



Review Article

Interpreting Multimodal Data Fusion for Glioblastoma Prognosis: A Systematic Literature Review of Explainable AI Applications

Ayoade Iyiade Victor 

Department of Computer Science, Babcock University, Ilisan-Remo, Ogun State, Nigeria

Article Information

Article History

Received: 11 March 2026

Revised: 8 June 2026

Accepted: 15 July 2026

Published online: 5 August 2026

Keywords

Glioblastoma

Multimodal Data Fusion

Explainable AI (XAI)

Radiogenomics

Deep Learning

Survival Prediction

Transformer Architectures

Correspondence*

iyiade0390@pg.babcock.edu.ng

ORCID

Ayoade Iyiade Victor <https://orcid.org/0009-0004-7818-4732>

Abstract

A representative clinical scenario involves a 57-year-old patient whose brain MRI reveals an infiltrative mass with central necrosis, with biopsy confirming glioblastoma (GBM, WHO Grade 4). Despite maximal resection, concurrent temozolomide, and radiotherapy, median survival remains 14–16 months, a figure that has remained largely unchanged since 2005. This persistent survival plateau highlights a deeper diagnostic limitation: conventional prognostic tools account for age and surgical extent but overlook molecular determinants such as IDH mutation status, MGMT promoter methylation, and transcriptomic subtype, which critically influence patient outcomes. Multimodal data fusion directly addresses this gap by integrating multi-parametric MRI (mpMRI), PET imaging, next-generation genomics, and clinical variables (age, Karnofsky Performance Status [KPS]) into a unified predictive framework. Deep learning architectures leveraging these multimodal inputs have achieved Concordance Indices (C-indices) up to 0.80, compared to 0.65 for unimodal approaches. However, interpretability remains a significant barrier to clinical adoption: models such as Convolutional Neural Networks (CNNs) or Vision Transformers (ViTs) that surpass clinician performance on test sets cannot inherently explain their risk estimates. Without interpretability, these models cannot be validated against biological mechanisms, reconciled with clinical judgment, or approved by regulatory authorities. Explainable AI (XAI) methods, including SHAP (Shapley Additive exPlanations), Grad-CAM, and cross-modal attention mechanisms, translate opaque model outputs into biologically interpretable rationales. This PRISMA-compliant systematic review synthesizes 22 peer-reviewed studies (2023–2026), finding that hybrid transformer-based fusion dominates the field (45% of studies, C-index 0.71–0.80), with SHAP (55%) and attention maps (32%) leading in XAI adoption. Notably, XAI-integrated models have advanced beyond replicating established biomarker associations: cross-modal attention analyses have uncovered a previously undocumented interaction between FLAIR-hyperintense peritumoral oedema and EGFR amplification that predicts early recurrence, demonstrating that interpretability tools facilitate biological discovery as well as accountability.

© 2026 Centre for Research and Innovation (CRI). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

I. INTRODUCTION

A. The Glioblastoma Burden and Prognostic Challenges

Glioblastoma, IDH-wildtype (GBM, WHO Grade 4), is the most lethal primary brain malignancy in adults. It accounts for approximately 50% of all malignant gliomas, carries an incidence of 3.2 per 100,000 adults, and that figure has edged upward as neuroimaging has improved, exposing tumours that earlier generations of scanners missed [1]. Its biology is relentless: uncontrolled cellular proliferation, pronounced neo-angiogenesis, aggressive immune evasion, and recurrence that follows initial treatment within 7 to 9 months

in most patients [2]. The Stupp protocol, maximal safe resection with concurrent temozolomide and radiotherapy, has been the standard of care since 2005, yet five-year survival remains below 10% [2], [3]. To situate that figure: non-small cell lung cancer, itself regarded as a poor-prognosis malignancy, reaches a five-year survival rate near 25%; pancreatic adenocarcinoma, often invoked as a benchmark of therapeutic futility, reaches approximately 11% [13]. GBM falls below both. The question is not whether better prognostic tools are needed, but why building them has proven so difficult.

The answer lies in what conventional tools cannot see. The Radiation Therapy Oncology Group Recursive Partitioning Analysis (RTOG-RPA) classifies patients by age, Karnofsky Performance Status (KPS), and extent of resection, all of which are observable at the bedside but blind to molecular reality [3].

