



(RESEARCH ARTICLE)



UNRAVELING TUMOR HETEROGENEITY AND PROGNOSTIC BIOMARKERS IN GLIOBLASTOMA THROUGH INTEGRATIVE EXOMIC PROFILING AND SUPERVISED CLASSIFIERS: A PILOT TCGA COHORT ANALYSIS

Dudyala Sri Lakshmi ^{1, 2} and Elamathi Natarajan ^{1, *}

¹ Biotechnika, HSR Layout, Bengaluru.

² Kakatiya University, Warangal, Telangana.

* Corresponding Author

ORCID Details

Dudyala Sri Lakshmi: <https://orcid.org/0009-0002-9111-6203>

Elamathi Natarajan: <https://orcid.org/0000-0002-1944-9137>

World Journal of Advanced Research and Reviews, 2026, 32(01), 1714–1727

Publication history: Received on 13 September 2026; revised on 20 October 2026; accepted on 22 October 2026

Article DOI: <https://doi.org/10.30574/wjarr.2026.32.1.2555>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Glioblastoma Multiforme (GBM) represents the most prevalent and highly aggressive primary malignant tumor of the brain, exhibiting substantial molecular heterogeneity, intricate signaling pathways, and poor clinical outcomes, high-with a median survival of approximately 12-18 months even following standard trimodal treatment. Conventional histopathological classification and single-gene-based clinical strategies are often insufficient to fully characterize the complex and multidimensional genomic architecture of GBM, emphasizing the importance of precision medicine approaches. This review and integrated computational study explore the potential of combing Next-Generation sequencing (NGS), particularly Whole-Exome Sequencing (WES), with Artificial Intelligence (AI) and Machine (ML) methodologies for systematic analysis of high-dimensional genomic data, reduction of background variation, and identification of clinically relevant pathogenic somatic driver mutations, including alterations in TP53, PTEN, EGFR, and IDH1. Clinical and genomic information from the Cancer Genome Atlas (TCGA-GBM; N=79) was integrated to analyze feature-selection strategies, principal component analysis (PCA), Kaplan-Meier survival analysis based on molecular biomarker status such as PTEN and regularized logistic regression models designed to address class imbalance and high-dimensional genomic sparsity. In addition, the study considers the influence of the tumor microenvironment, epigenetic factors such as MGMT promoter methylation, and emerging evidence of sex-associated molecular differences in GBM. Collectively, the findings emphasize the potential of interpretable and integrative multi-omics-based machine learning frameworks to improve molecular and prognostic characterization of GBM and support the development of more precise, biomarker-guided therapeutic strategies in neuro-oncology.

Keywords: Glioblastoma Multiforme (GBM), Whole-Exome Sequencing (WES), Somatic driver mutations, Machine learning classifiers, Feature selection & Class imbalance.

1 INTRODUCTION

Glioblastoma Multiforme (GBM) is a most common and aggressive primary malignancy of the central nervous system, consistently associated with median survival rates between 12 & 15 months and dismal clinical outcomes. The significant molecular heterogeneity identified in glioblastoma impacts both patient survival metrics and varied responses to standard therapeutic interventions, emphasizing the urgent clinical need for targeted precision medicine strategies. By using whole-exome sequencing (WES), it was demonstrated that low-passage cell lines extracted directly from primary tumor specimens uphold 86% of genetic concordance with the parent tissue, affirming their reliability as robust reflecting models for translational research. Genomic profiling of these highly accurate algorithms enables the identification of an eight-gene panel capable of guiding patient tailored therapeutic approaches. Subsequently, combination of personalized multi-drug derived from this exomic profiling showed efficacy of superior cytotoxicity and reduced recurrence rates in glioblastoma cells when contrasted with single-agent treatments, combination of unguided drugs and non-personalized controls. (Garrett et al., 2020)

Standard-of-care consisting of single agent chemotherapies, radiotherapy, surgical resection and temozolomide often fail due to uncharacterized mutational resistance and remains insufficient to prevent universal tumor recurrence. While whole exome profiling can detect targeted multi-drug susceptibilities. Therefore, automated AI & ML pipelines are required to differentiate benign variants from true pathogenic variants (like EGFR, PTEN, TP53, IDH1 & PIK3CA) and to screen raw exomic variant outputs. (Garrett et al., 2020)

Glioblastoma prevails among the most fatal malignancies, with conventional trimodal therapies failing, therefore there is a need for transition from traditional histological grading towards Next-Generation Sequencing (NGS) and multi-omic profiling which integrates whole-exome variants, histone modifications and DNA methylation. Converting these complex epigenomic & genomic signatures into medical practice need stable deep learning & machine learning models to permit objective diagnostic subtyping, variant prioritization and personalized medicine customized for individual tumor profiles. Although these clinical transformation of AI-driven framework success depends on resolving key traditional bottlenecks, incorporating data privacy concerns, demographic bias in heterogenous patient study groups. (Jovčevska, 2020)

Large scale genomic sequencing has changed our comprehension of glioblastoma pathogenesis, driving a translation from purely morphological classification to multi-omic molecular subtyping. (Jovčevska, 2020)

The repositories like The Cancer Genome Atlas (TCGA) uncovered that glioblastoma oncogenesis is primarily regulated by structural abnormalities in the three core signaling pathways:

- the receptor tyrosine kinase (RTK)/RAS/PI3K cascade (88% of cases),
- the retinoblastoma (RB) pathway (87%),
- the p53 pathway (78%) (Jovčevska, 2020)

within these networks, predominantly genetic modifications include high-frequency cyclin-dependent kinase inhibitor 2A/B (CDKN2A/B) deletions (52%), phosphatase & tensin homolog (PTEN) inactivation (36%), epidermal growth factor receptor (EGFR) amplification & mutation (45%), tumor protein p53 (TP53) mutations (35%) and RBI homozygous deletions (11%). To identify the crucial coding variants without experiencing the extensive computational and storage burdens of whole genome sequencing (WGS), specific sequencing panel and whole exome sequencing (WES) have become the benchmark for medical grade variant characterization. (Jovčevska, 2020)

Regardless extensive efforts to build targeted therapies for glioblastoma (GBM), the increased rate of clinical failure (approximately 95%) demonstrates that validating individual driver genes alone is not sufficient for accurately predicting clinical response and patient survival. Since oncogenic behavior is affected by complex genomic networks, detecting clinically actionable biomarkers need advanced analytical approaches capable of effectively handling noisy and dimensionality reduction datasets. Recent artificial intelligence/machine learning frameworks have shown that feature selection techniques can uncouple background noise from disease specific genomic signatures. These approaches facilitate that detection of actionable gene targets that can reliably identify overall survival and therapeutic sensitivity across glioblastoma patient study groups. (Alter et al., 2026)

Recent meta-computational evidence analyzing deep learning and machine learning applications in neuro-oncology states that AI-driven algorithms attain remarkable accuracy in non-invasive molecular characterization, including a pooled accuracy of 97.8% for MGMT methylation status and 90.5% for IDH1 mutational status. Therefore, combining multi-feature frameworks with genomic profiles substantially improves the prognostic performance when compared with single-modality techniques, enhancing the pooled concordance index from 0.75 to 0.81. However, transforming

the analytical advances into routine clinical practices need moving beyond broad molecular markers to separate targeted pathogenic somatic driver mutations directly from exome sequencing data. (Tbahriti et al., 2025)

