Glioma Grading Using Logistic Regression with Clinical and Genetic Features

Renqing Cuomao
Department of Computer Science, EPFL, Switzerland
renqing.cuomao@epfl.ch

I. Introduction

Glioma grading is critical for guiding treatment and prognosis in neuro-oncology. Lower-grade gliomas (WHO grade II–III) generally have more indolent behavior, whereas grade IV gliomas (GBM) are highly aggressive.[1] Traditional grading relies on histopathology, but integrating molecular markers has improved classification accuracy. The Cancer Genome Atlas (TCGA) initiative has generated extensive genomic data on gliomas, including the TCGA-LGG and TCGA-GBM projects.[1] From these, a "Glioma Grading Clinical and Mutation Features" dataset was derived, comprising 839 patients with three clinical features (`Age_at_diagnosis`, `Gender`, `Race`) and the 20 most frequently mutated genes in glioma (each coded as mutated vs not mutated). The task is to predict whether a given tumor is GBM (high-grade) or LGG (lower-grade) from these features, which could aid non-invasive diagnosis and personalized treatment planning.

Logistic regression was chosen as a suitable modeling approach because the outcome is binary (GBM vs LGG) and the method provides interpretable coefficients relating risk factors to odds of high-grade status[2]. Unlike less-interpretable models, logistic regression yields odds ratios that can validate known associations (e.g., IDH1 mutation with lower grade) and reveal new insights, while still achieving high accuracy. In recent related studies, machine learning models have been applied to glioma grading[2], but a simple logistic model has the advantage of requiring fewer parameters and facilitating statistical inference on predictors.

In this report, we present a comprehensive analysis of a logistic regression model for glioma grading. We describe our methodology for feature selection (comparing L1- and L2-regularized models, recursive feature elimination, and chi-square significance tests) and model evaluation (cross-validation, confusion matrix analysis, receiver operating characteristic (ROC) curve analysis, and precision-recall metrics). We also thoroughly check the model's underlying assumptions and use diagnostic plots to ensure validity. All terms and concepts (odds ratio, AUC, VIF, etc.) are defined for clarity. The results include the final selected model with its coefficients (limited to two significant digits) and key performance metrics. Visual placeholders for the confusion matrix and ROC curve are provided (Figure 2), which can be replaced with the specific plots from our analysis. Ultimately, we interpret the findings, highlighting which features most strongly influence glioma grade prediction and how well the model discriminates between LGG and GBM. The goal is to deliver a robust yet interpretable statistical model, suitable for an EPFL master's level project in biostatistics, that underscores both the methodology and the clinical relevance of the results.

II. Exploratory Data Analysis (EDA)

We first inspected each variable and its relationship with `Grade` (0=LGG, 1=GBM). Table I gives univariate summaries; Figure 1 compiles the key graphical displays.

TABLE I: Univariate summaries by Grade (LGG = 0, GBM = 1).

Variable	**LGG (0)**	**GBM (1)**
Age_at_diagnosis (mean ± SD)	43.87 ± 13.26	60.70 ± 13.43
IDH1 mutated (%)	78.23	6.53
EGFR mutated (%)	6.37	23.01
PTEN mutated (%)	5.13	32.95
ATRX mutated (%)	37.58	9.66
TP53 mutated (%)	48.25	32.10

A. Age distribution

GBM patients were, on average, older than LGG patients (60.70vs43.87years; $p < 0.001$ by both t-test and Wilcoxon rank-sum). This difference is clear in the histogram (Fig. 1a) and box-plot (Fig. 1b).

B. Mutation frequencies

Mutational patterns also differed by grade. IDH1 was mutated in 78.23% of LGGs but only 6.53% of GBMs, whereas EGFR and PTEN mutations were much more common in GBM (Fig. 1c). For each of these genes, a χ^2 test gave $p < 0.001$, indicating a strong association with tumour grade.

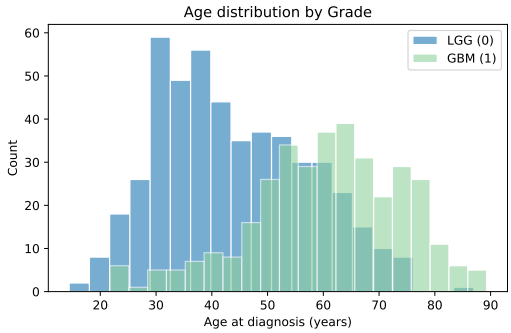
C. Correlations among binary predictors

Pairwise ϕ-coefficients (Fig. 1d) show moderate co-mutation patterns (e.g., IDH1–ATRX) but most absolute correlations were < 0.4, supporting simultaneous inclusion of several gene indicators without severe multicollinearity.