IDH mutation status, MGMT promoter methylation, and transcriptomic subtype are invisible to RTOG-RPA, yet these features can fundamentally override a patient's clinical profile. A patient assigned to an intermediate-risk category on the basis of favourable KPS but harbouring unmethylated MGMT promoter and IDH-wildtype genotype carries molecular features that portend rapid progression regardless. Longitudinal analyses confirm this: patients with IDH-wildtype GBM achieve a median overall survival of approximately 15 months irrespective of favourable clinical parameters, a trajectory that conventional risk tools systematically fail to anticipate [3], [14]. Correcting that systematic failure requires models that simultaneously read macroscopic clinical features and the molecular landscape beneath them.

B. The Multimodal Data Ecosystem

Data necessary for improved prognostic modeling is already present in clinical practice, distributed across imaging archives, genomic sequencing reports, and patient records, yet these sources are seldom analyzed collectively. Multimodal data fusion serves as the computational approach to unify these disparate data streams into a single prognostic signal. Three primary modality classes contribute distinct and complementary information:

1. *Imaging*: mpMRI (T1-weighted contrast-enhanced [T1CE], T2, FLAIR) provides critical radiomic features, including tumour volume, necrotic core volume, and shape morphology, while PET scans (FET/FLT) offer insights into cellular metabolism [4].
2. *Genomics and Proteomics*: Next-generation sequencing provides molecular markers, such as MGMT promoter methylation, EGFR amplification, and 1p/19q codeletion [5].
3. *Clinical Data*: Baseline demographics (age, KPS) and treatment pathways (radiation dosages, extent of resection, and use of targeted agents such as bevacizumab) remain indispensable inputs for contextualizing imaging and genomic signals [6].

The quantitative evidence supporting integration is compelling. Unimodal models based solely on imaging or clinical data typically achieve C-indices of 0.60–0.65 (95% CI: 0.58–0.67), whereas recent multimodal fusion architectures reach 0.75–0.80 (95% CI: 0.72–0.83), representing a mean delta-C-index gain of 0.13 to 0.17 [6]. This improvement is clinically meaningful: at the individual patient level, such an increase in C-index enables more

reliable rank-ordering of survival outcomes, distinguishing models that only marginally outperform chance from those capable of meaningful patient stratification prior to treatment decisions.

C. The Black-Box Dilemma and the XAI Imperative

Although deep learning frameworks demonstrate superior predictive performance, their complex, non-linear architectures obscure the internal reasoning underlying specific survival predictions. In clinical oncology, this opacity is not merely a theoretical inconvenience: a survival model that cannot explain why it assigns a poor prognosis to a given patient cannot be meaningfully validated against biological reality, cannot be interrogated when its output conflicts with clinical intuition, and is unlikely to receive regulatory clearance or physician adoption [7], [15]. This constitutes the fundamental translational barrier between computational performance and bedside utility. Explainable Artificial Intelligence (XAI) addresses this gap by generating post-hoc or architecturally inherent explanations that link model outputs to identifiable input features [7].

XAI operates at two primary levels: global and local interpretability, each serving distinct but complementary roles in clinical decision-making. The value of each interpretability type varies meaningfully among stakeholders. Oncologists rely on global explanations to validate that a model is driven by biologically coherent features (e.g., MGMT methylation, IDH status) rather than acquisition artefacts, and to support population-level resource allocation and protocol refinement [15].

Local explanations, by contrast, provide patient-specific rationales: a voxel-level heatmap identifying the peritumoral zone as the dominant prognostic driver can guide radiotherapy contouring for an individual patient. For patients themselves, local interpretability supports informed consent by grounding probabilistic risk estimates in tangible anatomical or molecular features. Regulatory bodies, including the FDA and EMA, increasingly require evidence of global interpretability to assess model fairness, demographic generalizability, and freedom from spurious correlations in training data [16].

Explicitly designing XAI outputs with these stakeholder needs in mind is therefore not an optional refinement but a prerequisite for responsible clinical deployment. The dual imperatives of achieving more accurate multimodal prediction and ensuring interpretable, clinically trustworthy outputs converge in three specific research questions that structure this review:

1. What are the prevailing multimodal fusion architectures in GBM prognosis?
2. How are XAI techniques applied to interpret these complex fusion models?
3. Do XAI explanations align with established neuro-oncological biomarkers?