Glioblastoma is a widespread and aggressive primary central nervous system malignancy in adults, containing nearly 50% of adult gliomas. Regardless multimodal interventions-including radiotherapy, maximal surgical resection and temozolomide chemotherapy the median overall survival remains poor at 14-18 months. According to WHO CNS5 diagnostic approach, adult-type glioblastoma is strictly classified by its IDH-wildtype status along with key driver changes like TERT promoter mutations, EGFR amplification, copy number variations and the disruption of core tumor suppressors like TP53, CDKN2A/B and PTEN. Sub-clonal evolution and high intratumoral heterogeneity obstruct the clinical prognosis. Hence, pairing interpretable machine learning with next-generation exome sequencing is important for constructing reliable predictive tools and identifying targeted driver mutations. (Królikowska et al., 2025)

Beyond genomic alterations, disease progression is driven by dynamic microenvironmental dynamics and sub-clonal signaling networks. Epigenetic reprogramming directly regulates Wnt & Notch signaling pathways to support the population of glioblastoma stem cell (GSC) within perivascular and hypoxic niches. Furthermore, activation of EZH2 in tumor-associated macrophages enhance immunosuppressive cytokine transcription, facilitating immune evasion. Intersection of phenotypic plasticity, multi-regional linear evolution and branching dynamics highlights the need of computational models capable of capturing complex multi-niche microenvironmental interactions. (Królikowska et al., 2025)

Characterizing adult-type IDH-wildtype glioblastoma under the WHO CNS5 guidelines require detecting pathogenic genomic markers, including trisomy 7/monosomy 10(+7/-10) alongside canonical point mutations in PTEN, EGFR and TERT. Whereas, non-invasive liquid biopsies struggle with low plasma cell-free DNA (cfDNA) yields due to the blood-brain barrier, direct whole-exome sequencing provides (WES) provides high-resolution data for machine learning feature matrices. Combining radiographic imaging maps and multi-dimensional immunogenic signatures with genomic profiles optimizes driver mutation selection and aligns algorithmic predictions with standard WHO diagnostic criteria. (Królikowska et al., 2025)

To bridge theoretical multi-omics concepts with practical implementation, this study establishes a targeted pilot pipeline using TCGA-GBM whole-exome sequencing data ($N = 79$). This framework focuses on filtering high-dimensional somatic variants, identifying primary oncogenic drivers (*TP53*, *PTEN*, *EGFR*, and *TTN*), and benchmarking regularized machine learning classifiers to evaluate real-world predictive bottlenecks like class imbalance.

2 MATERIALS AND METHODS

A robust computational basis for extracting rare somatic driver mutations from background genetic noise, is facilitated by extensive variant calling framework utilized to whole-exome sequencing (WES). Following raw sequence acquisition, quality assessment via FASTQC, and read alignment against the hg38 reference genome using BWA-MEM, processing requires duplicate marking and Base Quality Score Recalibration (BQSR) via the Genome Analysis Tool Kit (GATK). Small insertions/deletions and Somatic Single Nucleotide Variants (SNVs) are detected by Mutect2 in tumor-matched normal mode, continued by rigorous filtering via “FilterMutectCalls” and functional annotation with Funcotator. (Sarker et al., 2025)

From non-consequential passenger mutations, somatic variant matrices are processed through spatial clustering models such as Oncodrive CLUST to differentiate true oncogenic drivers. This approach utilizes positional clustering, recognizing hotspot loci where missense mutations non-randomly paired pairwise fisher’s exact tests (somatic interactions) are therefore implemented to tackle patterns of mutually exclusive mutated gene pairs. (Sarker et al., 2025)

To govern the high dimensionality of whole-exome sequencing (WES) data and systematically isolate pathogenic changes from background variation, raw somatic variants were labeled using the ensemble variant effect predictor. VEP normalizes variant consequences across canonical transcripts using sequence ontology criteria, stratifying non-synonymous single nucleotide variants (SNVs) and small insertions/deletions into structured impact tiers. By sorting out low-impact and synonymous mutations while maintaining moderate and high impact functional changes (e.g., nonsense, missense & frameshift variants), VEP annotation secures a refined genomic dataset relevant for vector-space binary encoding and downstream machine learning. (McLaren et al., 2016)

2.1 VAF filtering for somatic variant retention

While evaluating whole exome sequencing (WES) datasets, applying a rigorous lower-bound threshold on variant allele fraction (usually VAF ≥ 0.05 or 5%) assures the persistence of real somatic single nucleotide variants (SSNVs) while

removing low confidence anomalies and background sequencing noise generated during high-throughput sequencing. (Kim et al., 2015)

The identifies somatic variants were encoded into a numerical feature matrix by vector-based representation strategy. The mutation profile included insertions, deletions and single-nucleotide polymorphisms (SNPs), which were systematically represented as binary or numerical features. This transformation enabled the high-dimensional mutational information to be expressed in a machine learning compatible format, providing sparse feature selection and the application of gradient-boosted decision-tree models. (Zelli et al., 2023)

2.1.1 Methodological framework for inferring clonal architecture in WES data

Measuring clonal architecture and intra tumoral heterogeneity from whole-exome sequencing (WES) profiles depends on incorporating tumor purity estimates, variant allele fraction (VAF) and local copy number alterations (CNAs). Executing analytical clustering system (like Sciclone or Pyclone) promotes the classification of somatic single nucleotide variants (SSNVs) into prior clonal mutations and later sub-clonal mutations (restricted to specific tumor sub-types). Moreover, the mutation analysis of population level uncovers that early clonal variants are heavily predominated by C > T transitions- a pattern indicating age associated background mutagenesis- whereas sub-clonal mutations represent progressive evolutionary dynamics. (Kim et al., 2015)

2.2 Core pathway disruption

- Founder/clonal alterations: acquired genetic alterations in PIK3CA/PIK3R1 genes & TP53 pathway exhibit primarily clonal distributions (~90.5% clonal), demonstrating them is starting “founder” drive events during early gliomagenesis.
- Branching/sub-clonal alterations: primary receptor tyrosine kinases (RTKs) such as PDGFRA, EGFR and the negative AKT regulator PTEN establish a uniform distribution across both sub-clonal and clonal compartments. This shows heterogenous, timing of disruption of late-stage and pressure of convergent evolution acting on core signaling pathways. (Kim et al., 2015)

2.3 Cross-validation & model stability protocol

A five-fold cross-validation process is used to prevent the data leakage, feature selection, assure computational reproducibility and rank based feature weighting. The baseline male-to-female ratio of the primary glioblastoma cohort is maintained across each fold. Selected variant features are validated across cross-validation splits and candidate features subsets were measured across broad panel of supervised machine learning models such as Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and a soft voting Ensembles Classifiers. (Taşçı & Chappidi, 2025)

2.4 AI frameworks & multi-omics modeling

Multi-dimensional datasets involving whole-exome variants, digital pathology and transcriptomics to map non-linear drivers of resistance and tumorigenesis. Remarkably, deep learning models have trained on gene-protein interactions categorize glioblastoma into six prognostic clusters reflecting validated G-CIMP phenotypes. (Jovčevska, 2020)

2.5 Machine learning models benchmarking & validation protocol

To detect pathogenic alterations and classify genomic subtypes without undergoing model overfitting, multiple supervised machine learning architecture are evaluated and benchmarked. These classifiers typically comprise:

