Explainable machine learning models for glioma subtype classification and survival prediction

Olga Vershinina[1,2,*], Victoria Turubanova[1,2,3], Mikhail Krivonosov[1,2], Arseniy Trukhanov[4] and Mikhail Ivanchenko[1,2]

[1] Research Center in Artificial Intelligence, Institute of Information Technologies, Mathematics and Mechanics, Lobachevsky State University, Nizhny Novgorod 603022, Russia
[2] Institute of Biogerontology, Lobachevsky State University, Nizhny Novgorod 603022, Russia
[3] Department of Genetics and Life Sciences, Sirius University, Sochi 354340, Russia
[4] Mriya Life Institute, National Academy of Active Longevity, Moscow 124489, Russia
[*] Correspondence: olya.vershinina@itmm.unn.ru

Abstract: Gliomas are complex and heterogeneous brain tumors characterized by an unfavorable clinical course and a fatal prognosis, which can be improved by an early determination of tumor kind. Here, we develop explainable machine learning (ML) models for classifying three major glioma subtypes (astrocytoma, oligodendroglioma, and glioblastoma) and predicting survival rates based on RNA-seq data. Thirteen key genes (*TERT*, *NOX4*, *MMP9*, *TRIM67*, *ZDHHC18*, *HDAC1*, *TUBB6*, *ADM*, *NOG*, *CHEK2*, *KCNJ11*, *KCNIP2*, and *VEGFA*) proved to be closely associated with glioma subtypes as well as survival. The Support Vector Machine (SVM) turned out to be the optimal classification model with the balanced accuracy of 0.816 and the area under the receiver operating characteristic curve (AUC) of 0.896 for the test datasets. The Case-Control Cox regression model (CoxCC) proved best for predicting survival with the Harrell's C-index of 0.809 and 0.8 for the test datasets. Employing SHapley Additive exPlanations (SHAP) method we reveal the gene expression influence on the outputs of both models, thus enhancing the transparency of the prediction generation process. The results indicate that the developed models could serve as a valuable practical tool for clinicians, assisting them in diagnosing and determining optimal treatment strategies for patients with glioma.

Keywords: glioma; gene expression data; machine learning; explainable artificial intelligence; subtype classification; overall survival prognosis

1. Introduction

Gliomas constitute a group of primary tumors of the central nervous system (CNS) that are widespread in the human population and represent approximately 24% of all primary brain and CNS tumors, as well as 80.9% of malignant tumors [1]. The latest classification by the World Health Organization (WHO) of CNS tumors [2] assigns diffuse gliomas of adult type according to their molecular and genetic characteristics into three main subtypes: astrocytomas (IDH-mutant; II-IV grades), oligodendrogliomas (IDH-mutant and 1p/19q-codeleted; II and III grades), and glioblastomas (IDH-wildtype; IV grade). The typical treatment for glioma involves surgical removal of the tumor, followed by a combination of chemotherapy and radiation therapy. Existing studies show significant differences in overall survival (OS) and response to treatment among molecular subtypes of glioma [3–7].

Accurate determination of the glioma subtype is crucial to personalize treatment strategy, but the ambiguity of histomorphological analysis and the high heterogeneity of the tumors complicate the task [8]. For example, there are well-documented cases of disagreement between different specialists on the histological diagnosis of glioma [9]. In addition, the mutational status may vary in different parts of tumor [10,11], and between patients with the same tumor subtype [12,13]. Therefore, additional strategies and approaches for the identification of new biological markers and therapeutic targets, as well as for clinical diagnostics are required.

Recent advances in high-throughput sequencing have revolutionized the study of molecular changes at the genome-wide level. Sequencing technologies have shown their high performance in diagnosing and classifying tumors, leading to the inclusion of molecular features in the WHO classification in addition to established tumor characterization

approaches [2]. Both statistical methods and machine learning (ML) algorithms, including artificial intelligence (AI), are promising tools to analyze massive sequencing data, in particular, for various classification and regression tasks.

ML models based on gene expression data have been developed to classify glioblastomas into proneural, classical, neural, and mesenchymal types [14]; to discriminate glioblastoma, diffuse astrocytoma, and anaplastic astrocytoma [15]; to subtype and grade low grade gliomas (LGG), such as astrocytoma, oligodendroglioma and oligoastrocytoma [16]; to predict OS in patients with glioblastoma multiforme (GBM) [17], LGG [18], glioma of any grade of malignancy [19]; and many others. In [20], multi-omics data including mRNA expression and DNA methylation were used to identify biomarkers and classify three subtypes of glioma (astrocytoma, oligodendroglioma, and glioblastoma). A similar combination of RNA sequencing and DNA methylation data was used to develop deep learning classifiers for LGG and GBM subtypes [21].

Often, ML models have complicated and "black-box" architectures that impede the interpretation and explanation of their predictions, undermining end-user trust and application at bedside [22]. Developing eXplainable AI (XAI) systems [23] is particularly necessary in healthcare because such models make critical predictions that can significantly impact treatment and disease outcomes [24]. However, the current number of explainable AI models for glioma classification remains limited, and most of them perform MRI image analysis, to name classifying gliomas into two classes (low and high grade) [25,26], classifying astrocytoma grades [27], predicting OS in patients with glioblastoma [28], and a few others.

In this paper we develop an explainable ML model based on SHapley Additive exPlanations (SHAP) that classifies three main glioma subtypes — astrocytoma, oligodendroglioma, and glioblastoma — based on RNA-seq data. The analysis of the obtained classification model produces an informative set of biological markers (genes) associated with these subtypes. Furthermore, we identify that this set of genes also associated with patients survival, and train the second explainable model for predicting OS in patients with glioma, universal with respect to subtype. Gene expression patterns contributing to models glioma subtype classification and survival prediction are descibed and analyzed.

2. Materials and Methods

2.1. Data Collection

The bulk RNA-seq data of glioma patients were obtained from two independent data portals. Datasets mRNAseq_693 and mRNAseq_325 were retrieved from the the Chinese Glioma Genome Atlas (CGGA) [29] (http://www.cgga.org.cn/), whereas datasets TCGA-LGG and TCGA-GBM were sourced from The Cancer Genome Atlas (TCGA) [30] (https://portal.gdc.cancer.gov/). For each dataset, STAR-count files with the number of mapped RNA-seq reads per gene, along with corresponding clinical information files, were downloaded. Primary tumor samples were collected for three glioma subtypes: astrocytoma (class 0), oligodendroglioma (class 1) and glioblastoma (class 2). Oligoastrocytoma (mixed glioma) samples were excluded from consideration because this glioma subtype is notoriously difficult to differentiate [31]. The mRNAseq_693 dataset was selected as the training set, and the mRNAseq_325 dataset was randomly divided equally into validation and test sets using class stratification. For additional external testing of the models, a dataset obtained by combining TCGA-LGG and TCGA-GBM data was used.

2.2. Data Preprocessing

All datasets were checked for missing values and duplicate samples. Genes common across datasets and identified as protein-coding based on the genome annotations used by the data authors (GENCODE v19 for CGGA data and GENCODE v36 for TCGA data) were selected for further analysis. Genes whose expression was zero in all samples of the training set, as well as genes whose variance in the training set was zero, were removed from the data. The resulting number of genes after filtering was 17588. Since RNA-seq

data have specific properties such as extreme values and dependence of the mean on variance (heteroscedasticity), all samples were logarithmic transformed $log_2(counts + 1)$ to make the data more suitable for ML algorithms. It is worth noting that no special methods of sequencing data normalization, such as FPKM or TPM, were used, since such normalization often worsens the results in disease diagnostics and classification tasks [32]. To eliminate technical variations and batch effects between datasets, batch correction was performed using the *ComBat* function [33] from the *sva* R package v.3.50.0 [34], where the mRNAseq_693 dataset (training set) was chosen as the reference batch. Finally, z-score normalization (standardization) to zero mean and unit standard deviation was applied to the training data. The validation and test sets were normalized using the means and standard deviations of the training set.

2.3. Feature Selection

To reduce the dimensionality of the input data space, filter-based feature selection methods were employed. These methods determine the importance of input features by evaluating their relationship with the target variable—-specifically, in our experiments, we examine how gene expression is associated with the class label. Based on the obtained importance ratings, features can be ranked, enabling the selection of the most significant ones for classification. We applied two methods: Mutual Information (MI) [35] and Tuned ReliefF (TuRF) [36]. To compute MI, we used the *mutual_info_classif* function from the *scikit-learn* Python package v.1.5.1 [37], which uses an entropy estimate based on the distances to the nearest neighbors. To calculate feature importance scores using the TuRF method, we trained the *TuRF* model from the *scikit-rebate* Python package v.0.62 [38]. This model recursively identifies the most important features by examining differences between nearest neighbors, increasing scores when feature values vary between pairs of nearest neighbors that belong to different classes. Feature selection was conducted on the training set, with the number of nearest neighbors kept at the default value for both methods.