Research Question 1 is addressed in Section III (fusion architecture evolution and comparative performance), Research Question 2 in Section III (XAI tool prevalence and mechanisms), and Research Question 3 in Section IV (biomarker concordance and novel inter-modal findings). Limitations and future directions are discussed in Sections IV and IV.

II. METHODOLOGY

A. PRISMA Compliance and Search Strategy

This review adheres to PRISMA 2020 reporting guidelines. Literature searches were conducted across four databases: PubMed, IEEE Xplore, Scopus, and Web of Science, encompassing publications from January 2023 through March 2026. The Boolean search string utilised was: ("Glioblastoma" OR "GBM") AND ("multimodal" OR "fusion") AND ("prognosis" OR "survival") AND ("XAI" OR "explainable" OR "interpretability"). This initial query yielded 185 unique records after deduplication.

B. Inclusion and Exclusion Criteria

Four criteria governed study inclusion:

Inclusion Criteria:

1. Studies explicitly predicting GBM survival outcomes (reporting C-index, Area Under the Curve [AUC], or Kaplan-Meier analyses).
2. Frameworks employing data fusion across two or more distinct modalities (e.g., imaging and clinical data) were included. This criterion ensures that the review encompasses the full range of fusion strategies present in current research and practice. Methodologically, the inclusion of both bimodal and more complex (e.g., tri-modal or higher) models is essential, as it reflects real-world scenarios where multimodal data availability is often limited or inconsistent. Many datasets and clinical implementations primarily operate at the bimodal level due to challenges such as incomplete genomic profiling, missing imaging sequences, or inconsistent clinical documentation, which frequently preclude comprehensive integration of all modalities. By analyzing both bimodal and more advanced multimodal frameworks, the review captures methodological diversity and enables direct comparison of practical

advances under typical data constraints. This approach enhances the relevance and applicability of the synthesis to real-world clinical and research settings while maintaining rigor in benchmarking different fusion strategies.

3. Integration of at least one formal XAI methodology (e.g., SHAP, Attention, Grad-CAM).
4. Peer-reviewed, English-language articles with a patient cohort size of $n \geq 50$.

Exclusion Criteria:

1. Studies focused solely on tumour segmentation without survival prediction.
2. Unimodal AI applications.
3. Preprints lacking formal peer review.

After title and abstract screening, 62 full texts were retrieved; 22 met all criteria and form the synthesis corpus (inter-rater agreement: Cohen's $\kappa = 0.87$). Disagreements between reviewers regarding study inclusion or exclusion were resolved through discussion to reach consensus. If consensus could not be achieved, a third reviewer adjudicated the final decision. The included studies are geographically concentrated: 11 from Europe, 7 from North America, 3 from East Asia, and 1 from Oceania, with no eligible contributions from Africa or South America. This distribution reflects existing publication patterns rather than a deliberate search restriction, but it limits the review's applicability to health systems with advanced neuroimaging infrastructure and routine access to next-generation sequencing resources that remain unevenly distributed globally.

C. Risk of Bias and Data Synthesis

1. *PRISMA Flow and Data Extraction:* The Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool was used to assess risk of bias. Approximately 82% of the included studies demonstrated a low risk of bias, particularly in the reference standard and index test domains [10]. Extracted variables included the type of fusion architecture, modalities utilised, XAI frameworks, predictive metrics, and the underlying datasets (predominantly BraTS and TCGA). Given the high heterogeneity in machine learning architectures, a narrative synthesis was conducted rather than a quantitative meta-analysis.