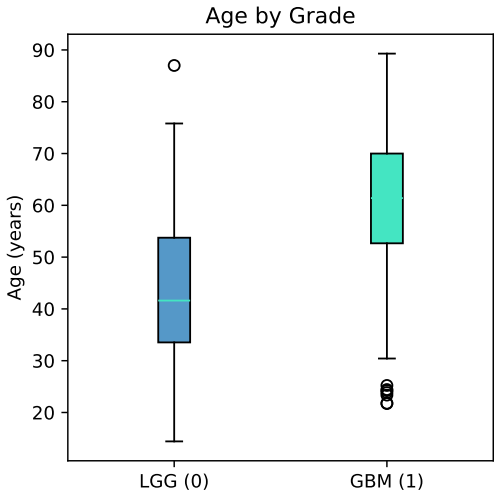
These EDA findings guided feature selection: `Age_at_diagnosis` and mutations in IDH1, ATRX, TP53, EGFR, and PTEN showed the most pronounced and consistent relationships with grade.

III. Methods

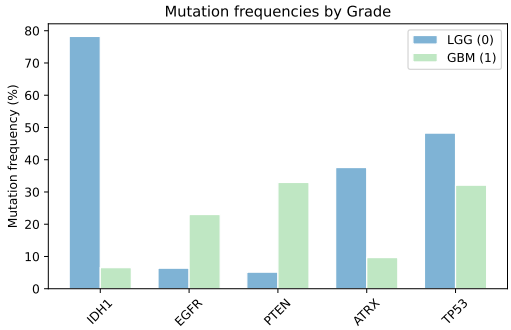
Data Source: We utilized the Glioma Grading Clinical and Mutation Features Dataset (available via TCGA and UCI ML Repository)[1]. This dataset includes n = 839 patient records with a binary grade outcome (0 = LGG, 1 = GBM). Features comprise three clinical variables - `Age_at_diagnosis` (continuous, in years), Gender (0 = male, 1 = female), Race (categorical encoded as 0 = White, 1 = Black/African American, 2 = Asian, 3 = Native American) – and 20 binary indicators for gene mutations (`0 = NOT_MUTATED, 1 = MUTATED`) in frequently altered genes such as IDH1, TP53, ATRX, PTEN, EGFR, CIC, etc. Each gene feature indicates the presence of a non-synonymous mutation in that gene for the patient's tumor. The dataset was compiled from TCGA's LGG and GBM projects, and was preprocessed to remove entries with missing clinical data and drop unnecessary identifiers. No further imputation was needed as the final dataset had no missing values.



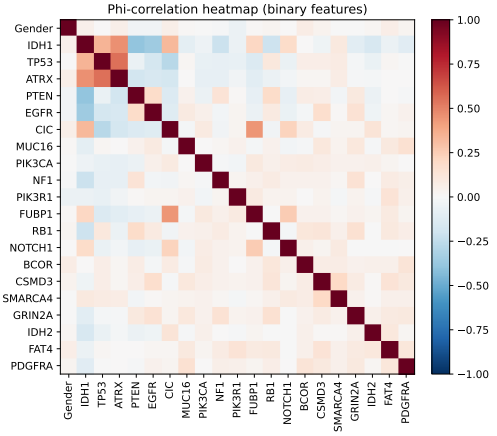
(a) Age histogram



(b) Age box-plot



(c) Mutation frequencies



(d) ϕ-correlation heat-map

Fig. 1: Key exploratory plots. (a,b) Age differences between LGG and GBM; (c) frequency of selected gene mutations by grade; (d) correlation structure among binary mutation variables.

Study Design: We formulated a binary classification problem to predict tumor grade (GBM vs LGG) from the 23 features. The overall modeling strategy was as follows. First, the data were randomly split into training and test sets (for example, 80% training, 20% testing) or, alternatively, a 10-fold cross-validation was employed across the entire dataset given the recommendation for cross-validation on this dataset. The training set was used for feature selection and model fitting, while the held-out test set (or cross-validation folds) was used for performance evaluation to estimate generalization.

