

Advancing Glioblastoma Prognosis: A Review of Machine Learning, Radiomics, and Multi-Omics Integration for Survival Prediction and Subtyping

Muna Awel, Rushit Dave, Mansi Bhavsar

Department of Computer Information Science, Minnesota State University at Mankato, Mankato, USA
Email: Rushit.dave@mnsu.edu

How to cite this paper: Awel, M., Dave, R. and Bhavsar, M. (2025) Advancing Glioblastoma Prognosis: A Review of Machine Learning, Radiomics, and Multi-Omics Integration for Survival Prediction and Subtyping. *Journal of Computer and Communications*, 13, 29-45.

<https://doi.org/10.4236/jcc.2025.135003>

Received: April 8, 2025

Accepted: May 20, 2025

Published: May 23, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Glioblastoma (GBM) is known for its poor prognosis and aggressive nature, driving the need for advanced models that provide survival prediction to improve patient prognosis. This literature review synthesizes 20 studies employing machine learning (ML), deep learning (DL), radiomics, and clustering to enhance GBM prognosis using clinical, imaging, and molecular data. The review is grouped into five thematic groups based on their methodology and data type. Studies within these groups explore the predictive accuracy of models, evaluation metrics, and data types of implementations. The research also highlights the limitations which include small datasets, sparse clinical variables, computational complexity, and hindering scalability and generalizability. This review provides a clear understanding of the various studies that are aimed at enhancing patient prognosis in glioblastoma through advanced predictive and subtyping methodologies.

Keywords

GBM, Survival Prediction, Machine Learning (ML), Deep Learning (DL), Radiomics, Multi-Omics, Feature Engineering Clustering

1. Introduction

Glioblastoma Multiforme (GBM) is the most aggressive brain tumor with the median overall survival rate at only 15 months [1]. GBM is known for its poor patient prognosis due to its intra-tumor heterogeneity [2]. This heterogeneity reduces the value of biopsy of the analysis of genomics providing insight for exploring medical

imaging that can illustrate the whole genome [2]. These cancer cells originate from Glial cells, and its malignant glioma is grade (IV) [3]. Statistical data shows that GBM malignancy accounts for 45.2 % primarily in the brain and central nervous system reflecting on how deadly and aggressive cancer type it is [3]. The risk of developing GBM increases significantly with age, with the highest incidence observed in older adults [1]. **Figure 1** depicts various factors influencing the occurrence of glioblastoma multiforme (GBM).

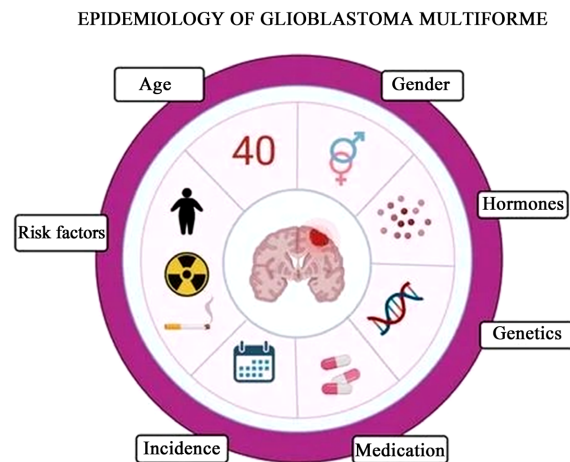


Figure 1. Key epidemiological factors of glioblastoma multiforme (GBM) [1].

Mechanisms for treatment strategy include surgical removal, followed by radiotherapy. After this stage of application, a key therapeutic drug like TMZ is mostly used. This drug is an oral alkylating agent, and it has demonstrated anti-tumor activity as a single agent in the treatment of recurrent glioma [4]. This means that the drug mainly interferes with the tumor's DNA to halt the process of its growth. Despite these treatment strategies key genetic biomarkers like IDH mutations, MGMT promoter methylation, and EGFR amplification have been identified as critical factors influencing survival outcomes. These genetic biomarkers are essential as they play a role in identifying the effectiveness of treatment options and provide a clear insight into the overall patients' survival. Research has shown that 73.4% of IDH mutations were associated with prolonged progression and higher response to TMZ [5]. Additionally, patients with both IDH mutations and MGMT promoter methylation showed the best response rate to TMZ, as both genetic features are essential for patient effective response to drug and prolonged prognosis [5]. EGFR is the other genetic biomarker that is a transmembrane receptor that mutations and overexpression shows the aggressive nature of the glioma cell [6]. Its amplification and overexpression showed a homogeneous distribution across the tumor tissue and were barely detectable in normal brain tissue, which gives an advantage to EGFR-targeted therapy [6]. Thus, this means that a higher level of EGFR expression indicates poor prognosis and shorter survival of patients with GBM. Overall, this interaction between biomarkers and patient characteristics is

essential to provide an effective personalized treatment plan. However, the heterogeneous nature of GBM makes it difficult to predict patient prognosis or overall patient survival rate. Research conducted on building a predictive model using machine learning techniques to analyze complex high dimensional data to uncover patterns and have improved prognosis.

This literature review examines the role of ML models and Deep Learning in predicting overall survival for patients with GBM. It explores how each model is built with multi-omics and multimodal datasets including genomic, transcriptomic, radiomic, and histopathological imaging data, to extract prognostic insights and improve patient outcomes. Additionally, it dives into the accuracy and effectiveness of each model, evaluating their strengths and limitations when applied to diverse datasets.

