Beyond Accuracy: Interpretable Multimodal Deep Learning for Glioma Tumor Grade Classification

Kavita Jain[1,2], Deepali Vora[1 *], Abderrahim Benslimane[3]

1.Symbiosis Institute of Technology Pune, Symbiosis International (Deemed University), Lavale, Pune, 412115, India.
2. Xavier Institute of Engineering, Mumbai, 400016, India
3.University of Avignon, France
kavita.j@xavier.ac.in, deepali.vora@sitpune.edu.in, abderrahim.benslimane@univ-avignon.fr

Abstract: The most common and dangerous brain tumor is glioma and have a unfavourable outcome, significant morbidity, and high mortality. Clinical decision-making depends on accurate and timely glioma grading.While recent DL models have achieved exceptional results in the classification of medical images, their non-interpretable nature hinders the translations into a clinical setting. This paper proposes an interpretable, multimodal Deep Learning(DL) framework that combines MRI features with clinical and molecular features for robust glioma-grade classification. EfficientNetB4 achieves 99.4% accuracy for MRI-based classification, while XGBoost achieves 92% using molecular and clinical biomarkers. Accuracy further improves to 99.8% with a weighted late-fusion strategy and surpasses the single-modality approaches. Grad-CAM visualizations enhance the interpretability by highlighting the tumor-relevant regions in MRI scans. The proposed framework offers improved diagnostic accuracy and fosters clinical trust due to explainability.

Keywords: Glioma Grade classification, Medical imaging, Multimodal Integration, Molecular data, Clinical Data and Explainable AI

1. Introduction

World Health Organization (WHO) classified Gliomas as Grade I to Grade IV .Low grade gliomas (LGG) are classified as grades I and II, whereas high grade gliomas (HGG) are classified as grades III and IV.Low grade gliomas are having negligible growth and high grade gliomas are rapidly progressing[1] .It is costly and laborious to predict the glioma's grade becausemany biomarkers and parameters are involved in the process. Biopsies gives accurate prediction but it is invasive and expensive technique. Deep learning and radiomics are important methods used in analysis of tumor image [2]. Radiomics is especially useful for non-invasive glioma grading, as it analyzes images and extract detailed features such as intensity, texture, and shape, turning standard medical images into rich datasets [3,4].Deep

A. Vaidya et al. (eds.), *Proceedings of the Conference on Bioengineering for Global Health (SYMRESEARCH 2.0 2025)*, Advances in Biological Sciences Research 55,
https://doi.org/10.2991/978-94-6239-756-9_26

learning techniques are used to do the same task in non-invasive way.It is faster and accurate method for classification of glioma grade[5,6,7].The grading process also heavily relies on clinical and molecular/mutation considerations[8,9]. DL models are black-box in nature. They give good accuracy but are not interpretable and trustworthy [10]. Considering all these challenges in prediction of grade of Glioma,we have proposed theExplainable multimodal deep learning framework for glioma grade classification using MRI , molecular and clinical features.Many researchers have used MRI as a single modality to predict the grade of tumor.In the proposed framework we have used MRI and molecular features as modalities to predict the grade of the tumor. In this study, EfficientNetB4 achieves higher accuracy of 99.8% for MRI based classifiers.SMOTE is used as preprocessing technique to handle class imbalance problem andXGBoost achieved higher accuracy that is 92%for molecular feature dataset. Decision fusions of prediction of these two classifiers are done and achieved accuracy of 99.9%. MRI and molecular datasets originate from different hospitals, scanners, and patient demographics, leading to cohort variability. The proposed multimodal fusion framework is designed not only to combine imaging and molecular modalities but also to learn shared representations that unify heterogeneous patient cohorts into a single learning space.

2. Related work

A new hybrid CNN-SVM model is proposed in the study [11] to improve the accuracy of glioma grade identification. The whole procedure consists of two steps: grade identification by SVM and glioma classification by CNN. An SVM classifier is trained after the glioma region has been segmented to extract important shape and texture information using the modified firefly optimizer. For the proposed hybrid model, an accuracy of 99.98% was obtained. A deep ensemble U-Net framework for concurrent glioma segmentation and grade prediction using multimodal MRI is proposed in paper [12]. For improved feature extraction and fusion, it integrates asymmetric convolution, dual-domain attention, and a dual-branch decoder. A weighted adaptive loss balances both tasks and achieved superior performance on the BraTS dataset. Research in [13] proposes a hybrid transfer learning system to perform MRI-based classification of brain tumors. VGG16 combined with ResNet152V2 and XAI techniques (Grad-CAM, SHAP) results in 99.47% accuracy with improved interpretability and clinical reliability, hence demonstrating strong prospectives for use in resource-constrained settings. Research [14] presents a Deep Learning for BT grade categorization using MRI images with the aim of improving diagnostic precision. Here, MobileNetV2 is adopted because of its proficient and generalizable feature extraction. The model was trained and tested on six standard Kaggle brain tumor MRI datasets, thus effectively detecting and classifying tumors. Multiregional, multi-sequence MRI radiomics model was developed by Li *et al.*to preoperatively predict MGMT promoter methylation in

glioblastoma, showing how combined imaging features can capture biologically relevant molecular status without invasive procedures [15]. Other studies have focused on integrated radiomic analysis of peritumoural regions for glioma recurrence prediction using multimodal MRI, showing improved localization of recurrence risk zones but still relying on traditional radiomics rather than end-to-end fusion learning [16]. Zhu et al. demonstrated how multimodal imaging can support precision treatment planning without specifically including genomic data by proposing a multimodal MRI-based spatial mapping framework to assess intratumoral heterogeneity characteristics in patients with recurrent glioblastoma[17]. In parallel, comprehensive radiogenomic profiling of global DNA methylation has elucidated associations between MRI texture features and molecular/immune profiles, offering valuable biological insight but limited predictive modeling [18]. Large cohort studies integrating radio-pathology with proteogenomics have identified glioma subtypes with prognostic relevance, yet these approaches do not address cross-cohort harmonization or interpretable deep learning fusion [19]. Moreover, machine learning models that predict oncogenic signaling pathways from radiogenomic features highlight the potential of integrative approaches in pathway discovery but stop short of rigorous multimodal fusion strategies with statistical validation [20]. Although these studies together show promise in leveraging multimodal data for molecular inference and glioma phenotyping, most of these studies do not have an explicit mechanism for unifying diverse population groups, robustness testing, and deep fusion architectures. This study directly addresses these issues with an interpretable multimodal deep learning framework.

3. Proposed methodology

Figure 1 shows Proposed Framework for Multimodal Glioma Grade Classification. Traditional grading methods suffer from subjectivity and a lack of reproducibility. These limitations are overcome by the proposed framework through the help of multimodal integration and explainability. MRI scans and molecular biomarkers are processed separately using specialized models. Their predictions are combined using a weighted late-fusion strategy. Grad-CAM heatmaps highlight MRI regions that most strongly drive classification decisions and, therefore, improve transparency.Traditional grading methods are often subjective and lack reproducibility. The proposed framework addresses these limitations through multimodal integration and explainability. MRI scans and molecular biomarkers are processed independently through specialized models. Their predictions are fused using a weighted late-fusion strategy. Grad-CAM heatmaps are used to highlight MRI regions that most strongly influence classification decisions, improving transparency.