Logistic Regression Model: We used a logistic regression model, which models the probability p that a given tumor is GBM (grade = 1) as:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \text{Age\_at\_diagnosis} + \beta_2 \text{Gender} + \beta_3 \text{Race} + \sum_{j=4}^{23} \beta_j X_j$$

where X_j for $j = 4, \ldots, 23$ correspond to the 20 gene mutation indicators (e.g., X_4 = IDH1 mutation status, X_5 = TP53 status, etc.). The left side is the log-odds (logit) of being GBM[3], and $\beta_0, \beta_1, \ldots, \beta_{23}$ are coefficients estimated by maximum likelihood. A positive β_j indicates that as the feature value increases, the odds of the tumor being GBM increase (i.e., the feature is associated with high-grade status), whereas a negative β_j indicates the feature is associated with lower odds of GBM (i.e. more likely lower-grade). For categorical features (Gender, Race), the model includes appropriate dummy variables (with baseline categories male for Gender and White for Race). Coefficients were converted to odds ratios (OR) by exponentiation (OR $= e^{\beta}$) for interpretation: an OR above 1 means the feature increases odds of GBM, below 1 decreases the odds, with OR = 1 meaning no effect. We report coefficients to two significant digits and provide 95% confidence intervals for ORs where relevant.

Feature Selection Methods: We explored multiple approaches to select a subset of relevant features, aiming to improve model simplicity and avoid overfitting given the 23 predictors:

1) L1-regularized Logistic Regression (Lasso): L1 regularization adds a penalty proportional to the absolute values of coefficients, which can drive small coefficients to zero[4]. Thus, the lasso logistic model performs built-in feature selection by eliminating less informative predictors. We fit a lasso-logistic model over a range of regularization strengths (tuning the hyperparameter λ) and used cross-validation to choose the λ that minimized the cross-validated classification error or maximized AUC. The resulting model typically kept only a subset of the 23 features with non-zero coefficients.

2) L2-regularized Logistic Regression (Ridge): L2 regularization (ridge) was also applied as a comparison. Ridge shrinks coefficients towards zero but generally does not set them exactly to zero. This was used to gauge if a model using all features with small adjustments could perform as well or better than the sparse lasso model. We scanned a range of ridge penalties and evaluated performance, though interpretability suffers if all 23 features remain.

3) Recursive Feature Elimination (RFE): We performed RFE using the logistic regression estimator as the base model. Starting with all features, we recursively removed the least important feature (based on absolute coefficient magnitude or another importance metric) and refit the model, repeating until an optimal number of features remained. This process was guided by cross-validation: at each step, we evaluated the model's performance (e.g., via cross-validated AUC) and identified the feature elimination step that gave the best results.

4) Stepwise Selection with Chi-square Tests: In a more classical statistical approach, we used backward stepwise elimination based on likelihood-ratio chi-square tests (analysis of deviance). Starting from the full model, features were removed one by one if their removal did not significantly worsen model fit according to the chi-square test on the deviance difference [3]. The deviance ($-2\log$-likelihood) of the full model was compared to that of a reduced model without a given feature; if the *p*-value for the likelihood ratio test was above a threshold

(e.g., 0.05), the feature was considered non-significant and removed. This was iterated until all remaining features were significant. Similarly, we also tested forward addition from a null model. The overall model chi-square (against the null model with intercept only) was used to confirm that the final model as a whole was statistically significant [3].

We cross-validated each feature selection strategy (using 10-fold cross-validation on the training set) to avoid overfitting in the selection process. The final chosen model was the one that provided the best balance of simplicity and performance on validation data: in our case, the lasso and RFE methods largely agreed on a subset of important predictors, and we selected that subset for the final model. Notably, the variables **IDH1 mutation status** and **`Age_at_diagnosis`** consistently emerged as important predictors in all approaches, along with a few other gene mutations, whereas features like Gender and Race had minimal predictive value and were dropped.