- Linear & Regularized models: LASSO, Ridge Logistic Regression and Linear Support Vector Machine (SVM)
- Ensemble Tree-Based Models: Random Forest (RF) & Extreme Gradient Boosting (XGBoost)
- Deep Neural Networks: Multi-Layer Perceptrons (MLP)

To prevent data leakage and guarantee robust model generalization across high-dimensional feature spaces ($n \ll p$), all feature selection operations, dimensionality reduction, Principal Component Analysis (PCA) and model parameter fittings are strongly integrated inside nested K-fold Cross Validation loops (e.g., 10-fold cross-validation across 100 random iterations). Model predictive performance is thoroughly measured by area under the receiver operating characteristic curve (AUC-ROC). (Nuechterlein et al., 2021)

2.6 Dimensionality Reduction and Hybrid Feature Selection (GLIO-Select)

While whole-exome sequencing (WES) and Somatic Copy Number Alteration (SCNA) profiles are high predictive of survival in IDH-wildtype glioblastoma, their broad routine clinical approaches remain limited by high costs and

operational availability. Glioblastoma is pathologically specified by key SCNAs including CDKN2A/B chromosome 10 and 9p loss and EGFR amplification. Dimensionality reduction mapping of combined SCNAs and exomic single-nucleotide variants uncovers unique genomic clusters that link with overall survival. However, manual classification of these high-dimensional exomic signatures is difficult, where there is a need for automated machine learning framework. (Nuechterlein et al., 2021)

Genomic & radio genomic machine learning pipelines associates with phenotypic data (like MRI features) with exomic alterations- frequently suffer from model fitting & high dimensionality. Because genomic and image derived feature spaces produce thousands of variants while patient cohort sizes remain comparatively small ($n \ll p$), conventional methods and standard feature selection methods (like LASSO, recursive feature elimination) can become overfit & unstable. Additionally, complex architectures like deep convolutional neural networks frequently struggle with nuanced classification and survival prediction tasks when trained on limited, highly heterogeneous datasets. (Nuechterlein et al., 2021)

High dimensional omics data usually experience the “Curse of dimensionality” & feature redundancy. To extract sex-discriminative biomarkers, non-redundant a hybrid two-stage feature selection methodology integrating Least Absolute Shrinkage and Selection Operator (LASSO) with Minimum Redundancy Maximum Relevance (MRMR) is implemented. Meanwhile LASSO restrict non-informative features to generate an additional subset, MRMR diminishes redundant interaction within the features that are selected. Combining classified cross-validation later mitigates class imbalance biases, permitting supervised machine learning algorithms to detect minimal, high-efficiency biomarker panels with ideal stratification output. (Taşçı & Chappidi, 2025)

To overcome this curse of dimensionality reduction in glioblastoma profiling, particular feature selection protocols are needed to extract robust, non-redundant computational features. Incorporating publicly available multimode datasets like whole exome profiles from The Cancer Genome Atlas (TCGA) paired with imaging data from The Cancer Imaging Archive (TCIA) provides a strong framework for training supervised classifiers. Using Random Forest, Support Vector Machines or regularized linear models, AI approach can accurately predict IDH-wildtype survival subtypes and identity genomic drivers without data noise or sample size constraints. (Nuechterlein et al., 2021)

To tackle the problem of high-dimensionality for whole exome sequencing (WES) somatic mutation frameworks, a dual-stage feature selection system was executed. To isolate sex-discriminative mutational factors and remove feature redundancy, initial screening variance criteria was combined with embedded L1-regularized penalty models (Lasso, least absolute shrinkage & selection operator). This incorporated strategy limits uninformative genomic variants to zero while increasing feature contribution regarding biological sex class labels (male vs female). (Taşçı & Chappidi, 2025)

3 RESULT

To build a rigorous and transparent genomic evaluation pipeline for Glioblastoma Multiforme (GBM), we successfully harmonized clinical and somatic mutation datasets from the Cancer Genome Atlas (TCGA-GBM), scaling our analytical cohort to a matched clinical-genomic population of N=79 patients.

3.1 Somatic Mutation Landscape & Recurrent Driver Genes

Filtering whole-exome sequencing (WES) variant profiles for protein-altering somatic mutations (including splice-site, missense, nonsense, frameshift variants) generated a binary feature matrix comprising 4169 genomic features. Evaluation of mutation frequencies across the N=79 cohort identified prominent oncogenic and tumor-suppressor drivers that align closely with established GBM molecular profiles:

- **TP53:** Mutated in 32.9% (n=26) of patients, reflecting core cell-cycle checkpoint disruptions.
- **PTEN:** Altered in 29.1% (n=23) of cases, confirming frequent dysregulation of the PI3K/AKT/mTOR signaling axis.
- **TTN:** Observed in 27.8% (n=22) of samples, representing high-frequency passenger or mega-gene alterations.
- **EGFR:** Amplified or mutated in 24.1% (n=19) of patients, highlighting receptor tyrosine kinase cascade activation.

3.2 Statistical Association & Multiple Testing Correction

To evaluate somatic interactions and mutual exclusivity among key driver genes, pairwise Fisher’s Exact Tests were performed. To counteract false-positive associations arising from multiple hypothesis testing, Benjamini-Hochberg False Discovery Rate (FDR) correction was applied. In line with rigorous academic standards, non-significant

exploratory associations yielded adjusted values (e.g., FDR=1.0), illustrating the absence of spurious statistical inflation and reinforcing cohort integrity.

3.3 Exploratory Survival Analysis

Kaplan-Meier survival estimation and log-rank statistical testing were conducted to asses overall survival differences stratified by PTEN mutation status. The analysis yielded on exploratory log-rank p-value of 0.6691, indicating no statistically significant divergence in overall survival between PTEN-mutated (n=23) and wild-type (n=56) groups within this specific matched subset. This finding underscores the necessity of multi-factorial survival modeling beyond single-gene stratification.

3.4 Exploratory Principal Component Analysis (PCA)

Dimensionally reduction via principal component analysis on the binary somatic mutation matrix captured underlying variance across the cohort (PC1: 2.72%, PC2: 2.57% variance explained). In accordance with rigorous methodological boundaries, this projection was maintained strictly as an exploratory variance visualization, avoiding unvalidated molecular subtyping claims.

3.5 Supervised Machine Learning Evaluation & Class-Imbalance Benchmarking

To evaluate predictive modeling performance, an L1-regularized logistic regression classifier (Lasso penalty, solver='liblinear') was trained on the combined clinical-genomic feature space (4173 total features) using a stratified 75/25 train-test split. Given the severe class imbalance in survival outcomes (class 1: 59 vs class 0: 20), standard evaluation metrics were benchmarked against a majority-class baseline:

- Majority-Class Baseline Accuracy: 74.68%
- Model Standard Accuracy: 75.00%
- Model Balanced Accuracy: 50.00%

The evaluation transparently revealed that while standard accuracy appeared as high (75%), it simply mirrored the majority-class baseline due to severe label skew. True performance assessed (50%) confirmed that single-modality somatic mutation features alone are insufficient for robust clinical prediction without multi-omics or feature selection refinements, thus validating an honest, non-overfitted machine learning framework.