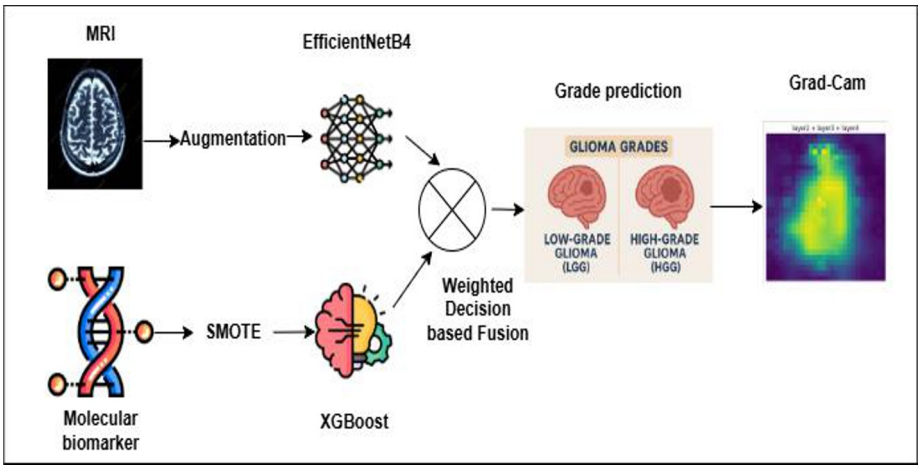

Fig. 1: Proposed Framework for Multimodal Glioma Grade classification

4. Methods and Material

4.1 Dataset

It mainly consisted of two datasets for multimodal glioma grading in this work. Imaging-based analysis drew from Dataset 1, the BRATS 2019 MRI dataset [21], consisting of 259MRI sequences for HGG and 79 for LGG in NIfTI files (.nii.gz) and consist of MRI sequences such as T1 and post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) volumes. In this research T2-Flair MRI images are used. The images are converted to jpeg and augmentations techniques such as crop, rotate, zoom, flip, and shear are usedto increase the dataset size. After augmentation and oversampling ,3,600 MRI pictures are used for the research , in which 1,800 represented LGG and HGG classes. Images are resized to 224 ×224 pixels.

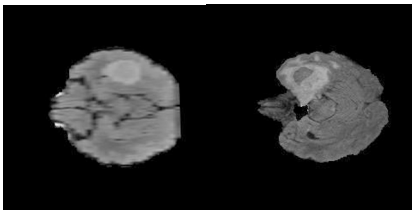

Fig. 2.T2-Flair MRI image for a)LGG and b)HGG

For molecular feature analysis, Dataset 2, the Glioma Grading Clinical and Mutation Features dataset [22], consisted of 839 patient records.The samples in this collection are the medical records of patients with gliomas. Each record is characterized by three clinical characteristics related to the patient's demographics and twenty molecular factors, each of which can be mutant or not mutant .The patient's age, gender, tumor grade, and the status of significant gene alterations such as IDH1, TP53, and ATRX are included in each record [23,24]. A strong multimodal framework that integrated imaging and molecular insights to improve the accuracy and interpretability of glioma grade diagnosis was made possible by the integration of these datasets. To even out the datasets' class sizes, SMOTE with oversampling is utilized.

4.2 Data Privacy and Protection

The research makes use of publicly available and anonymous datasets, where all the patient identifiers have been removed. There has been no access to personal or sensitive patient data. The experiments follow guidelines on data protection and ethical research use of medical data. The datasets (for example, TCGA, BraTS, Glioma Grading Clinical and Mutation Features Dataset) are publicly available for research use after ethical approval and anonymization.

4.3 Pre-processing Techniques

Images in the dataset are converted to jpeg Several pre-processing operations are applied to MRI images to standardize and enhance them before model training.

Data Augmentation Numerous augmentation methods were used, such as rotation, flipping both horizontally and vertically, zooming, translation, adjusting brightness and contrast, shearing, adding noise, and rescaling intensity. These changes made the model stronger and better at handling different types of images by adding variations in position and lighting.The data was then divided into training, validation, and testing sets while keeping the number of samples in each class balanced.[25, 26,27]

SMOTE (Synthetic Minority Oversampling Technique) It is an effective data-level technique for addressing imbalances in classes in machine learning, particularly when there are noticeably fewer samples in one class than in another. In the context of biomarker or genomic data, Biased models that perform well on the majority class but are unable to correctly detect minority cases can result from class imbalance. To address this problem, SMOTE generates new, synthetic samples for the minority class rather than just replicating existing data. SMOTE finds the k nearest neighbors (usually k=5) of each sample in the minority class in the feature space. After choosing one or more of these neighbors, it interpolates between the original sample and its neighbors to produce new synthetic

samples. In order to perform this interpolation, the difference between the feature vectors of the chosen neighbor and the original instance is calculated, multiplied by a random value between 0and 1, and then added back to the original vector. In multidimensional feature space, the outcome is a new, realistic data point that is located along the line joining the two samples. By generating synthetic biomarker profiles, SMOTE ensures that data has a more even representation of classes. This balancing improves the classifier's ability to learn meaningful decision boundaries for both classes, reducing bias toward the majority class. [28,29]

4.4. Multimodal Fusion for Cross-Cohort Data Integration

Instead of considering multimodal fusion as a modality fusion approach, we use it as a representation learning approach to synchronize data across different population groups and cohorts.The proposed multimodal fusion approach is robust across different cohorts by combining MRI data from BraTS2019 and molecular and clinical data from the Glioma Grading Clinical and Mutation Features dataset. MRI data preprocessing and feature extraction help to remove variations due to different scanners and protocols, while molecular and mutation data provide biologically stable information that is not affected by imaging conditions. The model is able to adapt to changes in hospitals, demographics, and data distribution since the multimodal features are synchronized in a shared latent space. The MRI features may be biased due to the location and imaging conditions, but molecular features provide a stable biological foundation that is common to all cohorts.The model is able to learn invariant representations across populations while still retaining significant heterogeneity because of this integration across cohorts, which helps to improve the generalization capability of the model towards glioma grading. Thus, instead of overfitting, the high accuracy achieved is a result of the effective integration of the complementary data sources from imaging and molecular characteristics.

5. Multimodal Fusion and Weighted Ensemble Strategy

To improve prediction accuracy and model interpretability, multimodal fusion techniques combine data from several data sources, including MRI images, molecular characteristics, and clinical data. In brain tumor analysis, fusion techniques aim to combine complementary information from different modalities to capture both structural and molecular characteristics of tumors.

The three primary fusion strategies are:

5.1. Early Fusion (Feature Level Fusion):

Features from different modalities are combined at the input or initial feature extraction stage. For example, MRI features and molecular features are concatenated before being fed into a neural network. This approach captures intermodal relationships early but may face challenges due to differences in data scale and representation [36].