1.1. Recent Trends on Survival Prediction Analysis

Several recent research on survival predictions focusses on utilizing multimodal data which includes clinical features, imaging data, genomic features and histopathological information. The development of deep learning methods like CNN has shown improved predictive analysis for medical images like MRI scans [7]. CNN tends to have better performances in identifying complex patterns with 3D medical imaging data, with better segmentation of tumors [7]. Another study explores deep learning to classify tumor types based on their gene expressions [8]. This is to help in identifying patterns of gene expressions that can have potential in diagnosis, prognosis, and predictive value in treatment plan [7]. Additionally, research has been rapidly increasing utilizing multi-omics datasets. These datasets include genomic features like gene expression, transcriptomics and proteomics in which they get combined with demographic data like age, gender, and ethnicity. The integration of multi-omics data enhances the accuracy and performance of survival prediction models by providing a comprehensive view of tumor biology [9]. This allows researchers to uncover biological pathways and molecular mechanisms that improve patient prognosis. Key aspect in integrating datasets is to determine which model suits better. Studies introduce unsupervised learning to uncover hierarchical structure within cancer gene expression data [10]. Recent studies show that the advancement of ML models such as Random Forests (RF), Support Vector Machines (SVM), and deep learning architectures, including Convolutional Neural Networks (CNNs) and Multi-Layer Perceptron's (MLPs), have been applied to multi-omics datasets to improve the accuracy of survival predictions [11]. The advancement of ML models provides insight into patient prognosis and aids in identifying novel therapeutic targets by uncovering hidden patterns across different biological levels, contributing to the development of more effective treatment plans for glioblastoma patients [12].

1.2. Research Questions

- 1) How have machine learning models been applied to multi-omics datasets for

predicting survival in glioblastoma patients?

2) What types of machine learning techniques have been most successful in analyzing multi-omics data for survival prediction in glioblastoma patients?

3) How do deep learning approaches compare to traditional machine learning methods in terms of accuracy and interpretability when applied to multi-modal datasets for survival prediction?

2. Literature Review

To provide a structured analysis, each study is categorized into five thematic groups based on their focus, methods, and data types. This category reflects the evolution of survival prediction from traditional ML to advanced DL, while highlighting their contributions and performance of models in survival analysis.

2.1. ML Models with Omics Data for GBM Survival Prediction

Studies within this group utilize ML to predict GBM survival using omics data, which includes genomics, transcriptomic, and epigenomic features over traditional imaging. Research [13] explores SVM, RF, and linear regression using TCGA datasets of 577 samples. Evaluation metrics of C-index was used to evaluate the predictive performance of overall survival and correlated features. SVM outperformed the other models with a higher C-index value [13]. **Figure 2** illustrates a flowchart methodology of utilizing omics datasets and building ML models.

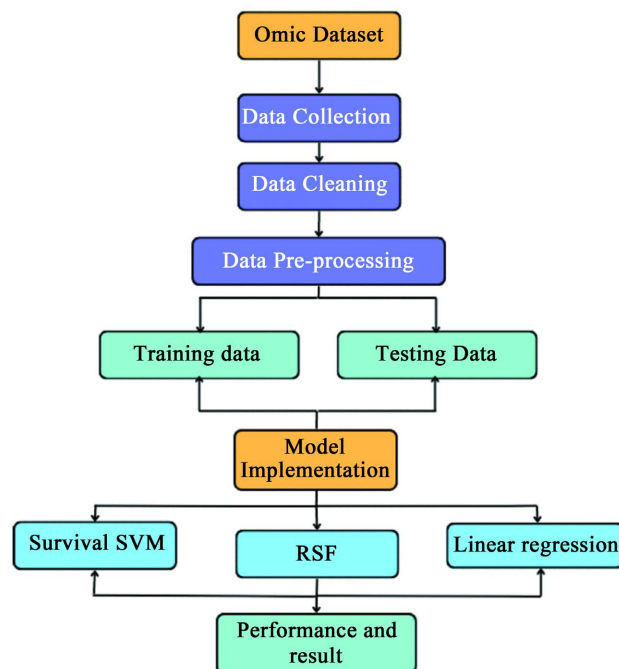


Figure 2. Methodology of survival predictions with multi-omics datasets [13].

In a similar way [14] evaluates seven ML models which include Survival Forest, Random Forest Regressor, Gradient Boosting Regressor, XGBRegressor, Survival

Support Vector Machine (SVM), Ridge Regressor, and Lasso Regressor. A larger set of cohorts with 619 TCGA samples were used, and Gradient Boosting achieves a standout C-index, underscoring its ability to handle complex omics interactions [14]. In [15], a Bayesian Neural Network (BNN) is employed against traditional ML models on TCGA and CGGA mRNA datasets (252 and 315 samples). This model achieved 70% accuracy in survival classification and identifying seven prognostic genes tied to high-risk profiles. Additional research studies focus on utilizing LASSO regression on microRNA expression data from 475 TCGA samples, developing a 9-microRNA signature. This effectively stratified patients into high-risk (median survival: 9.5 months) and low-risk (13.1 months) groups, validated across GBM subtypes and treatment cohorts [16]. The study in [17] extends the scope with deep learning via SurvivalNet, applied to multi-omics TCGA data, demonstrating superior performance over Cox Elastic Net and RF, particularly for GBM, with key prognostic features like IDH1/IDH2 mutations and PTEN deletions emerging as critical predictors.

Comprehensive analysis within this group shows that the performance of models with omics datasets is better than imaging based evidenced by [13] and [14] which explains on the strength of omics data's ability to capture GBM's molecular heterogeneity. The studies were also able to determine key biomarkers for longer survival rates in the expression of IDH mutations and G-CIMP methylation which correlates with longer survival rates [13]. The methodology within the study explores the use of preprocessing techniques like imputation [13] and feature selection [14] [16] which enhances the data quality. Additionally, the use of the C-index as a primary evaluation metric, favored over accuracy for its suitability in continuous survival outcomes [14], reflecting on a methodological consensus that enhances comparability. Moreover, advanced techniques like BNN and Survival Net [17] introduce probabilistic and deep learning frameworks, offering improved handling of high-dimensional data compared to traditional ML approaches like RF or Linear Regression [13].