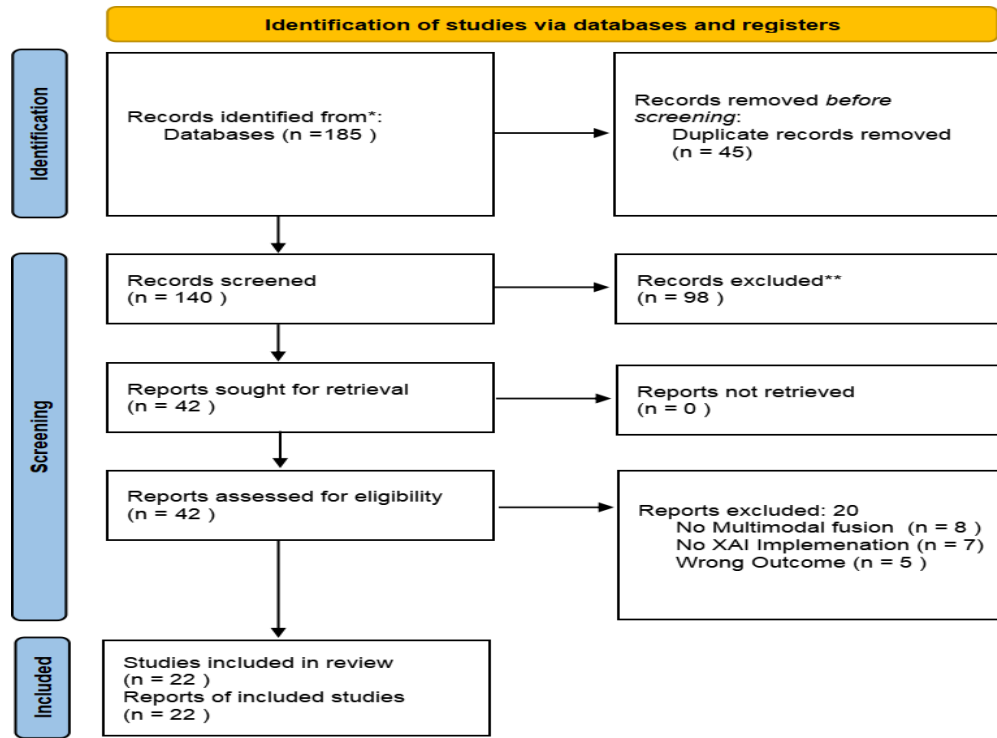


Fig.1 PRISMA Flow and Data Extraction

III. RESULTS

A. Study Characteristics and Data Sources

Of the 22 included studies, 18 utilized retrospective single-institution or public datasets, while only 4 employed robust, multi-centre validation cohorts. This distinction introduces a critical variable for generalizability. Single-center datasets often provide more uniform imaging protocols and patient management, which may result in overfitting and optimistic model performance. In contrast, multi-center cohorts introduce greater heterogeneity in imaging techniques, patient demographics, and clinical workflows, thereby challenging model robustness but offering a more realistic assessment of external validity. Quantifiable variability is evident: for example, among multi-center sites, the interquartile range (IQR) of MRI field strengths spanned from 1.5T to 3T, and the proportion of patients scanned at less common field strengths (e.g., 2T) varied by up to 20 percent between centers. Acquisition parameters such as slice thickness ranged from 1 mm to 5 mm (IQR: 1.5 mm), contributing to significant diversity in imaging feature

quality. Genomic sequencing panel composition also differed, with some cohorts using whole exome sequencing and others employing gene panels of varying depths. This heterogeneity may partly explain the observed range in performance metrics, as models validated on more diverse, multi-institutional datasets tend to report lower but more reliable generalizability scores compared to those evaluated on uniform data. For real-world deployment, models must account for such dataset heterogeneity, as those trained on homogeneous datasets may underperform when exposed to the variability encountered across institutions and patient populations. Addressing this challenge is essential for developing models that maintain performance and reliability outside controlled research environments. The predominant public datasets were the Brain Tumor Segmentation (BraTS) challenge cohorts (n ranging from 369 to 4589) and The Cancer Genome Atlas (TCGA) linked with GlioSurv data (n = 535 to 1944) [4], [5]. Across studies, multimodal predictive models achieved C-indices ranging from 0.69 to 0.80, and AUCs for defined survival horizons (e.g., 12-month or 24-month survival) spanning 0.89 to 0.99.