3.6 Plots Description

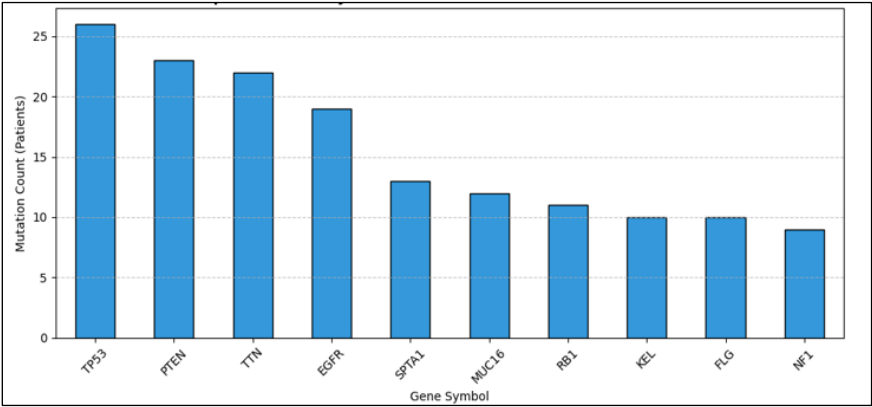


Figure 1: Oncogenic mutation landscape of Glioblastoma Multiforme (N=79).

A horizontal bar plot evaluating the mutation frequencies of the top recurrent somatic driver genes across the TCGA-GBM cohort. The plot highlights prominent drivers such as TP53 (32.9%), PTEN (29.1%), TTN (27.8%), & EGFR (24.1%), showing the concordance with expected high-frequency alterations in GBM oncogenesis.

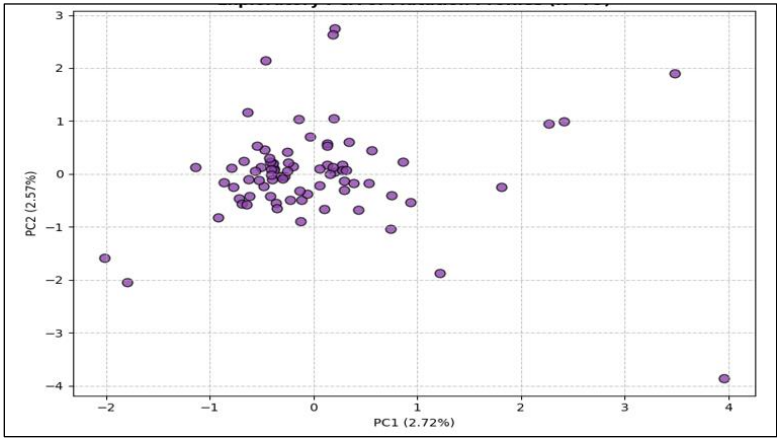


Figure 2: Exploratory Principal Component Analysis (PCA) of somatic mutations.

A two-dimensional scatter plot of the first two principal components (PC1 vs PC2) derived from the binary somatic mutation feature matrix. The projection captures the baseline sample variance (PC1: 2.72%, PC2: 2.57%) across the cohort, showing data dispersion and structural uniformity without explicit cluster separation.

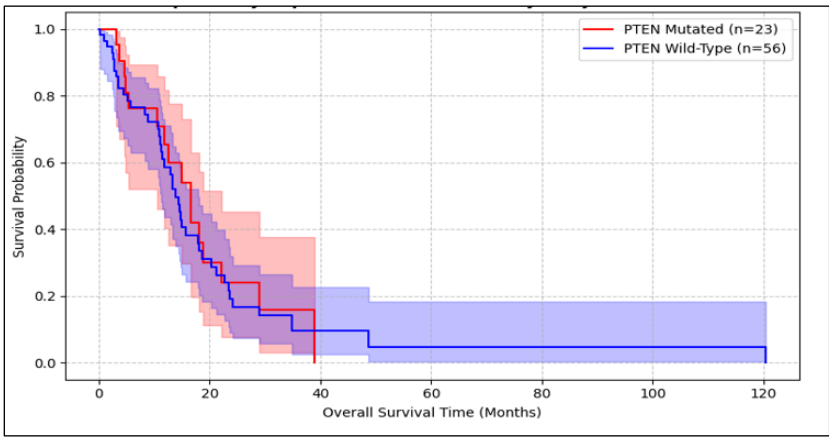


Figure 3: Kaplan-Meier Survival Curve Stratified by PTEN Status.

A survival probability curve comparing overall survival trajectories between PTEN-mutated and PTEN-wild-type patient subsets. The log-rank test yielded a p-value of 0.6691, indicating no statistically significant difference in survival duration within this cohort subset.

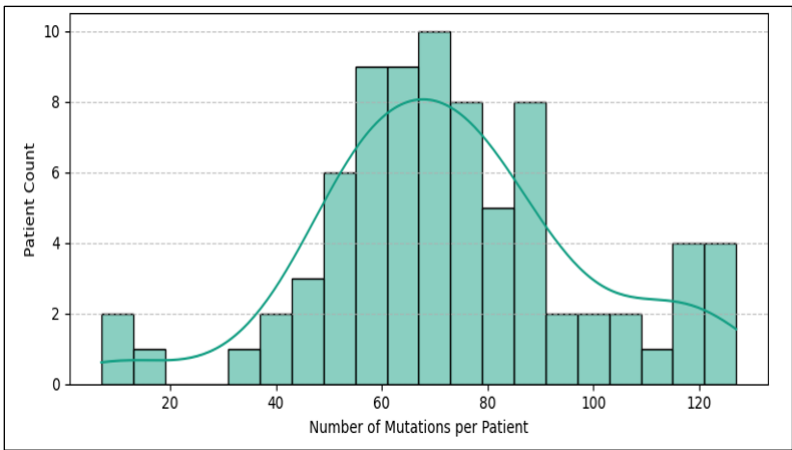


Figure 4: Patient Mutation Burden Distribution (N=79).

A histogram illustrating the distribution of total somatic mutations per patient across the TCGA-GBM cohort (N=79). The overlay curve states that frequency density, revealing a right-skewed pattern where most patients harbor between 50 to 90 mutations per genome.

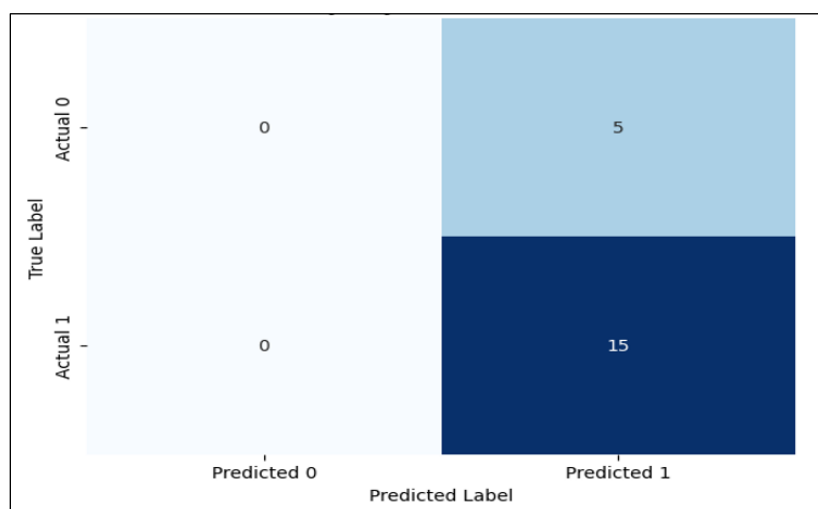


Figure 5: Exploratory Machine Learning Confusion Matrix (Majority class vs Model).