2.4. ML Models for Classification

Various ML models, both simple and advanced, were evaluated to classify glioma samples into three classes. The k-Nearest Neighbors (kNN) algorithm [39] assigns objects to the class of the majority of their k-nearest neighbors in the training dataset, based on distance metrics like Euclidean distance. The Support Vector Machine (SVM) [40] constructs an optimal hyperplane maximizing the margin between classes, using an "one-vs-rest" strategy for multi-class classification. Decision tree ensembles based on the idea of bagging, including Random Forest (RF) [41] and Extremely Randomized Trees (ERT) [42], train independent decision trees on random bootstrap samples and average their predictions. Decision trees can be combined into an ensemble through boosting, which involves sequentially training base estimators to correct prior errors. Notable Gradient Boosted Decision Trees (GBDT) models include eXtreme Gradient Boosting (XGBoost) [43], Light Gradient Boosting Machine (LightGBM) [44], and Categorical Boosting (CatBoost) [45]. Deep Neural Networks (DNNs) are complex AI systems composed of interconnected layers of nodes, inspired by the structure of the biological neural network. We examined lightweight DNN architectures such as the Attentive Interpretable Tabular Learning Neural Network (TabNet) [46] and the Gated Adaptive Network for Deep Automated Learning of Features (GANDALF) [47].

Software implementations of the models were sourced from the following Python packages: *scikit-learn* v.1.5.1 [37] (kNN, SVM, RF, and ERT), *xgboost* v.2.0.3 [43] (XGBoost), *lightgbm* v.4.3.0 [44] (LightGBM), *catboost* v.1.2.5 [45] (CatBoost), and *pytorch_tabular* v.1.1.0 [48] (TabNet and GANDALF).

To evaluate the performance of classification models on imbalanced data, balanced accuracy (BA) [49] served as the primary metric. The Area Under the Receiver Operating Characteristic (ROC) curve (AUC) was used as an additional performance metric, and the LogLoss metric, which measures the divergence between predicted probabilities and actual classes, was employed to gauge the calibration of the models. Higher values of the BA and

AUC, along with a lower LogLoss, indicate a better performing model. All metrics were calculated using the *scikit-learn* Python package v.1.5.1 [37].

2.5. ML Models for Survival Prediction

A widely used model for examining the association between survival times and predictor variables, as well as for estimating mortality risk scores at given time points, is the Cox Proportional Hazards (CoxPH) regression model [50]. Since Cox regression depends on a linear combination of covariates, it may struggle to capture nonlinear effects of risk factors on survival, prompting the increasing adoption of more advanced techniques in survival analysis. Random Survival Forest (RSF) [51] and Extra Survival Trees (EST) are ensemble methods that combine multiple survival trees constructed using bootstrap samples and operate similarly to RF and ERT but are adapted for censored survival data. There is also a generalization of the gradient boosting model XGBoost, called XGBoost Survival Embeddings (XGBSE) [52], where an ensemble of trees determines the feature transformation (embedding) of the input data, enabling training of models like logistic regression (the XGBSEDebiasedBCE model, abbreviated as XGBSE-DBCE) or kNN (the XGBSEKaplanNeighbors model, abbreviated as XGBSE-KN). DNN architectures are also used to predict survival. Neural networks such as the Cox proportional hazards deep neural network (DeepSurv) [53] and the Case-Control Cox regression model (CoxCC) [54] are nonlinear extensions of Cox regression and differ in their loss functions, which are minimized. The Piecewise Constant Hazard model (PC-Hazard) [55] assumes that the continuous-time hazard function is constant within predefined intervals. The Neural Multi-Task Logistic Regression model (N-MTLR) [56] is a discrete-time approach that models the hazard function by combining multiple locally dependent logistic regression models, leveraging nonlinear data transformations from a multilayer perceptron.

Software implementations of the models were sourced from the following Python packages: *scikit-survival* v.0.23.0 [57] (CoxPH, RSF, and EST), *xgbse* v.0.3.1 [52] (XGBSE-DBCE and XGBSE-KN), and *pycox* v.0.3.0 [54] (DeepSurv, CoxCC, PC-Hazard, and N-MTLR).

The Harrell's concordance index (C-index) [58], which measures the rank correlation between predicted and observed outcomes, was used as the primary metric to evaluate the performance of survival prediction models. We also analyzed the time-dependent ROC curve [59], a modified version of the traditional ROC curve for time-to-event data, and the associated AUC values. Furthermore, we evaluated model calibration using the Brier score [60], calculating the Integrated Brier Score (IBS) for a comprehensive assessment across all time points. Higher values of the C-index and AUC, along with a lower IBS, indicate a better performing model. The C-index and IBS were computed using the respective Python packages for models. Time-dependent ROC analyses for all models were performed using the *survivalROC* R package v.1.0.3.1 [59].

2.6. Tuning and Training ML Models

The search for optimal hyperparameters of the classification and survival models was performed using the multivariate Tree-structured Parzen Estimator (TPE) algorithm with the group decomposed search space [61,62] (the software implementation was sourced from the *optuna* Python package v.3.6.1 [63]). Supplementary Table S1 lists the tunable hyperparameters for each model, including their descriptions and search distributions. The total number of optimization trials for each experiments was set to 500. The best trial was defined as the trial with the maximum metric value (BA for classification problem, C-index for survival prediction) obtained using the stratified 5-fold cross-validation procedure on the training dataset. For the classification task, stratification was based on the class label, while for the survival prediction task, stratification was based on the survival status. The selected models with optimal hyperparameters were then tested on an independent internal test dataset and an external test dataset to determine the model with the best performance.

A pre-allocated validation dataset (half of mRNAseq_325 dataset) was used to prevent overfitting of GBDT and DNN models through early stopping: the maximum number of

rounds of the training process was set to 100, and the number of early stopping rounds was set to 10 (the training process ends if the evaluation metric in the validation dataset was not improving for 10 rounds). For training DNN models, the Adam optimization algorithm [64] was used and the batch size was fixed at 32. For the TabNet and GANDALF models, the option of automatic selection of the optimal learning rate was enabled, whereas for the DeepSurv, CoxCC, PC-Hazard and N-MTLR models, the learning rate was determined through hyperparameter tuning.

2.7. Functional Enrichment Analysis

Gene Ontology (GO) enrichment analysis of the gene set was performed using the g:Profiler online tool [65] (https://biit.cs.ut.ee/gprofiler/gost). GO terms were considered statistically significantly enriched if the adjusted p-value after Benjamini–Hochberg FDR correction was less than 0.001. In addition, redundant terms were eliminated using the greedy search strategy implemented in g:Profiler.

2.8. Statistical Analysis of Survival

All statistical analyses were performed using the *lifelines* Python package v.0.28.0 [66]. A univariate Cox regression analysis was conducted to evaluate the association between gene expression levels and overall survival (OS) of the patients. An adjusted p-value (after Benjamini–Hochberg FDR correction) less than 0.05 was considered statistically significant. Kaplan-Meier survival curves and the log-rank test were used to compare differences in OS between the two groups. The low- and high-mortality risk groups were defined based on the median risk score from the training dataset.

2.9. eXplainable Artificial Intelligence

The SHAP (SHapley Additive exPlanations) explainability method is a game theory-based approach that quantifies the impact of features on predictions by assigning Shapley values to them [67]. The SHAP method is particularly useful for interpreting complex "black-box" ML models. SHAP values are derived by averaging the contributions of a feature to the model's predictions, considering all possible permutations of other features within the model. SHAP values show how changing a specific feature value for a given data point affects the model's base prediction (which is usually the average prediction over the training dataset). Using SHAP method, both global and local explainability can be obtained. Global explainability refers to the analysis of aggregate contributions from features in a dataset. This helps to understand the overall patterns of influence of various features on the model's predictions as a whole. Local explainability focuses on explaining the contribution of features to individual predictions in order to provide an understanding of the reasons behind the specific decisions made by the model. To calculate the SHAP values, the *Explainer* function from the *shap* Python package v.0.45.1 was used with the training dataset as the background data.

3. Results

3.1. Patient Characteristics

ML models for the diagnosis of glioma are based on the publicly accessible RNA sequencing data from primary glioma samples that were acquired from the CGGA and TCGA portals. The clinical features of the assembled patient cohorts are summarized in Table 1.

3.2. Selection of Features Associated With Glioma Subtypes

RNA-seq data represent high-dimensional, low-sample-size (HDLSS) data that increase the computational expenses during training of ML models and frequently result in lowering prediction accuracy due to the presence of redundant and irrelevant features (genes) in the data. To overcome the challenge, we apply two filtering methods to reduce the dimensionality of the feature space. Mutual Information (MI) criterion and the TuRF

Table 1. Clinical features of glioma patient cohorts included in the study.