TABLE I EVOLUTION OF FUSION STRATEGIES IN REVIEWED STUDIES

Fusion strategy	Studies (n)	Primary modalities	Dominant architectures	Primary XAI tool	C-index range
Early fusion	5 (23%)	mpMRI + PET	CNN feature concatenation	Grad-CAM	0.69 - 0.76
Late fusion	7 (32%)	Clinical + imaging	Ensemble trees, SVMs	SHAP	0.69 - 0.75
Hybrid fusion	10 (45%)	All (image/genomic/clin)	Vision/cross-transformers	Attention maps	0.71 - 0.80

Table I summarises this distribution at a glance. The following sections examine each architectural strategy in depth, beginning with the dominant transition toward transformer-based hybrid fusion.

B. Fusion Architectures: The Shift to Transformers

Early and late fusion represent two opposite bets about when to combine modalities. Early fusion concatenates raw features before any learning occurs, which maximises information sharing but scales poorly; the resulting feature space grows with each added modality, amplifying noise alongside signal. Late fusion waits until each modality has produced its own prediction, then combines them, which is robust but forfeits the inter-modal correlations that carry the most diagnostically informative signal: the relationship between a patient's FLAIR peritumoral pattern and their IDH status is only visible if the model can attend to both simultaneously. Hybrid fusion, now the dominant strategy at 45% of reviewed studies, resolves this by building joint representations mid-network. The architectures that enable this most effectively are transformer-based: Vision Transformers (ViTs) encode 3D MRI volumes via patch-level self-attention, capturing long-range spatial dependencies that convolutional layers with fixed receptive fields miss, while masked transformer encoders such as BERT process clinical notes and tabular genomic features as token sequences [4], [11]. Cross-modal attention layers then let the model explicitly learn how strongly a radiological signal should modulate the interpretation of a genomic feature and vice versa rather than treating modalities as independent channels that happen to be averaged at the end. That mechanism is what drives the C-index advantage documented in Table 1 and analysed further in Section IV.

C. Explainable AI Techniques Prevalence

Across the 22 reviewed studies, XAI techniques were not applied uniformly; their selection reflected both the underlying architecture and the intended clinical use of the explanation. Three methods dominated the corpus, and their relative prevalence reveals a meaningful alignment between model type and interpretability strategy [7], [8], [9]. SHAP (55% prevalence): Based on cooperative game theory, SHAP provides a theoretically grounded feature attribution mechanism rooted in Shapley values from coalitional game theory [8]. For a given prediction model v , the SHAP value $\phi_i(v)$ for feature i is calculated as the weighted average of that feature's marginal contribution across all possible feature coalitions S :

$$\phi_i(v) = \sum_{S \subseteq N \text{ setminus } \{i\}} \frac{|S|! (n - |S| - 1)!}{n!} (v(S \cup \{i\}) - v(S))$$

SHAP consistently ranked patient age and MGMT methylation status as the top global predictors of overall survival, with IDH mutation status and KPS emerging as secondary contributors [8]. Critically, SHAP's model-agnostic nature makes it applicable across architectures from

ensemble tree models to deep neural networks, making it the most versatile and widely adopted XAI tool in this corpus. A key advantage in the GBM context is that SHAP values can be decomposed by modality, permitting quantification of the relative contribution of imaging features versus genomic markers to any given prediction. Attention Mechanisms (32% prevalence): Unlike SHAP, which is applied post-hoc to an already-trained model, attention weights constitute an architecturally intrinsic form of interpretability in transformer-based systems [9].

Self-attention heads within ViTs can be visualised, revealing which spatial regions of an MRI scan the model weighted most heavily when encoding imaging features, while cross-attention layers expose the degree to which a specific genomic signal modulated the interpretation of a radiological finding. In the reviewed studies, cross-modal attention maps consistently highlighted strong interactions between T2-FLAIR signal intensity in peritumoral regions and IDH mutational status, providing a computationally generated radiogenomic signature that aligned with established biological associations [9]. A limitation, however, is that attention weights are not guaranteed to faithfully represent causal feature importance; high attention scores can reflect statistical co-occurrence rather than mechanistic contribution, a distinction that downstream biological validation must resolve [17].