2.2. Deep Learning and Multi-Modal Approaches for GBM Survival Prediction

Studies with this category explore deep learning (DL) and multi-modal data integration (imaging, omics, demographic) to predict GBM survival which includes a more advanced perspective compared to utilizing only omics data. Study in [18] employs a comparative analysis between multi-path CNN versus traditional ML models (SVM, RF). These models were employed on TCGA dataset of 272 samples, preprocessing single nucleotide polymorphism (SNP) and demographic data (age, gender) with normalization and one-hot encoding, followed by 10-fold cross-validation, achieving higher accuracy than base-lines [18]. In a similar way [19] this study also explores multi-task CNN, compared to mono-task CNN and RF using MRI and genomic biomarker. This study involves a thorough step of image segmentation and feature extractions. The model was evaluated with root

mean square error. The multi-task CNN leverages multi-modal MRI (T1c, B0, FA, MD) and predicts genomic biomarkers (MGMT, IDH, 1p/19q, TERT) along-side overall survival (OS), demonstrating significant improvements over single-task models [19]. Additionally, in [20] a novel Histopathological Integrating Multiple Kernel Learning (HI-MKL) method integrates histopathological images and multi-omics TCGA data. The methods involve preprocessing images with stain normalization and omics with principal component analysis (PCA), optimizing kernels via a genetic algorithm [20]. The Kaplan-Meier (K-M) estimate plot in **Figure 3** illustrates survival probabilities, incorporating a 2-year survival timeline [20]. The K-M plot effectively differentiates between long-term and short-term survivors, highlighting the method's ability to stratify patients based on survival outcomes [20]. These findings underscore the HI-MKL method's robust performance in GBM prognosis prediction. The integration of histopathological and multi-omics data, combined with advanced preprocessing and optimization techniques, positions HI-MKL as a promising approach for improving prognostic accuracy in GBM, with superior AUC performance compared to existing MKL methods.

Research in [21] enhances prognosis with Multiple Kernel Learning (MKL) and minimum Redundancy Maximum Relevance (mRMR). Feature selection on a TCGA cohort of 276 samples, was performed by integrating gene expression, methylation, miRNA, copy number, and age [21]. It employed a radial basis function kernel and 5-fold cross-validation, optimizing AUC with 130 features. Finally, [22] introduces comparative analysis on a GraphSurv, a Graph Convolutional Network (GCN), versus DeepSurv and Random Survival Forest (RSF) on TCGA multi-omics data across 11 cancer types (including GBM), constructing graph structures with have a better C-index performance in survival analysis.

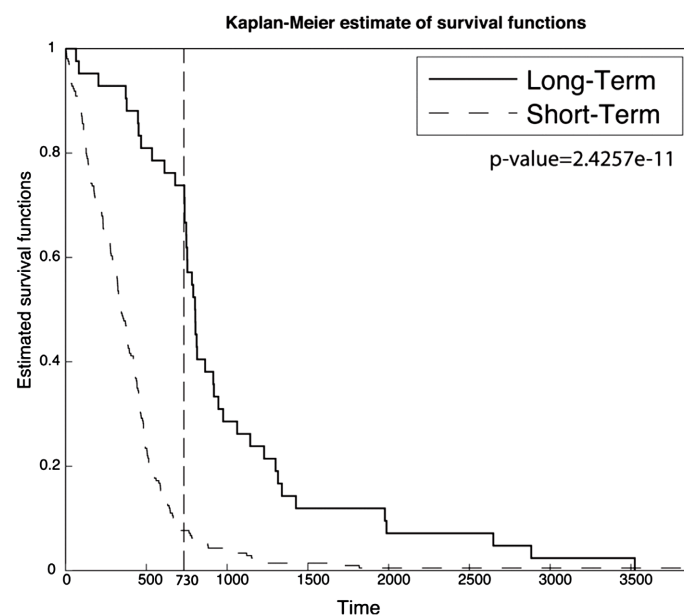


Figure 3. Kaplan-Meier plot survival analysis of GBM [20].

Comprehensive analysis within these studies explores multi-modal datasets of TCGA and imaging. The major strengths within the findings of the studies are the superior performance of DL and hybrid models over traditional ML, as evidenced by [18]’s CNN outperforming SVM/RF and [19]’s multi-task CNN reducing RMSE with MGMT/IDH integration. Multi-modal fusion enhances predictive power, with [20] and [21] leveraging MKL to combine histopathological and omics data, identifying critical features like methylation profiles and age [21]. The GCN approach in [22] innovates by incorporating biological pathway knowledge, boosting prognostic accuracy across cancers, including GBM.

2.3. Clinical and Large-Scale Data Applications in GBM Survival Prediction

This group of study focuses on real-world clinical utility, using datasets like cancer registries and integrated clinical cohorts (see **Table 1** for details). The study demonstrates the power of applying ML extensively to cancer registry, achieving a higher predictive performance of SVM which excels at modeling non-linear relationships among clinical features such as age and resection extent [23]. This approach yields reliable survival predictions, emphasizing scalability and practicality over the complexity of multi-modal DL, making it a potential blueprint for routine clinical tools where rapid, interpretable outcomes are critical [23]. **Figure 4** presents ROC (Receiver Operating Characteristics) curves to assess the sensitivity and specificity of ML models. For the entire glioma patient cohort (**Figure 4(a)**), the AUC values are 0.854 (CPH), 0.849 (SVM), and 0.844 (RF), indicating similar performance in distinguishing 1-year survivors from non-survivors. For the GBM subset (**Figure 4(b)**), the AUC values are 0.823 (CPH), 0.821 (SVM), and 0.825 (RF), reflecting comparable but slightly lower performance due to the complexity of GBM. The close AUC values across models suggest that CPH, SVM, and RF have similar predictive capabilities for 1-year survival, with minimal differences in sensitivity and specificity.