Characteristic	CGGA, mRNAseq_693 (training set, n = 398)	CGGA, mRNAseq_325 (validation/test set, n = 229)	TCGA, TCGA-LGG and TCGA-GBM (external test set, n = 536)
Histology			
Astrocytoma	167	84	193
Oligodendroglioma	91	60	190
Glioblastoma	140	85	153
Age, years			
Range	11–76	10–79	17–89
Mean ± std	44.22 ± 12.97	43.79 ± 12.29	48.19 ± 15.25
Gender			
Male	231	142	311
Female	162	87	225
Survival status			
Alive	179	85	313
Dead	204	139	221
OS, days			
Range	27–4725	19–4809	1–6423
Median	1283	1118	544

OS: overall survival.

algorithm give feature importance scores that quantify the relationship between feature values and class labels (class 0—astrocytoma, class 1—oligodendroglioma, and class 2—glioblastoma). In each case, the features were sorted in descending order of importance score, and 1% of top features (176 features) was selected. The intersection of the two resulting lists contained 84 features that were left for further analysis.

3.3. Classifier for Glioma Subtypes

We applied nine ML models to classify glioma subtypes based on the normalized expression of the 84 selected genes. For each model, hyperparameter optimization was performed, followed by training the models with their optimal combinations. The performance of the models was evaluated using the balanced accuracy (BA) metric, which was estimated through stratified 5-fold cross-validation and also assessed on both the internal and external test datasets (Table 2).

Table 2. Results of glioma subtype prediction based on 84 genes. Models are ranked based on their balanced accuracy computed on the test dataset in descending order.

Model	BA, train	BA, CV	BA, test	BA, external test
GANDALF	0.793	**0.824**	**0.808**	0.785
kNN	**0.802**	**0.805**	**0.805**	0.794
SVM	**0.843**	**0.823**	**0.805**	**0.821**
LightGBM	**0.852**	**0.833**	0.778	0.781
TabNet	**0.830**	**0.826**	0.777	0.784
XGBoost	**0.843**	**0.826**	0.775	0.775
RF	**0.934**	**0.821**	0.772	0.780
ERT	**0.917**	**0.827**	0.767	0.788
CatBoost	0.783	**0.842**	0.714	0.728

BA: balanced accuracy, CV: cross-validation. Values greater than 0.8 are highlighted in bold.

All models exhibit high cross-validation accuracy (above 0.8), and only three models (GANDALF, kNN, and SVM) demonstrated accuracy above 0.8 on an internal test dataset that was not included in the training process. However, only the SVM model achieved

a performance greater than 0.8 on the external test dataset, manifesting a superior level of generalization. Besides, the SVM model has one of the highest AUC values among the others (see Supplementary Table S2). Furthermore, the SVM model demonstrates one of the lowest LogLoss scores, suggesting good calibration and a strong agreement between its predicted probabilities and the actual class labels.

To reduce the number of variables in the SVM model and improve its applicability, an additional round of feature selection was conducted. We sorted the 84 features in descending order according to their importance for the global explainability, as determined by SHAP values. The global importance of features was estimated by averaging the absolute SHAP values computed for the training dataset. We then examined a series of feature subsets comprising the top $1, 2, \ldots, 84$ most important features. For each subset, hyperparameter tuning, training and testing of the SVM model were conducted. We achieved the highest BA on the test dataset, $BA_{test} = 0.816$, when the top 13 genes (*TERT*, *NOX4*, *MMP9*, *TRIM67*, *ZDHHC18*, *HDAC1*, *TUBB6*, *ADM*, *NOG*, *CHEK2*, *KCNJ11*, *KCNIP2*, and *VEGFA*) were used to train the SVM classifier with polynomial kernel (Figure 1A). The classification accuracy calculated on the training dataset via cross-validation and on the external test dataset was also high: $BA_{CV} = 0.837$ and $BA_{external\ test} = 0.816$, respectively. Therefore, the 13-gene SVM model was selected as the optimal model for glioma classification.

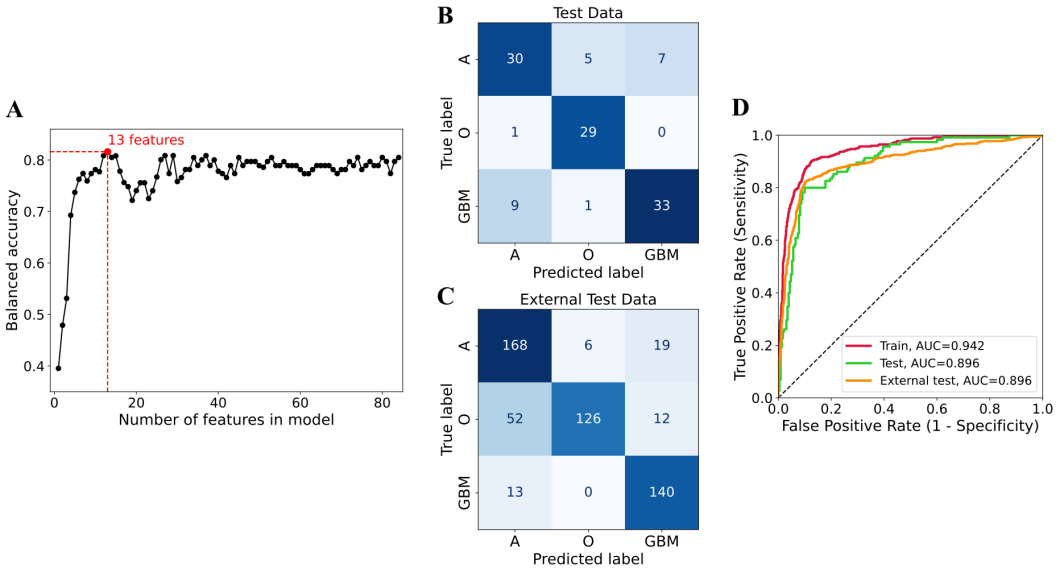

Figure 1. Establishment and validation of the optimal classification SVM model for three glioma subtypes. (**A**) Dependence of the balanced accuracy calculated on the internal test dataset on the number of features in the model. Dotted lines correspond to the optimal small model with an accuracy 0.816 and 13 features. (**B**) Confusion matrix for the optimal 13-gene model on the internal test dataset. (**C**) Confusion matrix for the optimal 13-gene model on the external test dataset.(**D**) ROC curves for the optimal 13-gene model. ROC curves were generated using training, internal test and external test datasets. A: Astrocytoma, O: Oligodendroglioma, GBM: Glioblastoma.

The confusion matrices for the resulting classifier model are shown in Figure 1B,C. In both test datasets, the tumor subtypes of most patients are predicted correctly, as indicated by the high values of diagonal elements. We observed misclassifications between astrocytomas and glioblastomas, with astrocytoma sometimes being classified as glioblastoma and vice versa too. Some oligodendrogliomas were misclassified as astrocytomas by our model on the external test dataset.

The relationship between sensitivity and specificity of the classification model is presented in Figure 1D by the ROC curves. The AUC values for the test datasets are high

($AUC_{test} = AUC_{external\ test} = 0.896$), indicating that the 13-gene SVM model can classify glioma subtypes with high specificity and sensitivity, and has a potential at bedside.

3.4. Functional Analysis of Prognostic Genes

The GO term enrichment analysis revealed that in the "biological processes" category, the 13 prognostic genes were significantly enriched in positive regulation of biological process (GO:0048518), regulation of cell communication (GO:0010646), system development (GO:0048731), apoptotic process (GO:0006915), response to abiotic stimulus (GO:0009628), positive regulation of cell population proliferation (GO:0008284), cardiac muscle tissue development (GO:0048738), and muscle cell proliferation (GO:0033002).

3.5. Model for Survival Prediction

Next, we examined the relationship between the 13 identified genes and overall survival (OS) rates for the patients. The maximum follow-up time in the training dataset was 4725 days (12.9 years). Patients from the test datasets whose survival times were outside the ranges of the training dataset were excluded. According to the results of the univariate Cox regression analysis, all genes demonstrated significant associations with OS in both the training and test cohorts, cf. Supplementary Table S3. Hazard ratios (HR) less than 1 for covariates *TRIM67*, *NOG*, *KCNJ11*, and *KCNIP2* indicate the negative association with the probability of death in patients with glioma, while HR greater than 1 for the remaining covariates correspond to the positive association with mortality.

Led by these observations we made use of 13 previously isolated genes to construct a predictive model for survival. We tuned hyperparameters and trained nine ML models, encompassing both traditional multivariate Cox regression and more sophisticated algorithms like survival forests and neural networks. Models performance evaluation is presented in Table 3.

Table 3. Results of glioma patient survival prediction based on 13 genes. Models are ranked in descending order according to their C-index scores computed on the test dataset.