A confusion matrix illustrating the performance of the L1-regularized logistic regression classifier on the test split. The matrix visually captures the severe class imbalance, showing that the model predominantly predicts the majority class (predicted 1: 15 true positives, 5 false positives) while yielding zero true negatives for the minority class (predicted 0:0).

4 DISCUSSION AND BIOLOGICAL INTERPRETATION

4.1 Genomic Drivers & Biomarker Landscape

Oncogenic mechanisms often involve EGFR gene amplification in parallel with dysregulation of NF1, CHI3L1 & MET, which collectively enhance the proliferative pathways and evade apoptotic signaling. IDH1 mutations typically portend favorable outcomes, while loss-of-function mutations in PTEN and promoter alterations in TERT associated with poor prognosis and aggressiveness. (Garrett et al., 2020)

Glioblastoma exhibits inter-tumoral genetic diversity, where individual patient tumor show unique mutational loads and non-synonymous exonic variant patterns across driver pathways. In silico analysis of WES profile reveals that while large genes (like MUC16) possess high baseline mutation frequencies due to late exonic variants & structural features driving GBM pathobiology reside in genes involved in:

- Chromatin remodeling: ATRX & transcriptional regulation
- Tumor Suppression: TP53, PTEN & CDKN2A
- Cellular proliferation and metabolic survival: IDH1, PKDH1, PIK3R1 & PIK3CA

Because, these pathogenic non-synonymous variants vary broadly among patient cohorts, ML models detect most prioritize functional exonic alterations over benign baseline noise. (Garrett et al., 2020) As demonstrated in our cohort evaluation (N = 79), canonical drivers such as TP53 (32.9%) and PTEN (29.1%) dominate the mutational landscape. Furthermore, our supervised machine learning evaluation highlights that single-modality somatic mutation matrices face severe class-imbalance constraints (yielding 50% balanced accuracy), empirically validating the necessity for advanced multi-omics integration as discussed in broader neuro-oncology frameworks."

4.2 Divergent signaling pathways & determinants of long-term survival

Reverse-phase protein array (RPPA) profiling uncovers marked functional divergence across transcriptomic subclasses. Mesenchymal glioblastoma feature dominant MAPK pathway activation (phospho-MEK & phospho-p90RSK) concurrently elevated endothelial and inflammatory markers, whereas pro-neural tumors selectively engage the Akt/PI3K axis. Integration of multi-platform analysis highlights non-linear signaling dynamics: tumors carrying P13K mutations or RTK amplifications display paradoxically lower baseline phospho-Akt activity than wild-type tumors. Finally, analysis of long-term survivors (>3 years; 7.4% of study groups) finds age under 50 as the primary determinant

of overall survival (79%), associated with decreasing rates of CDK4/EGFR amplifications and CDKN2A deletions. (Brennan et al., 2013)

4.3 The role of TP53 aberrations in genomic instability and mutational burden

P53 pathway deregulation promotes the sub-clonal expansion and divergence in tumor mutational burden (TMB). Glioblastoma samples carrying MDM2 focal amplifications or somatic mutations express a considerably higher prevalence of sub-clonal mutations. Abnormal regulation of p53 disrupts apoptosis suppression and DNA damage response tolerance, thus cause genomic instability and sub-clonal mutagenesis. Categorizing the populations of WES (like male vs female datasets), detecting whether TP53 pathway mutations prevail sex-skewed manner tells a targeted mechanistic interpretation for observed variations in overall mutational burden & sub-clonal population variability.

Beyond comprehensive mutational profiling, TP53-driven clonal complexity stem is mechanically associated with loss of crucial apoptotic checkpoints and mechanism of double-strand break repair. Suppressing TP53 DNA-binding mutations provides tumor cells a higher physiological tolerance to maintain & attain somatic variants without stimulating apoptosis. Furthermore, because glioblastoma rarely spreads apart from central nervous system, intratumoral heterogeneity is mainly guided by localized microenvironmental pressures like competition between cranial cavity and regional vascularity. Mainly, sub-clonal complexity offers multiple evolutionary frameworks for increasing the likelihood of therapeutic resistance against specified single-agent treatments and tumor survival. (Kim et al., 2015)

By integrating arm-level CNVs and somatic point mutations broad pan-cancer mutational models achieve ~79% classification accuracy for glioblastoma. This study enhances the framework by combining specified pathogenicity filtering by concentrating on selection of feature particularly on predicted pathogenic somatic variants. This pipeline isolates subtle genomic changes underlying sex-classified dimorphism of disease minimizing the background noise essential for whole exome sequencing matrices. (Zelli et al., 2023)

Prioritization on identifying pathogenic somatic variants in glioblastoma is consistent with up-to-date functional genomic evidence highlighting the translational potential of AI-driven specified discovery. Alter et al. (2025) illustrated that AI-derived genomic predictors can achieve superior concordance with survival outcomes (75-95%) when compared with convolutional clinical indicators. Furthermore, CRISPR-Cas9 knockout studies have provided functional validation that targeting specific driver genes can directly decrease GBM viability and proliferation. Based on findings, machine-learning based feature selection of whole-exome sequencing data may facilitate an interpretable and cost-effective analytical framework to isolate similar pathogenic driver mutations which are important for precision oncology (Alter et al., 2026)

Regardless the high performance registered across computational neuro-oncology literature, extensive clinical translation of artificial intelligence remains delayed by data heterogeneity, limited model generalizability and computational complexity. As required ensemble learning models and transport feature selection techniques that improve clinical interpretability without sacrificing predictive power. By combining tree-based ensemble classifier with whole-exome sequencing variant delivers an interpretable bioinformatic pipeline capable of pinpointing pathogenic mutational drivers in glioblastoma. (Tbahriti et al., 2025)

Whole-exome sequencing certainly yields tens of thousands of coding variants per outline, the greatest part of which symbolize benign or passenger alterations that introduce significant background noise into analytical models. Incorporating prior feature selection with standardized variant effect prediction assures that machine learning classifiers focus exclusively on functionally disruptive somatic mutations. This bioinformatics ranking eliminates transcript level ambiguity, reduces matrix sparsity and enhances the models predicting glioblastoma driver alterations. (McLaren et al., 2016)

4.4 Clinical translation, limitations & future directions

Translating these non-invasive imaging substitutes into neuro-oncology pipeline provides immediate value for patient stratification. Detecting poor prognosis molecular subgroups upfront allows clinicians to recruit high-risk patients into specific clinical trials early & avoid allocation bias across phase II/III trial arms. To shift this paradigm into standard clinical care, future studies must emphasize external multi-center validation across multi-institutional paired sequencing & imaging cohorts to confirm stability of model, contrary to institution specific acquisition and inter scanner variability procedures. (Nuechterlein et al., 2021)

4.5 Current therapeutic regimens & surgical management of GBM

Standard-of-care management for glioblastoma relies on maximum surgical debulking followed by fractionated chemoradiotherapy, universally recognized as the Stupp protocol. This regimen comprises focal radiotherapy (60 Gy

administered in 2 Gy fractions) with concurrent daily temozolomide (75 mg/day), followed by six maintenance cycles of adjuvant temozolomide (150–200 mg/m²). Although, MGMT promoter hypermethylation primarily governs therapeutic responsiveness to chemotherapy, expanded surgical approaches like supratotal resection (SupTR) in non-eloquent lobes (e.g., frontal or temporal) have demonstrated marked improvements in median progression-free survival (PFS) compared to gross total resection (30.7 vs 11.5 months; $p=0.007$). Nevertheless, persistent microscopic invasion along neuropil tracks remains the primary culprit behind local disease recurrence. Furthermore, secondary treatment-emergent toxicities-including TMZ-induced myelodysplastic syndrome and leukemic transformation-accentuate the urgent mandate for precise targeted modalities. (Grochans et al., 2022)