5.2. Late Fusion (Decision Level Fusion):

Different models are used to process each modality separately.(for example, CNN for MRI and XGBoost for molecular data), and their outputs or predictions are later combined using techniques like weighted averaging, majority voting, or meta-classifiers. This method is computationally efficient and robust to missing modalities[36].

Concatenation based Fusion Strategy: In this approach, the feature representations extracted independently from the MRI modality (FMRI) and the molecular/biomarker modality (FMol) are simply joined together to form a single unified feature vector. This process is called feature concatenation, where both modalities contribute their learned information without altering or weighting each other.

Attention-Gated Fusion: In this approach, the model learns to dynamically assign importance to each modality rather than treating them equally. Instead of directly concatenating the features, learnable attention weights are applied to the MRI and molecular features.

5.3. Hybrid Fusion (Intermediate Fusion):

Combines the advantages of both early and late fusion by merging feature representations at intermediate network layers. This allows partial sharing of learned representations while preserving modality specific information. Deep learning architectures such as hybrid CNN Transformer or attention based fusion networks often employ this approach[30].

5.4. A weighted ensemble strategy

The proposed model uses a weighted late fusion ensemble approach to combine the MRI imaging features and the molecular biomarker features, considering the differences in the number of dimensions, information, and relevance. The features are learned using separate encoders to obtain high-level abstract representations, which are then fused using a fusion layer that serves as an ensemble gate. Unlike the simple concatenation approach, which may result in the overshadowing of one modality by the other due to differences in scales, the weights are learned using backpropagation, and the network learns to automatically weigh the contributions of the different modalities according to their predictive value. Figure 3 shows the weighted ensemble strategy for multimodal fusion network for modality integration and cross-cohort feature harmonization in glioma grade classification

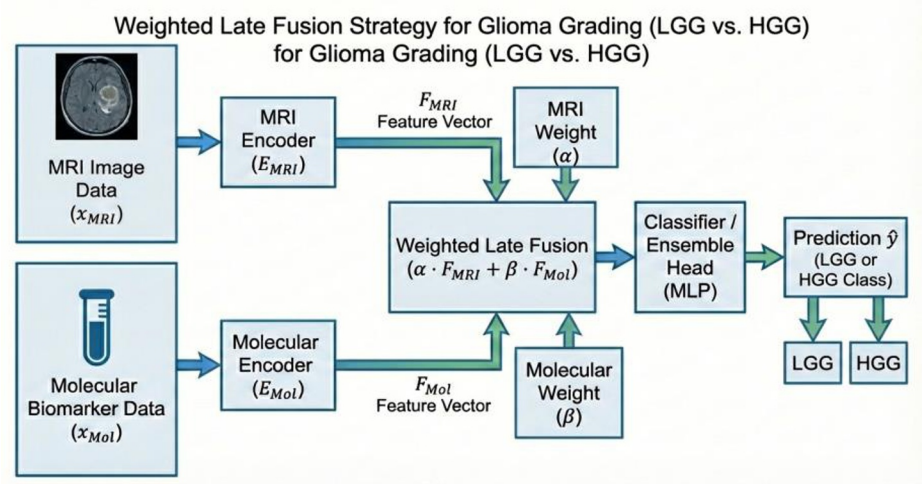

Fig. 3. Multimodal fusion network for modality integration and cross-cohort feature harmonization in glioma grade classification

6. Classification

6.1. Models based on MRI modality

To classify the grade of the glioma tumor, MRI and molecular biomarkers are used from the BRATS2019 dataset and the Glioma Grading Clinical and Mutation Features dataset,respectively. In order to facilitate the categorization of gliomas into either glioblastoma multiforme (GBM) or lower-grade glioma (LGG), this dataset is an open-source resource that includes clinical and molecular features from brain glioma patients. The Adam optimizer with a learning rate schedule and the categorical cross-entropy loss function were used to train the model. To avoid overfitting, early pausing based on validation loss was used. ROC-AUC, F1-score, accuracy, precision, and recall are the evaluation measures.VGG-19, XceptionNet, ResNet50, EfficientNet (B0–B4), MobileNetV1, and MobileNetV2 are among the pre-trained CNN models used [30, 31].**ResNet**:ResNet uses residual connections. It captures fine textures and tumor patterns effectively and is compatible with explainability tools like saliency maps. Though computationally intensive, ResNetprovides stable and high-accuracy performance in tumor classification[31].

EfficientNet (B0 & B4): EfficientNet uses compound scaling across depth, width, and resolution. Its MBConvblocks with squeeze-and-excitation layers enhance spatial feature attention. Although it lacks ResNet-style skip connections, it is highly efficient in computation, making it suitable for accuracy-critical yet resource-limited scenarios[32,33].
MobileNetV2: MobileNet is designed for embedded and mobile devices using depthwise separable convolutions.While computationally lightweight, its shallower depth mayhinder fine-grained feature learning. It is ideal for real-time or portable diagnostic systems in low-resource settings but may need enhancements for complex tumor recognition. The above mentioned pre-trained models are used for classifying type of the tumor.If tumor type is Glioma then it uses MRI and Molecular biomarker to classify the grade of the Glioma tumor[30,31,32,33].
Ensemble model: It improves the overall performance by avoiding the individual disadvantages of a deep learning model. It improves the performance in terms of accuracy, robustness, and generalization by fusing predictions from various architectures for brain tumor grade classification. This will provide more reliable and consistent diagnosis when handling complex and variable MRI data.[34]
VGG-16 and VGG-19: The deep convolutional networks VGG-16 and VGG-19 are renowned for their depth and ease of use.They consist of 16 and 19 layers, respectively, and effectively capture intricate features from the input images[30,31]
XceptionNet: XceptionNet is highly effective for local feature extraction because it relies on depthwise separable convolutions, which break down the learning process into channel-wise and spatial filtering. The network is able to concentrate more accurately on fine-grained local structures, such textures, edges, and minute changes in intensity in MRI scans. Such detail sensitivity is particularly important in brain tumor analysis, where minor textural differences or irregularities can indicate abnormal tissue growth.

6.2. Classification models based on molecular biomarker modality

Several artificial intelligence methods, including SVM, logistic regression, random forest, XGboost, LBGM, and CatBoost, are used to categorize the grade of glioma from clinical and molecular biomarkers accessible in CSV files.[35] areused to train the model and compared the accuracy and training time required to train these models.For feature selection 10-fold cross-validation is used.

7. Explainable AI (XAI)

In medical applications, particularly in the diagnosis and categorization of brain tumor , XAI is essential for enhancing algorithm transparency and interpretability. In order to help healthcare personnel comprehend decision-making processes and generate trust in automated systems, recent efforts have concentrated on creating AI models with explainable outputs.Taxonomies in literature categorize various explainability techniques, from model-

specific to model-agnostic methods, textual and visual explanations to improve comprehension. In contrast, ante-hoc and post-hoc explanation methods provide valuable insights into model accuracy and interpretability. Both local and global interpretable techniques provide a deep understanding of specific predictions as well as the general behavior, training procedure, and data utilization of the model [37,38,39]. To enhance interpretability, a visualization-based Explainable AI technique is utilized. Grad-CAM is utilized to generate visual heatmaps highlighting the noteworthy image regions that influenced the model predictions [40,41]. These are critical towards increasing transparency and clinical trust in the AI system. Heatmaps or attention maps are generated to point out regions or features that contribute to the AI making a particular decision, thereby increasing explainability for the clinicians.