Model	C-index, train	C-index, CV	C-index, test	C-index, external test
RSF	**0.851**	0.781	**0.815**	**0.816**
PCHazard	**0.811**	0.787	**0.814**	0.783
EST	**0.810**	0.779	**0.813**	**0.815**
CoxPH	0.783	0.765	**0.812**	**0.803**
CoxCC	**0.802**	0.790	**0.809**	**0.800**
XGBSE-DBCE	**0.919**	0.778	**0.804**	0.799
XGBSE-KN	**0.858**	0.784	**0.800**	**0.812**
N-MTLR	**0.808**	0.787	0.796	0.784
DeepSurv	**0.800**	0.787	0.787	0.786

CV: cross-validation. Values greater than 0.8 are highlighted in bold.

RSF, EST, CoxPH, CoxCC, and XGBSE-KN models consistently achieved high C-index values exceeding 0.8 across both internal and external test datasets. However, the models with the highest C-index in the test dataset (RSF, EST, and CoxPH) also exhibited a high IBS value, suggesting that they are poorly calibrated (see Supplementary Table S4). In result, the neural network CoxCC was selected as the best model based on its high C-index values on the test datasets and the highest cross-validation score for the metric ($C\text{-}index_{CV} = 0.790$), as well as displaying low IBS values.

The model was then used to predict the mortality risk score for each patient. According to the median risk score calculated on the training dataset, patients were divided into low-risk and high-risk groups. To assess the model's discriminative ability, we performed the survival analysis. It confirmed that OS in the low-risk group was significantly higher compared to the high-risk group (log-rank p-value < 0.05) in both test datasets (Figure 2A,B).

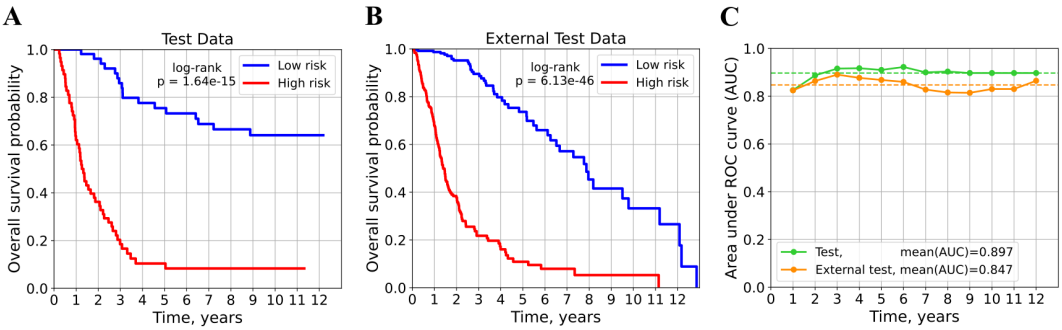

Figure 2. Validation of the CoxCC model for predicting overall survival in patients with glioma. (**A**) Kaplan–Meier survival curves for the low- and high-risk groups in the internal test dataset. (**B**) Kaplan–Meier survival curves for the low- and high-risk groups in the external test dataset. (**C**) Time-dependent ROC analysis of the prognostic model in the test datasets. The dotted lines represent the mean AUC value calculated over 12 years.

To assess the predictive performance of the model further, we conducted a time-dependent ROC analysis by calculating the areas under the ROC curves for predicting 1- to 12-year risk scores (Figure 2C). This approach allowed us to estimate how accurately the risk score determined by the model can discriminate patients who died by a specified period (e.g., 1, 2, ..., 12 years) from those who remained alive. In the internal test dataset, the AUC values were 0.910 and 0.897 for predicting 5- and 10-year OS, respectively, and the mean AUC value over 12 years was 0.897. In the external test dataset, the values were slightly lower but remained high: the AUCs for 5- and 10-year predictions were 0.868 and 0.830, respectively, and the mean AUC value was 0.847.

The obtained results demonstrate the reliability and strong predictive power of the neural network CoxCC trained on 13 genes in forecasting the survival outcomes of glioma patients.

3.6. Explainability of ML models

To understand how the resulting models make predictions, we used the SHAP explainability method to assess the importance of features and their contributions to diagnostics. By examining SHAP values, one can determine which features have the greatest impact on particular outcomes and infer the rationale behind specific predictions. In our study, we illustrate the global and local explainability of the constructed models using samples from the internal test dataset. To calculate the SHAP values for the glioma subtype classification model, the class probabilities predicted by the SVM model were utilized. In the context of multi-class classification, the SHAP values are calculated for each individual class. For the survival analysis model, the 12-year survival probability predicted by the CoxCC model was used.

The global explainability of the models is illustrated in Figure 3. In the plots, the prognostic genes are ordered by their average contribution to the model's output. For the classification model of three glioma subtypes (Figure 3A), the contribution of features to the prediction of each class is indicated by different colors. For example, the *NOX4* gene has the greatest influence on classification, and its expression levels are more critical for predicting astrocytomas and glioblastomas than for oligodendrogliomas. For the survival prediction model (Figure 3B), gene *NOG* is the most important. Notably, the order of importance of genes for the two models differs significantly. For instance, gene *MMP9* exhibits high importance for classification but low importance for survival prediction, whereas gene *KCNJ11* demonstrates the opposite trend, showing high importance for survival prediction and low importance for classification.

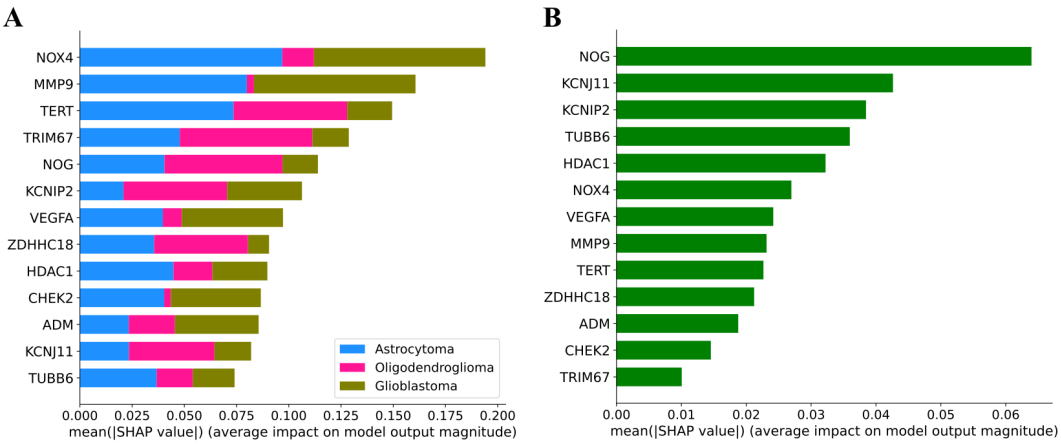

Figure 3. Average impact (the mean absolute SHAP values) of features on model predictions. Features with higher values exert stronger influence. (**A**) Global importance of features in the classification model for three glioma subtypes. For each feature, its contribution to certain classes is indicated by the corresponding color. (**B**) Global importance of features in the model for predicting overall survival of patients with glioma.

Then we constructed scatterplots to visualize the direction of the relationship between model predictions and the expression levels of both the most important genes (Figure 4) and the remaining genes (Supplementary Figures S1 and S2).

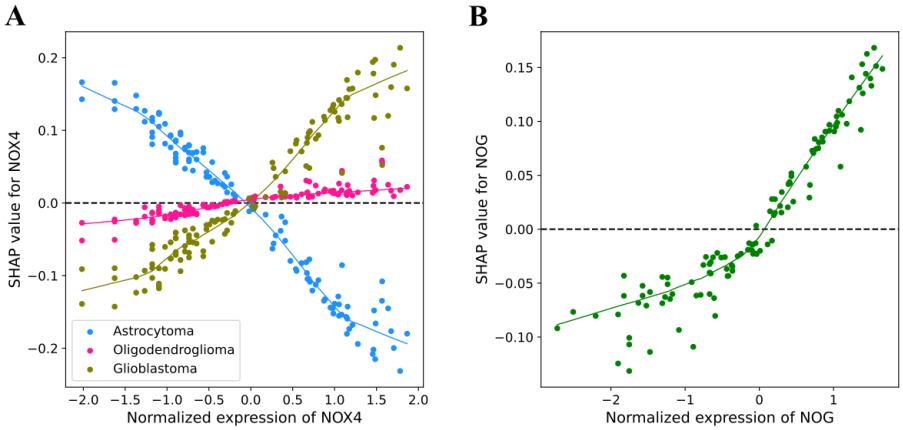

Figure 4. Scatterplots for the SHAP values and expression level of the most important genes in the models. (**A**) Dependence of SHAP values on the expression of the *NOX4* gene in the classification model for three glioma subtypes. Each patient in the test dataset corresponds to three dots, which indicate the SHAP values associated with the three glioma subtypes. Positive SHAP values for a particular class indicate that the feature value increases the probability of predicting that class, while negative values decrease it. (**B**) Dependence of SHAP values on the expression of the *NOG* gene in the model for predicting overall survival of patients with glioma. Positive SHAP values indicate that the feature value increases overall survival, while negative values decrease it.