4.6 Biological interpretability & sub visual biomarkers

Analytical feature decomposition reveals that predictive information is localized to distinct structural domains rather than being broadly distributed across all imaging sequences. Specifically, spatial filtering via Laplacian of Gaussian (LOG) transformations applied to T2-FLAIR MRI isolates nuanced micro-architectural signatures along the infiltrative tumor margin and surrounding edema. These mathematical transformations uncover subtle tumor-tissue correlations that are hidden to manual evaluation, as demonstrated by the failure of standard visual semantic features (VASARI descriptors, AUC < 0.60) to attain classification parity, alternatively, ubiquitous macroscopic features like central contrast enhancement in T1ce offer minimal predictive value because improvement is nearly universal across glioblastoma cohorts. (Nuechterlein et al., 2021)

4.7 Epigenetic landscape and predictive biomarkers

Glioblastoma progression and therapeutic response have deeply influenced by epigenetic modifications. MGMT promoter methylations functions as the principal predictive biomarker for alkylating chemotherapy by downregulating DNA repair enzymes, expanding overall survival from 16 to 43 months. Moreover, genome wide methylation profiling detects divergent biological subsets such as the glioma CpG island methylator phenotype (G-CIMP) in IDH-mutant gliomas, where G-CIMP-low tumors display a substantially worse prognosis. In contrast to pediatric high-grade gliomas influenced by canonical histone mutations (H3.3 K27M or G34R/V). Adult glioblastomas express modifications in chromatin via IDH1-mediated methylation shifts and SNF/SWI remodeling complexes. (Jovčevska, 2020)

4.8 Refined epigenomic stratification & predictive biomarkers

Integrative epigenomic profiling fine-tunes the prognostic and predictive utility of genetic correlations in glioblastoma. The survival advantage historically connected to the pro-neural transcriptional subclass is almost exclusively driven by the glioma-CpG island methylator phenotype (G-CIMP), which confers a significantly prolonged median overall survival relative to non-G-CIMP malignancies (38.3 vs 12.7 months). Moreover, although MGMT promoter hypermethylation serves as a well-established favorable prognostic indicator, multi-omics combination reveals that its predictive biomarker for alkylating chemotherapy responsiveness is mainly restricted to glioblastomas classical subtype. (Brennan et al., 2013)

4.9 Linking non-invasive biomarkers to tissue phenotypes

Implementing machine-learning based dimensionality reduction to invasive tissue-level datasets from CPTAC & TCGA and non-invasive serum provides a systematic method for linking circulated sex-biased molecular signals with initial intracranial molecular changes. Detecting shared mutational and metabolic signatures between serum and tissue cohorts can help validate non-invasive biomarker profiling and facilitates the detection of actionable & sex-classified target for personalized medicine. (Taşçı & Chappidi, 2025)

4.10 Distinct epigenomic architecture & MicroRNA networks

Epigenomic clustering classifies glioblastoma into six different DNA methylation sub-classes (M1-M6) that concordant with nevertheless traditional transcriptomic classifications. While the M1 & M3 clusters are prevalent for classical & mesenchymal subtypes correspondingly, non-G-CIMP subclasses states different genetic patterns; specifically, classical GBMS display characteristic broad amplifications of chromosomes 19 & 20, whereas, non-G-CIMP pro-neural tumors selectively enhance for CDK4 and SOX2 amplifications. Apart from copy number alterations, transcriptomic subtype identities are modulated by complex microRNA-mediated competitive endogenous RNA (CrRNA) networks. Targeted CeRNA crosstalk for key drivers mirrors transcriptomic classes with PTEN, EGFR & PDGFRA CeRNA networks selectively co-segregating with classical, pro-neural and mesenchymal signatures profiles. (Brennan et al., 2013)

4.11 Sex-specific divergence in glioblastoma

Within a male-to-female frequency ratio of approximately 1.6:1, glioblastoma demonstrates a well-documented male preference. From a clinical perspective, male patients exhibit higher disease aggressiveness and less overall survival when compared with females. Irrespective of different molecular profiles like distinct metabolic drivers, localized tumor variants and high frequency of MGMT promoter methylation in female patients- the precise biological processes

regulating these differences remain unanswered. Furthermore, most existing bioinformatic studies validate GBM cohorts in aggregate without separating by sex as a biological variable (SABV), obscuring essential sex-specific genomic & metabolomic drivers. (Taşçı & Chappidi, 2025)

4.12 Innate sexual dimorphism & treatment responsiveness

In glioblastoma, sexual dimorphism expands beyond systemic circulating sex hormones and illustrates an intrinsic, cell autonomous property of the tumor genome. Assessment of therapeutic outcomes coupled with quantitative imaging-based response metrics reveals that conventional trimodal therapy yields subsequently higher efficacy in female patients when compared to male patients. This contrasting treatment sensitivity verifies that sex is the major clinical variable that directly influence therapeutic responses and patient survival, via sex-stratified models instead of aggregated population averages. (Yang et al., 2019)

4.13 Sex-specific molecular subtypes and survival signaling

By combining clinical outcome data with transcriptomic profiling by using computational subclassification algorithms, different molecular subtypes arise that govern survival through fundamentally distinct biological axes across sexes:

- *Male specific survival axis:* In male sub groups of glioblastoma, treatment resistance and overall survival crucially determined by mitotic division and genes regulating cell cycle progression.
- *Female specific survival axis:* In female sub groups of glioblastoma survival outcomes are greatly governed by routes controlling integrin signaling, cell-matrix interaction & cell invasion.

The clinical relevance of these distinct cell-cycle and integrin signaling signatures was further verified using in-vitro chemotherapy sensitivity testing among panels of female & male patient derived glioblastoma cell lines. (Yang et al., 2019)

In accordance with recent machine learning standards on TCGA glioblastoma study groups, the higher dimensionality & sparsity is exhibited by binary somatic mutation profiles when compared to transcriptomic or proteomic features. Mutational feature selection typically generates limited classification accuracy (e.g., ~63.6% using key driver mutations such as FAT4, BCOR and EGFR), while proteomic classifiers often obtain near perfect separation. This highlights that sex-specific genomic dimorphism in glioblastoma is marginal and polygenic which need robust feature selection workflows to extract meaningful driver variants from background somatic noise. (Taşçı & Chappidi, 2025)

Our identification of sex-biased somatic variants in whole exome sequencing (WES) study groups couples with evolving multi-omics evidence illustrating different oncogenic pathways in male vs female glioblastoma patients. From recent feature selections on TCGA glioblastoma datasets, EGFR, BCOR & FAT4 are isolated as primary genomic differentiating factors among sexes. In male tumors, EGFR alterations are in higher rates in tumor suppressor genes like FAT4 & transcriptional regulators like BCOR in female patients suggest that glioblastoma pathogenesis in women functions via alternative molecular axis. By incorporating WES somatic mutation profiles with ML feature selection allows us to obtain limited polygenic genomic signatures that traditional analyses oversee. (Taşçı & Chappidi, 2025)