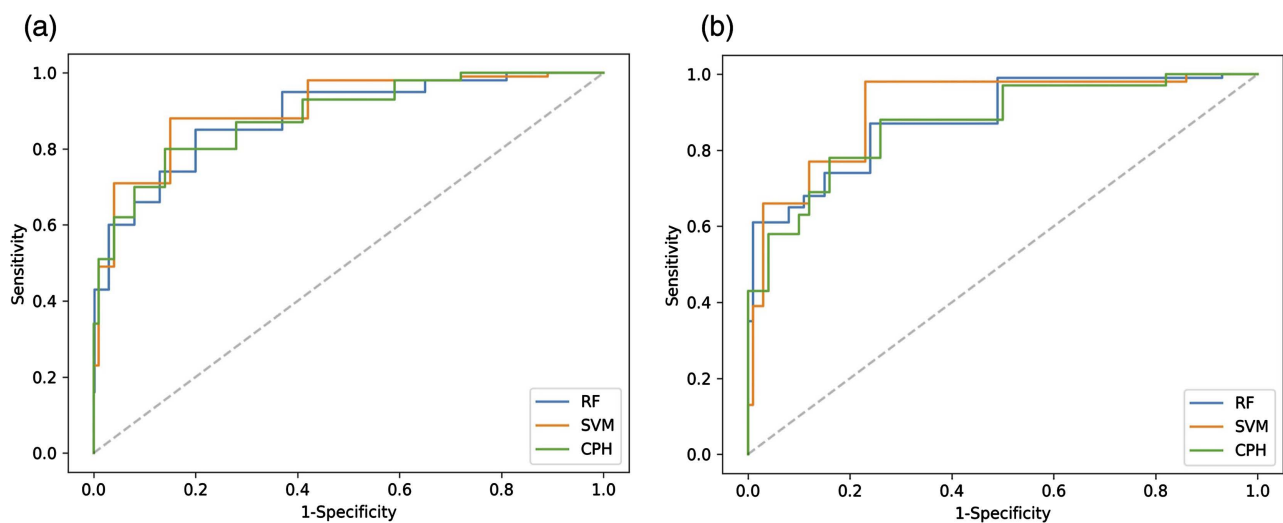


Figure 4. ROC Curve for ML models to assess the sensitivity and specificity [23].

Additionally, the study within [24] takes a different approach as it highlights the potential of combining radiomics with clinical and genomic features, integrating diverse data types to enhance prognostic accuracy beyond what clinical variables alone can achieve. By extracting quantitative imaging features (e.g., texture, shape) alongside genomic markers like MGMT methylation, this study employs a regularized Cox Proportional Hazards (Cox-PH) model to balance complexity and clinical relevance [24]. To enhance GBM prognosis beyond predictive modeling, clustering clinically relevant subtypes offer molecular insights within a clinical framework [25]. This provided insight on molecular perspective within a clinical framework by analyzing gene expressions to segment patients into survival-linked categories [25]. It employed hierarchical clustering with agglomerative average linkage to detect robust clusters from a dataset of 200 GBM samples and two normal samples based on gene expression profiles [25]. The findings included the identification of subtype-specific signature genes, significant gene ontology terms, and correlations with genomic alterations such as copy number variations and mutations. Additionally, the study in [26] predicts overall survival (OS) in glioblastoma (GBM) patients using radiomic features extracted from multi-modal MRI (FLAIR and T1ce) data. The study's approach takes extracting 679 radiomic features (first-order, shape-based, textural, wavelet-decomposed) from segmented tumor regions, reduces them to 54 via correlation and significance filtering [26]. It employs multilayer perceptron (MLP) and random forest (RF) to predict OS and classify patients into survival groups [26]. The outcome of the study shows MLP outperforming RF tending to achieve better performance with radiomics features.

2.4. Review and Comparative Studies in Survival Prediction

This group studies reviews that provides broad overviews or empirical comparisons of survival prediction methods, often spanning multiple cancer types or techniques. The review in [27] analyzes 107 studies (2015-2024) on ML and DL for GBM survival prediction. It identifies Random Forest as a widely used model and the combination of radiomics and clinical data as a common input [27]. Evaluated metrics include accuracy and C-index, with findings emphasizing personalized decision-making potential. It highlights the potential of machine learning and deep learning in personalized clinical decision-making, given the complexity and heterogeneity of GBM [27]. Additionally, the study in [28] provides comparative analysis study of statistical and machine learning techniques used in cancer survival analysis. It aims to assess the role of machine learning in predicting cancer patient survival compared to traditional statistical methods [28]. It mainly compares statistical (Cox PH with Ridge penalty) and ML methods (Random Survival Forests, Gradient Boosting with Cox PH loss, SVM) across 10 cancer survival datasets, including GBM [28]. This literature review provides a broader context of offering benchmarks illustrating how Machine learning model perform better than statistical model with their interpretability and application to GBM and other cancer types.

2.5. Graph-Based and Clustering Approaches for Survival and Subtyping

The study within this group explores cancer survival predictions and subtyping using graph based and clustering methodologies. The approaches span on CNN with support vector machines modified SVM regression, supervised graph clustering (S2GC), and regularized unsupervised multiple kernel learning (rMKL-LPP) which offers insights to capture structural relationships and molecular heterogeneity, addressing limitations of traditional machine learning (ML) and deep learning (DL) models highlighted in Groups 2.1 and 2.2. The study explores methodology to predict 5-year survival rates in rectal cancer patients using RhoB-stained immunohistochemistry (IHC) images [29]. This research leverages CNNs for feature extraction, capturing intricate patterns from preprocessed IHC images, followed by SVM classification to enhance predictive accuracy over traditional manual analysis [29]. The study contributes significantly to cancer prognosis by showcasing AI's potential to optimize treatment strategies, proposing future exploration of multi-layer CNN feature fusion to further refine predictions. Another study explores onto the SVM regression approach applied to censored survival data, validated across datasets such as “alco”, “breast”, and “myel” [30]. This study mainly utilizes traditional SVM by optimizing key parameters like penalty parameter (C) and kernel choice (Polynomial vs. RBF)—to predict survival times [30]. This is then evaluated with conservative root mean square error (RMSE) compared to Cox Proportional Hazards (Cox-PH) models [30]. Although the model isn't on GBM cancer data, it shows how models like SVM potential in additional regression-based survival analysis. The study in [31] introduces S2GC, a novel approach for cancer subtype identification that integrates multi-omics data with survival analysis. It integrates miRNA, mRNA, and methylation profiles with survival analysis, constructing a graph-based framework to identify survival-linked subtypes. The results show that S2GC outperforms established methods like iCluster and Similarity Network Fusion (SNF), achieving superior subtype differentiation validated by Cox log-rank tests and Kaplan-Meier survival curves [31]. This superiority stems from its unified optimization framework, where survival analysis model learning and patient similarity graph learning reinforce each other [31].