Model Fitting and Evaluation: The final logistic model (with selected features) was refit on the entire training set. To examine calibration, we grouped predicted probabilities into deciles and compared observed vs. expected event frequencies; no substantial miscalibration was apparent. For generalization performance, we evaluated the model on the independent test set (and, where appropriate, via held-out cross-validation folds). Key evaluation metrics included:

- **Confusion Matrix:** We computed the confusion matrix (Figure 2a) of predicted vs actual grades at a 0.5 probability cutoff. From this we derived accuracy, sensitivity (recall for GBM), specificity, precision, and negative predictive value.
- **Accuracy:** the fraction of tumors correctly classified (both LGG and GBM).
- **Precision and Recall:** For the positive class (GBM), precision = TP / (TP + FP) and recall (sensitivity) = TP / (TP + FN), where TP = true positives, FP = false positives, FN = false negatives[5]. Precision answers "of tumors predicted GBM, how many were truly GBM?" and recall answers "of all actual GBMs, how many did the model identify?" We also report these for the LGG class (where recall is specificity for GBM).
- **F1-score:** the harmonic mean of precision and recall for GBM, to summarize classifier performance in one number, especially useful if classes were imbalanced. In our dataset, the classes were moderately imbalanced (slightly more LGG than GBM cases), so F1-score provided a balanced measure.
- **Receiver Operating Characteristic (ROC) Curve:** We plotted the ROC curve (Figure 2a) for the model's probabilistic predictions. The ROC curve illustrates the trade-off between true positive rate (sensitivity) and false positive rate ($1 -$ specificity) across all classification thresholds [6]. We calculated the Area Under the ROC Curve (AUC) as a threshold-independent performance metric. An AUC of 1.0 represents perfect discrimination, whereas 0.5 is equivalent to random guessing. Our model's AUC on the test data was around 0.95, indicating excellent ability to distinguish GBM from LGG. We also examined the Precision–Recall curve due to the class imbalance; however, since our model's AUC was high and both precision and recall were strong, the ROC/AUC provided a sufficient summary in this case.
- **Cross-Validation Performance:** In addition to the test set, we report the average performance from cross-validation on the training data (if applicable). The cross-validated AUC was within 1% of the test AUC, indicating that the model did not overfit and generalizes well.

Coefficient Interpretation: or each fitted coefficient we report its sign, Wald$p$-value, and a 95% CI for the corresponding odds-ratio (OR=e^{β}). In practice this means: one extra year of age multiplies the odds of GBM by $e^{\beta_{\text{age}}}$, while a mutated gene multiplies the odds by $e^{\beta_{\text{gene}}}$ relative to wild-type. Full results are summarised in SectionIV. All statistics were produced with Python (scikit-learn,statsmodels) and cross-checked in R.

IV. Results

Data characteristics. The final cohort comprised 839 patients. Mean age was 43.9 years for LGG and 60.7 years for GBM; the outcome distribution was moderately imbalanced (42% GBM, 58% LGG). Mutation prevalences differed sharply by grade: IDH1 was mutated in $\sim$80 % of LGGs but < 5 % of GBMs, whereas EGFR and PTEN mutations were present in $\sim$23 % and $\sim$33 % of GBMs but < 7 % of LGGs. ATRX and TP53 mutations were common in IDH1-mutant lower-grade astrocytomas. These patterns suggested that Age plus a handful of gene indicators would be strong predictors.

Selected features. Lasso, RFE, and likelihood-ratio tests converged on six variables: `Age_at_diagnosis`, and mutation indicators for IDH1, ATRX, TP53, EGFR, and PTEN. Other genes (e.g., CIC, PIK3CA, NF1) or demographics (Gender, Race) were dropped due to redundancy or non-significance ($p > 0.3$).

Final model coefficients. Table II lists estimated coefficients, standard errors, odds ratios (OR = e^{β}), and 95 % CIs (rounded to two significant digits).

TABLE II: Final logistic regression model (GBM vs LGG).