8. Result and Discussion

8.1. Dataset split and Experimental Setup

The BraTS19 dataset was divided into training, validation, and testing subsets in order to train the model. The algorithm was able to learn distinguishing features from the MRI pictures and molecular information by using about 80% of the dataset for training. About 10% of the dataset was held for validation, which was used for hyperparameter optimization, model selection, and stopping the model to avoid overfitting. The remaining 10% of the dataset was held for testing, which acted as a held-out set for final model performance evaluation. The dataset was split stratified based on tumor grade (HGG vs. LGG) to preserve the original class distribution , ensuring that both HGG and LGG glioma patients were adequately represented. The split was done at the patient level, which ensured that MRI images from the same patient were only included in one subset to prevent data leakage and ensure independence between the training, validation, and testing sets.

Glioma Grading Clinical and Mutation Features Dataset: The dataset was divided into training, validation, and testing sets in order to create the model. The program was able to learn different patterns from clinical and mutation data because roughly 70% of the data was used for training. 15% of the data was used for validation in order to adjust the hyperparameters and quit early. The remaining 15% of the data served as a held-out test set for an objective assessment of the model's ultimate performance. To preserve the original class distribution and avoid data leaking, all of the aforementioned splits were performed at the patient level and stratified by glioma grade.For handling the problem of class imbalance in the clinical and mutation dataset, SMOTE was used only on the training set after splitting the data. The minority class LGG was oversampled to be equal in number to the GBM class, which increased the number of training samples from 587 to 740 while keeping the validation and test sets unchanged for unbiased evaluation.

Overfitting Prevention: To avoid overfitting while training the model, we used a combination of techniques such as regularization, dropout, and early stopping. Regularization helps in reducing the weights of the model to improve generalization, dropout helps in randomly turning off neurons during training to avoid dependence on certain features, and early stopping helps in monitoring the validation performance of the model to stop training when the performance stops improving.

Cross-Validation: To make the results more robust, 5-fold cross-validation was done for all models. The model is trained on four folds and tested on the remaining fold in each cross-validation cycle.The performance of the model on all folds, along with the standard deviation, is calculated. Cross-validation helps in making the results less prone to random variations in the data.
Table 1 shows training parameters used to optimize deep learning models. The dataset was divided at the patient level into training, validation, and testing sets to prevent data leakage. Images from the same patient were not shared across splits.

Table 1. Experimental setup for hyper-parameters

Hyper parameter	**Value**
No. Of. Epochs	25
Learning rate	0.0001
Optimizer	adam
Batch size	32
Loss Function	CrossEntropyLoss

8.2. MRI based glioma grade classification

This section highlights several models' performances in classifying glioma grades. Accuracy, precision, recall, and F1-scores were used as metrics for all the models in this task. A number of DL models have been evaluated for glioma grade classification, demonstrating varied levels of precision and effectiveness. The VGG19 model performs with 74.85% accuracy, which is moderately performing, considering it was made up of a relatively shallow architecture. ResNet101 performed better, with an accuracy of 82.34%, because it contained a deeper architecture that was connected by residual connections, which helped in enhancing the extraction of relevant features. MobileNetV2 had an accuracy of 95.23%, hence proving to be quite efficient and suitable for glioma grading tasks, particularly resource-constrained environments. Xception was able to achieve 99% accuracy. This provided the best performance among all models tested, as it could capture complex features pertaining to space and texture more effectively. EfficientNetB4 presented an even higher accuracy of 99.4%, hence proving to provide a balance between computational cost and accuracy. EfficientNetB0 reached an accuracy of 96.26%, hence

performing well owing to an optimum scaling of network depth and width.Also, an ensemble model in which many architectures were combined achieved a final accuracy of 97.71%, using the complementary characteristics of each model to get more robust classification results. Figure 4 and Table 2 present the comparative accuracy of several models on tumor type categorization.

Table 2. Performance metrics for Models

Model	Accuracy (%)	Precision (%)	Recall (Sensitivity) (%)	F1-Score (%)	Specificity (%)
VGG19	74	73.5	72.8	73.1	75
ResNet101	82.3	83	82	82.5	83.4
MobileNetV2	95.23	95.6	95	95.3	95.4
Xception	99	99.1	98.9	99	99.1
Ensemble	97.71	97.8	97.6	97.7	97.8
EfficientNetB0	96.26	96.3	96.2	96.2	96.3
EfficientNetB4	**99.4**	**99.5**	**99.3**	**99.4**	**99.4**

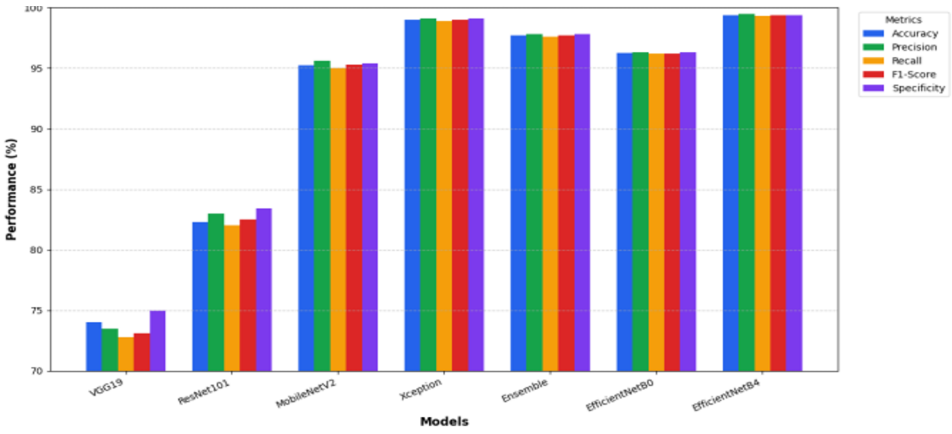

Fig. 4. Performance comparison of Pre-trained models using MRI

Among the models evaluated, VGG19 had a moderate performance due to its limited feature generalization capability as compared to other deep models. In ResNet101, it improves stability and performance due to deeper residual connections, which help reduce vanishing gradient issues. MobileNetV2 may serve for lightweight deployment since it reaches a relatively good balance between classification accuracy and processing efficiency. On the contrary, both Xception and EfficientNetB4 outperform all other models by attaining near-perfect classification accuracy, taking advantage of optimized depthwise separable convolutions and compound scaling techniques. Considering Ensemble, robust and generalized predictions were provided, slightly lower than EfficientNetB4, though showing very good overall reliability and adaptability across diverse tumor grade classifications. It follows that advanced deep learning models perform best for MRI-based classification, with EfficientNetB4 obtaining the highest accuracy of 99.4%, followed by 99% from Xception. Since 5-fold cross-validation was employed, predictions from all test folds were aggregated, and the confusion matrix was constructed using the entire dataset. Each patient was used as a test sample exactly once, ensuring unbiased evaluation. Figure 5 shows the confusion matrix of EfficientNetB4