Another study within this sub-group introduces the rMKL-LPP method for integrating multi-type cancer data (gene expression, DNA methylation, miRNA expression) to improve patient clustering and survival prediction [32]. This study employs regularized multiple kernel learning with locality-preserving projections (LPP), testing two scenarios—3 K (one kernel per data type) and 15 K (five kernels per type) to enhance clustering and survival prediction [32]. The study employs k-means clustering with optimal cluster numbers. The outcome of the study illustrates on how rMKL-LPP (15 K) outperforms Similarity Network Fusion (SNF), achieving a median log-rank p-value of 2.4×10^{-4} versus SNF's 0.0011, alongside greater stability in leave-one-out cross-validation. Additionally, rMKL_LPP identifies clinically relevant subtypes, including clusters aligned with known gene ex-

pression and methylation profiles, and those showing differential Temozolomide responses, enhancing personalized treatment insights [32]. Overall, this study provides a deep understanding of using multiple kernels per data type captures intricate data patterns, improving survival differentiation and robustness over unregularized MKL-LPP and SNF.

3. Discussion

The studies within the literature review explore the methods of ML and DL models mainly utilizing multi-omics, radiomics, and clinical data to predict overall survival (OS) in GBM patients. **Table 1** shows an overview of the methodology, datasets, result and contribution of each study. Support Vector Machines (SVM), as demonstrated in studies [13] and [23], excel in handling high-dimensional omics data, achieving higher concordance indices (c-index) than RF and Cox-PH models by leveraging kernel functions to capture non-linear relationships, such as those involving IDH1 mutations and G-CIMP methylation. However, SVM's reliance on hyperparameter tuning, as noted in [30], can complicate model optimization, and its limited interpretability hinders clinical translation. Random Forests, widely used in [14] and [26], offer resilience to noisy data and provide feature importance scores that highlight prognostic biomarkers like MGMT methylation,

Table 1. ML models, datasets, and findings in glioblastoma prognostic research.

Title	Methods	Datasets	Contributions
Survival Rate Prediction in Glioblastoma Patients Using Machine Learning [13]	SVM, RF, Linear Regression; C-index	TCGA (577 samples); Omics (gene expression, methylation, IDH1 mutation)	SVM outperformed others (high C-index); G-CIMP methylation and IDH1 mutation strongly correlated with survival
Machine Learning and Omics Data for Predicting Overall Survival in Glioblastoma Patients [14]	Seven ML models (e.g., Gradient Boosting Regressor, SVM, RF); C-index	TCGA (619 samples); Omics (genomic, clinical features)	Gradient Boosting Regressor achieved the highest C-index (0.81); Omics integration improved prognosis prediction
Survival Prediction and Risk Estimation of Glioma Patients Using mRNA Expressions [15]	Bayesian Neural Network (BNN) vs. ML (RF, SVM); Accuracy, Precision, F1 scores	TCGA (252 samples), CGGA (315 samples); mRNA expression	BNN outperformed ML models (70% accuracy); Identified 7 prognostic genes linked to survival and risk
Multi-path convolutional neural network for glioblastoma survival group prediction with point mutations and demographic feature [18]	Multi-path CNN vs. SVM, RF; 10-fold cross-validation, accuracy	TCGA (272 samples); SNP data, demographic features (age, gender)	CNN outperformed SVM/RF; Combining SNP and demographic data enhanced predictive accuracy

Continued

A Novel MKL Method for GBM Prognosis Prediction by Integrating Histopathological Image and Multi-Omics Data [20]	Histopathological Integrating Multiple Kernel Learning (HI-MKL); AUC	TCGA via cBioPortal; Multi-omics (genomic) and histopathological images	HI-MKL improved accuracy over Simple MKL; Effective integration of histopathological and omics data enhanced prognosis
Improve Glioblastoma Multifome Prognosis Prediction by Using Feature Selection and Multiple Kernel Learning [21]	MKL with mRMR feature selection; AUC, ROC curves	TCGA (276 samples); Multi-omics (gene expression, methylation, miRNA, copy number), age	MKL outperformed single-kernel methods; 130 features optimized AUC, showing multi-omics superiority
Prediction of Clinical Outcome in Glioblastoma Using a Biologically Relevant Nine-microRNA Signature [16]	LASSO regression; Cox regression, log-rank tests	TCGA (475 samples); MicroRNA expression	9-microRNA signature stratified high/low-risk groups Linked toMAPK/WNT pathways
Deep Learning of Imaging Phenotype and Genotype for Predicting Overall Survival Time of Glioblastoma Patients [19]	Multi-task CNN vs. mono-task CNN, RF; RMSE, Pearson CC	120 patients; MRI (T1c, DWI), genomic biomarkers (MGMT, IDH, etc.)	Multi-task CNN achieved lower RMSE; MGMT/IDH and B0 MRI were key predictors.
Machine Learning Based Survival Prediction in Glioma Using Large-Scale Registry Data [23]	SVM, Cox PH, RF; C-index, Kaplan-Meier curves	BC Cancer Registry (2000-2018, 3462 patients); Clinical features (age, resection status, etc.)	SVM had the highest C-index; Clinical features drove robust predictions in a large cohort
Survival Prediction of Glioblastoma Patients Using Machine Learning and Deep Learning: A Systematic Review [27]	Literature review (107 studies, 2015-2024); Analyzed trends	Mostly TCGA (107 studies); Various (omics, radiomics, clinical)	RF and radiomics-clinical data prevalent; Gaps in interpretability and validation identified
Machine Learning for Survival Analysis in Cancer Research: A Comparative Study [28]	Cox-PH, RF, Gradient Boosting, SVM; Literature review, empirical comparison	10 cancer datasets (not specified); Various (clinical, genomic)	Cox-PH comparable to ML; ML offers flexibility but lacks interpretability for high-dimensional data.
Clinical Measures, Radi-omics, and Genomics Offer Synergistic Value in AI-Based Prediction of Overall Survival in Patients with Glioblastoma [24]	Cox-PH with LASSO; C-index, Integrated Brier Score	516 patients; Multi-omics (clinical, radiomics, MGMT methylation, genomic)	Integrated model achieved C-index of 0.75; RB1/NOTCH2 mutations and radiomics enhanced risk stratification
Supervised Graph Clustering for Cancer Sub-typing Based on Survival Analysis and Integration of Multi-Omic Tumor Data [31]	S2GC (graph clustering with survival analysis); Cox log-rank test	TCGA (9 cancer types, incl. GBM); Multi-omics (miRNA, mRNA, methylation)	Outperformed iCluster/SNF; Identified survival-linked GBM sub-types with distinct profiles