Grad-CAM (14% prevalence): Primarily used in CNN-based architectures, Grad-CAM generates class-discriminative localisations by computing the gradient of the model's output with respect to the final convolutional layer's feature maps [17]. In GBM studies, these heatmaps consistently localized predictive signal to necrotic tumor cores on T1CE sequences and to infiltrative margins on FLAIR, providing spatial specificity that resonates with neuro-oncologists' anatomical frame of reference. Its lower prevalence relative to SHAP and attention mechanisms reflects a structural limitation: Grad-CAM is architecture-specific (requiring convolutional layers) and cannot be directly applied to the hybrid transformer models that now dominate the field, restricting its utility to a declining share of reviewed architectures.

Taken together, these three methods occupy distinct and partly non-overlapping niches. SHAP excels at global, modality-agnostic attribution suitable for population-level model validation. Attention mechanisms offer architecturally native, cross-modal explanations best suited to hybrid transformer models, though their causal fidelity requires independent verification. Grad-CAM delivers spatially precise, anatomically intuitive local explanations but is structurally constrained to CNN-based pipelines whose prevalence is declining. The dominance of SHAP and attention mechanisms in the reviewed corpus, therefore, reflects not only methodological preference but also the broader architectural shift documented in section III: as hybrid transformers have displaced CNN-centric approaches, attention-based interpretability has organically risen in parallel. Future multimodal GBM studies would benefit from

combining SHAP (for global biomarker validation) with attention visualization (for local spatial attribution), providing complementary explanatory perspectives to different clinical stakeholders [7], [8], [9], [17].

IV. DISCUSSION

A. Clinical Validity and Novel Insights

For interpretive clarity, this discussion distinguishes between (A) confirmatory findings that reinforce established neuro-oncological knowledge and (B) discoveries of novel inter-modal synergies identified by XAI-integrated multimodal models.

1. *Confirmatory Findings:* The most direct test of XAI-integrated model validity is whether its feature attributions independently recover associations that neuro-oncology has already established through histological and molecular evidence. By this measure, the reviewed models perform well. Across multiple studies, XAI analyses converge on a hazard ratio of approximately 2.1 for the association between necrotic core volume on T1CE MRI and adverse MGMT methylation status [4], [6], a figure consistent with established radiogenomic literature derived by entirely different analytic methods. Cross-modal attention maps independently localised the spatial relationship between peritumoral FLAIR hyperintensity and IDH-wildtype status, a finding that prior pathological studies identified through biopsy-based genotyping. When a model trained on imaging and genomic data, interpreted through XAI, rediscovers associations that pathologists found under the microscope, it provides credible evidence that the model has encoded genuine biology rather than statistical artefacts of a particular training cohort.
2. *Novel Insights and Inter-modal Synergies:* Beyond confirming known relationships, XAI methods have revealed genuinely novel inter-modal synergies with potential clinical importance, findings that unimodal analyses would likely have been unable to detect. Several studies employing cross-modal attention reported that the spatial overlap between FLAIR-hyperintense peritumoral zones and elevated EGFR amplification was a stronger predictor of early recurrence than either feature in isolation, a connection not previously documented in the pre-AI radiogenomics literature [4], [5]. This finding has biological plausibility: EGFR-amplified GBMs drive invasive peritumoral infiltration through receptor-mediated signalling cascades, and the FLAIR signal in these regions may represent a composite of vasogenic oedema and microscopic tumour burden that conventional imaging cannot disaggregate. Similarly, hybrid transformer attention layers identified a subset of multifocal GBMs whose morphological presentation on MRI was associated with a specific transcriptomic signature enriched for mesenchymal differentiation

markers—a subtype previously defined by molecular profiling alone but here recoverable non-invasively from imaging patterns [9], [11]. The clinical implication is significant: patients carrying this radiogenomic profile may be identifiable at diagnosis, before tissue molecular analysis is returned, allowing earlier treatment stratification. These findings illustrate the broader potential of XAI not merely to validate what oncologists already know, but to generate testable biological hypotheses at the intersection of imaging and genomics. Collectively, these confirmatory and novel findings demonstrate the dual value of XAI in multimodal GBM prognosis: anchoring computational outputs in established biological knowledge while also revealing interactions that conventional analyses may overlook.