In a classification problem, the SHAP value indicates how much the probability of predicting a certain class increases (if the SHAP value is positive) or decreases (if the SHAP value is negative), depending on the observed value of the considered feature. As shown in Figure 4A, high levels of *NOX4* expression increase the probability of predicting glioblastoma, as evidenced by the positive correlation between the SHAP values for the glioblastoma class and the gene expression values. Low *NOX4* expression values increase the probability of predicting astrocytoma, as a negative correlation is observed between the

SHAP values for the astrocytoma class and gene expression. The SHAP values for the oligodendroglioma class are close to zero, indicating a minimal impact of gene expression on this tumor subtype. Analysis of the SHAP values for the remaining genes (Supplementary Figure S1) revealed that increased expression of *ZDHHC18*, *HDAC1*, and *TUBB6* enhances the probability of predicting astrocytoma, while elevated expression of *TERT*, *TRIM67*, *NOG*, *KCNIP2*, and *KCNJ11* increases the probability of predicting oligodendroglioma. Additionally, higher expression levels of *MMP9*, *TERT*, *VEGFA*, *ZDHHC18*, *CHEK2*, and *ADM* elevate the probability of predicting glioblastoma.

For the survival prediction model, SHAP values determine the amount of increase (or decrease) in the predicted survival probability due to the given feature value. We can observe that as the expression of the *NOG* gene increases, the SHAP values rise monotonically, indicating improved patient survival (Figure 4B). Regarding the remaining genes (Supplementary Figure S2), increased expression of *KCNJ11* and *KCNIP2* also associated with increased survival, while increased expression of *TUBB6*, *HDAC1*, *NOX4*, *VEGFA*, *MMP9*, *TERT*, *ZDHHC18*, *ADM*, and *CHEK2* reduces survival. The relationship between the risk of mortality and the expression of *TRIM67* is highly nonlinear.

Using SHAP values, individual model predictions can be explained locally. SHAP waterfall plots in Figure 5 exemplify local explainability of our classification model, taken for a glioblastoma sample that was correctly classified by the SVM model. The graphs demonstrate how the model adjusts its probability prediction for each class based on patient-specific normalized expression values. The genes in the graphs are arranged from bottom to top according to their increasing importance to refining the prognosis. The blue (red) lines correspond to features whose values decrease (increase) the probability of the corresponding class. For the sample considered, the expression levels of 11 of 13 genes increase the probability of the "Glioblastoma" class (Figure 5C), resulting in the prediction of this tumor subtype with a probability of 0.948.

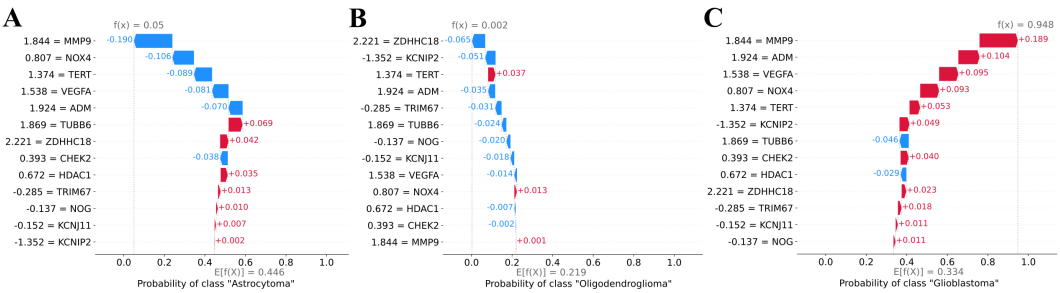

Figure 5. SHAP waterfall plots of an individual sample from the test dataset for a model classifying the glioma subtype. (**A**) Waterfall plot for class "Astrocytoma". (**B**) Waterfall plot for class "Oligodendroglioma". (**C**) Waterfall plot for class "Glioblastoma". The lower part of the graphs displays the average (baseline) probability of predicting the corresponding class in the training dataset, $E[f(X)]$. From bottom to top, features are presented in order of their increasing contribution to the prediction, with each line indicating how the feature value modifies the prediction probability. Blue lines represent negative SHAP values and show how much the feature value decreases the probability of the class, whereas red bars represent positive SHAP values and show how much the feature value increases the probability of the class. The resulting $f(x)$ values are the probabilities of the classes predicted by the model, with the highest one determining the class to which the tested sample will be assigned.

Figure 6 displays local explainability graphs of survival predictions generated by the CoxCC model for three distinct patients with glioma. The blue (red) lines correspond to features whose values decrease (increase) the survival probability. For patients who died after 450 and 1133 days, the model predicted low 12-year survival probabilities of 0.019 (Figure 6A) and 0.45 (Figure 6B), respectively. Meanwhile, for the patient who survived just over 12 years (4427 days), normalized expression levels of all genes were indicative

of prolonged survival, which led the model to predict a high survival probability of 0.84 (Figure 6C).

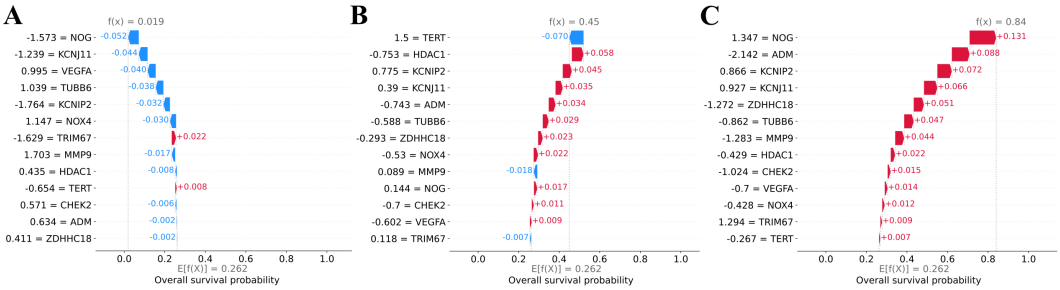

Figure 6. SHAP waterfall plots of individual glioma patients from the test dataset for a model predicting survival. (**A**) Waterfall plot for a patient who died after 450 days. (**B**) Waterfall plot for a patient who died after 1133 days. (**C**) Waterfall plot for a patient who was alive at least 4427 days later. The lower part of the graphs displays the average (baseline) predicted survival probability in the training dataset, $E[f(X)]$. From bottom to top, features are presented in order of their increasing contribution to the prediction, with each line indicating how the feature value modifies the prediction probability. Blue lines represent negative SHAP values and show how much the feature value decreases the survival probability, whereas red bars represent positive SHAP values and show how much the feature value increases the survival probability. The resulting $f(x)$ values represent the 12-year survival probabilities predicted by the model.

4. Discussion

In this study, we developed the ML model based on the expression of protein-coding genes to classify three subtypes of glioma with varying degrees of severity: astrocytoma, oligodendroglioma, and glioblastoma. Feature selection identified the most informative genes for accurate classification. Among the nine classification algorithms evaluated, the SVM model trained on 13 key genes manifested the highest performance with the balanced accuracy of 0.816 and AUC of 0.896 in both test datasets. Despited the existence of the more advanced ML algorithms, such as gradient boosting and neural networks, the SVM model has shown a high eficiency in classifying gliomas before [15,16,68]. Implementation of SHAP, an explainable AI technique, enabled visualization of the magnitude and direction of specific input variables contribution to the model predictions, both for the cohort and individual patients. Some existing explainable classification models focus mainly on categorizing gliomas into two classes based on MRI images [25,26,69]. We are aware of only one study that classifies gliomas into the same three classes as we consider and utilizes various libraries to visualize the explainability [70]. However, that work relies on gene mutation data, whereas we use RNA-seq data.

In addition, we examined the relationship between the expression levels of the 13 identified genes and patient survival, and developed the ML model to forecast the probability of survival. The resulting model based on the CoxCC neural network demonstrated high predictive capability. Thus, the C-index achieved values of 0.809 and 0.800 for the internal and external datasets, respectively. Furthermore, the 5- and 10-year AUC were 0.910 and 0.897 for the internal dataset, and 0.868 and 0.830 for the external dataset. To date, numerous studies have explored prediction glioma patient survival using RNA-seq data, although most rely primarily on the Cox regression model [17–19,71–73]. The advantage of our model lies in its exceptional predictive quality, outperforming most of the existing models, combined with the global and local explainability for the predictions it generates.

According to literature, the identified genes exhibit a strong association with the glioma subtypes studied and OS. mRNA expression levels of *NOX4* in glioblastomas (WHO grade IV) were significantly higher compared to other astrocytomas (WHO grades II and III) [74], and OS was significantly lower in patients with high *NOX4* expression [75].