Continued

Integrating Different Data Types by Regularized Unsupervised Multiple Kernels Learning with Application to Cancer Subtype Discovery [32]	rMKL-LPP; Log-rank test, k-means clustering	TCGA (GBM, 213 samples); Multi-omics (gene expression, methylation, miRNA)	Identified GBM subtypes with treatment response differences (e.g., Temozolomide); Outperformed SNF.
The Application of Support Vector Machine in Survival Analysis [30]	Modified SVM regression vs. Cox PH; Conservative RMSE	“Alco” (29 samples), “Breast”, “Myel”; Clinical and survival data	Modified SVM rivaled Cox-PH; Kernel choice and penalty parameter(C) impacted accuracy.
Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1 [25]	Hierarchical clustering; SAM, ROC curves	TCGA (200 samples, validated with 260 samples); Gene expression	Identified GBM subtypes linked to survival via PDGFRA, IDH1, EGFR, NF1; Validated with 260 independent samples
Convolutional Neural Networks and Support Vector Machines for Five-Year Survival Analysis of Metastatic Rectal Cancer [29]	CNN + SVM; Accuracy metrics	Rectal cancer tissue samples; RhoB-stained IHC images	CNN-SVM improved accuracy over manual analysis; Small sample size limited generalizability
Multi-omics Cancer Prognosis Analysis Based on Graph Convolution Network [22]	GraphSurv (GCN) vs. EN-Cox, RSF, DeepSurv; C-index	TCGA (11 cancer types, incl. GBM); Multi-omics	GCN improved C-index by 4% over DeepSurv; KEGG pathways enhanced multi-omics predictions
Predicting Clinical Outcomes from Large Scale Cancer Genomic Profiles with Deep Survival Models [17]	SurvivalNet (deep learning), Cox Elastic Net, RSF; C-index	TCGA (LGG/GBM, others); Multiomics	SurvivalNet excelled for GBM; Key features (e.g., IDH1/2 mutations, PTEN deletions) improved prognosis accuracy.
Overall Survival Prediction in Glioblastoma With Radiomic Features Using Machine Learning [26]	Multilayer Perceptron (MLP): 2 hidden layers, sigmoid activation Random Forest (RF)	BraTS 2018: 163 training, 53 validation, 130 test GBM samples	Provides a clinically applicable radiomics-ML framework for GBM OS prediction

enhancing interpretability. Yet, RF struggles with very high-dimensional omics data, underperforming compared to DL models like SurvivalNet in [17]. Additionally, the studies also demonstrate on how integrating diverse datasets such as gene expression, DNA methylation, microRNA, histopathological images, and MRI-derived features enhances predictive accuracy, and improved risk stratification. Furthermore, advanced techniques like Multiple Kernel Learning (MKL) methods, as in [20] and [21], optimize multi-omics and histopathological integration, achieving superior areas under the curve (AUC), but their complexity and reliance on feature selection can reduce generalizability across diverse cohorts.

This analysis underscores the trade-offs between predictive power, interpretability, and computational feasibility, guiding the selection of algorithms to enhance GBM prognosis and risk stratification.

3.1. Limitations

The limitations that are observed across the literature reviews provide insights into improving future studies. One of the limitations relies on the generalization of datasets that is available limiting the robustness and interpretability of the models. Some studies incorporate minimal clinical variables, such as age or resection status, restricting their ability to capture the full prognostic picture and reducing relevance for comprehensive clinical use. A primary challenge is data heterogeneity across datasets like TCGA, CGGA, and BraTS, where differences in sequencing platforms, imaging protocols, and patient demographics introduce variability that hinders model comparability. For instance, studies [13] and [14] report C-index values based on TCGA cohorts of 577 and 619 samples, respectively, but variations in sample size and feature distributions may skew performance metrics. Small dataset sizes, such as the 252 samples in [15], further limit the detection of rare molecular subtypes, reducing generalizability. This narrow scope, compounded by potential biases in data collection protocols, restricts their applicability to the diverse GBM patient landscape. High computational complexity, particularly in graph-based models like S2GC [31] and HI-MKL [20], poses barriers to scalability in resource-limited clinical environments. Furthermore, methodological assumptions, such as reliance on fixed kernel parameters or proportional hazards, often oversimplify GBM's non-linear progression, while inconsistent validation. This ranges from insufficient external testing to overfitting training data undermines model reliability. Thus, these recurring challenges highlight the need for larger, more diverse datasets, richer clinical integration, and streamlined, adaptable approaches to enhance the practical impact of these predictive models in clinical settings.