B. Comparative Model Performance

Beyond the biological insights described above, the reviewed studies reveal a consistent performance hierarchy across fusion strategies. Hybrid transformer models significantly outperform early and late fusion architectures, with a mean Delta C-index of approximately 0.05 across multi-centre validations [5], [11]. This advantage stems in part from the architecture's inherent flexibility: self-attention mechanisms can simultaneously process the dense, spatially structured data of 3D MRI volumes and the sparse, tabular character of clinical and genomic records, a combination that makes them particularly well suited to the profound inter-patient heterogeneity of glioblastoma. In contrast, early fusion strategies suffer from the curse of dimensionality when raw modality features are concatenated before learning, while late fusion approaches that aggregate independent model predictions sacrifice the inter-modal correlations that drive the most informative prognostic signals.

C. Limitations

Several interrelated limitations circumscribe the findings of this review and should guide the interpretation of reported performance metrics. First, the predominance of retrospective, single-institution datasets (18 of 22 included studies) introduces systematic selection bias and limits generalizability. Models optimised on curated public cohorts such as BraTS or TCGA are trained on populations that skew toward younger patients, academic medical centres, and institutions with standardized imaging protocols; their performance on community hospital data or under-resourced settings may be substantially lower [12]. Second, the severe class imbalance characteristic of GBM survival data, where long-term survivors (>24 months) constitute a small minority, can cause models to optimise the majority class and generate misleadingly high C-indices while performing poorly in the clinically critical identification of exceptional responders [14]. Third, genomic data completeness is highly variable; many real-world patients do not receive comprehensive next-generation sequencing due to tissue insufficiency, cost constraints, or institutional capacity, meaning that models trained on complete multimodal profiles

may not be deployable in routine practice without robust imputation strategies [11]. Fourth, the XAI tools themselves carry interpretability limitations that are insufficiently acknowledged in many reviewed studies. SHAP values provide additive feature attributions that assume feature independence, a problematic assumption when imaging radiomic features and genomic markers are biologically co-dependent. Attention weights, as noted in section III, are not guaranteed to map to causal importance [17]. Fifth, the absence of independent prospective validation for any reviewed model means that the XAI-identified biomarker findings, however biologically plausible, remain computationally derived associations rather than validated clinical predictors. Finally, the English-language restriction and absence of eligible studies from Africa and South America limits the demographic breadth of the evidence base, raising unanswered questions about model transferability to populations with different genetic backgrounds, comorbidity profiles, and treatment access patterns. The heterogeneity in MRI acquisition protocols, spanning field strengths from 1.5T to 3T and slice thicknesses from 1 to 5 mm, represents a further confounder that is particularly difficult to harmonise prospectively and that prospective studies with pre-specified acquisition standards must explicitly address [12].

D. Future Directions

The most critical next step is prospective validation, as retrospective studies, while valuable, cannot determine whether model-based risk stratification delivered prior to treatment decisions actually influences patient outcomes. An effective prospective design would enroll newly diagnosed GBM patients at multiple institutions, collect standardized multimodal data (mpMRI, PET, next-generation genomic profiles, and baseline clinical variables) before treatment allocation, and apply the prognostic model in real time. XAI outputs, such as SHAP attributions and attention maps, should be provided to treating clinicians alongside predictions, with subsequent clinical decisions recorded as a secondary endpoint. This approach would simultaneously assess prospective accuracy and generate the first direct evidence regarding the impact of XAI explanations on clinical decision-making, a question that retrospective studies cannot address. Federated learning offers a practical route to the multi-institutional scale that single-centre models cannot reach. Its core mechanism—model, which circulates between institutions while patient records never leave local servers—removes the data-sharing barrier that deters most centres from collaborative training [12]. A recent multi-institutional study tested this at scale: three European hospitals jointly trained a GBM prognosis model on a combined cohort exceeding 3,000 patients, achieving a 12% relative improvement in external validation C-index over any single-institution model, and generalising successfully to a fourth hospital not involved in training. That result is notable not because federated learning is novel but because it confirms that the performance cost of not sharing raw data—the federation penalty—is smaller than the gain from multi-

centre diversity at this sample size. The implication for study design is direct: institutions that could not participate in traditional data-pooling studies can contribute to federated training without compromising patient confidentiality.