MMP9 expression levels are increased in high-grade gliomas and are overexpressed in glioblastomas [76,77], while lower *MMP9* expression correlates with improved patient survival [76–78]. *KCNIP2* gene expression was elevated in stage II and III gliomas compared to stage IV gliomas, and its high expression was significantly associated with prolonged OS [79]. *KCNJ11* expression was related to the degree of tumor malignancy and was an independent factor that affected the OS, and survival times of patients with high expression of KCNJ11 are longer than those of patients with low expression [80]. Increased *VEGFA* expression is significantly associated with decreased OS in patients with GBM [81]. Interestingly, 4 genes *MMP9*, *KCNIP2*, *KCNJ11* and *VEGFA* were identified by the authors of [82] as prognostic markers for glioma-related epilepsy. Increased expression of *NOG*, included in a number of prognostic models, leads to an increase in OS in patients with glioma [83,84]. *TRIM67* is known to be a highly expressed gene in oligodendrogliomas [85]. In work [86], it was mentioned that increased expression of *ZDHHC18* leads to decreased OS in LGG. With increasing glioma grade, *ADM* mRNA expression levels were observed to rise, and this increase was associated with poorer OS [87]. *HDAC1* expression was significantly associated with the clinical stage of glioma and high *HDCA1* levels were significantly associated with reduced OS [88]. Increased *TUBB6* gene expression is associated with decreased OS in patients with primary glioblastoma [89]. We found no information on the association between the expression levels of the *TERT* and *CHEK2* genes and glioma subtypes in the scientific literature. However, it's well-established that the *TERT* mutation, commonly seen in glioblastomas and oligodendrogliomas, is included in the classification of CNS tumors [2]. Mutations in the *CHEK2* gene have been identified in glioblastomas [90,91] and glial CNS tumors in children [92].

Using an explainable AI approach (SHAP), we established relationships between gene expression and glioma subtype, as well as between expression and OS (Figure 4, Supplementary Figures S1 and S2). The identified associations are consistent with the reported scientific studies, confirming the accuracy and high biological interpretability of our models. Moreover, the ability to provide local explanations for model predictions proves valuable not only for accurate forecasts, but also for incorrect ones. In the latter case, it enables the identification of deviant features that deserve further study at the individual and molecular levels.

The presented results indicate that the use of explainable ML models can significantly improve the discrimination of glioma subtypes. However, our study has some limitations. First, the search for optimal hyperparameters of the ML models was conducted within a limited range of values close to default values, which means that the resulting models might not be globally optimal with respect to their quality metrics. Second, although there are numerous feature selection algorithms, we utilized only a select few. Third, we classified glioma subtypes solely on the basis of their histological types, disregarding the grade. However, in another study focusing on low-grade glioma subtypes, it was shown that further subdividing subtypes by grades improves the classification accuracy [16]. Fourth, to develop a model to predict patient survival in glioma, the same 13 genes used to classify the three subtypes of glioma were used. Since survival rates already differ substantially across these subtypes, future research aimed at developing survival prediction models should prioritize the identification of survival markers unique to each individual subtype. Finally, although we tested the models on independent datasets, including one from an external repository, additional large-scale datasets from multiple centers are necessary to further evaluate the robustness and generalizability of our models.

5. Conclusions

In summary, we developed and tested two explainable ML models: one to classify three glioma subtypes and the other to predict survival probability. Both models make predictions based on the expression of 13 genes. The results of the quality assessment showed that the models have high classification and discrimination accuracy and are well calibrated. The application of the SHAP approach allowed us to determine the general

contribution of the expression levels of selected genes on the predictions of both models, as well as producing explanations for individual model outputs. The resulting classification of glioma subtypes and the evaluation of patient survival could be useful to select a personalized treatment strategy and to improve prognosis at bedside.

Author Contributions: Conceptualization, O.V., V.T. and M.I.; formal analysis, O.V.; investigation, O.V. and V.T.; methodology, O.V. and M.K.; software, O.V.; supervision, A.T. and M.I.; validation, O.V.; visualization, O.V.; writing—original draft preparation, O.V.; writing—review and editing, O.V., V.T., M.K., A.T. and M.I. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Economic Development of the Russian Federation (grant No 139-15-2025-004 dated 17.04.2025, agreement identifier 000000C313925P3X0002).

Data Availability Statement: All the data used in the current study are publicly available in the CGGA (http://www.cgga.org.cn/) repository (accessed on 20 June 2022) and the TCGA (https://portal.gdc.cancer.gov/) repository (accessed on 2 August 2022).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ostrom, Q.T.; Price, M.; Neff, C.; Cioffi, G.; Waite, K.A.; Kruchko, C.; Barnholtz-Sloan, J.S. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2015–2019. *Neuro-Oncology* **2022**, *24*, v1–v95.
2. Louis, D.N.; Perry, A.; Wesseling, P.; Brat, D.J.; Cree, I.A.; Figarella-Branger, D.; Hawkins, C.; Ng, H.K.; Pfister, S.M.; Reifenberger, G.; Soffietti, R.; von Deimling, A.; Ellison, D.W. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro-Oncology* **2021**, *23*, 1231–1251.
3. Wen, P.Y.; Kesari, S. GMalignant gliomas in adults. *New England Journal of Medicine* **2008**, *359*, 492–507.
4. Omuro, A.; DeAngelis, L.M. Glioblastoma and other malignant gliomas: a clinical review. *Jama* **2013**, *310*, 1842–1850.
5. Eckel-Passow, J.E.; Lachance, D.H.; Molinaro, A.M.; Walsh, K.M.; Decker, P.A.; Sicotte, H.; Pekmezci, M.; et al. Glioma groups based on 1p/19q, IDH, and TERT promoter mutations in tumors. *New England Journal of Medicine* **2015**, *372*, 2499–2508.
6. Wijnenga, M.M.J.; French, P.J.; Dubbink, H.J.; Dinjens, W.N.M.; Atmodimedjo, P.N.; Kros, J.M.; Smits, M.; Gahrmann, R.; Rutten, G.-J.; Verheul, J.B.; Fleischeuer, R.;Dirven, C.M.F.; Vincent, A.J.P.E.; van den Bent, M,J. The impact of surgery in molecularly defined low-grade glioma: an integrated clinical, radiological, and molecular analysis. *Neuro-Oncology* **2018**, *20*, 103–112.
7. Patel, S.H.; Bansal, A.G.; Young, E.B.; Batchala, P.P.; Patrie, J.T.; Lopes, M.B.; Jain, R.; Fadul, C.E.; Schiff, D. Extent of Surgical Resection in Lower-Grade Gliomas: Differential Impact Based on Molecular Subtype. *American Journal of Neuroradiology* **2019**, *40*, 1149–1155.
8. Korshunov, A.; Sycheva, R.; Golanov, A. Molecular stratification of diagnostically challenging high-grade gliomas composed of small cells: the utility of fluorescence in situ hybridization. *Clinical cancer research* **2004**, *10*, 7820–7826.
9. Van Den Bent, M.J. Interobserver variation of the histopathological diagnosis in clinical trials on glioma: a clinician's perspective. *Acta neuropathologica* **2010**, *120*, 297–304.
10. Sottoriva, A.; Spiteri, I.; Piccirillo, S.G.; Touloumis, A.; Collins, V.P.; Marioni, J.C.; Curtis, C.; Watts, C.; Tavaré, S. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proceedings of the National Academy of Sciences* **2013**, *110*, 4009–4014.
11. Kumar, A.; Boyle, E.A.; Tokita, M.; Mikheev, A.M.; Sanger, M.C.; Girard, E.; Silber, J.R.; Gonzalez-Cuyar, L.F.; Hiatt, J.B.; Adey, A.; Lee, C.; Kitzman, J.O.; Born, D.E.; Silbergeld, D.L.; Olson, J.M.; Rostomily, R.C.; Shendure, J. Deep sequencing of multiple regions of glial tumors reveals spatial heterogeneity for mutations in clinically relevant genes. *Genome biology* **2014**, *15*, 1–9.
12. Hartmann, C.; Meyer, J; Balss, J.; Capper, D.; Mueller, W.; Christians, A.; Felsberg, J.; Wolter, M.; Mawrin, C.; Wick, W.; Weller, M.; Herold-Mende, C.; Unterberg, A.; Jeuken, J.W.M.; Wesseling, P.; Reifenberger. G.; von Deimling, A. Type and frequency of IDH1 and IDH2 mutations are related to astrocytic and oligodendroglial differentiation and age: a study of 1,010 diffuse gliomas. *Acta Neuropathol* 2009, *118*, 469–474.
13. Puchalski, R.B.; Shah, N.; Miller, J.; Dalley, R.; Nomura, S.R.; Yoon, J.G.; et al. An anatomic transcriptional atlas of human glioblastoma. *Science* 2018, *360*, 660–663.
14. Crisman, T.J.; Zelaya, I.; Laks, D.R.; Zhao, Y.; Kawaguchi, R.; Gao, F.; Kornblum, H.I.; Giovanni Coppola, G. Identification of an Efficient Gene Expression Panel for Glioblastoma Classification. *PLoS ONE* 2016, *11*, e0164649.
15. Cai, Y.D.; Zhang, S.; Zhang, Y.H.; Pan, X.; Feng, K.; Chen, L.; Huang, T.; Kong, X. Identification of the gene expression rules that define the subtypes in glioma. *Journal of clinical medicine* 2018, *7*, 350.
16. Munquad, S.; Si, T.; Mallik, S.; Li, A.; Das, A.B. Subtyping and grading of lower-grade gliomas using integrated feature selection and support vector machine. *Briefings in Functional Genomics* 2022, *21*, 408–421.
17. Wang, Z.; Gao, L.; Guo, X.; Feng, C.; Lian, W.; Deng, K.; Xing, B. Development and validation of a nomogram with an autophagy-related gene signature for predicting survival in patients with glioblastoma. *Aging (Albany NY)* 2019, *11*, 12246.