3.2. Conclusion

In conclusion, the review highlights 20 studies that focus on GBM survival prediction using machine learning, deep learning, radiomics, and clustering techniques across clinical and molecular datasets. The outcomes of the research provide significant insights of the prognostic accuracy, for personalized treatments for patients. The diverse study ranging from SVM and Random Forest outperforming traditional models in omics-driven survival prediction, to multi-task CNNs and graph-based methods like S2GC and rMKL-LPP integrating imaging, genomic, and clinical data. These studies achieve notable prognostic accuracy and reveal molecular signatures tied to patient outcomes. Additionally, the review highlights RF's prevalence and radiomics' synergy with clinical data, while novel methods like HI-MKL and GraphSurv enhance prediction by fusing histopathological and multi-omics inputs. Notably, multi-task CNNs [19] and

graph-based S2GC [31] excel in leveraging multi-modal data for superior prognostic accuracy, while IDH1 mutations and MGMT methylation [13] [24] consistently emerge as pivotal biomarkers for risk stratification. These findings within the research provide a strong foundation for understanding GBM's complexity, emphasizing the need for integrated, scalable solutions to improve personalized prognosis and treatment strategies. In fact, these advancements empower clinicians to tailor therapies, such as TMZ, to patients' molecular profiles, enhancing treatment efficacy. By unraveling GBM's molecular complexity, these findings establish a robust foundation for data-driven tools to optimize patient outcomes in clinical practice.

3.3. Future Works

Despite some limitations that have been discussed, the study provides a solid foundation for future research to overcome GBM's heterogeneity and clinical challenges. For future work the proposed study will address these gaps by integrating diverse data types and feature engineering techniques to enhance predictive performance. Specifically, this approach will use comprehensive multi-omics datasets which include gene expression, DNA methylation, microRNA expression, copy number variations, and mutation to capture the intricate molecular landscape of GBM. Feature engineering will focus on reducing dimensionality to eliminate redundancy, transforming raw omics data into biologically relevant features, and creating composite multi-omics biomarkers to amplify predictive signals. Machine learning models, potentially including kernel-based methods or deep learning architectures, will be trained on these engineered features, targeting improvements over existing metrics like C-index or accuracy.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Grochans, S., Cybulska, A.M., Simińska, D., Korbecki, J., Kojder, K., Chlubek, D., *et al.* (2022) Epidemiology of Glioblastoma Multiforme-Literature Review. *Cancers*, **14**, Article 2412. <https://doi.org/10.3390/cancers14102412>
- [2] Lao, J., Chen, Y., Li, Z., Li, Q., Zhang, J., Liu, J., *et al.* (2017) A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme. *Scientific Reports*, **7**, Article No. 10353. <https://doi.org/10.1038/s41598-017-10649-8>
- [3] Thakkar, J.P., *et al.* (2014) Epidemiologic and Molecular Prognostic Review of Glioblastoma. *Cancer Epidemiology, Biomarkers & Prevention*, **23**, 1985-1996.
- [4] Lee, C.Y. (2017) Strategies of Temozolomide in Future Glioblastoma Treatment. *Oncotargets and Therapy*, **10**, 265-270. <https://doi.org/10.2147/ott.s120662>
- [5] Wang, K., Wang, Y., Fan, X., Wang, J., Li, G., Ma, J., *et al.* (2015) Radiological Features Combined with IDH1 Status for Predicting the Survival Outcome of Glioblastoma Patients. *Neuro-Oncology*, **18**, 589-597. <https://doi.org/10.1093/neuonc/nov239>

- [6] Saadeh, F.S., Mahfouz, R. and Assi, H.I. (2018) EGFR as a Clinical Marker in Glioblastomas and Other Gliomas. *The International Journal of Biological Markers*, **33**, 22-32. <https://doi.org/10.5301/ijbm.5000301>
- [7] Nie, D., Zhang, H., Adeli, E., Liu, L. and Shen, D. (2016) 3D Deep Learning for Multi-Modal Imaging-Guided Survival Time Prediction of Brain Tumor Patients. In: Ourselin, S., Joskowicz, L., Sabuncu, M., Unal, G. and Wells, W., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 212-220. https://doi.org/10.1007/978-3-319-46723-8_25
- [8] Lyu, B. and Haque, A. (2018) Deep Learning Based Tumor Type Classification Using Gene Expression Data. *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, Washington, 29 August 2018-1 September 2018, 89-96. <https://doi.org/10.1145/3233547.3233588>
- [9] Yuan, Y., Qi, P., Xiang, W., Yanhui, L., Yu, L. and Qing, M. (2020) Multi-Omics Analysis Reveals Novel Subtypes and Driver Genes in Glioblastoma. *Frontiers in Genetics*, **11**, Article 565341. <https://doi.org/10.3389/fgene.2020.565341>
- [10] Young, J.D., Cai, C. and Lu, X. (2017) Unsupervised Deep Learning Reveals Prognostically Relevant Subtypes of Glioblastoma. *BMC Bioinformatics*, **18**, Article No. 381. <https://doi.org/10.1186/s12859-017-1798-2>
- [11] Mao, X., Xue, X., Wang, L., Lin, W. and Zhang, X. (2022) Deep Learning Identified Glioblastoma Subtypes Based on Internal Genomic Expression Ranks. *BMC Cancer*, **22**, Article No. 86. <https://doi.org/10.1186/s12885-022-09191-2>
- [12] Tunthanathip, T. and Oearsakul, T. (2021) Machine Learning Approaches for Prognostication of Newly Diagnosed Glioblastoma. *International Journal of Nutrition, Pharmacology, Neurological Diseases*, **11**, 57-63. https://doi.org/10.4103/ijnpnd.ijnpnd_93_20
- [13] Kabelma, D., Benamar, N., Echcharef, C., Hajji, N. and El Hachimy, I. (2023) Survival Rate Prediction in Glioblastoma Patients Using Machine Learning. 2023 *International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, Sakheer, 20-21 November 2023, 221-226. <https://doi.org/10.1109/3ict60104.2023.10391721>
- [14] Johnson, F.T.D., et al. (2023) Machine Learning and Omics Data for Predicting Overall Survival in Glioblastoma Patients. *IEEE Journal of Biomedical and Health Informatics*, **27**, 567-575.
- [15] Wijethilake, N., Meedeniya, D., Chitraranjan, C. and Perera, I. (2020) Survival Prediction and Risk Estimation of Glioma Patients Using mRNA Expressions. 2020 *IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, Cincinnati, 26-28 October 2020, 35-42. <https://doi.org/10.1109/bibe50027.2020.00014>
- [16] Hayes, J., Thygesen, H., Tumilson, C., Droop, A., Boissinot, M., Hughes, T.A., et al. (2014) Prediction of Clinical Outcome in Glioblastoma Using a Biologically Relevant Nine-microRNA Signature. *Molecular Oncology*, **9**, 704-714. <https://doi.org/10.1016/j.molonc.2014.11.004>
- [17] Yousefi, S., Amrollahi, F., Amgad, M., Dong, C., Lewis, J.E., Song, C., et al. (2017) Predicting Clinical Outcomes from Large Scale Cancer Genomic Profiles with Deep Survival Models. *Scientific Reports*, **7**, Article No. 11707. <https://doi.org/10.1038/s41598-017-11817-6>
- [18] Aljouie, A. and Roshan, U. (2019) Multi-Path Convolutional Neural Network for Glioblastoma Survival Group Prediction with Point Mutations and Demographic Features. 2019 *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, San Diego, 18-21 November 2019, 1274-1279.