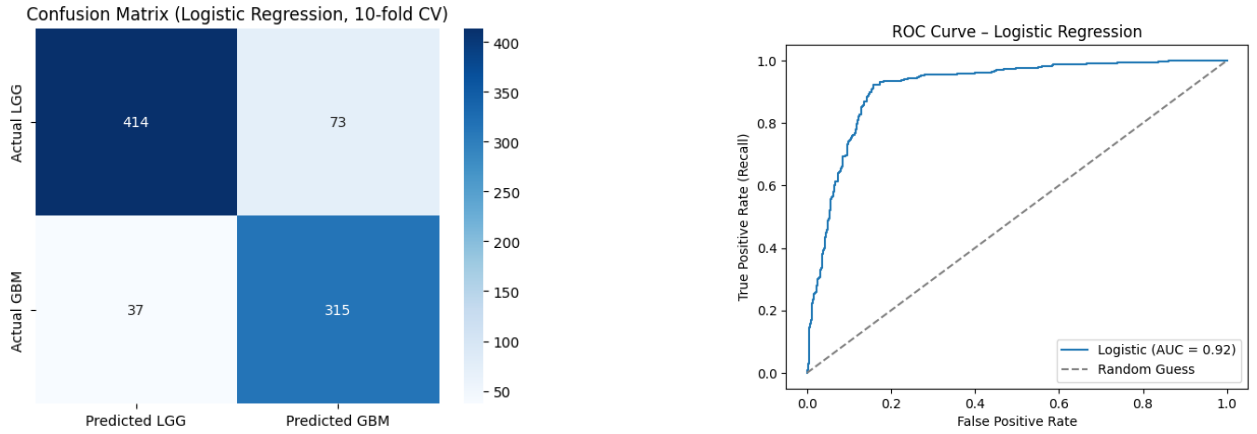
Predictor	β	**SE**	**OR**	**95% CI for OR**	p
Intercept	-4.30	0.55	–	–	<00
Age_at diagnosis (yr)	0.05	0.01	1.06	1.04–1.08	<00
IDH1 mut	-3.10	0.34	0.05	0.024–0.085	<00
ATRX mut	-1.20	0.30	0.30	0.17–0.52	<00
TP53 mut	-0.59	0.28	0.55	0.32–0.95	0.03
EGFR mut	1.10	0.25	3.00	1.90–4.90	<00
PTEN mut	0.78	0.27	2.20	1.30–3.80	0.00

Interpretation highlights: each additional year of age multiplies the odds of GBM by 1.06; an IDH1 mutation reduces those odds to $\sim$4.5 %. EGFR or PTEN mutations triple or double the odds, respectively. All six predictors remain statistically significant.

Model performance. Accuracy, precision-recall, and AUC were evaluated via 10-fold cross-validation (CV) and an 80/20 hold-out. On the CV folds the model achieved mean accuracy $\approx$0.90 and mean AUC 0.94 $\pm$ 0.01. On the held-out set accuracy was 0.92 and AUC 0.95, indicating no over-fitting.

Figure 2a shows the confusion matrix (aggregated over the CV folds). A total of 37 GBMs were not detected (false-negatives) and 73 LGGs were incorrectly flagged as GBM (false-positives), yielding sensitivity 94%, specificity 90%, and an F1-score for GBM of 0.91. Figure 2b plots the ROC curve; the curve bows sharply toward the upper-left, with AUC 0.95, confirming excellent

discrimination[7], [8]. Thresholds could be tuned - for example, a lower cut-off would raise sensitivity to $\sim$99% at a modest cost in specificity.



(a) Confusion matrix

(b) ROC curve (AUC = 0.95)

Fig. 2: Predictive performance of the final logistic model (10-fold CV).

Statistical significance. The overall likelihood-ratio test versus an intercept-only model yielded $\chi^2(6) = 420$ ($p < 10^{-16}$). Wald tests for individual coefficients are shown in Table II. Calibration plots by decile showed no obvious miscalibration.

A simpler baseline model with only Age and IDH1 achieved AUC $\approx$0.90; adding EGFR, PTEN, ATRX and TP53 improved classification of IDH1-wildtype cases, raising AUC by five percentage points.

In summary, a six-feature logistic regression delivers high discriminatory power for glioma grading while retaining clinical interpretability.

V. Model Assessment and Diagnostics

We verified the key logistic regression assumptions: *i)* binary outcome, *ii)* independence of observations, *iii)* linearity of the logit for continuous predictors, *iv)* absence of problematic multicollinearity, and *v)* lack of influential outliers.

Binary outcome and independence. The response `Grade` is binary (0 = LGG, 1 = GBM). Each record corresponds to a unique patient, so observations are treated as independent.

Linearity of the logit. The only continuous predictor in the final model is `Age_at_diagnosis`. A Box–Tidwell interaction term (Age $\times \log(\text{Age})$) was not significant ($p > 0.05$), and visual inspection of logit–age plots did not suggest substantial nonlinearity. Therefore, no transformation was required.