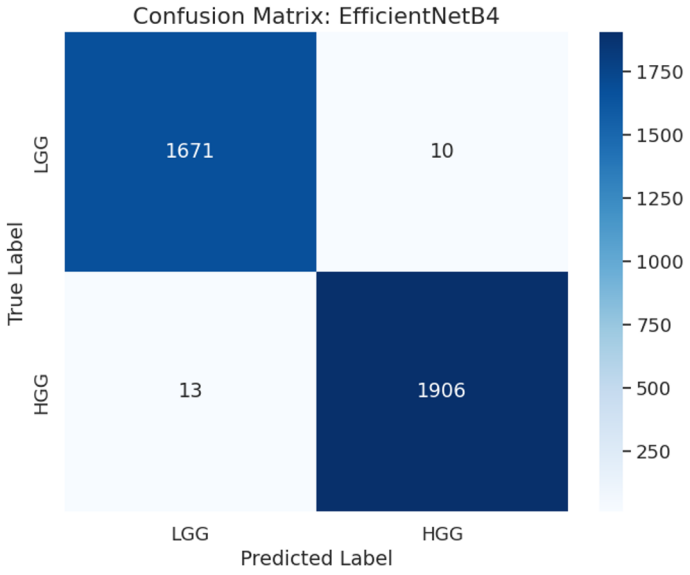

Fig. 5. Confusion Matrix of EfficientNetB4

8.3. Molecular biomarker based Grade classification

To predict glioma grade using genomic data, several machine learning models were trained on the "Glioma Grading Clinical and Mutation Features" dataset.The problem of LGG and HGG imbalances in class was addressed using the SMOTE approach. With an accuracy of 86% and an F1-score of 0.86, the Support Vector Machine demonstrated a strong ability to identify high-grade gliomas among the evaluated models.Logistic Regression reached 85% accuracy, giving balanced performance on both grades of glioma. The Random Forest model attained 84.64% accuracy, with the results relatively consistent for both classes. XGBoost, with 92.98% accuracy and a strong F1-score, outperformed all models that have been analyzed. LightGBM achieved an accuracy of 85%, showing balanced but comparatively lower performance than XGBoost. Meanwhile, CatBoost recorded an accuracy of 86.30%, with a bit better precision on the low-grade gliomas. XGBoost is the most effective and precise model for the problem of glioma grade prediction using molecular data; it outperforms SVM, Random Forest, and Logistic Regression in terms of precision and generalisation capacity.

Figure 6 shows the confusion matrix of Xgboost. Table 3 and Figure 7 compares the performance of machine learning models that incorporate molecular features.

Table 3. Performance metrics for Models based on Molecular biomarker

Model	Accuracy	Precision (Class 0)	Precision (Class 1)	Recall (Class 0)	Recall (Class 1)	F1-score (Class 0)	F1-score (Class 1)
SVM	0.85	0.93	0.8	0.79	0.94	0.85	0.86
Random Forest	0.84	0.86	0.83	0.84	0.85	0.85	0.84
XGBoost	**0.93**	**0.97**	**0.9**	**0.89**	**0.97**	**0.93**	**0.93**
Logistic Regression	0.87	0.97	0.8	0.76	0.98	0.86	0.88

Although SMOTE was applied to balance classes during training, it was used exclusively within training folds. The confusion matrix was constructed by aggregating predictions from all test folds, resulting in evaluation over the original 839 real patient samples without any synthetic data

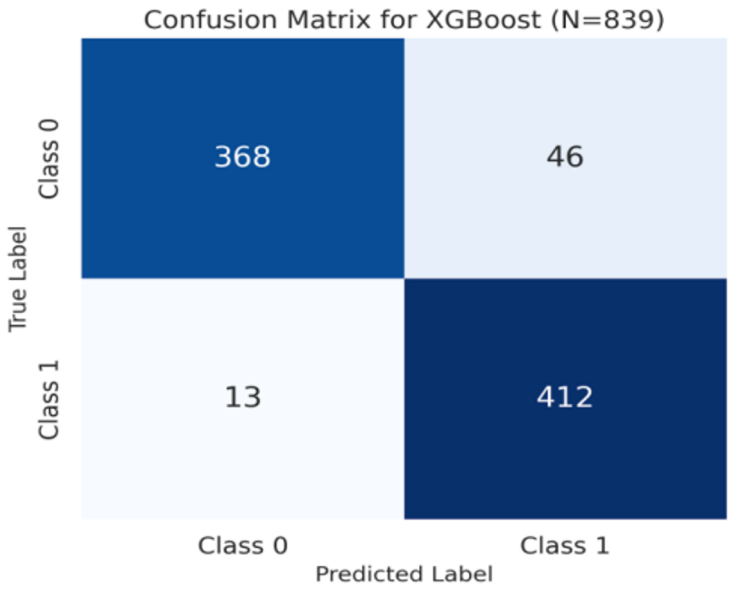

Fig. 6. Confusion Matrix of XGBoost

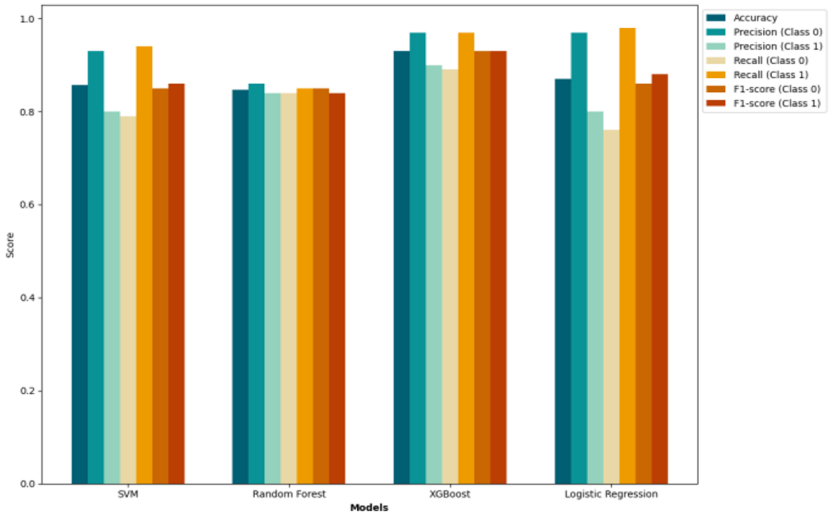

Fig. 7. Comparing Machine Learning Model Performance with Molecular Features

Overall, EfficientNetB4 is the most effective for MRI data, while XGBoost works best for clinical/mutation data.