- <https://doi.org/10.1109/bibm47256.2019.8983109>
- [19] Tang, Z., Xu, Y., Jin, L., et al. (2020) Deep Learning of Imaging Phenotype and Genotype for Predicting Overall Survival Time of Glioblastoma Patients. *IEEE Trans Med Imaging*, **39**, 2100-2109. <https://doi.org/10.1109/TMI.2020.2964310>
- [20] Zhang, Y., Li, A., He, J. and Wang, M. (2020) A Novel MKL Method for GBM Prognosis Prediction by Integrating Histopathological Image and Multi-Omics Data. *IEEE Journal of Biomedical and Health Informatics*, **24**, 171-179. <https://doi.org/10.1109/jbhi.2019.2898471>
- [21] Zhang, Y., Li, A., Peng, C. and Wang, M. (2016) Improve Glioblastoma Multiforme Prognosis Prediction by Using Feature Selection and Multiple Kernel Learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **13**, 825-835. <https://doi.org/10.1109/TCBB.2016.2551745>
- [22] Wang, Y., Zhang, Z., Chai, H. and Yang, Y. (2021) Multi-Omics Cancer Prognosis Analysis Based on Graph Convolution Network. 2021 *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Houston, 9-12 December 2021, 1564-1568. <https://doi.org/10.1109/bibm52615.2021.9669797>
- [23] Zhao, R., Zhuge, Y., Camphausen, K. and Krauze, A.V. (2022) Machine Learning Based Survival Prediction in Glioma Using Large-Scale Registry Data. *Health Informatics Journal*, **28**, 1-17. <https://doi.org/10.1177/14604582221135427>
- [24] Fathi Kazerooni, A., Saxena, S., Toorens, E., Tu, D., Bashyam, V., Akbari, H., et al. (2022) Clinical Measures, Radiomics, and Genomics Offer Synergistic Value in AI-Based Prediction of Overall Survival in Patients with Glioblastoma. *Scientific Reports*, **12**, Article No. 8784. <https://doi.org/10.1038/s41598-022-12699-z>
- [25] Verhaak, R.G.W., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., et al. (2010) Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*. *Cancer Cell*, **17**, 98-110. <https://doi.org/10.1016/j.ccr.2009.12.020>
- [26] Baid, U., Rane, S.U., Talbar, S., Gupta, S., Thakur, M.H., Moiyadi, A., et al. (2020) Overall Survival Prediction in Glioblastoma with Radiomic Features Using Machine Learning. *Frontiers in Computational Neuroscience*, **14**, Article 61. <https://doi.org/10.3389/fncom.2020.00061>
- [27] Poursaeed, R., Mohammadzadeh, M. and Safaei, A.A. (2024) Survival Prediction of Glioblastoma Patients Using Machine Learning and Deep Learning: A Systematic Review. *BMC Cancer*, **24**, Article No. 1581. <https://doi.org/10.1186/s12885-024-13320-4>
- [28] Tizi, W. and Berrado, A. (2023) Machine Learning for Survival Analysis in Cancer Research: A Comparative Study. *Scientific African*, **21**, e01880. <https://doi.org/10.1016/j.sciaf.2023.e01880>
- [29] Suliman, W., Ravi, V., Luo, B., Sun, X. and Pham, T.D. (2022) Convolutional Neural Networks and Support Vector Machines for Five-Year Survival Analysis of Metastatic Rectal Cancer. 2022 *International Joint Conference on Neural Networks (IJCNN)*, Padua, 18-23 July 2022, 1-8. <https://doi.org/10.1109/ijcnn55064.2022.9892935>
- [30] Ding, Z. (2011) The Application of Support Vector Machine in Survival Analysis. 2011 *2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)*, Deng Feng, 8-10 August 2011, 6816-6819. <https://doi.org/10.1109/aimsec.2011.6011384>
- [31] Liu, C., Cao, W., Wu, S., Shen, W., Jiang, D., Yu, Z., et al. (2022) Supervised Graph Clustering for Cancer Subtyping Based on Survival Analysis and Integration of Multi-Omic Tumor Data. *IEEE/ACM Transactions on Computational Biology and Bioin-*

formatics, **19**, 1193-1202. <https://doi.org/10.1109/tcbb.2020.3010509>

- [32] Speicher, N.K. and Pfeifer, N. (2015) Integrating Different Data Types by Regularized Unsupervised Multiple Kernel Learning with Application to Cancer Subtype Discovery. *Bioinformatics*, **31**, i268-i275. <https://doi.org/10.1093/bioinformatics/btv244>