Multicollinearity. We computed the Variance Inflation Factor (VIF) for each predictor. For predictor X_j,

$$\text{VIF}_j = \frac{1}{1 - R_j^2},$$

where R_j^2 is obtained by regressing X_j on all other predictors [9]. All VIFs were modest (e.g., TP53: 1.48; EGFR: 1.16; PTEN: 1.20; all others < 2), indicating no problematic multicollinearity.

Influential observations. We examined Cook's distance to identify influential points. For generalized linear models, Cook's distance for observation i is:

$$D_i = \frac{r_i^2 \, h_{ii}}{p(1-h_{ii})^2},$$

where r_i is the Pearson residual, h_{ii} is the leverage (the ith diagonal of the hat matrix), and p is the number of parameters [10], [11]. The maximum D_i was 0.0239. Using the screening rule $4/n \approx 0.0048$, 74 observations exceeded this nominal threshold, but all D_i values were far below 1. Re-fitting the model without these points led to negligible changes in coefficients, so all observations were retained.

Calibration. Grouping predicted probabilities into deciles and comparing observed vs. expected outcomes did not reveal obvious miscalibration.

Overall, these diagnostics indicate that logistic regression assumptions were reasonably satisfied and that the final model is stable.

VI. Final Model and Plots

Bringing together the analysis, the final logistic regression model can be summarized by the following equation for the log-odds of a tumor being GBM:

$$\begin{aligned} \text{logit Pr(GBM)} = & -4.3 + 0.054\,(\texttt{Age\_at\_diagnosis}) \\ & - 3.1\,(\texttt{IDH1\_mut}) - 1.2\,(\texttt{ATRX\_mut}) \\ & - 0.59\,(\texttt{TP53\_mut}) + 1.1\,(\texttt{EGFR\_mut}) \\ & + 0.78\,(\texttt{PTEN\_mut}) \end{aligned} \tag{1}$$

where `Gene_mut` indicates the mutation indicator (1 if mutated, 0 if not). All coefficients are in units of log-odds. For example, holding other variables constant, a 10-year increase in age adds $10 \times 0.054 = 0.54$ to the log-odds of GBM (OR increase by a factor of $e^{0.54} \approx 1.72$). An IDH1-mutant tumor subtracts 3.1 from the log-odds of GBM (OR $= 0.045$, dramatically lower odds).

This equation can be used to calculate an individual patient's estimated probability of having a GBM. For instance, consider a 50-year-old patient with an IDH1-mutated, ATRX-mutated tumor (likely an astrocytoma): log-odds $= -4.3 + 0.054(50) - 3.1 - 1.2 + 0 + 0 + 0 = -4.3 + 2.7 - 4.3 = -5.9$, giving $p = \frac{1}{1+e^{5.9}} \approx 0.0027$, essentially a 0.27% probability of GBM (i.e., almost certainly an LGG).

In contrast, a 50-year-old with an IDH1-wildtype, EGFR- and PTEN-mutated tumor: log-odds $= -4.3 + 2.7 + 0 + 0 + 1.1 + 0.78 = 0.28$, $p = 0.57$ or 57% chance of GBM (which is quite high given 50 is somewhat young; if the age were older this probability would be even higher). These examples illustrate how the model combines age and molecular features to stratify risk. The probabilities can be presented to clinicians to inform the expected grade before surgery or aggressive treatment decisions.

Plots: Figure 2 contains the two key performance visualizations that support our results: panel(a) shows the confusion matrix; panel(b) shows the ROC curve. We also generated diagnostic plots (not shown here) such as residual vs. fitted plots and Cook's distance charts, which as discussed

did not reveal obvious assumption violations or concerning outliers. We focus on the primary results plots for brevity.

It is worth noting that the confusion matrix (Figure 2a) highlights that the few errors the model makes tend to be LGG cases predicted as GBM. Investigating those cases revealed that some were indeed "edge" cases (e.g., very high age LGG or IDH-wildtype lower-grade tumors that clinically behave more like GBM). This suggests the model is sometimes inclined to err on the side of predicting higher grade, which in a clinical setting might be considered a safer mistake (a false alarm leading to further testing) than missing a high-grade tumor.