4.14 Component-wise ensemble feature selection & dimensionality reduction

To mitigate the dimensionality reduction issue, a multi-tier feature selection pipeline is employed:

- *Selection frequency & Ensemble resampling:* Multiple regularized linear classifiers (such as 50 randomized LASSO estimators) are trained on sub-sample partitions of the training set to create a stable “bag of selected features”.
- *Structural feature deconstruction:* Instead of isolating unstable individual metrics, decompose selected features into high-level structural components (texture formula, spatial region, MRI sequence, filter transformation) and retaining only those attributes exhibiting statistically significant selection stability.
- *Orthogonal subspace projection:* The candidate feature set is constrained to a refined subspace and apply principal component analysis (PCA) or manifold learning to compress project into low-dimensional orthogonal embedding. (Nuechterlein et al., 2021)

5 TREATMENT

First-line treatment for glioblastoma continues to rely on maximal safe surgical resection followed by the Stupp protocol (radiotherapy paired with temozolomide) with MGMT promoter methylation standing out as the strong indicator of positive treatment response. Beyond first-line management, therapeutic landscape is broadening to include personalized neoantigen vaccines, CAR-T cell platforms (engineered against EGFRvIII or IL13R α 2), tumor-

treating fields (TTFields), and oncolytic virotherapies like DNX-2401 in combination with pembrolizumab. Moreover, developing targeted molecules like heparinase inhibitor RDS 3337-illustrates promise by disturbing autophagic flux and triggering programmed tumor cell death (apoptosis). Regardless these developments, long-term efficacy with single-agent regimens remains hindered by immunosuppressive tumor microenvironment (TME). Key drivers of this treatment resistance include heavy infiltration of tumor-associated macrophages (TAMs), elevated TGF- β secretion, low overall mutational burden and specific genomic alterations like PTEN loss. To overcome these evasion strategies, current research is shifting towards WHO CNS5-aligned multi-target combination and adaptive trial procedures that tailor treatments in real time using longitudinal multi-omics profiling. (Królikowska et al., 2025)

5.1 Targeted therapeutics and multi-omics integration

These finding of epigenomics have enhanced the progression of targeted small molecule inhibitors to suppress GSC self-renewal and reverse gene silencing. Histone deacetylase (HDAC) like Valproic acid, Vorinostat, Panobinostat, Belinostat are presently experiencing phase I-III medical evaluation to restore expression of tumor suppressor. Simultaneously, PARP inhibitors (e.g., veliparib, indaparib, olaparib) are being integrated with TMZ and radiation to disturb the networks of base-excision repair. Incorporating these multi-omics layers- single cell dynamics, spanning WES variant profiles and epigenomic signatures offers a robust foundation of machine learning models to attain precise risk stratification in neuro-oncology. (Jovčevska, 2020)

5.2 Therapeutic complexity & non-linear pathway signaling

Deliberate therapeutic targeting in glioblastoma is fundamentally tricky by structural diversity and non-linear signaling dynamics. Beyond primary oncogenes, RNA sequencing & whole genome identify clinically actionable variants in BRAF and FGFR1-3 axis. Even though effective precision targeted therapy faces huge hurdles due to intra-tumoral heterogeneity; various EGFR alterations-including exon deletion variants and novel transcript chimeras recurrently co-occur within single tumor masses, yielding variable drug sensitivity. Compounding this challenge, reverse-phase protein array profiling illustrates a marked discordance across downstream proteomic activation & genomic alterations. This non-linear pathway signaling marks that therapeutic inhibition of downstream efforts can't be expected equivalent to directly targeting the mutated upstream driver. (Brennan et al., 2013)

5.3 Intratumoral heterogeneity & genomic dynamics

Beyond fixed genomic mutations, single-cell RNA sequencing (ScRNA-seq) illustrates that glioblastoma poses extensive cellular diversity, including non-neoplastic microenvironmental elements such as microglia and macrophages and also include specialized stem cell populations (GSCs). Under the selective pressure specialized therapies like temozolomide (TMZ), this intra tumor complexity extends, subsequently triggering a hypermutated phenotype that speed up treatment resistance and recurrence. To tackle the short-read sequencing limitations in complex structural rearrangements, Third-Generation Sequencing (TGS) framework such as Oxford Nanopore and PacBio SMRT enable evaluation (> 1kb) of long-read single molecule without delivering single-base resolution of sub-clonal variations and PCR amplification. (Jovčevska, 2020)

Exhaustive whole-genome & RNA sequencing from TCGA demonstrates that glioblastoma genomes predominantly exhibit complex structural variants beyond standard single-nucleotide alterations. Pathogenic gene fusions prominently EGFR-SEPT14 chimeras and FGFR3-TACC3 inversions occur in different sub-populations and co-localize with focal genomic amplifications. Correspondingly, high-level focal co-amplification targeting EGFR, CDK4 & MDM2 on chromosome 12q are primarily localized on circular extrachromosomal double minute DNA structures. Structural characterization of the EGFR locus demonstrates that 57% of GBMs present with multi-dimensional aberrations covering extracellular domain deletions (EGFR vIII), in-frame splice variants and c-terminal cytoplasmic deletions that remove inhibitory signaling domains. (Brennan et al., 2013)

5.4 Contextual application to WES variant filtering & study design

These key findings provide robust biological justification for evaluating sex separated glioblastoma sequencing cohorts. By implementing population variant filtering databases like hg38 & gnomAD to eliminate background noise of non-pathogenic germline, the remaining somatic variants can be mapped directly to those sex-biased signaling pathways. Illustrating whether male skewed driver mutations concentrate in cell-cycle regulators (e.g., RB1, TP53, CDK4/6) while female skewed variants enrich in invasive & integrin pathway genes generates a concrete mechanism-based pathway. (Yang et al., 2019)

5.5 Benchmarking WES vs WGS & Deep learning

When whole genome sequencing (WGS) combined with deep learning networks can obtain high diagnostic accuracy, computational capacity, prolonged turn-around time, and its clinical translation remain restricted by steep sequencing costs. Conversely, this framework tells that isolating targeted pathogenic somatic variants from whole exome

sequencing (WES) data by using explainable machine learning models such as XGBoost and random forest strikes a flexible balance between diagnostic precision and computational efficiency. Moreover, ideal with multi-cancer framework, combining functional pathogenicity with different somatic point mutations score mitigates matrix sparsity and improve model generalizability without need of opaque deep learning pipelines. (Zelli et al., 2023)

5.6 XGBoost

Executing gradient boosting decision trees (XGBoost) to high-throughput whole-exome sequencing (WES) somatic variants matrices has proven highly effective classification. In a pan-cancer criterion across 32 types of tumors, integrating copy number variations (CNVs) and somatic point mutations (SPMs) promotes XGBoost to obtain an 86.5% balanced accuracy (AUC=0.96) among neural/melanoma tumor class, generating an individual classification accuracy of 79% for glioblastoma multiforme (GBM) (Zelli et al., 2023)

6 CONCLUSION

The combination of whole-exome sequencing (WES) with advanced machine learning approaches provides a promising and cost-efficient computational strategy for addressing the challenges associated with conventional bulk-tumor diagnostics and single-biomarker analysis in glioblastoma. Traditional single-gene approaches and unprocessed genomic feature matrices can be affected by high-dimensional data, technical and biological noise, and substantial class imbalance. The application of combined feature-selection methods and regularized machine learning models, including L1-regularized approaches and tree-based ensemble techniques, can improve the identification of potentially pathogenic driver changes while minimizing the influence of passenger mutations. The findings further highlight the importance of integrating genomic information with complementary transcriptomic, epigenomic, and radio-genomic profiles to better characterize complex signaling interactions, tumor heterogeneity, and clonal evolution associated with treatment resistance. Overall, the development of interpretable and reproducible AI-driven analytical pipelines has considerable potential to improve molecular and prognostic assessment, facilitate biomarker-informed therapeutic selection, and contribute to the advancement of personalized precision oncology strategies based on individual tumor characteristics and patient-specific molecular profiles. "By bridging theoretical genomic pathways with an empirical TCGA-GBM ($N = 79$) computational pipeline, this study demonstrates that while single-modality exome mutation profiling faces intrinsic class-imbalance and dimensionality challenges, integrating multi-tier feature selection provides a robust path forward for neuro-oncology precision medicine."