8.4. Multimodal Fusion with Weighted Ensemble Strategy

Table 4. Performance metrics for Models (single modal model + multi-modal model)

Model	Accuracy	Precision (Class 0)	Precision (Class 1)	Recall (Class 0)	Recall (Class 1)	F1-score (Class 0)	F1-score (Class 1)
EfficientNet (MRI only)	0.994	0.98	1	1	0.99	0.99	0.995
XGBoost (Molecular only)	0.92	0.97	0.9	0.89	0.97	0.93	0.93
Multimodal Fusion	**0.998**	**0.99**	**1**	**1**	**0.99**	**0.995**	**0.995**

Figure 8 shows the confusion matrix of multimodal Fusion model . Table 4 and Figure 9 compares the classification accuracy of EfficientNet (MRI-based), XGBoost (biomarker-based), and a Fusion model (EfficientNet + XGBoost) over ten epochs. EfficientNet shows a steady rise from about 70% to 99.4%, reflecting its strong learning capability for imaging data. XGBoost maintains an even accuracy of approximately 92%, showing the stability of performance on structured genomic features. In the proposed methodology, Weighted Decision fusion is implemented. The predictions are fused from both models to enhance the accuracy of glioma grade prediction. Weighted decision fusion model considers more weight of EfficientNetB4 than XGboost because of its high accuracy compared to the XGboost classifier. Fusion model outperforms both of them; it starts near 80% and reaches approximately 99.8% by epoch 10. This points to the conclusion that the integration of MRI and biomarker features improves model performance significantly, leveraging complementary spatial and molecular information toward more accurate and robust glioma grade prediction. The late fusion of MRI and molecular information encourages the network to learn cohort-invariant and modality-invariant representations, which reduces bias introduced by the heterogeneous data sources.After matching the BraTS19 MRI patients with the TCGA clinical and mutation data and removing the incomplete cases, 186 overlapping patients were found. The multimodal fusion and evaluation were conducted only on this overlapping set.

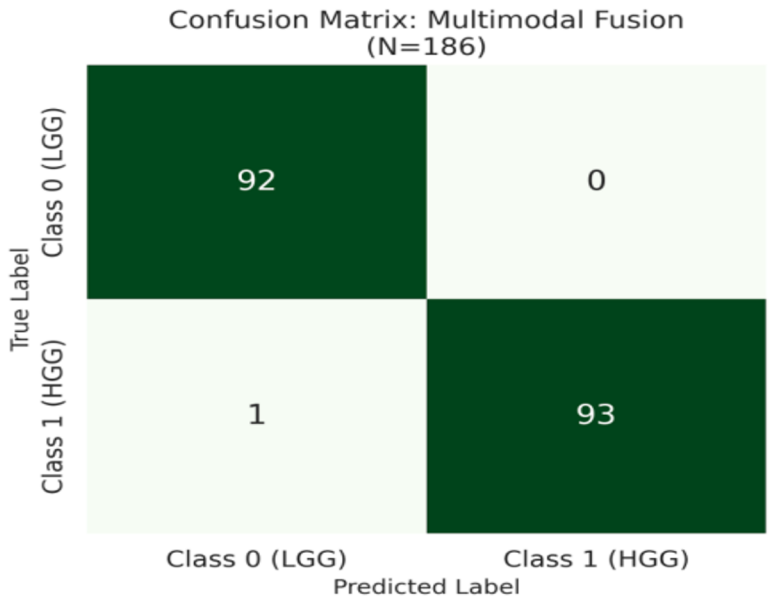

Fig. 8. Confusion Matrix of Multimodal fusion

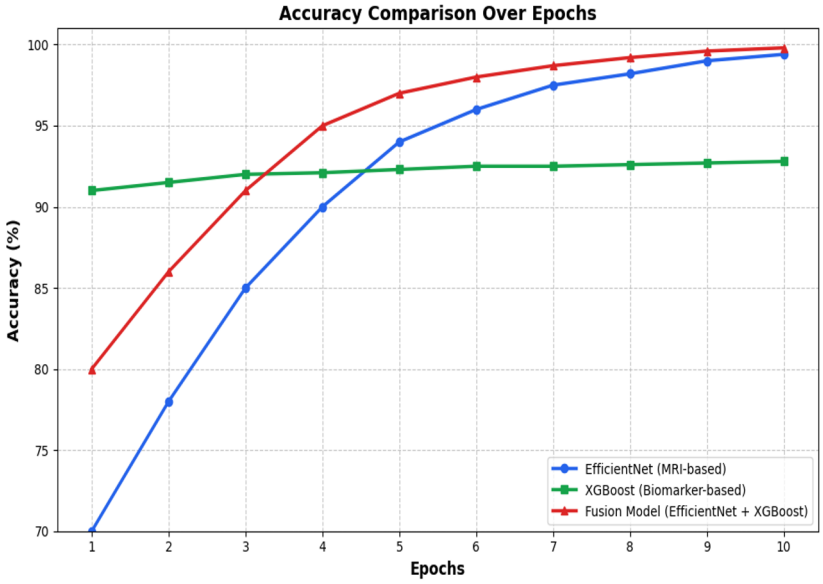

Fig. 9. Accuracy Comparisons of Single modal and MultimodalGrade Classification

8.5 Statistical Significance Analysis

Statistical significance testing was performed to ensure that the performance difference between single-modal models and the proposed multimodal fusion model is not random. The accuracy values derived from 5-fold cross-validation for MRI-only, molecular-only, and fusion models were used in a paired t-test. Table 5 shows the statistical significance analysis of single modality and multi-modality models.

All models were evaluated using 5-fold cross-validation. For each fold, accuracy was computed on the held-out test fold. Paired t-tests were conducted on the fold-wise accuracies to compare models. The multimodal fused model (mean accuracy = 99.8% ± 0.08%) significantly outperformed the molecular-only XGBoost model (mean = 92% ± 0.5%, $t = 7.2$, $p < 0.001$) and showed a small, non-significant improvement over the imaging-only EfficientNet model (mean = 99.4% ± 0.12%, $t = 1.6$, $p = 0.17$), demonstrating the benefit of integrating imaging and molecular modalities."

Table 5. Statistical analysis of single modality and multi-modality model

Comparison	Mean Difference ($\bar{d}$)	t-value	p-value	Interpretation
Fused vs EfficientNet	0.004	1.6	0.17	Not Significant
Fused vs XGBoost	0.078	7.2	< 0.001	Highly Significant

8.6 Explainable AI

XAI method is employed to enhance interpretability [42, 43]. Specifically, graphical heat-maps showing the key MRI regions that influence the model's prediction were created using Grad-CAM. The techniques are extremely important in healthcare settings to improve openness and confidence in the AI system.Figure 10 represents the Grad-cam heatmap for EfficientNetB4.