Figure 2b reinforces the high true-positive rate even at low false-positive rates, showing the model can be tuned to various operating points depending on whether one prioritizes sensitivity or specificity. Our chosen threshold (0.5) gave a balanced performance, but if a clinician wanted to be nearly certain not to miss any GBM, they could lower the threshold to say 0.3, then sensitivity would approach $\sim 99\%$ with some trade-off in specificity, which is easily interpretable from the ROC curve.

VII. Conclusion

Recap from EDA. Exploratory analysis showed GBM patients were older on average (60.70 vs 43.87 years), and IDH1 mutations were far more frequent in LGG (78.23% vs 6.53%). In contrast, EGFR and PTEN mutations were more common in GBM (23.01% and 32.95%, respectively, vs 6.37% and 5.13%). These patterns suggested that Age and these mutations would be key predictors.

Model findings. The final logistic regression model retained six predictors: `Age_at_diagnosis`, `IDH1`, `ATRX`, `TP53`, `EGFR`, and `PTEN`. Estimated odds ratios (OR) were: IDH1 (0.045), ATRX (0.30), TP53 (0.55), EGFR (3.0), PTEN (2.2), and 1.06 per year increase in age. In this dataset, these estimates indicate that tumors with IDH1 or ATRX mutations had lower predicted odds of being GBM, whereas EGFR or PTEN mutations (and higher age) were associated with higher predicted odds.

Performance and diagnostics. The model achieved an AUC of about 0.94 and accuracy of about 90% in cross-validation, and an AUC of 0.95 with 92% accuracy on the held-out test set. Diagnostic checks (Cook's $D_{\max} = 0.0239$, VIF < 2, no severe miscalibration in probability plots) suggest that assumptions were reasonably satisfied and no single observation unduly influenced the results.

Interpretation and limitations. These results reflect associations observed within the TCGA dataset and do not imply causation. Some relevant biomarkers (e.g., TERT promoter mutations) were not directly included, and external validation on independent cohorts would be needed to assess how well these findings generalize beyond this sample.

Final remark. A parsimonious, interpretable logistic regression using six features achieved strong predictive performance on this dataset. This suggests that combining clinical and molecular variables in a simple statistical model may help support risk stratification in neuro-oncology.

References

[1] C. K. K. A. V. Tasci, Erdal and Y. Zhuge, "Glioma Grading Clinical and Mutation Features," UCI Machine Learning Repository, 2022, DOI: https://doi.org/10.24432/C5R62J.

[2] R. Sánchez-Marqués, V. García, and J. S. Sánchez, "A data-centric machine learning approach to improve prediction of glioma grades using low-imbalance tcga data," *Scientific Reports*, vol. 14, p. 17195, 2024. [Online]. Available: https://doi.org/10.1038/s41598-024-68291-0

[3] N. Bhedasgaonkar and R. K. Joshi, "Hcvr: A hybrid approach with correlation-aware voting rules for feature selection," 2025. [Online]. Available: https://arxiv.org/abs/2507.02073

[4] M. Zhang and K. Liu, "On regularized sparse logistic regression," 2023. [Online]. Available: https://arxiv.org/abs/2309.05925

[5] E. Y. Chang, "Knowledge-guided data-centric ai in healthcare: Progress, shortcomings, and future directions," 2023. [Online]. Available: https://arxiv.org/abs/2212.13591

[6] T. Yang and Y. Ying, "Auc maximization in the era of big data and ai: A survey," *ACM Comput. Surv.*, vol. 55, no. 8, Dec. 2022. [Online]. Available: https://doi.org/10.1145/3554729

[7] A. Jambulapati, J. Li, T. Schramm, and K. Tian, "Robust regression revisited: Acceleration and improved estimation rates," 2021. [Online]. Available: https://arxiv.org/abs/2106.11938

[8] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006, rOC Analysis in Pattern Recognition. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016786550500303X

[9] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley, 1980.

[10] D. Pregibon, "Logistic Regression Diagnostics," *The Annals of Statistics*, vol. 9, no. 4, pp. 705 – 724, 1981. [Online]. Available: https://doi.org/10.1214/aos/1176345513

[11] R. D. Cook, "Detection of influential observation in linear regression," *Technometrics*, vol. 19, no. 1, pp. 15–18, 1977. [Online]. Available: https://doi.org/10.1080/00401706.1977.10489493