18. Zhang, M.; Wang, X.; Chen, X.; Zhang, Q.; Hong, J. Novel immune-related gene signature for risk stratification and prognosis of survival in lower-grade glioma. *Frontiers in genetics* 2020, *11*, 363.
19. Tu, Z.; Shu, L.; Li, J.; Wu, L.; Tao, C.; Ye, M.; Zhu, X.; Huang, K. A novel signature constructed by RNA-binding protein coding genes to improve overall survival prediction of glioma patients. *Frontiers in Cell and Developmental Biology* 2021, *8*, 588368.
20. Vieira, F.G.; Bispo, R.; Lopes, M.B. Integration of Multi-Omics Data for the Classification of Glioma Types and Identification of Novel Biomarkers. *Bioinformatics and Biology Insights* 2024, *18*, 11779322241249563.
21. Munquad, S.; Das, A.B. DeepAutoGlioma: a deep learning autoencoder-based multi-omics data integration and classification tools for glioma subtyping. *BioData Mining* 2023, *16*, 32.
22. Castelvecchi, D. Can we open the black box of AI? *Nature News* 2016, *538*, 20.
23. Adadi, A.; Berrada, M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access* 2018, *6*, 52138–52160.
24. Loh, H.W.; Ooi, C.P.; Seoni, S.; Barua, P.D.; Molinari, F.; Acharya, U.R. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Computer Methods and Programs in Biomedicine* 2022, *226*, 107161.
25. Ma, L.; Xiao, Z.; Li, K.; Li, S.; Li, J.; Yi, X. Game theoretic interpretability for learning based preoperative gliomas grading. Future Generation Computer Systems. *Future Generation Computer Systems* 2020, *112*, 1–10.
26. Chakraborty, S.; Sudhakar, S.; Swaminathan, R. An Explainable Machine Learning Model for Differentiation of Glioma Sub-types using MR Image Texture Analysis of Cerebral Edema. *IFAC-PapersOnLine* 2024, *58*, 227–232.
27. Ayaz, H.; Oladimeji, O.; McLoughlin, I.; Tormey, D.; Booth, T.C.; Unnikrishnan, S. An eXplainable Deep Learning Model for Multi-Modal MRI Grading of IDH-Mutant Astrocytomas. *Results in Engineering* 2024, *24*, 103353.
28. Alahakoon, A.M.H.H.; Walgampaya, C.K.; Walgampaya, S.; Ekanayake, I.U.; Alawatugoda, J. Overall Survival predictions of GBM Patients using Radiomics: An Explainable AI Approach using SHAP. *IEEE Access* 2024, *12*, 145234–145253.
29. Zhao, Z.; Zhang, K-N.; Wang, Q.; Li, G.; Zeng, F.; Zhang, Y.; Wu, F.; Chai, R.; Wang, Z.; Zhang, C.; Zhang, W.; Bao, Z.; Jiang, T. Chinese Glioma Genome Atlas (CGGA): A Comprehensive Resource with Functional Genomic Data from Chinese Glioma Patients. *Genomics, Proteomics & Bioinformatics* **2021**, *19*, 1–12.
30. Cancer Genome Atlas Research Network; Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet.* **2013**, *45*, 1113–1120.
31. Maintz, D.; Fiedler, K.; Koopmann, J.; Rollbrocker, B.; Nechev, S.; Lenartz, D.; Stangl, A.P.; Louis, D.N.; Schramm, J.; Wiestler, O.D.; von Deimling, A. Molecular Genetic Evidence for Subtypes of Oligoastrocytomas. *J Neuropathol Exp Neurol.* **1997**, *56*, 1098–1104.
32. Han, H.; Men, K. How does normalization impact RNA-seq disease diagnosis? *J Biomed Inform.* **2018**, *85*, 80–92.
33. Zhang, Y.; Jenkins, D.F.; Manimaran, S.; Johnson, W.E. Alternative empirical Bayes models for adjusting for batch effects in genomic studies. *BMC Bioinformatics* **2018**, *19*, 262.
34. Leek, J.T.; Johnson, W.E.; Parker, H.S.; Jaffe, A.E.; Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **2012**, *28*, 882–883.
35. Ross, B.C. Mutual Information between Discrete and Continuous Data Sets. *PLoS ONE* **2014**, *9*, e87357.
36. Moore, J.H.; White, B.C. Tuning ReliefF for genome-wide genetic analysis. In European conference on evolutionary computation, machine learning and data mining in bioinformatics, Berlin, Heidelberg, 2007; 166–175.
37. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *JMLR* **2011**, *12*, 2825–2830.
38. Urbanowicz, R.J.; Olson, R.S.; Schmitt, P.; Meeker, M.; Moore, J.H. Benchmarking Relief-Based Feature Selection Methods for Bioinformatics Data Mining. *arXiv* **2017**, arXiv:1711.08477.
39. Peterson, L.E. K-nearest neighbor. *Scholarpedia* **2009**, *4*, 1883.
40. Vapnik, V.N. *Statistical learning theory*; Wiley, New York, 2008; 1780 p.
41. Breiman, L. Random forests. *Machine learning* 2001, *45*, 5–32.
42. Geurts, P.; Ernst, D., Wehenkel, L. Extremely randomized trees. *Machine learning* 2006, *63*, 3–42.
43. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, 2016; 785–794.
44. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017; 30.
45. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems (NeurIPS 2018)* **2018**, *31*.
46. Arik, S.O., Pfister, T. TabNet: Attentive Interpretable Tabular Learning. *arXiv* **2019**, arXiv:1908.07442.
47. Joseph, M.; Raj, H. GANDALF: Gated Adaptive Network for Deep Automated Learning of Features. *arXiv* **2022**, arXiv:2207.08548.
48. Joseph, M. PyTorch Tabular: A Framework for Deep Learning with Tabular Data. *arXiv* **2021**, arXiv:2104.13638.
49. Mosley, L. A balanced approach to the multi-class imbalance problem. Dissertation, Iowa State University, Ames, Iowa, 2013.
50. Cox, D.R. *Analysis of Survival Data*, 1st ed.; Chapman & Hall/CRC, 1984; p. 212.
51. Ishwaran, H.; Kogalur, U B.; Blackstone, E.H.; Lauer, M.S. Random survival forests. *Ann. Appl. Stat.* **2008**, *2*, 841–860.