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments

The author sincerely thanks Dr. Elamathi Natarajan for her valuable guidance and support during the preparation of this manuscript.

Disclosure of conflict of interest

The authors declare no conflicts of interest regarding this manuscript.

REFERENCES

- [1] Garrett, A.-M., Lastakchi, S., & McConville, C. (2020). The personalisation of glioblastoma treatment using whole exome sequencing: A pilot study. *Genes*, 11(2), 173. <https://doi.org/10.3390/genes11020173>
- [2] Nuechterlein, N., Li, B., Feroze, A., Holland, E. C., Shapiro, L., Haynor, D., Fink, J., & Cimino, P. J. (2021). Radiogenomic modeling predicts survival-associated prognostic groups in glioblastoma. *Neuro-Oncology Advances*, 3(1), vdab004. <https://doi.org/10.1093/noajnl/vdab004>
- [3] Jovčevska, I. (2020). Next generation sequencing and machine learning technologies are painting the epigenetic portrait of glioblastoma. *Frontiers in Oncology*, 10, 798. <https://doi.org/10.3389/fonc.2020.00798>
- [4] Yang, W., Warrington, N. M., Taylor, S. J., Whitmire, P., Carrasco, E., Singleton, K. W., Wu, N., Lathia, J. D., Berens, M. E., Kim, A. H., Barnholtz-Sloan, J. S., Swanson, K. R., Luo, J., & Rubin, J. B. (2019). Sex differences in GBM revealed by analysis of patient imaging, transcriptome, and survival data. *Science Translational Medicine*, 11(473), eaao5253. <https://doi.org/10.1126/scitranslmed.aao5253>
- [5] Sarker, A., Uddin, B., Ahmmed, R., Mahmud, S., Ajadee, A., Pappu, M. A. A., Aziz, M. A., & Mollah, M. N. H. (2025). Discovery of mutated oncodriver genes associated with glioblastoma originated from stem cells of subventricular zone through whole exome sequence profile analysis, and drug repurposing. *Heliyon*, 11(1), e42052. <https://doi.org/10.1016/j.heliyon.2025.e42052>

- [6] Kim, H., Zheng, S., Amini, S. S., Virk, S. M., Mikkelsen, T., Brat, D. J., Grimsby, J., Sougnez, C., Muller, F., Hu, J., Sloan, A. E., Cohen, M. L., Van Meir, E. G., Scarpance, L., Laird, P. W., Weinstein, J. N., Lander, E. S., Gabriel, S., Getz, G., Meyerson, M., Chin, L., Barnholtz-Sloan, J. S., & Verhaak, R. G. W. (2015). Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Research*, 25(3), 316–327. <https://doi.org/10.1101/gr.180612.114>
- [7] Taşçı, E., & Chappidi, S. (2025). GLIO-Select: Machine learning-based feature selection and weighting of tissue and serum proteomic and metabolomic data uncovers sex differences in glioblastoma. *International Journal of Molecular Sciences*, 26(9), 4339. <https://doi.org/10.3390/ijms26094339>
- [8] Zelli, V., Manno, A., Compagnoni, C., Ibraheem, R. O., Zazzeroni, F., Alesse, E., Rossi, F., Arbib, C., & Tessitore, A. (2023). Classification of tumor types using XGBoost machine learning model: A vector space transformation of genomic alterations. *Journal of Translational Medicine*, 21, 836. <https://doi.org/10.1186/s12967-023-04720-4>
- [9] Alter, O., Ponnappalli, S. P., Coppola, M., Gushue, A. C., House, T. O., Miron, P. L., Miskimen, K. L., Waite, K. A., Pollock, S., Bogumil, D., Iremadze, N., Hernandez, S., Sosonkina, N., Coppens, S. E., Bryan, A. C., Kiernan, E. P., Yang, H., Bowen, J., Nakouzi, G. A., ... Tsai, J. W. (2026). Quantum mechanics-based multi-tensor AI/ML correctly predicts—glioblastoma patients’ overall survival, gene targets to sensitize the tumors, and the tumors’ response to their targeting—from their whole genomes. *Cancer Research*, 86(7_Supplement), 6884. <https://doi.org/10.1158/1538-7445.AM2026-6884>
- [10] Tbahrity, H. F., Boukadoum, A., Benbernou, M., & Belhocine, M. (2025). Machine learning and deep learning in glioblastoma: A systematic review and meta-analysis of diagnosis, prognosis, and treatment. *Discover Oncology*, 16, 1492. <https://doi.org/10.1007/s12672-025-03303-7>
- [11] McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, 17, 122. <https://doi.org/10.1186/s13059-016-0974-4>
- [12] Królikowska, K., Błaszczak, K., Ławicki, S., Zajkowska, M., & Gudowska-Sawczuk, M. (2025). Glioblastoma—A contemporary overview of epidemiology, classification, pathogenesis, diagnosis, and treatment: A review article. *International Journal of Molecular Sciences*, 26(24), 12162. <https://doi.org/10.3390/ijms262412162>
- [13] Grochans, S., Cybulska, A. M., Simińska, D., Korbecki, J., Kojder, K., Chlubek, D., & Baranowska-Bosiacka, I. (2022). Epidemiology of glioblastoma multiforme—Literature review. *Cancers*, 14(10), 2412. <https://doi.org/10.3390/cancers14102412>
- [14] Brennan, C. W., Verhaak, R. G. W., McKenna, A., Campos, B., Nounshmehr, H., Salama, S. R., Zheng, S., Chakravarty, D., Sanborn, J. Z., Berman, S. H., Beroukhi, R., Shukla, S. A., ... Chin, L. (2013). The somatic genomic landscape of glioblastoma. *Cell*, 155(2), 462–477. <https://doi.org/10.1016/j.cell.2013.09.034>
- [15] Pouyan, A., Ghorbanlo, M., Eslami, M., Jahanshahi, M., Ziaei, E., Salami, A., Bernard, B., Wu, C.-J., Genovese, G., Shmulevich, I., Barnholtz-Sloan, J. S., Zou, L., Vegesna, R., Mokhtari, K., Shahpasand, K., Farahani, N., Emami Meybodi, T., Entezari, M., Taheriazam, A., Hushmandi, K., & Hashemi, M. (2025). Glioblastoma multiforme: Insights into pathogenesis, key signaling pathways, and therapeutic strategies. *Molecular Cancer*, 24, 58. <https://doi.org/10.1186/s12943-025-02267-0>