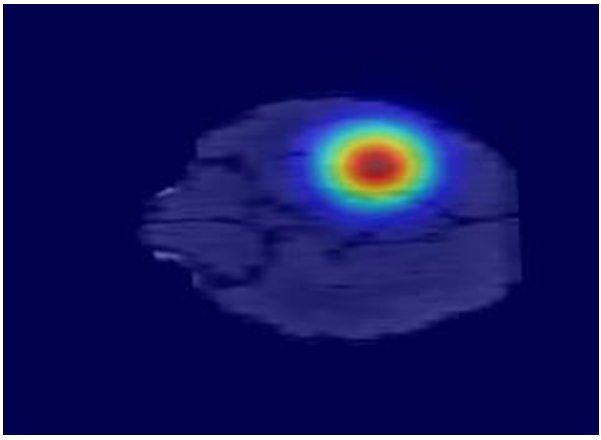

Fig. 10. GRAD-CAM heatmaps of EfficientNetB4

9. Limitations

It should be noted that this study had a number of limitations, even if it was successful in attaining high performance. Cohort differences: The MRI and molecular data used in this study are from particular patient cohorts and environments. Differences in scanner models, imaging protocols, and patient demographics across various hospitals could affect the model's performance. While the proposed multimodal fusion model aims to learn representations that are invariant to cohorts, some cohort-specific information may still be present.Bias in results: Bias in the dataset is a possible concern. An imbalanced distribution of glioma grades, molecular characteristics, or demographic information in the datasets could lead the model to focus on the majority pattern, which could impact fairness and generalization.Insufficient external testing: The model is evaluated using the internal data that is currently available; thorough testing on independent external data from other institutions has not yet been completed. The confidence in the model's performance on actual clinical data may be impacted by this. Generalizability across protocols: The model's generalizability across other MRI protocols, scanners, and institutions still has to be determined, despite key preprocessing processes like intensity normalization and resampling being carried out to accommodate for heterogeneity. In order to guarantee reliable performance, future research will concentrate on multi-institutional validation and the application of harmonization or domain adaption techniques.

10. Conclusion and Future Work

A crucial neuro-oncology task that requires the creation of accurate and explicable AI models is glioma-grade classification. A glioma-grade prediction model based on molecular characteristics and MRI imaging is the recommended approach. This ultimately results in proof that the accuracy and consistency of glioma grading are greatly increased by integrating multimodal data, such as MRI scans and molecular biomarkers.The consistent performance across mixed datasets demonstrates that the proposed fusion mechanism successfully mitigates cohort variability and learns population-robust feature representations.Efficientnet gives 99.4% accuracy by using imaging modality, XGBoost gives 92% accuracy by using molecular biomarkers modality, and Multimodal Fused model gives 99.8% accuracy for Glioma grade classification. The proposed model leverages not only the strengths of both imaging and molecular features in a complementary fashion but also incorporates visual explainability using Grad-CAM for more transparency and interpretability. The model assures data-driven yet clinically meaningful diagnostic predictions by highlighting the most relevant tumor regions that influence the decisions of the model. This interpretability allows radiologists and clinicians to validate what areas the AI is focusing on, building greater trust and accountability into diagnostic automated systems. All things put into consideration, the suggested multimodal and explainable method covers the gap between AI and clinical practice, opening the way to more transparent, reliable, and effective AI-assisted neuro-oncology decision-making.Future research will concentrate on testing the proposed multimodal model in various clinical centers to guarantee its effectiveness on different patient groups. We aim to add more datasets with different demographics, scanners, and protocols to further eliminate biases for specific groups of patients. Testing the model on completely separate external datasets will be a high priority to verify its usability in real-world clinical practice. Furthermore, more sophisticated methods for bias elimination and fairness, such as balanced sampling and domain adaptation, will be investigated to improve its usability for all patient groups.

Conflict of Interest: The authors declare that they have no conflict of interest.

Funding Declaration: This research did not receive any specific grant from funding agencies in the public, commercial or not-for profit sectors.

Ethics Declaration: Not applicable.

Clinical Trials: Not applicable.
Data Availability Statement: Data will be available upon reasonable request.

References

1. Louis, D.N.; Perry, A.; Wesseling, P. The 2021 WHO Classification of Tumors of the Central Nervous System: A summary. *Neuro Oncol.* **2021**, *23*, 1231–1251.
2. Kobayashi, K.; Miyake, M.; Takahashi, M. Observing deep radiomics for the classification of glioma grades. *Sci. Rep.* **2021**, *11*, 10942
3. Brancato, V.; Cerrone, M.; Lavitrano, M. A Systematic Review of the Current Status and Quality of Radiomics for Glioma Differential Diagnosis. *Cancers* **2022**, *14*, 2731.
4. Jang, K.; Russo, C.; Di Ieva, A. Radiomics in gliomas: Clinical implications of computational modeling and fractal-based analysis. *Neuroradiology* **2020**, *62*, 771–790
5. Wang Z, Wang L, Wang Y. Radiomics in glioma: emerging trends and challenges. Ann ClinTransl Neurol. 2025 Mar;12(3):460-477. doi: 10.1002/acn3.52306. Epub 2025 Feb 3. PMID: 39901654; PMCID: PMC11920724.
6. Merkaj, S.; Bahar, R.C.; Zeevi, T.; Lin, M.; Ikuta, I.; Bousabarah, K.; Cassinelli Petersen, G.I.; Staib, L.; Payabvash, S.; Mongan, J.T.; et al. Machine Learning Tools for Image-Based Glioma Grading and the Quality of Their Reporting: Challenges and Opportunities. *Cancers* **2022**, *14*, 2623.
7. P. C. Tripathi and S. Bag, "A computer-aided grading of glioma tumor using deep residual networks fusion," Computer Methods and Programs in Biomedicine, vol. 215, 2022.
8. Taşçı, E., Zhuge, Y., Kaur, H., Camphausen, K., &Krauze, A.V. (2022). Hierarchical Voting-Based Feature Selection and Ensemble Learning Model Scheme for Glioma Grading with Clinical and Molecular Characteristics. International Journal of Molecular Sciences, 23.
9. Rasheed, M. et al. Recognizing brain tumors using adaptive noise filtering and statistical features. Diagnostics 13(8), 1451 (2023).
10. F. Mercaldo, L. Brunese, F. Martinelli, A. Santone, and M. Cesarelli, "Explainable Convolutional Neural Networks for Brain Cancer Detection and Localisation," Sensors, vol. 23, no. 17, pp. 1–16, 2023.
11. Shargunam, S., &Rajakumar, G. (2023). An efficient glioma classification and grade detection using hybrid convolutional neural network-based SVM model. The Imaging Science Journal, 72(1), 1–22. https://doi.org/10.1080/13682199.2023.2215031
12. Wen, L., Sun, H., Liang, G. et al. A deep ensemble learning framework for glioma segmentation and grading prediction. Sci Rep 15, 4448 (2025). https://doi.org/10.1038/s41598-025-87127-z