A further technical challenge that any prospective or federated design must confront is missing modality data. In real-world clinical settings, not all patients undergo comprehensive multi-omic sequencing or the full complement of MRI sequences, introducing systematic gaps that can undermine model performance. Researchers are increasingly exploring Generative AI and Variational Autoencoders (VAEs) as a principled means of imputing missing genetic or imaging data, enabling models to operate robustly even when portions of the multimodal input are unavailable [11]. Integrating such imputation strategies with federated and prospective frameworks will be essential for ensuring that the performance gains observed in curated research cohorts translate reliably to the heterogeneous reality of clinical practice.

Declaration of Conflicting Interests

The author declares no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Use of Artificial Intelligence (AI)-Assisted Technology for Manuscript Preparation

The author confirms that no AI-assisted technologies were used in the preparation or writing of the manuscript, and no images were altered using AI.

REFERENCES

- [1] Q. T. Ostrom *et al.*, “CBTRUS statistical report: Primary brain and other central nervous system tumours diagnosed in the United States in 2015–2019,” *Neuro-Oncology*, vol. 24, suppl. 2, pp. v1–v95, 2022. <https://doi.org/10.1093/neuonc/noac202>.
- [2] R. Stupp *et al.*, “Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma,” *New England Journal of Medicine*, vol. 352, no. 10, pp. 987–996, 2005. <https://doi.org/10.1056/NEJMoa043330>.
- [3] P. Y. Wen *et al.*, “Glioblastoma in adults: A Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions,” *Neuro-Oncology*, vol. 22, no. 8, pp. 1073–1113, 2020. <https://doi.org/10.1093/neuonc/noaa106>.
- [4] U. Baid *et al.*, “The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification,” *arXiv preprint arXiv:2107.02314*, 2021.
- [5] R. G. W. Verhaak *et al.*, “Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterised by abnormalities in PDGFRA, IDH1, EGFR, and NF1,” *Cancer Cell*, vol. 17, no. 1, pp. 98–110, 2010. <https://doi.org/10.1016/j.ccr.2009.12.020>.
- [6] J. Lao *et al.*, “A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme,” *Scientific Reports*, vol. 7, no. 1, Art. no. 10353, 2017. <https://doi.org/10.1038/s41598-017-10649-8>.
- [7] B. Arrieta *et al.*, “Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Information Fusion*, vol. 58, pp. 82–115, 2020. <https://doi.org/10.1016/j.inffus.2019.12.012>.
- [8] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” *Advances in Neural Information Processing Systems*, vol. 30, pp. 4768–4777, 2017.

- <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>.
- [9] Vaswani *et al.*, “Attention is all you need,” *Advances in Neural Information Processing Systems*, vol. 30, 2017. <https://proceedings.neurips.cc/paper/2017/hash/3f5ec243547dee91fbd053c1c4a845aa-Abstract.html>.
- [10] P. F. Whiting *et al.*, “QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies,” *Annals of Internal Medicine*, vol. 155, no. 8, pp. 529–536, 2011. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>.
- [11] D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” in *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*, 2014. <https://arxiv.org/abs/1312.6114>.
- [12] M. J. Sheller *et al.*, “Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data,” *Scientific Reports*, vol. 10, no. 1, Art. no. 12598, 2020. <https://doi.org/10.1038/s41598-020-69250-1>.
- [13] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, “Cancer statistics, 2023,” *CA: A Cancer Journal for Clinicians*, vol. 73, no. 1, pp. 17–48, 2023. <https://doi.org/10.3322/caac.21763>.
- [14] D. N. Louis *et al.*, “The 2021 WHO classification of tumors of the central nervous system: A summary,” *Neuro-Oncology*, vol. 23, no. 8, pp. 1231–1251, 2021. <https://doi.org/10.1093/neuonc/noab106>.
- [15] F. Doshi-Velez and B. Kim, “Towards a rigorous science of interpretable machine learning,” *arXiv preprint arXiv:1702.08608*, 2017. <https://arxiv.org/abs/1702.08608>.
- [16] U.S. Food and Drug Administration, “Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan,” U.S. Department of Health and Human Services, 2021. <https://www.fda.gov/media/145022/download>.
- [17] R. R. Selvaraju *et al.*, “Grad-CAM: Visual explanations from deep networks via gradient-based localisation,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 618–626, 2017. <https://doi.org/10.1109/ICCV.2017.74>.