52. XGBoost Survival Embeddings. Available online: https://loft-br.github.io/xgboost-survival-embeddings/index.html (accessed on 30 January 2025).
53. Katzman, J.L.; Shaham, U.; Cloninger, A.; Bates, J.; Jiang, T.; Kluger, Y. Deepsurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol* **2018**, *18*, 24.
54. Kvamme, H.; Borgan, Ø; Scheel, I. Time-to-Event Prediction with Neural Networks and Cox Regression. *Journal of Machine Learning Research* **2019**, *20*, 1–30.
55. Kvamme, H.; Borgan, Ø. Continuous and discrete-time survival prediction with neural networks. *arXiv* **2019**, arXiv:1910.06724.
56. Fotso, S. Deep neural networks for survival analysis based on a multi-task framework. *arXiv* **2018**, arXiv:1801.05512.
57. Pölsterl, S. scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn. *Journal of Machine Learning Research* **2020**, *21*, 1–6.
58. Harrell Jr, F.E.; Lee, K.L.; Mark, D.B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine* **1996**, *15*, 361–387.
59. Heagerty, P.J.; Lumley, T.; Pepe, M.S. Time-dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. *Biometrics* **2000**, *56*, 337–344.
60. Graf, E.; Schmoor, C.; Sauerbrei, W.; Schumacher, M. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine* **18**, *56*, 2529–2545.
61. Bergstra, J.; Bardenet, R.; Bengio, Y.; Kégl, B. Algorithms for hyper-parameter optimization. In Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS 2011), Granada, Spain; 24.
62. Watanabe, S. Tree-Structured Parzen Estimator: Understanding Its Algorithm Components and Their Roles for Better Empirical Performance. *arXiv* **2023**, arXiv:2304.11127.
63. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. *arXiv* **2019**, arXiv:1907.10902.
64. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
65. Kolberg, L.; Raudvere, U.; Kuzmin, I.; Adler, P.; Vilo, J.; Peterson, H. g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Research* **2023**, *51*, W207–W212.
66. Davidson-Pilon, C. lifelines: survival analysis in Python. *Journal of Open Source Software* **2019**, *4*, 1317.
67. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017; 4768–4777.
68. Chen, S. C.-C.; Lo, C.-M.; Wang, S.-H.; Su, E. C.-Y. RNA editing-based classification of diffuse gliomas: predicting isocitrate dehydrogenase mutation and chromosome 1p/19q codeletion. *BMC Bioinformatics* 2019, *20*, 659.
69. Park, Y.W.; Eom, J.; Kim, D.; Ahn, S.S.; Kim, E.H.; Kang, S.-G.; Chang, J.H.; Kim, S.H.; Lee, S.-K. A fully automatic multiparametric radiomics model for differentiation of adult pilocytic astrocytomas from high-grade gliomas. *Eur Radiol* 2022, *32*, 4500–4509.
70. Gashi, M.; Vuković, M.; Jekic, N.; Thalmann, S.; Holzinger, A.; Jean-Quartier, C.; Jeanquartier, F. State-of-the-Art Explainability Methods with Focus on Visual Analytics Showcased by Glioma Classification. *BioMedInformatics* 2022, *2*, 139–158.
71. Hu, X.; Martinez-Ledesma, E.; Zheng, S.; Kim, H.; Barthel, F.; Jiang, T.; Hess, K.R.; Verhaak, R.G.W. Multigene signature for predicting prognosis of patients with 1p19q co-deletion diffuse glioma. *Neuro-Oncology* 2017, *19*, 786–795.
72. Liu, B.; Liu, J.; Liu, K.; Huang, H.; Li, Y.; Hu, X.; Wang, K.;Cao, H.; Cheng, Q. A prognostic signature of five pseudogenes for predicting lower-grade gliomas. *Biomedicine & Pharmacotherapy* 2019, *117*, 109116.
73. Song, L.R.; Weng, J.C.; Li, C.B.; Huo, X.L.; Li, H.; Hao, S.Y.; Wu, Z.; Wang, L.; Li, D.; Zhang, J.T. Prognostic and predictive value of an immune infiltration signature in diffuse lower-grade gliomas. *JCI Insight* 2020, *5*, e133811.
74. Shono, T.; Yokoyama, N.; Uesaka, T.; Kuroda, J.; Takeya, R.; Yamasaki, T.; Amano, T.; Mizoguchi, M.; Suzuki, S.O.; Niiro, H.; Miyamoto, K.; Akashi, K.; Iwaki, T.; Sumimoto, H.; Sasaki, T. Enhanced expression of NADPH oxidase Nox4 in human gliomas and its roles in cell proliferation and survival. *Cancer Cell Biology* 2008, *123*, 787–792.
75. Lin, L.; Li, X.; Zhu, S.; Long, Q.; Hu, Y.; Zhang, L.; Liu, Z.; Li, B.; Li, X. Ferroptosis-related NFE2L2 and NOX4 Genes are Potential Risk Prognostic Biomarkers and Correlated with Immunogenic Features in Glioma. *Cell Biochem Biophys* 2023, *81*, 7––17.
76. Xue, Q.; Cao, L.; Chen, X.-Y.; Zhao, J.; Gao, L.; Li, S.-Z.; Fei, Z. High expression of MMP9 in glioma affects cell proliferation and is associated with patient survival rates. *Oncology Letters* 2017, *13*, 1325–1330.
77. Zhou, W.; Yu, X.; Sun, S.; Zhang, X.; Yang, W.; Zhang, J.; Zhang, X.; Zheng Jiang, Z. Increased expression of MMP-2 and MMP-9 indicates poor prognosis in glioma recurrence. *Biomedicine & Pharmacotherapy* 2019, *118*, 109369.
78. Li, Q.; Chen, B.; Cai, J.; Sun, Y.; Wang, G.; Li, Y.; Li, R.; Feng, Y.; Han, B.; Li, J.; Tian, Y.; Yi, L.; Jiang, C. Comparative Analysis of Matrix Metalloproteinase Family Members Reveals That MMP9 Predicts Survival and Response to Temozolomide in Patients with Primary Glioblastoma. *PLoS ONE* 2016, *11*, e0151815.
79. Sun, F.; Liao, M.; Tao, Z.; Hu, R.; Qin, J.; Tao, W.; Liu, W.; Wang, Y.; Pi, G.; Lei, J.; Bao, W.; Dong, Z. Identification of PANoptosis-related predictors for prognosis and tumor microenvironment by multiomics analysis in glioma. *J Cancer* 2024, *15*, 2486–2504.
80. Zhou, K. 474P Potassium inward rectifier channel subfamily J member 11 mRNA expression in glioma and its significance in predicting prognosis and chemotherapy sensitivity. *Annals of Oncology* 2024, *35*, S418.
81. Luo, X.; Xu, S.; Zhong, Y.; Tu, T.; Xu, Y.; Li, X.; Wang, B.; Yang, F. High gene expression levels of VEGFA and CXCL8 in the peritumoral brain zone are associated with the recurrence of glioblastoma: A bioinformatics analysis. *Oncology Letters* 2019, *18*, 6171–6179.

82. Li, L.; Zhang, C.; Wang, Z.; Guo, Y.; Wang, Y.; Fan, X.; Jiang, T. Expression changes in ion channel and immunity genes are associated with glioma-related epilepsy in patients with diffuse gliomas. *Journal of Cancer Research and Clinical Oncology* 2022, *148*, 2793–2802.
83. Luo, W.; Quan, Q.; Jiang, J.; Peng, R. An immune and epithelial–mesenchymal transition-related risk model and immunotherapy strategy for grade II and III gliomas. *Front. Genet.* 2023, *13*, 1070630.
84. Pan, P.; Guo, A.; Peng, L. Establishment of glioma prognosis nomogram based on the function of meox1 in promoting the progression of cancer. *Heliyon* 2024, *10*, e29827.
85. Demirdizen, E.; Al-Ali, R.; Narayanan, A.; Sun, X.; Varga, J.P.; Steffl, B.; Brom, M.; Krunic, D.; Schmidt, C.; Schmidt, G.; Bestvater, F.; Taranda, J.; Turcan, S. TRIM67 drives tumorigenesis in oligodendrogliomas through Rho GTPase-dependent membrane blebbing. *Neuro-Oncology* 2023, *25*, 1031—1043.
86. Lu, F.; Shen, S.H.; Wu, S.; Zheng, P.; Lin, K.; Liao, J.; Jiang, X.; Zeng, G.; Wei, D. Hypomethylation-induced prognostic marker zinc finger DHHC-type palmitoyltransferase 12 contributes to glioblastoma progression. *Ann Transl Med* 2022, *10*, 334.
87. Ying, T.; Lai, Y.; Lu, S.; E, S. Identification and validation of a glycolysis-related taxonomy for improving outcomes in glioma. *CNS Neuroscience & Therapeutics* 2024, *30*, e14601.
88. Li, J.; Yan, X.; Liang, C.; Chen, H.; Liu, M.; Wu, Z.; Zheng, J.; Dang, J.; La, X.; Liu, Q. Comprehensive Analysis of the Differential Expression and Prognostic Value of Histone Deacetylases in Glioma. *Front. Cell Dev. Biol.* 2022, *10*, 840759.
89. Jiang, L.; Zhu, X.; Yang, H.; Chen, T.; Lv, K. Bioinformatics Analysis Discovers Microtubular Tubulin Beta 6 Class V (TUBB6) as a Potential Therapeutic Target in Glioblastoma. *Front. Genet.* 2020, *11*, 566579.
90. Sallinen, S.-L.; konen, T.; Haapasalo, H.; Schleutker, J. CHEK2 mutations in primary glioblastomas. *Journal of Neuro-Oncology* 2005, *74*, 93-–95.
91. Dong, Y.-S.; Hou, W.-G.; Li, X.-L.; Jin, T.-B.; Li, Y.; Feng, D.-Y.; Liu, D.-B.; Gao, G.-D.; Yin, Z.-M.; Qin, H.-Z. Genetic association of CHEK2, GSTP1, and ERCC1 with glioblastoma in the Han Chinese population. *Tumor Biology* 2014, *35*, 4937–4941.
92. Tlais, D.; Barros Guinle, M.I.; Wheeler, J.R.; Prolo, L.M.; Vogel, H.; Partap, S. CHEK2 mutations in pediatric brain tumors. *Neuro-Oncology Advances* 2023, *5*, vdad038.