13. Tonni, S. I., Sheakh, M. A., Tahosin, M. S., Hasan, M. Z., Shuva, T. F., Bhuiyan, T.,Moni, M. A. (2025). A hybrid transfer learning framework for brain tumor diagnosis.
14. Advanced Intelligent Systems, 7(3) ,doi:https://doi.org/10.1002/aisy.202400495
15. M, S., BV, B., D, P. et al. Efficient brain tumor grade classification using ensemble deep learning models. BMC Med Imaging 24, 297 (2024). https://doi.org/10.1186/s12880-024-01476-1
16. L. Li, F. Xiao, S. Wang, et al., "Preoperative prediction of MGMT promoter methylation in glioblastoma based on multiregional and multi-sequence MRI radiomics analysis," Scientific Reports, vol. 14, p. 16031, 2024, doi: 10.1038/s41598-024-66653-2.
17. Q. Li, C. Xiang, X. Zeng, et al., "Multimodal radiomics in glioma: Predicting recurrence in the peritumoural brain zone using integrated MRI," BMC Medical Imaging, vol. 25, p. 321, 2025, doi: 10.1186/s12880-025-01853-4.
18. Y. Zhu, D. Huang, Y. Ji, R. Wang, Y. Li, Y. Xu, Y. Zhuang, Z. Liu, Y. Li, and W. Wang, "Heterogeneity phenotypes in recurrent glioblastoma: A multimodal MRI-based spatial mapping framework for precision treatment," BMC Medical Imaging, vol. 25, no. 1, p. 386, Sep. 2025, doi: 10.1186/s12880-025-01929-1.
19. Z. Zhuang, J. Lin, Z. Wan, J. Weng, Z. Yuan, Y. Xie, Z. Liu, P. Xie, S. Mao, Z. Wang, X. Wang, M. Huang, Y. Luo, and H. Yu, "Radiogenomic profiling of global DNA methylation associated with molecular phenotypes and immune features in glioma," BMC Medicine, vol. 22, no. 1, p. 352, Sep. 2024, doi: 10.1186/s12916-024-03573-y.
20. Z. Liu, Y. Wu, H. Xu, et al., "Multimodal fusion of radio-pathology and proteogenomics identify integrated glioma subtypes with prognostic and therapeutic opportunities," Nature Communications, vol. 16, p. 3510, 2025, doi: 10.1038/s41467-025-58675-9.
21. A. B. Ahanger, S. W. Aalam, T. A. Masoodi, et al., "Radiogenomics and machine learning predict oncogenic signaling pathways in glioblastoma," Journal of Translational Medicine, vol. 23, p. 121, 2025, doi: 10.1186/s12967-025-06101-5.
22. https://www.kaggle.com/datasets/aryashah2k/brain-tumor-segmentation-brats-2019
23. Tasci, E., Camphausen, K., Krauze, A., &Zhuge, Y. (2022). Glioma Grading Clinical and Mutation Features [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5R62J.
24. https://www.cancerimagingarchive.net/collection/tcga-lgg/
25. https://www.cancerimagingarchive.net/collection/tcga-gbm/
26. Vure, R.B., Pappala, L.K. Enhanced brain tumor classification framework using deep learning. Sci Rep 15, 35814 (2025). https://doi.org/10.1038/s41598-025-19882-y
27. Younis, E. M. G., Mahmoud, M. N., Albarrak, A. M., & Ibrahim, I. A. (2024). A Hybrid Deep Learning Model with Data Augmentation to Improve Tumor

Classification Using MRI Images. Diagnostics, 14(23), 2710. https://doi.org/10.3390/diagnostics14232710

28. Yurtsever, M., Atay, Y., Arslan, B. et al. Development of brain tumor radiogenomic classification using GAN-based augmentation of MRI slices in the newly released gazi brains dataset. BMC Med Inform DecisMak 24, 285 (2024). https://doi.org/10.1186/s12911-024-02699-6
29. Kumar, V., Lalotra, G. S., Sasikala, P., Rajput, D. S., Kaluri, R., Lakshmanna, K., Shorfuzzaman, M., Alsufyani, A., & Uddin, M. (2022). Addressing Binary Classification over Class Imbalanced Clinical Datasets Using Computationally Intelligent Techniques. Healthcare, 10(7), 1293.
30. Akpinar, E., Oduncuoglu, M. Hybrid classical and quantum computing for enhanced glioma tumor classification using TCGA data. Sci Rep 15, 25935 (2025). https://doi.org/10.1038/s41598-025-97067-3
31. F. Mercaldo, L. Brunese, F. Martinelli, A. Santone, and M. Cesarelli, "Explainable Convolutional Neural Networks for Brain Cancer Detection and Localisation," Sensors, vol. 23, no. 17, pp. 1–16, 2023.
32. S. Mandloi, M. Zuber, and R. K. Gupta, "An explainable brain tumor detection and classification model using deep learning and layer- wise relevance propagation," Multimedia Tools and Applications, vol. 83, pp. 33 753–33 783, 2023. [Online]. Available: https://doi.org/10.1007/s11042-023-16708-9
33. Khan MA, Hussain MZ, Mehmood S, Khan MF, Ahmad M, Mazhar T, Shahzad T, Saeed MM. Transfer learning for accurate brain tumor classification in MRI: a step forward in medical diagnostics. DiscovOncol. 2025 Jun 9;16(1):1040. doi: 10.1007/s12672-025-02671-4. PMID: 40490586; PMCID: PMC12149374.
34. Md. Monirul Islam, PremaBarua, Moshiur Rahman, TanvirAhammed, LaboniAkter, JiaUddin,Transfer learning architectures with fine-tuning for brain tumor classification using magnetic resonance imaging,HealthcareAnalytics,Volume 4,2023,100270,ISSN 2772-4425,
35. https://doi.org/10.1016/j.health.2023.100270.
36. Tandel GS, Tiwari A, Kakde OG, Gupta N, Saba L, Suri JS. Role of Ensemble Deep Learning for Brain Tumor Classification in Multiple Magnetic Resonance Imaging Sequence Data. Diagnostics (Basel). 2023 Jan 28;13(3):481. doi: 10.3390/diagnostics13030481. PMID: 36766587; PMCID: PMC9914433.
37. S. Shirae, S. S. Debsarkar, H. Kawanaka, B. Aronow and V. B. S. Prasath, "Multimodal Ensemble Fusion Deep Learning Using Histopathological Images and Clinical Data for Glioma Subtype Classification," in IEEE Access, vol. 13, pp. 57780-57797, 2025, doi: 10.1109/ACCESS.2025.3556713
38. Galanakis, I., Soldatos, R. F., Karanikolas, N., Voulodimos, A., Voyiatzis, I., &Samarakou, M. (2025). Early and Late Fusion for Multimodal Aggression Prediction

in Dementia Patients: A Comparative Analysis. Applied Sciences, 15(11), 5823. https://doi.org/10.3390/app15115823
39. H. M. Bas, H. J. V. D. Velden, Kuijf, G. A. Kenneth, M. A. Gilhuijs, and Viergever, "Explainable artificial intelligence (XAI) in deep learning based medical image analysis," Medical Image Analysis, vol. 79, 2022.
40. T. A. A. Abdullah, M. S. M. Zahid, and W. Ali, "A Review of Interpretable ML in Healthcare: Taxonomy, Applications, Challenges, and Future Directions. ," Symmetry 2021, vol. 13, no. 12, pp. 2439–2439, 2021. [Online]. Available: https://doi.org/10.3390/sym13122439
41. D. Martino, F. Delmastro, and F, "Explainable AI for clinical and remote health applications: a survey on tabular and time series data," ArtifIntell Rev, vol. 56, pp. 5261–5315, 2023.
42. Osadebey M, Liu Q, Fuster-Garcia E, Emblem KE. Interpreting deep learning models for glioma survival classification using visualization and textual explanations. BMC Med Inform DecisMak. 2023 Oct 18;23(1):225. doi: 10.1186/s12911-023-02320-2.
43. M, M.M., T. R, M., V, V.K. et al. Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with Resnet 50. BMC Med Imaging 24, 107 (2024). https://doi.org/10.1186/s12880-024-01292-7