. (2022). Brain tumor and glioma grade classification using Gaussian convolutional neural network. *IEEE Access*, 10, 29731–29740. <https://doi.org/10.1109/ACCESS.2022.3153108>
- Rogers, J., & Gunn, S. (2006). Identifying feature relevance using a random forest. In C. Saunders, M. Grobelnik, S. Gunn, & J. Shawe-Taylor (Eds.), *Subspace, latent structure and feature selection* (Lecture Notes in Computer Science, Vol. 3940, pp. 173–184). Springer. https://doi.org/10.1007/11752790_12
- Sánchez-Marqués, R., García, V., & Sánchez, J. S. (2024). *A data-centric machine learning approach to improve prediction of glioma grades using low-imbalance TCGA data*. *Scientific Reports*, 14, 17195. <https://doi.org/10.1038/s41598-024-68291-0>
- Shuhendler, A. J., Ye, D., Brewer, K. D., Lee, S. C., Sun, Y., Bull, A., Singh, A., Morgan, A., Doria, S. C., Wright, K., Waisman, D. M., & Rutt, B. K. (2015). Molecular magnetic resonance imaging of tumor response to therapy. *Scientific Reports*, 5, Article 14759. <https://doi.org/10.1038/srep14759>
- Sivakumar, M., Parthasarathy, S., & Padmapriya, T. (2024). *Trade-off between training and testing ratio in machine learning for medical image processing*. *PeerJ Computer Science*, 10, e2245. <https://doi.org/10.7717/peerj-cs.2245>

- Sun, X., Jia, Q., Li, K., Tian, C., Yi, L., Yan, L., & Gu, M. (2024). Comparative genomic landscape of lower-grade glioma and glioblastoma. *PLOS ONE*, *19*(8), e0309536. <https://doi.org/10.1371/journal.pone.0309536>
- Taşçı, E., Camphausen, K., Krauze, A., & Zhuge, Y. (2022a). *Glioma grading clinical and mutation features* [Data set]. UCI Machine Learning Repository. <https://doi.org/10.24432/C5R62J>
- Taşçı, E., Zhuge, Y., Kaur, H., Camphausen, K., & Krauze, A. V. (2022b). Hierarchical voting-based feature selection and ensemble learning model scheme for glioma grading with clinical and molecular characteristics. *International Journal of Molecular Sciences*, *23*, 14155. <https://doi.org/10.3390/ijms232214155>
- Tseng, C. J., & Tang, C. (2023). An optimized XGBoost technique for accurate brain tumor detection using feature selection and image segmentation. *Healthcare Analytics*, *4*, 100217. <https://doi.org/10.1016/j.health.2023.100217>
- Vershinina, O., Turubanova, V., Krivonosov, M., Trukhanov, A., & Ivanchenko, M. (2025). Explainable machine learning models for glioma subtype classification and survival prediction. *Cancers*, *17*(16), 2614. <https://doi.org/10.3390/cancers17162614>
- Wang, Y., & Zhou, C. (2021). Feature selection method based on Chi-square test and minimum redundancy. In M. Tavana, N. Nedjah, & R. Alhajj (Eds.), *Emerging trends in intelligent and interactive systems and applications (IISA 2020)* (Advances in Intelligent Systems and Computing, Vol. 1304). Springer. https://doi.org/10.1007/978-3-030-63784-2_22
- Zhang, J., Jeng, X. J., & Liu, H. (2008). *Some two-step procedures for variable selection in high-dimensional linear regression* (arXiv:0810.1644). arXiv. <https://doi.org/10.48550/arXiv.0810.1644>